Topological order in a three-dimensional toric code at finite temperature

Claudio Castelnovo[1] and Claudio Chamon[2]
[1]*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford OX1 3NP, United Kingdom*
[2]*Department of Physics, Boston University, Boston, Massachusetts 02215, USA*
(Received 12 May 2008; revised manuscript received 31 July 2008; published 21 October 2008)

We study topological order in a toric code in three spatial dimensions or a 3+1D $\mathbb{Z}_2$ gauge theory at finite temperature. We compute exactly the topological entropy of the system and show that it drops, for any infinitesimal temperature, to half its value at zero temperature. The remaining half of the entropy stays constant up to a critical temperature T_c, dropping to zero above T_c. These results show that topologically ordered phases exist at finite temperatures, and we give a simple interpretation of the order in terms of fluctuating strings and membranes and how thermally induced point defects affect these extended structures. Finally, we discuss the nature of the topological order at finite temperature and its quantum and classical aspects.

DOI: 10.1103/PhysRevB.78.155120 PACS number(s): 03.65.Vf, 11.15.−q, 03.67.Mn

I. INTRODUCTION

Some quantum systems are characterized by a type of order which cannot be captured by a local order parameter that signals broken symmetries, but instead the order is topological in nature.[1] One of the ways in which this topological order manifests itself is in a ground-state (GS) degeneracy that cannot be lifted by any local perturbation, and that depends on the genus of the surface in which the system is defined. Recently, there have been efforts to find characterizations of topological order other than ground-state degeneracies, in particular, exploring the entanglement in the ground-state wave function.[2,3]

At zero temperature, topological order can be detected using the von Neumann entanglement entropy, more precisely a topological contribution to it that can be separated from the boundary contribution by appropriate subtractions of different bipartitions of the system.[2,3] Because the pure state density matrix is constructed from the ground state, it was argued in Ref. 2 that topological order is a property of the wave function, and not of the Hamiltonian, at absolute zero temperature.

An interesting question is what happens with topological order at finite temperature. The question is relevant because thermal fluctuations, no matter how small, are present in any laboratory system. To address this issue, it was proposed in Ref. 4 to use the topological entropy as a probe of topological order, but to compute it using an equilibrium mixed-state-density matrix $\hat{\rho}=Z^{-1}e^{-\beta\hat{H}}$. It becomes clear that, as opposed to zero temperature for which one can do away with the full information contained in the Hamiltonian and just use the ground-state wave function, topological order, if present at finite temperature, must be a property of the Hamiltonian.

The topological entropy was computed exactly for the two-dimensional (2D) Kitaev model[5] at finite temperature T, and it was shown that the infinite system size limit and the $T\to 0$ limit do not commute and that at finite T the topological entropy vanishes in the thermodynamic limit. Thus, it was argued that the topological order in the 2D system was fragile.[4,6,7]

Here we show that the situation in three dimensions is rather different using the three-dimensional (3D) version of Kitaev's model as an example.[8] In contrast to two dimensions, topological order survives up to a phase transition at a finite temperature T_c. The order can be probed through a nonvanishing topological entropy, as well as understood from a simple cartoon picture that we present in the paper, using the fact that in 3D strings can move around point defects (as opposed to two dimensions).

We prove in this paper that the von Neumann entropy of a subsystem $\mathcal{A}$ of a $\mathbb{Z}_2$ gauge model such as Kitaev's toric code, in any number of dimensions, can be always decomposed into two additive contributions from each of the two gauge structures (magnetic and electric),[9]

$$S_{\mathrm{vN}}(\mathcal{A};T)=S_{\mathrm{vN}}^{(S)}(\mathcal{A};T/\lambda_A)+S_{\mathrm{vN}}^{(P)}(\mathcal{A};T/\lambda_B), \tag{1.1}$$

where $S_{\mathrm{vN}}^{(S)}$ and $S_{\mathrm{vN}}^{(P)}$ are the separable contributions from the stars and plaquettes of the model and λ_A and λ_B the associated coupling constants for these two structures. Consequently, the same additive separability holds for the topological entropy, which is a sum of two independent contributions,

$$S_{\mathrm{topo}}(T)=S_{\mathrm{topo}}^{(S)}(T/\lambda_A)+S_{\mathrm{topo}}^{(P)}(T/\lambda_B). \tag{1.2}$$

One of the contributions, $S_{\mathrm{topo}}^{(S)}$, evaporates for any infinitesimal temperature in the thermodynamic limit, just as in two dimensions, but the other one, $S_{\mathrm{topo}}^{(P)}$, remains constant up to a finite-temperature phase transition at $T_c=1.313\,346(3)\lambda_B$, which occurs for the 3D case,

$$S_{\mathrm{topo}}^{\mathrm{3D}}(T)=\begin{cases}2\ln 2, & T=0\\ \ln 2, & 0<T<T_c\\ 0, & T>T_c.\end{cases} \tag{1.3}$$

As a consequence of these results, we argue that topological order can be well defined at finite temperatures in three dimensions.[10] This finding raises the following interesting question: is the finite T order classical or quantum? Perhaps another way to ask the question is the following: what kind of information can be robustly stored using the isolated topological sectors in phase space that cannot be connected by local moves (2^3 such states in three dimensions)? Classical (bits) or quantum (qubits) information? While we cannot argue that the system does not realize a full quantum memory,

 ©2008 The American Physical Society

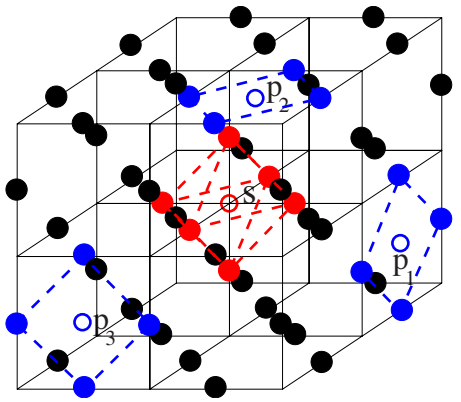

FIG. 1. (Color online) Illustration of the Kitaev model in three dimensions, with explicit examples of a star operator $A_s=\Pi\sigma_i^x$ at the lattice site s and of three plaquette operators $B_p=\Pi\sigma_i^z$ at the plaquette-dual sites p_1, p_2, and p_3. The σ spin index i labels, respectively, the six (red) spins around s and the four (blue) spins around p (connected by dashed lines).

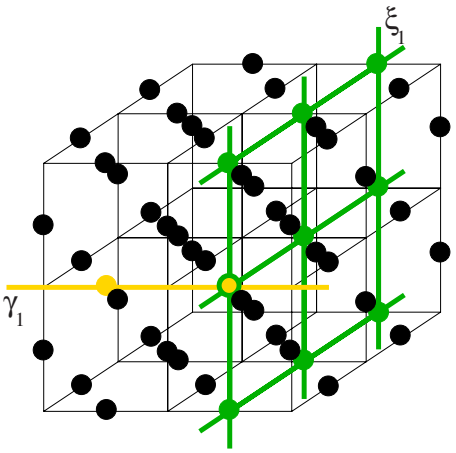

FIG. 2. (Color online) Two examples of the nonlocal operators needed to distinguish between the degenerate GS of the 3D Kitaev model.

we can at the least argue that it can store probabilistic information (pbits—probabilistic bits[11]) in the form of a quantum superposition of states in the different topological sectors, where the square amplitudes for all states in a given sector (a probability) do not fluctuate in the thermodynamic limit if the coupling to a thermal bath is local. However, the relative phases for all these amplitudes could be scrambled. This weak type of quantum superposition is not discernible from a classical probability distribution. Finally, this example shows that the notion of classical topological order, suggested for hard constrained models in two dimensions,[12] is well defined in three dimensions without resorting to any hard constraints.

II. MODEL

Consider a three-dimensional version of Kitaev's toric code,[8] defined on a simple-cubic lattice of size $N=L\times L\times L$, with periodic boundary conditions (BCs) and spin-1/2 degrees of freedom $\vec{\sigma}_i$ living on the bonds, $i=1,\ldots,3N$ (σ_i^x, σ_i^y, and σ_i^z being the three Pauli matrices). Let us label the centers of each single square plaquette in the lattice with $p=1,\ldots,3N$ and each site of the cubic lattice with $s=1,\ldots,N$.

Let us define the plaquette and star operators on the lattice,

$$B_p=\prod_{i\in p}\sigma_i^z,\quad A_s=\prod_{i\in s}\sigma_i^x, \tag{2.1}$$

as illustrated in Fig. 1. The Hamiltonian of the model can then be written in terms of these operators as

$$H=-\lambda_A\sum_s A_s-\lambda_B\sum_p B_p, \tag{2.2}$$

where λ_A and λ_B are two real positive constants.

Notice that all star and plaquette operators commute, but they are not all independent. While only the product of all star operators equals the identity, therefore leaving $N-1$ independent star operators, the product of the plaquette operators around each cubic unit cell gives the identity, therefore introducing $N-1$ constraints in the $3N$ total plaquette operators (the product of all but one cube is equivalent to that same cube, so we have one less constraint). Moreover, three additional constraints come from the fact that the product of all plaquette operators along any crystal plane (CP) in the cubic lattice (i.e., $\langle x,y\rangle$, $\langle x,z\rangle$, or $\langle y,z\rangle$) yields the identity, and we are finally left with $2N-2$ independent plaquette operators.

The GS manifold of the system is identified by having all plaquette and star quantum numbers equal to +1, and it is $2^{3N-(N-1)-(2N-2)}=2^3$ dimensional, assuming periodic boundary conditions in all three directions. Similarly to the 2D case, one can notice that this degeneracy has a topological nature, and the different sectors are distinguished by three nonlocal operators,

$$\Gamma_1=\prod_{i\in\gamma_1}\sigma_i^z,\quad \Gamma_2=\prod_{i\in\gamma_2}\sigma_i^z,\quad \Gamma_3=\prod_{i\in\gamma_3}\sigma_i^z \tag{2.3}$$

or

$$\Xi_1=\prod_{i\in\xi_1}\sigma_i^x,\quad \Xi_2=\prod_{i\in\xi_2}\sigma_i^x,\quad \Xi_3=\prod_{i\in\xi_3}\sigma_i^x, \tag{2.4}$$

which are diagonal in the σ^z and σ^x bases, respectively. Here the γ_i can be any winding paths along the edges of the cubic lattice in each of the three crystal directions (x, y, or z), and the ξ_i can be any winding planes perpendicular to each of the crystal directions and passing through the midpoints of the corresponding edges of the cubic lattice (i.e., crystal planes in the dual lattice whose sites sit at the centers of the elementary cubic cells). Two examples are shown in Fig. 2 for clarity.

In the σ^z basis and in the topological sector where all the Γ_i equal +1, the GS wave function of the system can be written as

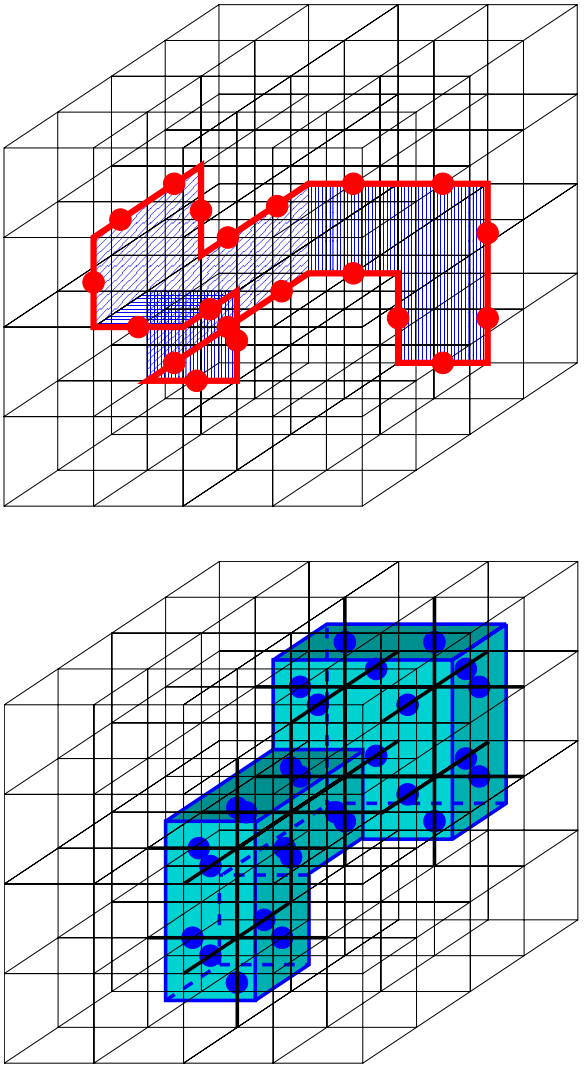

FIG. 3. (Color online) Two examples of the underlying structures of the 3D Kitaev model: the closed σ^z loops along the edges of the cubic lattice, which satisfy $\Pi_{\text{loop}}\sigma_i^z=1$ and the closed σ^x membranes in the body-centered dual lattice, satisfying $\Pi_{\text{membrane}}\sigma_i^x=1$.

$$|\text{GS}\rangle = \frac{1}{|G|^{1/2}} \sum_{g\in G} g|0\rangle, \tag{2.5}$$

where $|0\rangle$ is any state in the sector, say the state with all the $\sigma_i^z=+1$, and G is the Abelian group generated by all products of star operators (of dimension $|G|=2^{N-1}$). In the σ^x basis and in the topological sector where all the Ξ_i equal +1, the GS wave function of the system can be written as in Eq. (2.5), where now $|0\rangle$ is any state in the sector, say the state with all the $\sigma_i^x=+1$, and G is the Abelian group generated by all products of plaquette operators (of dimension $|G|=2^{2(N-1)}$).

Notice the two *different* underlying structures in the system: the closed σ^z *loops* along the edges of the cubic lattice, which satisfy $\Pi_{\text{loop}}\sigma_i^z=1$ identically, and the closed σ^x *membranes* in the body-centered dual lattice (locally perpendicular to the edges of the original lattice), satisfying $\Pi_{\text{membrane}}\sigma_i^x=1$ identically (see Fig. 3).

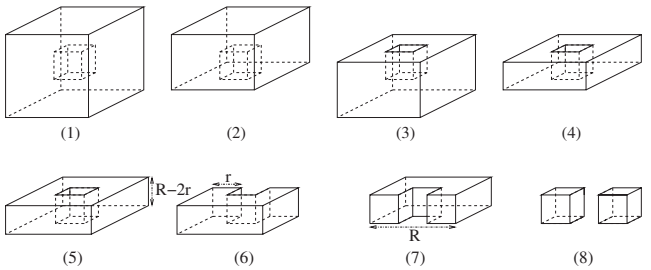

FIG. 4. Illustration of the two bipartition schemes used for the 3D Kitaev model: spherical (top) and donut shaped (bottom).

III. TOPOLOGICAL ENTROPY AT ZERO TEMPERATURE

Let us first compute the zero-temperature topological entropy of the system using a three-dimensional version of the bipartition scheme proposed by Levin and Wen[2] in two dimensions. Notice, however, that in three dimensions a bipartition can be topologically nontrivial with respect to closed loops but not with respect to closed membranes—e.g., a donut—and vice versa—e.g., a spherical shell. Thus, there is no unique way to generalize the 2D case. Two equally valid options are illustrated in Fig. 4 based on a "spherical" (1–4) and a "donut-shaped" (5–8) bipartition scheme, respectively.

In the σ^z basis,[13] where G is generated by the star operators, the calculation of the entanglement entropy S_{vN} proceeds as in the 2D case.[4,12,14] Using the group property of G in Eq. (2.5), one can show that

$$S_{\text{vN}}(\mathcal{A}) = -\ln\frac{d_{\mathcal{A}}d_{\mathcal{B}}}{|G|}, \tag{3.1}$$

where $d_{\mathcal{A}}$ is the dimension of the subgroup $G_{\mathcal{A}}\subset G$ containing all the elements of G that act as the identity on $\mathcal{B}$, $G_{\mathcal{A}}=\{g\in G | g=g_{\mathcal{A}}\otimes \mathbf{1}_{\mathcal{B}}\}$, and similarly for subsystem $\mathcal{B}$. As in the 2D case, these subgroup dimensions depend on the number $\mathcal{N}_{\mathcal{A}}^{(s)}$ ($\mathcal{N}_{\mathcal{B}}^{(s)}$) of star operators acting solely on spins in $\mathcal{A}$ ($\mathcal{B}$) and on the number $m_{\mathcal{A}}$ ($m_{\mathcal{B}}$) of connected components of $\mathcal{A}$ ($\mathcal{B}$),

$$d_{\mathcal{A}} = 2^{\mathcal{N}_{\mathcal{A}}^{(s)}+m_{\mathcal{B}}-1}, \tag{3.2}$$

$$d_{\mathcal{B}} = 2^{\mathcal{N}_{\mathcal{B}}^{(s)}+m_{\mathcal{A}}-1}. \tag{3.3}$$

The $m_{\mathcal{B}}$ contribution to $d_{\mathcal{A}}$ and the $m_{\mathcal{A}}$ contribution to $d_{\mathcal{B}}$ come from the so-called collective operations, i.e., elements of the groups $G_{\mathcal{A}}$ ($G_{\mathcal{B}}$) that cannot be expressed as products of star operators in $\mathcal{A}$ ($\mathcal{B}$). In the 3D case, such collective operations correspond to noncontractible closed membranes. In this respect, bipartitions 1 and 8 are special in that subsystems $1\mathcal{B}$ and $8\mathcal{A}$ are composed of two separate connected components ($m_{1\mathcal{B}}=m_{8\mathcal{A}}=2$), while all other subsystems have only one component.

We can then compute the topological entropy S_{topo} of the system in the σ^z basis from either the spherical or the donut-shaped bipartition scheme,

$$S_{\text{topo}}^{(z)} = \lim_{r,R\to\infty}[-S_{\text{vN}}^{1,\mathcal{A}} + S_{\text{vN}}^{2,\mathcal{A}} + S_{\text{vN}}^{3,\mathcal{A}} - S_{\text{vN}}^{4,\mathcal{A}}] = \ln 2,$$

$$S_{\text{topo}}^{(z)} = \lim_{r,R\to\infty} [-S_{\text{vN}}^{5,\mathcal{A}} + S_{\text{vN}}^{6,\mathcal{A}} + S_{\text{vN}}^{7,\mathcal{A}} - S_{\text{vN}}^{8,\mathcal{A}}] = \ln 2, \quad (3.4)$$

where we used the fact that all $\mathcal{N}^{(s)}$ contributions cancel out exactly. In fact, if we define $\mathcal{N}_{i,\mathcal{AB}}^{(s)} = \mathcal{N}^{(s)} - \mathcal{N}_{i,\mathcal{A}}^{(s)} - \mathcal{N}_{i\mathcal{B}}^{(s)}$ to be the number of star operators acting simultaneously on $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{N}^{(s)} = N$ being the total number of star operators in the system, one can show that

$$\mathcal{N}_{1,\mathcal{A}}^{(s)} + \mathcal{N}_{1\mathcal{B}}^{(s)} + \mathcal{N}_{4,\mathcal{A}}^{(s)} + \mathcal{N}_{4\mathcal{B}}^{(s)} - [\mathcal{N}_{2,\mathcal{A}}^{(s)} + \mathcal{N}_{2\mathcal{B}}^{(s)} + \mathcal{N}_{3,\mathcal{A}}^{(s)} + \mathcal{N}_{3\mathcal{B}}^{(s)}]$$
$$= 2N - \mathcal{N}_{1,\mathcal{AB}}^{(s)} - \mathcal{N}_{4,\mathcal{AB}}^{(s)} - 2N + \mathcal{N}_{2,\mathcal{AB}}^{(s)} + \mathcal{N}_{3,\mathcal{AB}}^{(s)} = 0. \quad (3.5)$$

This result relies on the fact that the total boundary in bipartitions 1 and 4 is the same—with the same multiplicity and with precisely the same edge and corner structure—as in bipartitions 2 and 3 by construction. Therefore, $\mathcal{N}_{1,\mathcal{AB}}^{(s)} + \mathcal{N}_{4,\mathcal{AB}}^{(s)} = \mathcal{N}_{2,\mathcal{AB}}^{(s)} + \mathcal{N}_{3,\mathcal{AB}}^{(s)}$, and similarly for bipartitions 5–8.

Let us also compute the topological entropy in the σ^x basis,[13] as it will be useful when we consider the finite-temperature case. The group G is now generated by the plaquette operators, which are highly redundant and require more involved calculations to obtain the von Neumann entropy S_{vN}. In fact, while Eq. (3.1) still holds, one needs to count the number of independent plaquette generators of subgroups $G_{\mathcal{A}}$ and $G_{\mathcal{B}}$ in order to obtain the equivalent of Eqs. (3.2) and (3.3). Notice that the collective operations are now given by closed loops, and only bipartitions 4 and 5 allow for nontrivial (i.e., noncontractible) loops.

As we discussed before, $|G| = 2^{2(N-1)}$. This arises from counting all independent generators of G as the total number of plaquettes in G (all possible generators) minus the number of independent constraints. These are all but one of the cubic unit cells plus three crystal planes. Similar arguments apply to the bipartitions 1–8. Notice that in all of the bipartitions, subsystem $\mathcal{A}$ does *not* contain any entire crystal plane, while subsystem $\mathcal{B}$ *always* contains all three crystal planes. Taking advantage of this simplification, in the following it will be understood that $G_{\mathcal{B}}$ has three less independent generators with respect to $G_{\mathcal{A}}$.

Let us proceed case by case. For bipartitions where both $\mathcal{A}$ and $\mathcal{B}$ have only one connected component without handles, such as bipartitions 2, 3, 6, and 7 in Fig. 4, the group $G_{\mathcal{A}}$ (equivalently $G_{\mathcal{B}}$) is generated by all the plaquette operators acting solely on $\mathcal{A}$, subject to the constraints given by all cubic unit cells entirely contained in $\mathcal{A}$. There are no collective operations in this case, and one obtains

$$d_{\mathcal{A}} = 2^{\mathcal{N}_{\mathcal{A}}^{(p)} - \mathcal{N}_{\mathcal{A}}^{(c)}}, \quad (3.6)$$

$$d_{\mathcal{B}} = 2^{\mathcal{N}_{\mathcal{B}}^{(p)} - \mathcal{N}_{\mathcal{B}}^{(c)}}, \quad (3.7)$$

where $\mathcal{N}_{\mathcal{A}}^{(p)}$ is the number of plaquette operators acting on spins in $\mathcal{A}$, $\mathcal{N}_{\mathcal{A}}^{(c)}$ is the number of cubic unit cells in $\mathcal{A}$, and similarly for $\mathcal{B}$.

Consider then the case of bipartition 4 (equivalently 5). Although both $\mathcal{A}$ and $\mathcal{B}$ are still connected, the presence of a handle allows now for collective operations. Take a crystal plane perpendicular to the largest surface of subsystem $\mathcal{A}$ and draw it so that it bisects the donut into two identical U-shaped portions [see Fig. 5 (top)]. The intersection of this plane with $\mathcal{A}$ gives two rectangles of size $r \times (R-2r)$, a distance $R-2r$ apart. Now take the product of all plaquettes belonging to one of the rectangles plus those at its boundary. The resulting operation acts on $\mathcal{B}$ alone, yet it cannot be constructed from plaquettes in $\mathcal{B}$ because the "outer boundary" of the rectangle cannot be the *sole* boundary of a surface in $\mathcal{B}$. Notice that this collective operation can be deformed at will and moved along the donut by appropriate products of plaquettes in $\mathcal{B}$; therefore there is only one such independent operation. Similar arguments apply if we repeat the construction starting from a plane parallel to the largest surface of the subsystem $\mathcal{A}$, again chosen so as to bisect the donut. This yields another independent collective operation acting now on $\mathcal{A}$ [see Fig. 5 (bottom)]. As a result,

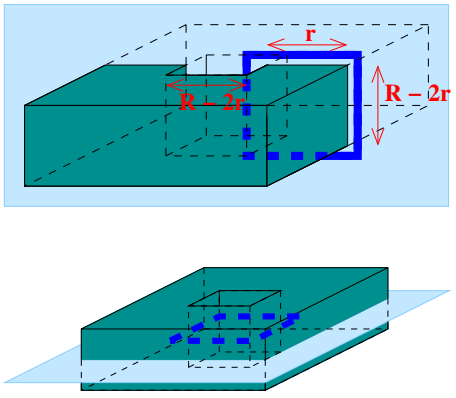

FIG. 5. (Color online) Illustration of the collective operations in bipartitions 4 and 5 in the σ^x basis, acting on subsystem $\mathcal{B}$ (top) and subsystem $\mathcal{A}$ (bottom), respectively.

$$d_{\mathcal{A}} = 2^{\mathcal{N}_{\mathcal{A}}^{(p)} - \mathcal{N}_{\mathcal{A}}^{(c)} + n_{\mathcal{A}}}, \quad (3.8)$$

$$d_{\mathcal{B}} = 2^{\mathcal{N}_{\mathcal{B}}^{(p)} - \mathcal{N}_{\mathcal{B}}^{(c)} + n_{\mathcal{B}}}, \quad (3.9)$$

where $n_{\mathcal{A}} = 1$ and $n_{\mathcal{B}} = 1$ are the number of collective operations in $\mathcal{A}$ and $\mathcal{B}$, respectively.

Finally, one can show that there are no collective operations in the σ^x basis in bipartitions 1 and 8. In fact, all closed loops are contractible to a point both in $\mathcal{A}$ and in $\mathcal{B}$ in these bipartitions. However, the disconnected nature of subsystem $\mathcal{B}$ in bipartition 1 (equivalently, subsystem $\mathcal{A}$ in bipartition 8) requires special care in the counting of the independent generators of $G_{\mathcal{A}}$ (respectively, $G_{\mathcal{B}}$). As in the previous cases, all plaquettes in $\mathcal{A}$ belong to $G_{\mathcal{A}}$, and all cubic unit cells in $\mathcal{A}$ act as independent constraints toward the counting of the independent generators of $G_{\mathcal{A}}$. However, in bipartition 1, there is a class of closed membranes in $\mathcal{A}$ that cannot be assembled as a product of cubic cells in $\mathcal{A}$. This is the case, for example, of the closed cubic membranes in $\mathcal{A}$ that surround entirely the inner component of $\mathcal{B}$. Any two such membranes can be obtained one from the other via multiplication by cubic unit cells in $\mathcal{A}$. Thus, they only give rise to one additional constraint in the counting of the independent generators. In general, the number of such constraints is given by $m_{\mathcal{B}} - 1$, where $m_{\mathcal{B}}$ is the number of connected com-

ponents of $\mathcal{B}$. Similarly for bipartition 8 and subsystem $\mathcal{B}$, one obtains $m_{\mathcal{A}}-1$ additional constraints, where $m_{\mathcal{A}}$ is the number of connected components of $\mathcal{A}$.

Combining all of the above considerations into a general expression for the dimensions of subgroups $G_{\mathcal{A}}$ and $G_{\mathcal{B}}$ in the σ^x basis, one obtains

$$d_{\mathcal{A}} = 2^{\mathcal{N}_{\mathcal{A}}^{(p)} - \mathcal{N}_{\mathcal{A}}^{(c)} + n_{\mathcal{A}} - (m_{\mathcal{B}}-1) - m_{\mathcal{A}}^{(\mathrm{CP})}}, \tag{3.10}$$

$$d_{\mathcal{B}} = 2^{\mathcal{N}_{\mathcal{B}}^{(p)} - \mathcal{N}_{\mathcal{B}}^{(c)} + n_{\mathcal{B}} - (m_{\mathcal{A}}-1) - m_{\mathcal{B}}^{(\mathrm{CP})}}, \tag{3.11}$$

where $m_{\mathcal{A}}^{(\mathrm{CP})}$ ($m_{\mathcal{B}}^{(\mathrm{CP})}$) is the number of crystal planes CPs entirely contained in $\mathcal{A}$ ($\mathcal{B}$). Recall that all bipartitions of interest have $m_{\mathcal{A}}^{(\mathrm{CP})}=0$ and $m_{\mathcal{B}}^{(\mathrm{CP})}=3$.

We can then use Eq. (3.1) to compute the topological entropy of the system using the spherical and the donut-shaped bipartition schemes in the σ^x basis,

$$\begin{aligned} S_{\mathrm{topo}}^{(x)} &= \lim_{r,R\to\infty} [-S_{\mathrm{vN}}^{1,\mathcal{A}} + S_{\mathrm{vN}}^{2,\mathcal{A}} + S_{\mathrm{vN}}^{3,\mathcal{A}} - S_{\mathrm{vN}}^{4,\mathcal{A}}] \\ &= (-1+2)\ln 2 = \ln 2, \\ S_{\mathrm{topo}}^{(x)} &= \lim_{r,R\to\infty} [-S_{\mathrm{vN}}^{5,\mathcal{A}} + S_{\mathrm{vN}}^{6,\mathcal{A}} + S_{\mathrm{vN}}^{7,\mathcal{A}} - S_{\mathrm{vN}}^{8,\mathcal{A}}] \\ &= (2-1)\ln 2 = \ln 2, \end{aligned} \tag{3.12}$$

where we used the fact that all $\mathcal{N}^{(p)}$ and $\mathcal{N}^{(c)}$ contributions cancel out exactly. In fact, if we define $\mathcal{N}_{i,\mathcal{AB}}^{(p)}=\mathcal{N}^{(p)}-\mathcal{N}_{i,\mathcal{A}}^{(p)}-\mathcal{N}_{i,\mathcal{B}}^{(p)}$ to be the number of plaquette operators acting simultaneously on $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{N}^{(p)}=3N$ being the total number of plaquette operators in the system, and we define $\mathcal{N}_{i,\mathcal{AB}}^{(c)}=\mathcal{N}^{(c)}-\mathcal{N}_{i,\mathcal{A}}^{(c)}-\mathcal{N}_{i,\mathcal{B}}^{(c)}$ to be the number of cubic unit cells simultaneously encompassing spins in $\mathcal{A}$ and in $\mathcal{B}$, $\mathcal{N}^{(c)}=N$ being the total number of cubic unit cells in the system, one can show that

$$\begin{aligned} &(\mathcal{N}_{1,\mathcal{A}}^{(p)} - \mathcal{N}_{1,\mathcal{A}}^{(c)}) + (\mathcal{N}_{1,\mathcal{B}}^{(p)} - \mathcal{N}_{1,\mathcal{B}}^{(c)}) + (\mathcal{N}_{4,\mathcal{A}}^{(p)} - \mathcal{N}_{4,\mathcal{A}}^{(c)}) + (\mathcal{N}_{4,\mathcal{B}}^{(p)} - \mathcal{N}_{4,\mathcal{B}}^{(c)}) \\ &\quad - [(\mathcal{N}_{2,\mathcal{A}}^{(p)} - \mathcal{N}_{2,\mathcal{A}}^{(c)}) + (\mathcal{N}_{2,\mathcal{B}}^{(p)} - \mathcal{N}_{2,\mathcal{B}}^{(c)}) + (\mathcal{N}_{3,\mathcal{A}}^{(p)} - \mathcal{N}_{3,\mathcal{A}}^{(c)}) + (\mathcal{N}_{3,\mathcal{B}}^{(p)} - \mathcal{N}_{3,\mathcal{B}}^{(c)})] \\ &= 4N - \mathcal{N}_{1,\mathcal{AB}}^{(p)} - \mathcal{N}_{4,\mathcal{AB}}^{(p)} + \mathcal{N}_{1,\mathcal{AB}}^{(c)} + \mathcal{N}_{4,\mathcal{AB}}^{(c)} - 4N + \mathcal{N}_{2,\mathcal{AB}}^{(p)} + \mathcal{N}_{3,\mathcal{AB}}^{(p)} - \mathcal{N}_{2,\mathcal{AB}}^{(c)} - \mathcal{N}_{3,\mathcal{AB}}^{(c)} = 0. \end{aligned} \tag{3.13}$$

This result relies on the fact that the total boundary in bipartitions 1 and 4 is the same—with the same multiplicity and with precisely the same edge and corner structure—as in bipartitions 2 and 3 by construction. Therefore, $\mathcal{N}_{1,\mathcal{AB}}^{(p)}+\mathcal{N}_{4,\mathcal{AB}}^{(p)}=\mathcal{N}_{2,\mathcal{AB}}^{(p)}+\mathcal{N}_{3,\mathcal{AB}}^{(p)}$, $\mathcal{N}_{1,\mathcal{AB}}^{(c)}+\mathcal{N}_{4,\mathcal{AB}}^{(c)}=\mathcal{N}_{2,\mathcal{AB}}^{(c)}+\mathcal{N}_{3,\mathcal{AB}}^{(c)}$, and similarly for bipartitions 5–8.

Clearly, both bipartition schemes capture the topological nature of the system and provide an equally valid measure of the topological entropy. In two dimensions the choice of bipartitions 1–4 in Ref. 2 is such that bipartition 1 is topologically equivalent to bipartition 4 upon exchange of subsystem $\mathcal{A}$ with subsystem $\mathcal{B}$, while bipartitions 2 and 3 are actually topologically invariant upon the same exchange. Hence, because the von Neumann entropy for the ground state is symmetric under the exchange of $\mathcal{A}$ and $\mathcal{B}$, the topological contribution measured in the 2D scheme is bound to be double counted, namely, $S_{\mathrm{topo}}=2\ln D=\ln D^2$, where D is the so-called quantum dimension of the system.[2,3] In three dimensions, both the scheme 1–4 and the scheme 5–8 isolate the topological contribution to the entanglement entropy without double counting. Notice that all the bipartitions are topologically invariant under the exchange of $\mathcal{A}$ and $\mathcal{B}$, except for bipartitions 1 and 8. If we want to recover the symmetry of the 2D scheme, a possible solution is to define

$$\begin{aligned} S_{\mathrm{topo}} = \lim_{r,R\to\infty} [&-S_{\mathrm{vN}}^{1,\mathcal{A}} + S_{\mathrm{vN}}^{2,\mathcal{A}} + S_{\mathrm{vN}}^{3,\mathcal{A}} - S_{\mathrm{vN}}^{4,\mathcal{A}} \\ &- S_{\mathrm{vN}}^{5,\mathcal{A}} + S_{\mathrm{vN}}^{6,\mathcal{A}} + S_{\mathrm{vN}}^{7,\mathcal{A}} - S_{\mathrm{vN}}^{8,\mathcal{A}}] = \ln D^2, \end{aligned} \tag{3.14}$$

with $D=2$. As we will see in the following, the symmetric 1–8 choice is actually required if we are interested in studying the finite-temperature case since the von Neumann entropy is no longer invariant upon exchange of $\mathcal{A}$ and $\mathcal{B}$, and a nontopologically symmetric choice of bipartitions would lead to different results depending on whether we work with subsystem $\mathcal{A}$ or subsystem $\mathcal{B}$.[15]

IV. FINITE TEMPERATURE BEHAVIOR

In this section we study the behavior of the entanglement and topological entropies at finite temperature via a generalization of the approach used for the 2D Kitaev model in Ref. 4. A qualitative picture of the effect of thermal fluctuations can be argued by comparison with the two-dimensional case. There the information about the topological sectors is stored in the eigenvalues of winding loop operators, namely, products of spin operators along winding loops. On a torus, there are infinitely many choices for such winding loop operators, but the absence of magnetic and electric charges (i.e., plaquettes and stars with eigenvalue −1) in the gauge structure at zero temperature reduces them to only two independent ones: the two noncontractible winding loops on the torus. Any other can be obtained from these two via multiplication by an appropriate set of plaquette or star operators, which have eigenvalue +1 at $T=0$. Clearly the presence of order 1 (deconfined) thermal defects destroys immediately all topological information stored in the system since the eigenvalues of two loops on opposite sides of a defect are no longer consistent with each other (see Fig. 6).

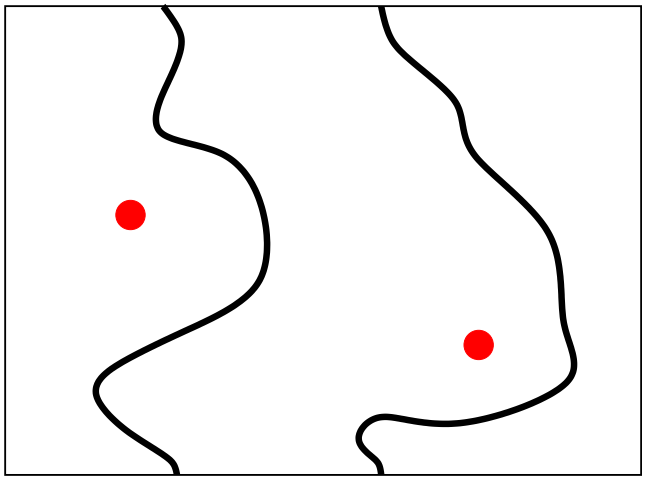

FIG. 6. (Color online) Qualitative illustration of the disruptive effect of two defects (solid red dots) in the 2D Kitaev model on a torus: two winding loops (black wavy lines) on either side of a defect (solid circle), *read off* opposite eigenvalues of the corresponding winding loop operator.

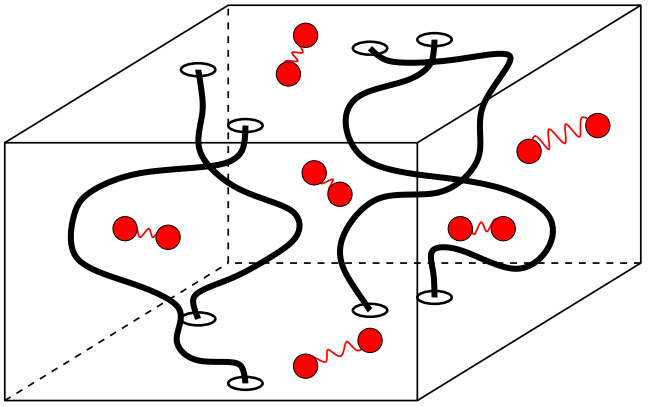

FIG. 7. (Color online) Qualitative illustration of the reason why the topological information stored in the underlying σ^z loop structure of the 3D Kitaev model is robust to thermal fluctuations: even in presence of sparse defects (solid red circles), any two winding loops (black wavy lines), with equal winding numbers, can be smoothly deformed one into the other without crossing any defects. (The wiggly lines represent qualitatively the confining strings between defect "pairs" discussed in the text.)

Let us now consider the case of the Kitaev model in three dimensions. First of all, we need to discuss the two gauge structures separately since they are no longer identical as in two dimensions. If we work in the σ^x basis, then the topological information is stored in the eigenvalues of winding membrane operators, given by the product of all σ^x operators belonging to a closed winding surface locally perpendicular to the bonds of the sites it crosses (see Fig. 2). All possible choices of these membranes yield the same result at zero temperature since the corresponding operators can be obtained one from the other by products of sets of star operators, which have all eigenvalue +1 in the GS. Thermal defects in this case play exactly the same role as in two dimensions since two membranes on opposite sides of a defect *read off* opposite eigenvalues of the corresponding winding membrane operator.

On the other hand, the situation is quite different for the loop operators defined in the σ^z basis. There the topological information is stored in winding loop operators—as in the 2D case—but they are now embedded in three dimensions. Clearly, localized defects have no disruptive effects on the topological information because any two winding loops (with equal winding numbers) can be smoothly deformed one into the other *without crossing any defects* at low enough temperatures (see Fig. 7). This is indeed the case here, where we learn from 3D lattice gauge theory that defective plaquettes are confined at low temperatures. They are created in quadruplets by a single spin-flip operation, and they can be pairwise separated only at the cost of creating a string of defective plaquettes in between the two pairs.[16,17] Therefore, the winding loop operators will keep carrying the same quantum information in presence of a low density of defects. If we were to read out the topological information from the system, we would be getting the correct result as long as the chosen loop does not pass directly *through* a defect.

However, can this information be accessed by means of the same expectation values of loop operators that are used at zero temperature [Eqs. (2.3) and (2.4)]? The answer to this question is negative, as it was recently shown using gauge theory arguments in Ref. 18. A simple reason as to why naively choosing a given loop operator and looking at its expectation value alone does not capture the order below T_c is that, typically, winding loops will pass through at least one defect in the thermodynamic limit [the probability of a loop not crossing any defect scales as $(1-\rho_{\mathrm{def}})^L$, where ρ_{def} is the equilibrium density of defects at a given temperature and L is the linear size of the system]. However, only those loops that avoid the defects contain the topological information. (Recall that in two dimensions the eigenvalues of loop operators, even when they do not pass through defects, differ on two sides of one defect, in contrast to the situation in three dimensions.) This implies that the average expectation value of loop operators is bound to vanish exponentially in system size for any finite density of defects, i.e., for any finite temperature, independently of the nature of the system. As we will show in the following, the topological entropy of the system is capable of capturing these physical differences, and it accurately reflects the topological properties of the different phases.

The physical meaning of the distinct sectors can be understood as follows. Consider preparing the system in a coherent superposition of different topological sectors at zero temperature. Raise the temperature to some value $T<T_c$ and then lower it again back to zero. If defects are confined, transitions between different loop sectors are forbidden throughout the process. We are thus bound to obtain a final state where the probability (magnitude of amplitude square) of finding the final state in each loop sector is the same as in the initial state. In this sense, the loop sectors are protected from thermal fluctuations at low temperatures, and topological order survives at finite temperature ($T<T_c$).

That the system does not change sectors during the time that it is in thermal equilibrium with the bath is a *dynamical* problem (broken ergodicity). This can be understood by con-

trasting the time scales for mixing sectors if defects are confined or deconfined. Deconfined thermal defects are free to randomly walk across the system and induce transitions between different topological sectors by means of creation, system-spanning propagation, and annihilation processes. The characteristic time for a sector-changing process scales therefore as some power of the system size, $\tau_{\text{deconfined}} \sim L^{\alpha}$. In contrast, confined defects will have to overcome an energy barrier of the order of L to be able to wind around the system and induce a change in the topological sector. As a result, their characteristic time scale is instead exponential in system size, $\tau_{\text{confined}} \sim e^{cL}$. Even for rather small systems, confined defects would require time scales larger than the age of the universe to transition between sectors.

An even more interesting situation occurs when both $\mathbb{Z}_2$ gauge defect types are confined, so that the σ^x and σ^z topological sectors are both protected. This case is briefly discussed in Appendix A, and it is related to error recovery that was argued to be realizable, for example, in a four-dimensional (4D) toric code.[6] What we argue here based on the finite-temperature studies is that the system can be *self-correcting*: if the system is prepared in a given superposition at zero temperature and its temperature is raised and again lowered to zero without ever going above T_c, the system returns to the same original quantum state (a "boomerang" effect).

The protection holds at low temperatures, but it is bound to vanquish as the density of defective plaquettes with eigenvalue -1 grows with temperature: once enough defects are in place, one can no longer deform paths around them. Therefore, we expect a loss of topological information as temperature is increased via a topological phase transition at finite temperature.

In analogy with 3D lattice gauge theory, we expect this transition to occur when plaquette defects deconfine at high enough temperature. This is captured by the expectation value of Wilson loop operators, which is exponentially suppressed with the length of the loop (perimeter law) at low temperatures, while it is suppressed with the area of the minimal enclosed surface (area law) at high temperatures.[16,17,19,20] In our notation, the transition temperature is set by the energy scale λ_B, and the transition is expected to occur at the critical point of the 3D lattice gauge theory.

The topological entropy is a nonlocal order parameter that detects the presence of topological order in a system. Any loss of topological information, e.g., whenever some topological sectors become ill-defined, should have a measurable effect on such entropy. Indeed, we show below that this is the case and that the qualitative picture inferred from the arguments above is confirmed by an exact calculation of the topological entropy at finite temperature.

A. Density matrix

Let us work for convenience in the σ^x (tensor product) basis, where the Hilbert space $\mathcal{H}$ is spanned by the whole set of orthonormal states $|\alpha\rangle$, labeled by the configurations α of a classical Ising model on the bonds of a 3D simple-cubic lattice (the value ± 1 of each Ising variable corresponds to the eigenvalue of the σ^x operator at the same site). Define G to be the group generated by all plaquette operators $B_p = \Pi_{i \in p} \sigma_i^z$. Recall that any two elements of the group differing by products of plaquettes around closed membranes are, in fact, the same element (i.e., they are defined modulo the identities $\Pi_{\text{closed membrane}} B_p = \mathbf{1}$), where we are assuming periodic boundary conditions, and full crystal planes are therefore closed membranes as well. Recall also that $|G| = 2^{2N-2}$, where N is the number of sites in the simple-cubic lattice. Every two elements of the group commute with each other and $g^2 = \mathbf{1}$, $\forall g \in G$. For later convenience, let us label with $\alpha = 0$ the fully magnetized state $\sigma^x = +1$.

The equilibrium properties of the system at finite temperature are captured by the density matrix,

$$\rho(T) = \frac{1}{Z} e^{-\beta \hat{H}} = \frac{\sum_{\alpha,\beta} \langle \alpha | e^{-\beta H} | \beta \rangle |\alpha\rangle\langle\beta|}{\sum_{\alpha} \langle \alpha | e^{-\beta H} | \alpha \rangle}. \tag{4.1}$$

For convenience of notation, let us rewrite Hamiltonian (2.2) as

$$H = -\lambda_B P - \lambda_A S,$$

$$P = \sum_p B_p,$$

$$S = \sum_s A_s.$$

Notice that $S|\alpha\rangle = M_s(\alpha)|\alpha\rangle$, where $M_s(\alpha)$ is the net "star magnetization," i.e., the difference between the number of stars with eigenvalue of $+1$ and with eigenvalue -1 in the state $|\alpha\rangle$. The action of any group element g is to flip plaquettes, which cannot change the sign of any star operator since they commute, and therefore $Sg|\alpha\rangle = M_s(\alpha) g|\alpha\rangle$, $\forall g \in G$.

Thus, the denominator of Eq. (4.1) becomes

$$\sum_{\alpha} \langle \alpha | e^{-\beta H} | \alpha \rangle = \sum_{\alpha} e^{\beta \lambda_A M_s(\alpha)} \langle \alpha | e^{\beta \lambda_B P} | \alpha \rangle. \tag{4.2}$$

Upon expanding

$$e^{\beta \lambda_B P} = \prod_p [\cosh \beta \lambda_B + \sinh \beta \lambda_B B_p], \tag{4.3}$$

as follows from the definition $P = \Sigma_p B_p$ and from the fact that $B_p^2 \equiv \mathbf{1}$, $\forall p$, one can explicitly compute the last term

$$\langle \alpha | e^{\beta \lambda_B P} | \alpha \rangle = \langle \alpha | \prod_p [\cosh \beta \lambda_B + \sinh \beta \lambda_B B_p] | \alpha \rangle. \tag{4.4}$$

All nonvanishing contributions in Eq. (4.4) come from products of plaquette operators that reduce to the identity (i.e., products around closed membranes). The above equation is therefore independent of α, which we set to the reference state 0 in the following.

The set of all possible closed membranes in a periodic 3D simple-cubic lattice is in one-to-two correspondence with all

possible configurations of an Ising model on the dual simple-cubic lattice (the membranes are, say, the antiferromagnetic domain boundaries), provided we allow for both periodic and antiperiodic boundary conditions in all three directions. In this language, the sum of all nonvanishing contributions can be written as

$$\langle 0|e^{\beta\lambda_B P}|0\rangle = \frac{1}{2}\sum_{\mathcal{C}} [\cosh \beta\lambda_B]^{3N-N_{\mathrm{AF}}(\mathcal{C})}[\sinh \beta\lambda_B]^{N_{\mathrm{AF}}(\mathcal{C})}, \quad (4.5)$$

where $\mathcal{C}$ is a generic configuration of the 3D Ising model with any type of boundary conditions, $3N$ is the total number of nearest-neighbor (NN) bonds, and $N_{\mathrm{AF}}(\mathcal{C})$ is the number of antiferromagnetic NN bonds. The factor 1/2 comes from the $\mathbb{Z}_2$ symmetry: a given membrane configuration corresponds to two equivalent but distinct Ising configurations. For convenience, let us introduce the simplified notation $c=\cosh \beta\lambda_B$, $s=\sinh \beta\lambda_B$, and $t=s/c=\tanh \beta\lambda_B$ and define $J>0$ such that $e^{-2J}=t$ (recall that $\lambda_B>0$). The above expression can then be further simplified to

$$\begin{aligned} 2\langle 0|e^{\beta\lambda_B P}|0\rangle &= c^{3N}\sum_{\mathcal{C}} t^{N_{\mathrm{AF}}(\mathcal{C})} \\ &= c^{3N}\sum_{\mathcal{C}} e^{-2JN_{\mathrm{AF}}(\mathcal{C})} = c^{3N}\sum_{\mathcal{C}} \exp\left(J\sum_{\langle i,j\rangle} S_iS_j - 3NJ\right) \\ &= (sc)^{3N/2}\sum_{\mathcal{C}} \exp\left(J\sum_{\langle i,j\rangle} S_iS_j\right) \equiv (sc)^{3N/2}Z_J^{\mathrm{tot}}, \quad (4.6) \end{aligned}$$

where Z_J^{tot} is the partition function of an Ising model on a simple-cubic lattice of size $N=L\times L\times L$ with reduced ferromagnetic coupling constant J, summed over all possible choices of (periodic or antiperiodic) boundary conditions.[21]

We can now move on to compute the numerator of Eq. (4.1),

$$\begin{aligned} \sum_{\alpha,\beta} \langle\alpha|e^{-\beta H}|\beta\rangle|\alpha\rangle\langle\beta| &= \sum_{\alpha,\beta} e^{\beta\lambda_A M_s(\alpha)}\langle\alpha|e^{\beta\lambda_B P}|\beta\rangle|\alpha\rangle\langle\beta| \\ &= \sum_{g\in G}\sum_{\alpha} e^{\beta\lambda_A M_s(\alpha)}\langle\alpha|e^{\beta\lambda_B P}g|\alpha\rangle|\alpha\rangle\langle\alpha|g, \end{aligned} \quad (4.7)$$

where we used the fact that all matrix elements $\langle\alpha|e^{\beta\lambda_B P}|\beta\rangle$ vanish identically unless $|\beta\rangle=g|\alpha\rangle$, $\exists g\in G$. Once again, the expectation value $\langle\alpha|e^{\beta\lambda_B P}g|\alpha\rangle$ is independent of α, and the above expression simplifies to

$$\sum_{g\in G}\sum_{\alpha} e^{\beta\lambda_A M_s(\alpha)}\langle 0|e^{\beta\lambda_B P}g|0\rangle|\alpha\rangle\langle\alpha|g. \quad (4.8)$$

The expectation value can be computed explicitly by expanding the exponential

$$\langle 0|e^{\beta\lambda_B P}g|0\rangle = \langle 0|\prod_p [\cosh \beta\lambda_B + \sinh \beta\lambda_B B_p]\prod_{p'\in g} B_{p'}|0\rangle. \quad (4.9)$$

Here, the notation $\Pi_{p'\in g}B_{p'}$ represents the decomposition of g in terms of the group generators $\{B_p\}$. Clearly this decomposition is highly nonunique since the group elements are defined modulo the identities $\Pi_{\text{closed membrane}}B_p=\mathbf{1}$, and Eq. (4.9) needs to be handled with care.

As before, all nonvanishing contributions come from products of plaquette operators that reduce to the identity. In this case, however, there are two options for every operator $B_{p'}$: (i) it can be multiplied out directly by $\sinh \beta\lambda_B B_p$, with $p=p'$ (recall that $B_{p'}^2=\mathbf{1}$), or (ii) it can be completed to an identity by an appropriate product of B_p terms so that $B_{p'}\Pi B_p$ forms a closed membrane. Notice that in the second case the product over p may not include p' itself.

All this can be expressed in more elegant terms in the Ising language defined previously. Case (i) corresponds to the two spins across the bond p' being ferromagnetically aligned in the Ising model and contributing a Boltzmann factor $\sinh \beta\lambda_B$. Case (ii) corresponds to the two spins across p' being antiferromagnetically aligned and contributing a Boltzmann factor $\cosh \beta\lambda_B$. Notice that the correlations between the different p' are automatically taken care of in the Ising language, and we obtain

$$\begin{aligned} 2\langle 0|e^{\beta\lambda_B P}g|0\rangle &= (sc)^{3N/2}\sum_{\mathcal{C}} \exp\left(J\sum_{\langle i,j\rangle} \eta_{ij}(g)S_iS_j\right) \\ &\equiv (sc)^{3N/2}Z_J^{\mathrm{tot}}(g), \end{aligned} \quad (4.10)$$

where

$$\eta_{ij}(g) = \begin{cases} +1 & \text{if } \langle i,j\rangle \notin g \\ -1 & \text{if } \langle i,j\rangle \in g. \end{cases} \quad (4.11)$$

Recall that a bond in the Ising model corresponds to a plaquette in the original system and $\langle i,j\rangle \in g$ means that the corresponding plaquette operator appears in the decomposition of g.

In order to derive Eq. (4.10), let us define $N_{\mathrm{F}}(\mathcal{C}|g)$ ($N_{\mathrm{AF}}(\mathcal{C}|g)$) to be the number of bonds with ferromagnetically (antiferromagnetically) aligned spins in the subset of bonds corresponding to g of a given Ising configuration $\mathcal{C}$. Define as well $N_{\mathrm{F}}(\mathcal{C}|\bar{g})$ ($N_{\mathrm{AF}}(\mathcal{C}|\bar{g})$) to be the number of bonds with ferromagnetic (antiferromagnetic) spin alignment within bonds in the subset complementary to g. Clearly, $N_{\mathrm{F/AF}}(\mathcal{C}) = N_{\mathrm{F/AF}}(\mathcal{C}|g)+N_{\mathrm{F/AF}}(\mathcal{C}|\bar{g})$.

We can then rewrite Eq. (4.9) in the Ising language as

$$\begin{aligned} 2\langle 0|e^{\beta\lambda_B P}g|0\rangle &= \sum_{\mathcal{C}} c^{N_{\mathrm{F}}(\mathcal{C}|\bar{g})}s^{N_{\mathrm{AF}}(\mathcal{C}|\bar{g})}s^{N_{\mathrm{F}}(\mathcal{C}|g)}c^{N_{\mathrm{AF}}(\mathcal{C}|g)} \\ &= \sum_{\mathcal{C}} c^{N_{\mathrm{F}}(\mathcal{C})}s^{N_{\mathrm{AF}}(\mathcal{C})}t^{N_{\mathrm{F}}(\mathcal{C}|g)}t^{-N_{\mathrm{AF}}(\mathcal{C}|g)} \\ &= c^{3N}\sum_{\mathcal{C}} e^{-2J[N_{\mathrm{AF}}(\mathcal{C})+N_{\mathrm{F}}(\mathcal{C}|g)-N_{\mathrm{AF}}(\mathcal{C}|g)]} \\ &= c^{3N}\sum_{\mathcal{C}} \exp\left[J\left(\sum_{\langle i,j\rangle} S_iS_j - 3N - 2\sum_{\langle i,j\rangle\in g} S_iS_j\right)\right] \\ &= (sc)^{3N/2}\sum_{\mathcal{C}} \exp\left(J\sum_{\langle i,j\rangle} \eta_{ij}(g)S_iS_j\right). \end{aligned}$$

In the following, it is convenient to introduce the convention that a bond $\langle ij\rangle$ belongs to or is inside a partition $\mathcal{A}$ of the system ($\langle ij\rangle \in \mathcal{A}$) if all the spins on the corresponding plaquette operator belong to $\mathcal{A}$, and the bond does not belong or is outside $\mathcal{A}$ ($\langle ij\rangle \notin \mathcal{A}$) otherwise. (Similarly, we will refer to a cubic unit cell in or not in $\mathcal{A}$ if its six composing plaquettes are all in $\mathcal{A}$ or not.)

In conclusion, the numerator of Eq. (4.1) can be mapped onto the partition function of a 3D random-bond Ising model on a simple-cubic lattice, where the randomness is controlled by the choice of g. Again, summation over all possible boundary conditions is understood.

Substituting Eqs. (4.6) and (4.10) into Eq. (4.1) gives

$$\rho(T) = \sum_{g\in G} \frac{Z_J^{\mathrm{tot}}(g)}{Z_J^{\mathrm{tot}}(\mathbf{1})} \sum_{\alpha} \frac{e^{\beta\lambda_A M_s(\alpha)}}{Z_s} |\alpha\rangle\langle\alpha| g, \tag{4.12}$$

where $J=-(1/2)\ln[\tanh(\beta\lambda_B)]$, $Z_s=\Sigma_\alpha e^{\beta\lambda_A M_s(\alpha)}$ is the partition function of a noninteracting Ising system in a magnetic field of reduced strength $\beta\lambda_A$, and $Z_J^{\mathrm{tot}}(\mathbf{1}) \equiv Z_J^{\mathrm{tot}}$.

In the limit of $T\to 0$ ($\beta\to\infty$), $J\to 0^+$, all g are equally weighed,

$$Z_0^{\mathrm{tot}}(g) = Z_0^{\mathrm{tot}}(\mathbf{1}) \quad \forall\, g \in G, \tag{4.13}$$

and only the states with maximal star magnetization $M_s(\alpha)=N$, i.e., those that are eigenstates of the star operators with eigenvalue +1 everywhere, survive,

$$\frac{e^{\beta\lambda_A M_s(\alpha)}}{Z_s} \to \frac{1}{2^3|G|}\delta_{M_s(\alpha),N}. \tag{4.14}$$

Such states are of the form $g|0_k\rangle$, where $k=1,\ldots,2^3$ labels the states obtained from $|0\rangle$ by the action of the nonlocal Γ operators in Eq. (2.3). Namely, the states $|0_k\rangle$ are of the form $\Gamma_1^{m_1}\Gamma_2^{m_2}\Gamma_3^{m_3}|0\rangle$ for all possible choices of $m_1,m_2,m_3=0,1$. The factor $1/2^3|G|$ in the above equation appears because there are precisely $2^3|G|$ states with maximal star magnetization. Thus, one recovers the density matrix of the zero-temperature Kitaev model, prepared with equal probability across all topological sectors,[14]

$$\rho(T=0) = \frac{1}{2^3}\sum_{k=1}^{2^3} \frac{1}{|G|} \sum_{g,g'\in G} g|0_k\rangle\langle 0_k|gg'. \tag{4.15}$$

In the limit $T\to\infty$ ($\beta\to 0$), $J\to\infty$, all g are exponentially suppressed except for $g=\mathbf{1}$, while all states α become equally weighed. In this case one obtains the mixed state density matrix,

$$\rho(T\to\infty) = \frac{1}{2^{3N}}\sum_{\alpha} |\alpha\rangle\langle\alpha|, \tag{4.16}$$

of a noninteracting Ising model defined on the bonds of a simple-cubic lattice.

Clearly from Eq. (4.12), one expects something to happen in the system when the value of the temperature T, i.e., the parameter J, is such that the 3D Ising model described by Z_J^{tot} becomes critical. In order to understand how this relates to the presence of topological order at zero temperature, we need to proceed with the calculations and compute the von Neumann entropy and the topological entropy as a function of temperature.

B. von Neumann entropy

Let us consider a generic bipartition of the original system $\mathcal{S}$ into subsystems $\mathcal{A}$ and $\mathcal{B}$ ($\mathcal{S}=\mathcal{A}\cup\mathcal{B}$). The von Neumann (entanglement) entropy of partition $\mathcal{A}$ is given by

$$S_{\mathrm{vN}}^{\mathcal{A}} \equiv -\mathrm{Tr}[\rho_{\mathcal{A}} \ln \rho_{\mathcal{A}}] = -\lim_{n\to 1} \partial_n \mathrm{Tr}[\rho_{\mathcal{A}}^n], \tag{4.17}$$

where $\rho_{\mathcal{A}}=\mathrm{Tr}_{\mathcal{B}}\,\rho$ is the reduced density matrix obtained from the full density matrix ρ by tracing out the degrees of freedom in subsystem $\mathcal{B}$; similarly for $S_{\mathrm{vN}}^{\mathcal{B}}$. $S_{\mathrm{vN}}^{\mathcal{A}}=S_{\mathrm{vN}}^{\mathcal{B}}$ holds if ρ is a pure state density matrix.

In order to compute the von Neumann entropy (4.17) from the finite-temperature density matrix (4.12), we first obtain the reduced density matrix of the system using an approach similar to the one in Ref. 14,

$$\begin{aligned}\rho_{\mathcal{A}}(T) &= \sum_{g\in G} \frac{Z_J^{\mathrm{tot}}(g)}{Z_J^{\mathrm{tot}}(\mathbf{1})} \sum_{\alpha} \frac{e^{\beta\lambda_A M_s(\alpha)}}{Z_s} |\alpha_{\mathcal{A}}\rangle\langle\alpha_{\mathcal{A}}| g_{\mathcal{A}} \langle\alpha_{\mathcal{B}}| g_{\mathcal{B}} |\alpha_{\mathcal{B}}\rangle \\ &= \sum_{g\in G_{\mathcal{A}}} \frac{Z_J^{\mathrm{tot}}(g)}{Z_J^{\mathrm{tot}}(\mathbf{1})} \sum_{\alpha} \frac{e^{\beta\lambda_A M_s(\alpha)}}{Z_s} |\alpha_{\mathcal{A}}\rangle\langle\alpha_{\mathcal{A}}| g_{\mathcal{A}},\end{aligned} \tag{4.18}$$

where we used the generic tensor decomposition $|\alpha\rangle=|\alpha_{\mathcal{A}}\rangle\otimes|\alpha_{\mathcal{B}}\rangle$, $g=g_{\mathcal{A}}\otimes g_{\mathcal{B}}$, and the fact that $\langle\alpha_{\mathcal{B}}|g_{\mathcal{B}}|\alpha_{\mathcal{B}}\rangle=1$ if $g_{\mathcal{B}}=\mathbf{1}_{\mathcal{B}}$ and 0 otherwise. As in Sec. III, we denoted by $G_{\mathcal{A}}=\{g\in G | g_{\mathcal{B}}=\mathbf{1}_{\mathcal{B}}\}$ the subgroup of G given by all operations g that act trivially on $\mathcal{B}$ (similarly for $G_{\mathcal{B}}$).

Notice that a plaquette operator B_p can either act solely on spins in partition $\mathcal{A}$ (represented in the following by the notation $p\in\mathcal{A}$), solely on spins in partition $\mathcal{B}$ ($p\in\mathcal{B}$), or simultaneously on spins belonging to $\mathcal{A}$ and $\mathcal{B}$ (which we will refer to as *boundary plaquette operators* and represent by $p\in\mathcal{AB}$). Recall from Sec. III that a complete set of generators for the subgroup $G_{\mathcal{A}}$ can be constructed by taking: (i) all plaquette operators that act solely on $\mathcal{A}$, i.e., $\{B_p | p\in\mathcal{A}\}$ ($\mathcal{N}_{\mathcal{A}}^{(p)}=|\{B_p|p\in\mathcal{A}\}|$); (ii) all possible (independent) collective operators constructed from plaquettes in $\mathcal{B}$ and at the boundary but acting solely on $\mathcal{A}$ (as illustrated in Sec. III, the number of such collective operators equals the number $n_{\mathcal{A}}$ of noncontractible loops in subsystem $\mathcal{A}$); and by (iii) accounting for all constraints given by the independent closed membranes in $\mathcal{A}$. That is, all $\mathcal{N}_{\mathcal{A}}^{(c)}$ cubic unit cells in $\mathcal{A}$, all possible $(m_{\mathcal{B}}-1)$ additional closed membranes if $\mathcal{B}$ is disconnected, and all independent entire crystal planes inside $\mathcal{A}$ ($m_{\mathcal{A}}^{(\mathrm{CP})}=0,1,2,3$). Again, for all bipartitions of interest in our study $m_{\mathcal{A}}^{(\mathrm{CP})}=0$ and $m_{\mathcal{B}}^{(\mathrm{CP})}=3$, and for simplicity we will restrict to this specific case.

The cardinalities of the subgroups $G_{\mathcal{A}}$ and $G_{\mathcal{B}}$ are thus given by

$$d_{\mathcal{A}} \equiv |G_{\mathcal{A}}| = 2^{\mathcal{N}_{\mathcal{A}}^{(p)}-\mathcal{N}_{\mathcal{A}}^{(c)}+n_{\mathcal{A}}-(m_{\mathcal{B}}-1)}, \tag{4.19a}$$

$$d_{\mathcal{B}} \equiv |G_{\mathcal{B}}| = 2^{\mathcal{N}_{\mathcal{B}}^{(p)}-\mathcal{N}_{\mathcal{B}}^{(c)}+n_{\mathcal{B}}-(m_{\mathcal{A}}-1)-3}. \tag{4.19b}$$

In particular, $n_{\mathcal{A}}=n_{\mathcal{B}}=1$ in bipartitions 4,5 and 0 otherwise, and $m_{\mathcal{A}}=2$ in bipartition 8, $m_{\mathcal{B}}=2$ in bipartition 1, and they equal 1 in all other cases.

Let us then use Eq. (4.18) to compute the trace of the nth power of $\rho_{\mathcal{A}}(T)$,

$$\mathrm{Tr}[\rho_{\mathcal{A}}^{n}(T)] = \sum_{g_1,\ldots,g_n\in G_{\mathcal{A}}} \left(\prod_{l=1}^{n} \frac{Z_J^{\mathrm{tot}}(g_l)}{Z_J^{\mathrm{tot}}(\mathbf{1})}\right) \sum_{\alpha_1,\ldots,\alpha_n} \left(\prod_{l=1}^{n} \frac{e^{\beta\lambda_A M_s(\alpha_l)}}{Z_s}\right) \langle\alpha_{1,\mathcal{A}}|g_{1,\mathcal{A}}|\alpha_{2,\mathcal{A}}\rangle\langle\alpha_{2,\mathcal{A}}|g_{2,\mathcal{A}}|\alpha_{3,\mathcal{A}}\rangle\cdots\langle\alpha_{n,\mathcal{A}}|g_{n,\mathcal{A}}|\alpha_{1,\mathcal{A}}\rangle. \tag{4.20}$$

Each expectation value above imposes that the two configurations α_{l+1} and α_l, $l=1,\ldots,n$ (with the identification $n+1\equiv 1$), can be mapped one onto the other over subsystem $\mathcal{A}$ via the plaquette flipping operation $g_{l,\mathcal{A}}$. This is possible only if the set $g_1,\ldots,g_n\in G_{\mathcal{A}}$ satisfies the condition $\Pi_{l=1}^{n} g_{l,\mathcal{A}}=\mathbf{1}_{\mathcal{A}}$, i.e., $\Pi_{l=1}^{n} g_l=\mathbf{1}$. Therefore, we can decompose each element g_l into a product $g_l=\tilde{g}_l\tilde{g}_{l+1}$, where $\tilde{g}_l\in G_{\mathcal{A}}$, $l=1,\ldots,n$ with periodic boundary conditions $n+1\equiv 1$ (the fact that this decomposition is highly nonunique is immaterial to the calculations below),

$$\begin{aligned}\mathrm{Tr}[\rho_{\mathcal{A}}^{n}(T)] &= \sum_{g_1,\ldots,g_n\in G_{\mathcal{A}}} \left(\prod_{l=1}^{n} \frac{Z_J^{\mathrm{tot}}(g_l)}{Z_J^{\mathrm{tot}}(\mathbf{1})}\right) \sum_{\alpha_1,\ldots,\alpha_n} \left(\prod_{l=1}^{n} \frac{e^{\beta\lambda_A M_s(\alpha_l)}}{Z_s}\right) \\ &\times \langle 0|\prod_{l=1}^{n} g_l|0\rangle\langle\alpha_{1,\mathcal{A}}|\tilde{g}_{1,\mathcal{A}}\tilde{g}_{2,\mathcal{A}}|\alpha_{2,\mathcal{A}}\rangle\langle\alpha_{2,\mathcal{A}}|\tilde{g}_{2,\mathcal{A}}\tilde{g}_{3,\mathcal{A}}|\alpha_{3,\mathcal{A}}\rangle\cdots\langle\alpha_{n,\mathcal{A}}|\tilde{g}_{n,\mathcal{A}}\tilde{g}_{1,\mathcal{A}}|\alpha_{1,\mathcal{A}}\rangle \\ &= \sum_{g_1,\ldots,g_n\in G_{\mathcal{A}}} \left(\prod_{l=1}^{n} \frac{Z_J^{\mathrm{tot}}(g_l)}{Z_J^{\mathrm{tot}}(\mathbf{1})}\right) \sum_{\alpha_1,\ldots,\alpha_n} \left(\prod_{l=1}^{n} \frac{e^{\beta\lambda_A M_s(\alpha_l)}}{Z_s}\right) \langle 0|\prod_{l=1}^{n} g_l|0\rangle\langle\alpha_{1,\mathcal{A}}|\alpha_{2,\mathcal{A}}\rangle\langle\alpha_{2,\mathcal{A}}|\alpha_{3,\mathcal{A}}\rangle\cdots\langle\alpha_{n,\mathcal{A}}|\alpha_{1,\mathcal{A}}\rangle, \end{aligned} \tag{4.21}$$

where we used the fact that the magnetization $M_s(\alpha)$ of state $|\alpha\rangle$ is the same as $M_s(g\alpha)$ of state $g|\alpha\rangle$, for any $g\in G$, to do away with the $\tilde{g}_l$ via relabeling of the states $|\alpha_l\rangle\to\tilde{g}_l|\alpha_l\rangle$.

We can further simplify the notation by defining the function $\delta(\alpha_{\mathcal{A}},\beta_{\mathcal{A}})=\langle\alpha_{\mathcal{A}}|\beta_{\mathcal{A}}\rangle$, and the above equation can be rewritten as

$$\mathrm{Tr}[\rho_{\mathcal{A}}^{n}(T)] = \sum_{g_1,\ldots,g_n\in G_{\mathcal{A}}} \left(\prod_{l=1}^{n} \frac{Z_J^{\mathrm{tot}}(g_l)}{Z_J^{\mathrm{tot}}(\mathbf{1})}\right) \langle 0|\prod_{l=1}^{n} g_l|0\rangle \times \sum_{\alpha_1,\ldots,\alpha_n} \left(\prod_{l=1}^{n} \frac{e^{\beta\lambda_A M_s(\alpha_l)}}{Z_s}\right) \prod_{l=1}^{n-1} \delta(\alpha_{l,\mathcal{A}},\alpha_{l+1,\mathcal{A}}) = \mathcal{Z}^{(P)}(n)\times\mathcal{Z}^{(S)}(n). \tag{4.22}$$

Notice that the product $\Pi_{l=1}^{n-1}\delta(\alpha_{l,\mathcal{A}},\alpha_{l+1,\mathcal{A}})$ implies $\delta(\alpha_{1,\mathcal{A}},\alpha_{n,\mathcal{A}})$, which is therefore redundant and has been omitted in the previous equation. In the notation of Eq. (4.22), it becomes evident that the *star* (S) contribution, i.e., involving only the star coupling constant λ_A, and the *plaquette* (P) contribution, i.e., involving only the plaquette coupling constant λ_B, decouple and factorize into two separate terms, $\mathcal{Z}^{(S)}(n)$ and $\mathcal{Z}^{(P)}(n)$. In particular, $\mathcal{Z}^{(P)}(n=1)=\mathcal{Z}^{(S)}(n=1)=1$.

Using the replica trick, we can compute the von Neumann entropy,

$$\begin{aligned} S_{\mathrm{vN}}(\mathcal{A};T) &= -\lim_{n\to 1}\partial_n \mathrm{Tr}[\rho_{\mathcal{A}}^{n}] \\ &= -\lim_{n\to 1}\partial_n[\mathcal{Z}^{(P)}(n)\mathcal{Z}^{(S)}(n)] \\ &= -\mathcal{Z}^{(S)}(1)\lim_{n\to 1}\partial_n\mathcal{Z}^{(P)}(n) - \mathcal{Z}^{(P)}(1)\lim_{n\to 1}\partial_n\mathcal{Z}^{(S)}(n) \\ &= -\lim_{n\to 1}\partial_n\mathcal{Z}^{(P)}(n) - \lim_{n\to 1}\partial_n\mathcal{Z}^{(S)}(n) \\ &= S_{\mathrm{vN}}^{(P)}(\mathcal{A};T/\lambda_B) + S_{\mathrm{vN}}^{(S)}(\mathcal{A};T/\lambda_A). \end{aligned} \tag{4.23}$$

Thus, from the factorizability in Eq. (4.22) above, it follows that the von Neumann entropy has two additive contributions from the star and plaquette terms that can then be computed separately.[9]

One can check that Eqs. (4.22) and (4.23) satisfy indeed the $T\to 0$ limit discussed in Sec. III, as well as the known $T\to\infty$ limit (see Appendix B). Notice that although in this paper we are concerned with 3D systems, the derivation is independent of the dimensionality, and this result holds true for $\mathbb{Z}_2$ models in *any number of dimensions*.

Because the von Neumann entropy is separable as the sum of the two independent contributions from star and plaquette terms, so is the topological entropy, which is a linear combination of the entanglement entropies for the partitions shown in Fig. 4,

$$S_{\mathrm{topo}}(T) = S_{\mathrm{topo}}^{(S)}(T/\lambda_A) + S_{\mathrm{topo}}^{(P)}(T/\lambda_B). \tag{4.24}$$

We now turn to the separate analysis of the two contributions.

1. Star contribution $S_{\mathrm{topo}}^{(S)}(T/\lambda_A)$

The computation of this contribution is very similar to the one in Ref. 4 for the 2D Kitaev model, where the limit $\lambda_B\to\infty$ was explicitly considered. In order to illustrate this analogy, let us define the following entropy differentials:

$$\begin{aligned}\Delta S_{\mathrm{vN}}(\mathcal{A};T) &\equiv S_{\mathrm{vN}}(\mathcal{A};T) - S_{\mathrm{vN}}(\mathcal{A};0) \\ &= \Delta S_{\mathrm{vN}}^{(S)}(\mathcal{A};T/\lambda_A) + \Delta S_{\mathrm{vN}}^{(P)}(\mathcal{A};T/\lambda_B) \end{aligned} \tag{4.25}$$

and

$$\Delta S_{\text{topo}}(T) \equiv S_{\text{topo}}(T) - S_{\text{topo}}(0) = \Delta S^{(S)}_{\text{topo}}(T/\lambda_A) + \Delta S^{(P)}_{\text{topo}}(T/\lambda_B), \quad (4.26)$$

where

$$\Delta S^{(S)}_{\text{vN}}(\mathcal{A};T/\lambda_A) \equiv S^{(S)}_{\text{vN}}(\mathcal{A};T/\lambda_A) - S^{(S)}_{\text{vN}}(\mathcal{A};0), \quad (4.27a)$$

$$\Delta S^{(S)}_{\text{topo}}(T/\lambda_A) \equiv S^{(S)}_{\text{topo}}(T/\lambda_A) - S^{(S)}_{\text{topo}}(0) \quad (4.27b)$$

and

$$\Delta S^{(P)}_{\text{vN}}(\mathcal{A};T/\lambda_B) \equiv S^{(P)}_{\text{vN}}(\mathcal{A};T/\lambda_B) - S^{(P)}_{\text{vN}}(\mathcal{A};0), \quad (4.27c)$$

$$\Delta S^{(P)}_{\text{topo}}(T/\lambda_B) \equiv S^{(P)}_{\text{topo}}(T/\lambda_B) - S^{(P)}_{\text{topo}}(0). \quad (4.27d)$$

Notice that for $\lambda_B \to \infty$, $\Delta S^{(P)}_{\text{vN}}(\mathcal{A};T/\lambda_B)=0$ and $\Delta S^{(P)}_{\text{topo}}(T/\lambda_B)=0$. Thus, one obtains that

$$\Delta S^{(S)}_{\text{vN}}(\mathcal{A};T/\lambda_A) = \Delta S_{\text{vN}}(\mathcal{A};T)|_{\lambda_B\to\infty}, \quad (4.28)$$

$$\Delta S^{(S)}_{\text{topo}}(T/\lambda_A) = \Delta S_{\text{topo}}(T)|_{\lambda_B\to\infty}. \quad (4.29)$$

Moreover, in the limit $\lambda_B \to \infty$ and choosing to work in the σ^z basis, one can show that both the group structure of G and the collective operations in $G_{\mathcal{A}}$ are very much the same in two dimensions and in three dimensions. For example, the group G is generated by all but one star operator, and the subgroup $G_{\mathcal{A}}$ is generated by all star operators in $\mathcal{A}$ with the addition of all but one collective operation that are obtained as products of star operators belonging to each component of $\mathcal{B}$ times the ones along the corresponding boundary. As a result, the topologically nontrivial bipartitions 1 and 4 in two dimensions correspond to bipartitions 1 and 8 in three dimensions. All calculations generalize straightforwardly to three dimensions, and one can derive the expressions for $\Delta S^{(S)}_{\text{vN}}$ and for $\Delta S^{(S)}_{\text{topo}}$ in a finite system at finite temperature. The actual values for $S^{(S)}_{\text{vN}}$ and $S^{(S)}_{\text{topo}}$ are then fixed by matching, say, the known $T\to 0$ limits.

From the 2D results in Ref. 4, we infer that the star contribution to the 3D topological entropy is fragile in the sense that it vanishes in the thermodynamic limit at any finite temperature. Namely, the behavior is singular in that the limits of $T\to 0$ and infinite size do not commute. If the thermodynamic limit is taken first,

$$\Delta S^{(S)}_{\text{topo}}(T/\lambda_A) = \begin{cases} 0, & T=0 \\ -\ln 2, & T>0. \end{cases} \quad (4.30)$$

Thus, in the thermodynamic limit, the star contribution to the topological entropy evaporates at any infinitesimal temperature. (The finite-temperature and finite-size expressions for the star contributions to the von Neumann and topological entropies are shown in Appendix C.)

2. Plaquette contribution $S^{(P)}_{\text{vN}}(\mathcal{A};T/\lambda_B)$

Similarly to the above, one obtains for the plaquette contribution,

$$\Delta S^{(P)}_{\text{vN}}(\mathcal{A};T/\lambda_B) = \Delta S_{\text{vN}}(\mathcal{A};T)|_{\lambda_A\to\infty}, \quad (4.31)$$

$$\Delta S^{(P)}_{\text{topo}}(T/\lambda_B) = \Delta S_{\text{topo}}(T)|_{\lambda_A\to\infty}. \quad (4.32)$$

Because of the very different nature of the 2D and 3D group structures when using the σ^x basis, the computation of the plaquette contribution in three dimensions is not a trivial extension of that in two dimensions, and it thus requires some work. The calculations are shown in detail in Appendix D, while only the results are summarized here for conciseness and clarity. The behavior of $\Delta S^{(P)}_{\text{topo}}(T/\lambda_B)$ as a function of temperature, in the thermodynamic limit, is

$$\Delta S^{(P)}_{\text{topo}}(T/\lambda_B) = \begin{cases} 0, & T<T_c \\ -\ln 2, & T>T_c, \end{cases} \quad (4.33)$$

where the critical temperature is associated with a 3D Ising transition and can be located at $T_c = 1.313\ 346(3)\lambda_B$.

V. DISCUSSION

We can now put all the pieces together and argue for the persistence of topological order at finite temperatures in the 3D Kitaev model. Adding the contributions from stars and plaquettes, which we have shown to be exactly separable, the topological entropy of the system is

$$S^{\text{3D}}_{\text{topo}}(T) = \begin{cases} 2\ln 2, & T=0 \\ \ln 2, & 0<T<T_c \\ 0, & T>T_c. \end{cases} \quad (5.1)$$

This is to be contrasted to the 2D case,[4]

$$S^{\text{2D}}_{\text{topo}}(T) = \begin{cases} 2\ln 2, & T=0 \\ 0, & T>0, \end{cases} \quad (5.2)$$

where the topological order is fragile, subsiding for any finite T (when the thermodynamic limit is taken first).

In three dimensions the order survives up to a transition temperature that is determined by the coupling constant λ_B associated with the plaquette degrees of freedom alone. The topological order in the system, as measured by the topological entropy, is thus the same as in the case where $\lambda_A=0$, that is, in a purely classical model. In this sense, the order at finite T is classical in origin.[22]

Our results show that the extension of the notion of topological order to classical systems applies beyond the hard constrained limit already discussed in Ref. 12 in two dimensions. In the 3D example discussed here, the order persists for noninfinite couplings λ_A and λ_B. Having obtained the result that topological order in the 3D toric code survives thermal fluctuations, in a classical sense, up to a finite critical temperature, we now turn to a discussion of what this type of order implies.

At zero temperature, topological sectors can be discerned according to the eigenvalues $I_\alpha = \pm 1$ of the loop operators Γ_α, where $\alpha=1,2,3$, as in Eq. (2.3). The eight ground states $|I\rangle$ in the different topological sectors can be labeled by integers $I=0,\ldots,2^3-1$ (made up of three bits, $I \equiv I_1 I_2 I_3$, $I_\alpha = 0,1$).

Suppose to prepare, at an initial time $t=t_i$, a superposition of states,

$$|\Psi(t_i)\rangle = \sum_{I=0}^{2^3-1} \sqrt{p_I}|I\rangle, \tag{5.3}$$

then raise the temperature to some value $0<T<T_c$, and bring it back to $T=0$ at some time t_f. The final $T=0$ state will again be, assuming thermodynamic equilibrium is reached, a superposition of the eight topologically degenerate ground states.

Following the discussion in Sec. IV, for temperatures below T_c, one can take a winding loop and deform it, past thermal defects, and read off the same eigenvalue of the topological operator as the path is deformed. The information stored in all winding loops that *do not* cross a thermal defect does not disappear as long as there is a way to pass a winding loop that avoids defects. Therefore, as long as the system temperature is not raised above T_c, upon returning to $T=0$ at t_f, the system should return to the same topological sector that it was originally prepared in at time t_i.

Thus, the state at t_f is a superposition,

$$|\Psi(t_f)\rangle = \sum_{I=0}^{2^3-1} \sqrt{p_I}e^{i\varphi_I}|I\rangle, \tag{5.4}$$

where phases φ_I are accumulated during the thermal cycle. These phases, unless locked together by some specific mechanism, shall be randomized by the thermal bath. However, the magnitude of the amplitudes remains $\sqrt{p_I}$, for $I=0,\ldots,2^3-1$, as there have been no transitions between different topological sectors if the system was never heated above T_c.

Hence, the only (accessible) information preserved under the time evolution from t_i to t_f is that the relative probability to find the state in sector I equals p_I. The state in Eq. (5.4) realizes a pbit or probabilistic bit.[11] It is not a qubit because of the thermal dephasing between the states $|I\rangle$. Although still a quantum superposition of a sort, in that it has probability p_I of being in sector I, it cannot be told apart by any type of measurement from a classical probabilistic system with the same probabilities p_I. The stability of the system against local measurements only tells us that the state is not projected onto a sector *until* a nonlocal measurement is carried out. This effect is a *nonmeasurable* difference between the state in Eq. (5.3) and a classical probabilistic state: whether the projection occurs *before* (as in the classical state) or *after* (as in the pbit) the measurement is not detectable.

VI. CONCLUSIONS

In this paper we have shown that topological order exists in the 3D toric code at finite temperatures up to a critical temperature $T_c=1.313\,346(3)\lambda_B$ which is set by one of the couplings (that associated to the plaquette terms in the Hamiltonian). This is in sharp contrast to what happens in the 2D toric code, where in the thermodynamic limit the order subsides for any infinitesimal temperature.

We first presented simple heuristic arguments for this result. These arguments are based on the observation that eigenvalues of operators defined as products of spin operators along winding loops can be used to determine the order even in the presence of (thermally activated) local defects because loops can be deformed around such obstacles in three dimensions, leaving unchanged the eigenvalues of such loop operators. This is to be contrasted to the 2D case, where one cannot move a loop around a point, and thus the eigenvalues of nonlocal loop operators are unequal on opposite sides of the point defect.

We subsequently substantiated the heuristic arguments by means of an exact calculation of the von Neumann and topological entropies in the system as a function of temperature. In carrying out this exact calculation, we derived a generic result that applies to toric codes defined in any number of spatial dimensions: that the von Neumann entropy is separable as a sum of two terms, one associated with stars alone (and a function of the dimensionless ratio T/λ_A) and another associated with plaquettes alone (and a function of the dimensionless ratio T/λ_B). The same separability follows naturally for the topological entropy, $S_{\text{topo}}(T)=S^{(S)}_{\text{topo}}(T/\lambda_A)+S^{(P)}_{\text{topo}}(T/\lambda_B)$. We then showed that, in the thermodynamic limit, the star contribution $S^{(S)}_{\text{topo}}(T/\lambda_A)$ vanishes for any $T\neq 0$, while the plaquette contribution $S^{(P)}_{\text{topo}}(T/\lambda_B)$ remains constant for $T/\lambda_B<1.313\,346(3)$ and vanishes for temperatures above this scale.

Because the critical temperature is set by λ_B and not λ_A, one can argue that the topological entropy remains nonzero when $\lambda_A\to 0$. The resulting Hamiltonian is purely classical, and thus one can argue that the nature of the finite T topological order must be classical as well.

Finally, we discussed the nature of the information that can be stored robustly in the system because of the topological order at finite T. We argued that the resilient information stored in the 3D system realizes a pbit.

We end with a note on an interesting situation that should occur in systems where both $\mathbb{Z}_2$ gauge defect types are confined. In three dimensions only one of the defect types is confined, the topological entropy drops from 2 ln 2 at $T=0$ to ln 2 for $0<T<T_c$, and only the probabilities of being in a given topological sector are preserved (magnitude square of the amplitudes but not the relative phases). If instead both defect types are confined, the notion of sectors in both the σ^x and σ^z bases is retained, and this implies (as discussed briefly in Appendix A) that, if the system is prepared in a given superposition at zero temperature and its temperature is raised and again lowered to zero without ever going above T_c, the system returns to the same original quantum state (a boomerang effect).

ACKNOWLEDGMENTS

We are indebted to Xiao-Gang Wen for attracting our attention toward the possibility of a finite-temperature topological phase transition in the three-dimensional Kitaev model and to Michael Levin, John Cardy, Eduardo Fradkin, and Roderich Moessner for several insightful discussions. This work was supported in part by EPSRC-GB under Grant No. GR/R83712/01 (C.Castelnovo).

APPENDIX A: THE CONFINED-CONFINED CASE

In this appendix, we briefly discuss how the nature of the topological protection at finite temperature changes when

both types of thermal defects in a $\mathbb{Z}_2$ gauge theory are confined at low temperature ($T<T_c$). For concreteness and simplicity, let us consider a modification of the 2D toric code, where some *ad hoc* energy terms have been introduced that confine both electric and magnetic thermal defects (without inquiring on the nature of these terms). (As mentioned in Sec. IV, this scenario should be realized in the 4D case without need of any additional term.)

The $T=0$ ground-state (GS) wave function in a given topological sector is uniquely specified by the ($\pm$) eigenvalues of two independent Wilson toric cycles, i.e., winding loop operators. In the σ^z basis, it is sufficient to consider the product of all $\hat{\sigma}_i^z$ operators along a horizontal ($\hat{\mathcal{T}}_h^z$) and a vertical ($\hat{\mathcal{T}}_v^z$) winding loop, respectively; similarly, in the σ^x basis, using loop operators in the dual lattice, $\hat{\mathcal{T}}_h^x$ and $\hat{\mathcal{T}}_v^x$. These loop operators satisfy the algebra $\{\hat{\mathcal{T}}_h^x,\hat{\mathcal{T}}_v^z\}=0$ and $\{\hat{\mathcal{T}}_v^x,\hat{\mathcal{T}}_h^z\}=0$.

Let us choose to work in the σ^z basis and define $|a,b\rangle$, $a=\pm$, to be the normalized GS wave functions that are also eigenvectors of $\hat{\mathcal{T}}_h^z$ and $\hat{\mathcal{T}}_v^z$,

$$\hat{\mathcal{T}}_h^z|a,b\rangle = a|a,b\rangle,$$

$$\hat{\mathcal{T}}_v^z|a,b\rangle = b|a,b\rangle.$$

Let us prepare the system in a given superposition of such basis states,

$$|\Psi_{\text{in}}\rangle = \sum_{a,b=\pm} \psi_{a,b}|a,b\rangle, \tag{A1}$$

where $\Sigma_{a,b=\pm}|\psi_{a,b}|^2=1$, and consider coupling the system to a thermal bath so that the temperature can be varied from $T_{\text{in}}=0$, via $0<T<T_c$, back to $T_{\text{fi}}=0$, as discussed in Sec. IV.

Trivially, the final state of the system must again be a ground state, and therefore it can be written as

$$|\Psi_{\text{fi}}\rangle = \sum_{a,b=\pm} \tilde{\psi}_{a,b}|a,b\rangle. \tag{A2}$$

Moreover, as long as the temperature was never raise beyond the deconfining transition at T_c, the coupling to the thermal bath cannot have transferred any amplitude between any of the topological sectors. Hence the following topological quantities must be conserved,

$$\langle\Psi_{\text{in}}|\hat{\mathcal{T}}_{h/v}^{z/x}|\Psi_{\text{in}}\rangle = \langle\Psi_{\text{fi}}|\hat{\mathcal{T}}_{h/v}^{z/x}|\Psi_{\text{fi}}\rangle. \tag{A3}$$

For simplicity, consider the case where

$$\psi_{+,+} = \cos(\theta/2),$$

$$\psi_{-,+} = \sin(\theta/2)e^{i\phi},$$

where $\theta\in(0,\pi)$ and $\phi\in(-\pi,\pi)$, and all others vanish. After a little algebra, one can show that the conditions in Eq. (A3) require that the only nonvanishing terms in the final GS wave function are

$$\tilde{\psi}_{+,+} = \cos(\tilde{\theta}/2),$$

$$\tilde{\psi}_{-,+} = \sin(\tilde{\theta}/2)e^{i\tilde{\phi}},$$

and they satisfy the relations

$$\cos(\theta) = \cos(\tilde{\theta}), \tag{A4}$$

$$\sin(\theta)\cos(\phi) = \sin(\tilde{\theta})\cos(\tilde{\phi}). \tag{A5}$$

That is, $\theta=\tilde{\theta}$ and $\phi=\pm\tilde{\phi}$.

The ambiguity in the sign of ϕ is immediately resolved if we further require, as expected below T_c, that also the expectation values of the products $i\hat{\mathcal{T}}_h^z\hat{\mathcal{T}}_v^x$ and $i\hat{\mathcal{T}}_v^z\hat{\mathcal{T}}_h^x$ are conserved, leading to the relation

$$\sin(\theta)\sin(\phi) = \sin(\tilde{\theta})\sin(\tilde{\phi}). \tag{A6}$$

Therefore, the quantum topological order in this system is fully protected from thermal fluctuations, as long as $T<T_c$, in the sense that the system is bound to come back to the same exact initial state upon cooling back to zero temperature.

APPENDIX B: CHECK AGAINST KNOWN LIMITS

As a check of the steps leading to Eqs. (4.22) and (4.23), let us verify that the known limits are indeed recovered. For $T=0$ (i.e., for $J=0$) we have that $e^{\beta\lambda_A M_s(\alpha_l)}/Z_s \to \delta_{M_s(\alpha_l),N}/2^3|G|$, while $Z_J^{\text{tot}}(g)=Z_J^{\text{tot}}(\mathbf{1})$, $\forall g$. In the notation introduced below Eq. (4.14), this restricts the summation over α_l to states of the form $|\alpha_l\rangle=g'|0_k\rangle$, with $g'\in G$ and $k=1,\ldots,2^3$ labeling the states obtained from $|0\rangle$ by the action of the nonlocal Γ operators in Eq. (2.3). Namely, the states $|0_k\rangle$ are of the form $\Gamma_1^{m_1}\Gamma_2^{m_2}\Gamma_3^{m_3}|0\rangle$ for all possible choices of $m_1,m_2,m_3=0,1$. Equation (4.22) reduces then to

$$\begin{aligned}
\text{Tr}[\rho_{\mathcal{A}}^n(T)] &= d_{\mathcal{A}}^{n-1}\times\frac{1}{2^{3n}|G|^n}\sum_{\alpha_1,\ldots,\alpha_n}\left(\prod_{l=1}^{n}\delta_{M_s(\alpha_l),N}\right)\prod_{l=1}^{n-1}\delta(\alpha_{l,\mathcal{A}},\alpha_{l+1,\mathcal{A}}) \\
&= d_{\mathcal{A}}^{n-1}\times\frac{1}{2^{3n}|G|^n}\sum_{g_1',\ldots,g_n'\in G}\sum_{k_1,\ldots,k_n}\prod_{l=1}^{n-1}\delta[(g_l'0_{k_l})_{\mathcal{A}},(g_{l+1}'0_{k_{l+1}})_{\mathcal{A}}] \\
&= d_{\mathcal{A}}^{n-1}\times\frac{1}{2^{3n}|G|^n}2^{3n}\sum_{g_1',\ldots,g_n'\in G}\prod_{l=1}^{n-1}\delta[(g_l'0)_{\mathcal{A}},(g_{l+1}'0)_{\mathcal{A}}]
\end{aligned}$$

$$= d_{\mathcal{A}}^{n-1} \times \frac{1}{|G|^n} |G| d_{\mathcal{B}}^{n-1} = d_{\mathcal{A}}^{n-1} \times \left(\frac{d_{\mathcal{B}}}{|G|}\right)^{n-1} = \left(\frac{d_{\mathcal{A}} d_{\mathcal{B}}}{|G|}\right)^{n-1}, \tag{B1}$$

where we used the fact that, for the cases of interest, subsystem $\mathcal{A}$ is finite and the nonlocal operators Γ can always be chosen so as to traverse only subsystem $\mathcal{B}$, $\delta[(g_l' 0_{k_l})_{\mathcal{A}}, (g_{l+1}' 0_{k_{l+1}})_{\mathcal{A}}] \equiv \delta[(g_l' 0)_{\mathcal{A}}, (g_{l+1}' 0)_{\mathcal{A}}]$. This in turn implies that $g_l' g_{l+1}' \in G_{\mathcal{B}}$, and the constrained summation over $g_1', \ldots, g_n' \in G$ can be replaced by an unconstrained summation over $g_1' \in G$, $g_2'', \ldots, g_n'' \in G_{\mathcal{B}}$ (where $g_{l+1}'' \equiv g_l' g_{l+1}'$ for $l = 2, \ldots, n$). Equation (B1) is indeed the same as in the 2D case at zero temperature.[4]

In this limit, the von Neumann entropy is given by

$$S_{\mathrm{vN}}(\mathcal{A}; T=0) = -\lim_{n\to 1} \partial_n \operatorname{Tr}[\rho_{\mathcal{A}}^n] = -\ln\left(\frac{d_{\mathcal{A}} d_{\mathcal{B}}}{|G|}\right) \tag{B2}$$

and the topological entropy by $S_{\mathrm{topo}} = 2 \ln 2$ as discussed in Sec. III (for the full bipartition scheme 1–8). For $T \to \infty$ (i.e., for $J \to \infty$), we have $Z_J^{\mathrm{tot}}(g)/Z_J^{\mathrm{tot}}(\mathbf{1}) \to \delta(g-\mathbf{1})$, all α are equally weighed, and Eq. (4.22) reduces to

$$\operatorname{Tr}[\rho_{\mathcal{A}}^n] = 1 \times \frac{1}{Z_s^n} \sum_{\alpha_1,\ldots,\alpha_n} \prod_{l=1}^{n-1} \delta(\alpha_{l,\mathcal{A}}, \alpha_{l+1,\mathcal{A}})$$

$$= 1 \times \frac{1}{2^{3Nn}} 2^{3N} 2^{\Sigma_{\mathcal{B}}(n-1)} = 1 \times \left(\frac{1}{2^{\Sigma_{\mathcal{A}}}}\right)^{n-1}, \tag{B3}$$

where $\Sigma_{\mathcal{A}}$ ($\Sigma_{\mathcal{B}}$) is the number of σ spin degrees of freedom in $\mathcal{A}$ ($\mathcal{B}$) and $\Sigma_{\mathcal{A}} + \Sigma_{\mathcal{B}} = 3N$. Here we used the fact that $\delta(\alpha_{l,\mathcal{A}}, \alpha_{l+1,\mathcal{A}})$ involves only subsystem $\mathcal{A}$, hence $\Sigma_{\mathcal{A}}$ spins are summed over only once, while there are n independent copies of the remaining $\Sigma_{\mathcal{B}}$ spins.

This result leads to

$$S_{\mathrm{vN}}(\mathcal{A}; T \to \infty) = -\lim_{n\to 1} \partial_n \operatorname{Tr}[\rho_{\mathcal{A}}^n]$$

$$= \ln(2^{\Sigma_{\mathcal{A}}}) = \Sigma_{\mathcal{A}} \ln 2, \tag{B4}$$

which is indeed the classical entropy of a collection of $\Sigma_{\mathcal{A}}$ free Ising spins. The topological entropy vanishes in this limit since the contributions from the different bipartitions cancel out exactly (recall that the total number of spins in $\mathcal{A}$ for bipartitions 2 and 3 is the same as that for bipartitions 1 and 4 and similarly for 6,7 and 5,8).

Notice that, in our chosen factorization scheme in Eq. (4.22), the plaquette term does not yield any contribution to the von Neumann entropy at infinite temperature, while at zero temperature the plaquette term contribution equals $-\ln d_{\mathcal{A}}$, and the star term contribution is $-\ln(d_{\mathcal{B}}/|G|)$.

APPENDIX C: THE STAR CONTRIBUTION

Here we present the expressions for the star contribution to the entropies for finite temperatures and finite system sizes. As we argued in the Sec. IV B 1, the star contribution to the entropies can be computed using Eqs. (4.28), which relate them to entropies evaluated for a hard constrained system where $\lambda_B \to \infty$. The calculation in this limit is done most conveniently in the σ^z basis, very much along the lines of the calculation carried out for 2D systems in Ref. 4. Paralleling the steps of the computation for 2D systems, one obtains for the 3D case that

$$\Delta S_{\mathrm{vN}}(\mathcal{A}; T)|_{\lambda_B \to \infty} = \ln\cosh\left(\frac{K_A}{2} N\right) - \mathcal{N}_{\mathcal{A}}^{(s)} (x \ln x) \frac{\cosh\left[\frac{K_A}{2}(N-1)\right]}{\cosh\left(\frac{K_A}{2} N\right)} - \mathcal{N}_{\mathcal{A}}^{(s)} (y \ln y) \frac{\sinh\left[\frac{K_A}{2}(N-1)\right]}{\cosh\left(\frac{K_A}{2} N\right)}$$

$$- \sum_i (\tilde{x}_i \ln \tilde{x}_i) \frac{\cosh\left[\frac{K_A}{2}(N - \mathcal{N}_{\mathcal{A}_i}^{(s)})\right]}{\cosh\left(\frac{K_A}{2} N\right)} - \sum_i (\tilde{y}_i \ln \tilde{y}_i) \frac{\sinh\left[\frac{K_A}{2}(N - \mathcal{N}_{\mathcal{A}_i}^{(s)})\right]}{\cosh\left(\frac{K_A}{2} N\right)}, \tag{C1}$$

where $K_A = -\ln[\tanh(\lambda_A/T)]$, $\mathcal{N}_{\mathcal{A}_i}^{(s)} \equiv \mathcal{N}_{\mathcal{B}_i}^{(s)} + \mathcal{N}_{\mathcal{A}\mathcal{B}_i}^{(s)}$ is the total number of star operators acting on the ith component of subsystem $\mathcal{B}$ (either entirely in $\mathcal{B}_i$ or at its boundary $\mathcal{A}\mathcal{B}_i$), and

$$x = \cosh\left(\frac{K_A}{2}\right) \quad y = \sinh\left(\frac{K_A}{2}\right), \tag{C2a}$$

$$\tilde{x}_i = \cosh\left(\frac{K_A}{2} \mathcal{N}_{\mathcal{A}_i}^{(s)}\right) \quad \tilde{y}_i = \sinh\left(\frac{K_A}{2} \mathcal{N}_{\mathcal{A}_i}^{(s)}\right). \tag{C2b}$$

Notice that only the last two terms in Eq. (C1) yield a topological contribution in our bipartition scheme since $\mathcal{N}_{1,\mathcal{A}}^{(s)} - \mathcal{N}_{2,\mathcal{A}}^{(s)} - \mathcal{N}_{3,\mathcal{A}}^{(s)} + \mathcal{N}_{4,\mathcal{A}}^{(s)} = 0$ and likewise for bipartitions 5–8. Therefore,

$$\Delta S^{(S)}_{\text{topo}}(T/\lambda_A,N) = \sum_{i=1}^{2} [\tilde{x}_i^{(1)} \ln \tilde{x}_i^{(1)}] \frac{\cosh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{1\bar{\mathcal{A}}_i})\right]}{\cosh\left(\frac{K_A}{2}N\right)} + \sum_{i=1}^{2} [\tilde{y}_i^{(1)} \ln \tilde{y}_i^{(1)}] \frac{\sinh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{1\bar{\mathcal{A}}_i})\right]}{\cosh\left(\frac{K_A}{2}N\right)}$$
$$- [\tilde{x}^{(2)} \ln \tilde{x}^{(2)}] \frac{\cosh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{2\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)} - [\tilde{y}^{(2)} \ln \tilde{y}^{(2)}] \frac{\sinh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{2\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)} - [\tilde{x}^{(3)} \ln \tilde{x}^{(3)}] \frac{\cosh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{3\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)}$$
$$- [\tilde{y}^{(3)} \ln \tilde{y}^{(3)}] \frac{\sinh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{3\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)} + [\tilde{x}^{(4)} \ln \tilde{x}^{(4)}] \frac{\cosh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{4\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)} + [\tilde{y}^{(4)} \ln \tilde{y}^{(4)}] \frac{\sinh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{4\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)}$$
$$+ [\tilde{x}^{(5)} \ln \tilde{x}^{(5)}] \frac{\cosh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{5\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)} + [\tilde{y}^{(5)} \ln \tilde{y}^{(5)}] \frac{\sinh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{5\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)} - [\tilde{x}^{(6)} \ln \tilde{x}^{(6)}] \frac{\cosh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{6\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)}$$
$$- [\tilde{y}^{(6)} \ln \tilde{y}^{(6)}] \frac{\sinh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{6\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)} - [\tilde{x}^{(7)} \ln \tilde{x}^{(7)}] \frac{\cosh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{7\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)} - [\tilde{y}^{(7)} \ln \tilde{y}^{(7)}] \frac{\sinh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{7\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)}$$
$$+ [\tilde{x}^{(8)} \ln \tilde{x}^{(8)}] \frac{\cosh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{8\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)} + [\tilde{y}^{(8)} \ln \tilde{y}^{(8)}] \frac{\sinh\left[\frac{K_A}{2}(N-\mathcal{N}^{(s)}_{8\bar{\mathcal{A}}})\right]}{\cosh\left(\frac{K_A}{2}N\right)}, \tag{C3}$$

where we used the fact that subsystem $\mathcal{B}$ has always one component except for bipartition 1, where it has two components.

With the expression above for $\Delta S^{(S)}_{\text{topo}}(T/\lambda_A,N)$, one can determine the topological entropy contribution from the star operators as a function of temperature and system sizes. In particular, let us look at two particular limits: that of the zero-temperature limit taken first and that of the thermodynamic limit taken first.

For $T\to 0$ first $K_A\to 0$, and one can easily check that all terms in Eq. (C3) vanish, which is expected as the difference $\Delta S^{(S)}_{\text{topo}}(T/\lambda_A,N)$ is, by definition, zero at $T=0$. Now, when the thermodynamic limit is taken first, i.e., when the sizes N and all of $\mathcal{N}^{(s)}_{1\bar{\mathcal{A}}_i}$ (for $i=1,2$) and $\mathcal{N}^{(s)}_{p\bar{\mathcal{A}}_i}$, $p=2-8$, are taken to infinity at fixed K_A, each term in the expression in Eq. (C3) gives $\mp\ln 2$ (with the sign determined by whether the partition is added or subtracted). Bipartition 1 gives $-2\ln 2$ (its contribution is doubled because $1\mathcal{B}$ has two disconnected components) and it is added to bipartitions 4, 5, and 8, which give $-\ln 2$ each; bipartitions 2, 3, 6, and 7 are subtracted and each of them gives $+\ln 2$. Altogether, we obtain $\Delta S^{(S)}_{\text{topo}}(T/\lambda_A,N\to\infty)=-\ln 2$ for any temperature T. Therefore, we obtain in the thermodynamic limit the result used in Eq. (4.30).

One can finally add the zero-temperature contributions to obtain

$$S^{(S)}_{\text{vN}}(T/\lambda_A) = \Delta S^{(S)}_{\text{vN}}(T/\lambda_A) - \frac{d_{\mathcal{B}}}{|G|} \tag{C4}$$

and

$$S^{(S)}_{\text{topo}}(T/\lambda_A) = \Delta S^{(S)}_{\text{topo}}(T/\lambda_A) + \ln\frac{d_{1\mathcal{B}}d_{4\mathcal{B}}d_{5\mathcal{B}}d_{8\mathcal{B}}}{d_{2\mathcal{B}}d_{3\mathcal{B}}d_{6\mathcal{B}}d_{7\mathcal{B}}}$$
$$= \Delta S^{(S)}_{\text{topo}}(T/\lambda_A) + \ln 2. \tag{C5}$$

APPENDIX D: THE PLAQUETTE CONTRIBUTION

As anticipated in Sec. IV B 2, the plaquette contribution in three dimensions is very different from the 2D case, and we need to carry out the calculations explicitly. Consider the expression for $\mathcal{Z}^{(P)}$,

$$\mathcal{Z}^{(P)}(n) = \sum_{g_1,\ldots,g_n\in G_{\mathcal{A}}} \left(\prod_{l=1}^{n} \frac{Z^{\text{tot}}_J(g_l)}{Z^{\text{tot}}_J(\mathbf{1})}\right) \langle 0|\prod_{l=1}^{n} g_l|0\rangle, \tag{D1}$$

where

$$Z_J^{\mathrm{tot}}(g) = \sum_{\{S_i\}} \exp\left(J \sum_{\langle ij \rangle} \eta_{ij}(g) S_i S_j\right) \tag{D2}$$

is the partition function of the 3D random-bond Ising model (summed over all possible boundary conditions), whose randomness is controlled by g according to Eq. (4.11). Namely, $\eta_{ij}(g) = \pm 1$ depending on whether the plaquette perpendicular to the bond $\langle ij \rangle$ is flipped in configuration g ($\eta_{ij} = -1$) or not ($\eta_{ij} = +1$).

Recall that the group G, and therefore its subgroup $G_{\mathcal{A}}$, is defined modulo the identities $\Pi_{\text{closed membrane}} B_p = \mathbf{1}$. In the language of the randomness realizations $\{\eta_{ij}\}$, this amounts to summing over gauge *inequivalent* configurations. In fact, any η_{ij} and η'_{ij} that differ by the product of plaquettes around closed surfaces are related by

$$\eta_{ij} = \eta'_{ij} \bar{S}_i \bar{S}_j, \quad \exists \{\bar{S}_i\}. \tag{D3}$$

Specifically, $\{\bar{S}_i\}$ corresponds to either of the two spin configurations that exhibit the closed surfaces in question as their only antiferromagnetic boundary (the two configurations are related by an overall Z_2 symmetry). Recall that the product of plaquettes belonging to an infinite crystal plane is also an allowed gauge transformation, and all possible boundary conditions (periodic or antiperiodic in each direction) should be taken into account when enumerating all configurations $\{\bar{S}_i\}$. In conclusion, every $\eta_{ij}(g)$ admits 2^{N+3} equivalent randomness realizations $\eta'_{ij} = \eta_{ij} \bar{S}_i \bar{S}_j$, labeled by all possible Ising configurations $\{\bar{S}_i\}_{i=1}^N$ (where $\{\bar{S}_i\}_{i=1}^N$ and $\{-\bar{S}_i\}_{i=1}^N$ yield the exact same η'_{ij}).

In the case of a summation over the whole group G, one has then the identity

$$\sum_{g \in G} \sum_{\{S_i\}} \exp\left(J \sum_{\langle ij \rangle} \eta_{ij}(g) S_i S_j\right) \equiv \frac{1}{2^{N+3}} \sum_{\{\eta_{ij}\}} \sum_{\{S_i\}} \exp\left(J \sum_{\langle ij \rangle} \eta_{ij} S_i S_j\right). \tag{D4}$$

For the subgroup $G_{\mathcal{A}}$, the situation is more convoluted. First of all, the operators $g \in G_{\mathcal{A}}$ correspond to randomness realizations $\{\eta_{ij}(g)\}$ where all the bonds outside $\mathcal{A}$ can be gauged to assume the value +1. Rather than considering all the equivalent configurations as for the whole group G, it is more convenient to introduce a restricted set of randomness realizations $\{\eta_{ij}^{(\mathcal{A})}\}$ where $\eta_{ij}^{(\mathcal{A})}$ is constrained to assume the value +1 whenever $\langle ij \rangle \notin \mathcal{A}$. Notice that we do not constrain the bonds inside $\mathcal{A}$, and we are therefore overcounting all the gauge equivalent configurations with respect to these bonds. The number of equivalent realizations in the restricted subgroup can be counted as seen in Sec. III and repeated hereafter for convenience. All cubic unit cells entirely contained in $\mathcal{A}$ are independent generators of gauge transformations. Also, if $\mathcal{A}$ contains crystal planes, there are up to three additional generators. Finally, we have one extra generator per connected component of $\mathcal{B}$ (i.e., entirely surrounded by $\mathcal{A}$) but for one of them. Thus, the total number of gauge equivalent configurations is now $2^{\mathcal{N}_{\mathcal{A}}^{(c)} + m_{\mathcal{A}}^{(\mathrm{CP})} + (m_{\mathcal{B}} - 1)}$, where again $\mathcal{N}_{\mathcal{A}}^{(c)}$ is the number of cubic unit cells entirely contained in $\mathcal{A}$, $m_{\mathcal{A}}^{(\mathrm{CP})}$ is the number of independent crystal planes in $\mathcal{A}$ ($m_{\mathcal{A}}^{(\mathrm{CP})} = 0$, $m_{\mathcal{B}}^{(\mathrm{CP})} = 3$ for all cases of interest), and $m_{\mathcal{B}}$ is the number of connected components of $\mathcal{B}$.

As a result, one obtains

$$\sum_{g' \in G_{\mathcal{A}}} \sum_{\{S_i\}} \exp\left(J \sum_{\langle ij \rangle} \eta_{ij}(g') S_i S_j\right) \equiv \frac{1}{2^{\mathcal{N}_{\mathcal{A}}^{(c)} + (m_{\mathcal{B}} - 1)}} \sum_{\{\eta_{ij}^{(\mathcal{A})}\}} \sum_{\{S_i\}} \left[\exp\left(J \sum_{\langle ij \rangle \in \mathcal{A}} \eta_{ij}^{(\mathcal{A})} S_i S_j\right) \times \exp\left(J \sum_{\langle ij \rangle \notin \mathcal{A}} S_i S_j\right)\right]. \tag{D5}$$

Having done so, the summation over $\{\eta_{ij}^{(\mathcal{A})}\}$ is now unconstrained, namely, the bond variables $\eta_{ij}^{(\mathcal{A})} = \pm 1$ are generated by freely flipping any of the plaquettes in $\mathcal{A}$, starting from the configuration with all $\eta_{ij}^{(\mathcal{A})} = +1$ (which we refer to in the following as $\eta^0 \equiv \{\eta_{ij}^0\}$, the ferromagnetic configuration). Notice that this accounts only for the bipartitions where the plaquette operators in $\mathcal{A}$ are sufficient to generate the whole group $G_{\mathcal{A}}$ (bipartitions 1, 2, 3 and 6, 7, 8). As discussed in Sec. III, this is not always the case and additional collective operations may be needed to generate $G_{\mathcal{A}}$ (bipartitions 4 and 5). The summation encompasses then all configurations obtained by flipping plaquettes in $\mathcal{A}$ starting from $\{\eta_{ij}^0\}$ *and* starting from the configurations derived from the ferromagnetic one via the action of each of the independent collective operations. For concreteness, in bipartitions 4 and 5 there is only one collective operation in $\mathcal{A}$, illustrated in the bottom panel of Fig. 5. In this case, the configurations $\{\eta_{ij}^{(\mathcal{A})}\}$ are obtained by flipping plaquettes in $\mathcal{A}$ starting from the ferromagnetic configuration η^0 and starting from the configuration with all $\eta_{ij}^{(\mathcal{A})} = +1$, except for those inside the blue thick line in the bottom panel of Fig. 5 (i.e., plaquettes in $\mathcal{B}$ or at the boundary), where $\eta_{ij}^{(\mathcal{A})} = -1$. (We will refer to this configuration in the following as $\eta^1 \equiv \{\eta_{ij}^1\}$.) If we label $\{\tilde{\eta}^{(\mathcal{A})} \equiv \{\tilde{\eta}_{ij}^{(\mathcal{A})}\}\}$ the set of all configurations obtained from the ferromagnetic one via the action of the plaquette operators in $\mathcal{A}$ alone, the summation in Eq. (D5) runs over $\eta^0\{\tilde{\eta}^{(\mathcal{A})}\} \cup \eta^1\{\tilde{\eta}^{(\mathcal{A})}\}$, where the product of two configurations represents the new configuration with variables given by the site-by-site product of the two original variables $\eta_{ij}^0 \tilde{\eta}_{ij}^{(\mathcal{A})}$ ($\equiv \tilde{\eta}_{ij}^{(\mathcal{A})}$) and $\eta_{ij}^1 \tilde{\eta}_{ij}^{(\mathcal{A})}$.

We can then apply the identity in Eq. (D5) to simplify our expression in Eq. (D1). The condition that a term is nonvanishing, namely, $\langle 0_{\mathcal{A}} | g'_{1,\mathcal{A}} \cdots g'_{n,\mathcal{A}} | 0_{\mathcal{A}} \rangle = 1$, translates into the condition that

$$\prod_{l=1}^{n} \eta_{ij}^{(\mathcal{A},l)}(g_l) = \tilde{S}_i \tilde{S}_j \quad \forall \langle ij \rangle, \quad \exists \{\tilde{S}_i\}, \tag{D6}$$

i.e., the product of all $\eta_{ij}^{(\mathcal{A},l)}(g_l)$, $l = 1, \ldots, n$, is gauge equivalent to η^0 (equivalently $g'_{1,\mathcal{A}} \cdots g'_{n,\mathcal{A}} = \mathbf{1}$). The very same nature of a collective operation in $\mathcal{A}$ requires that such operation cannot be completed to an identity (a closed membrane) by means of plaquette operators in $\mathcal{A}$ alone. Therefore the

above equation holds independently for the collective operations and for the $\tilde{\eta}^{(\mathcal{A})}$ configurations. Namely, it imposes that the number of collective operations appearing in $\{\eta_{ij}^{(\mathcal{A},l)}\}_{l=1}^{n}$ is even and that

$$\prod_{l=1}^{n} \tilde{\eta}_{ij}^{(\mathcal{A},l)}(g_l) = \tilde{S}_i \tilde{S}_j \quad \forall \langle ij \rangle, \quad \exists \{\tilde{S}_i\}. \tag{D7}$$

Trivially, Eqs. (D6) and (D7) become equivalent if no collective operations are present in $\mathcal{A}$.

Notice that $\tilde{S}_i \tilde{S}_j \equiv 1$ for all $\langle ij \rangle \notin \mathcal{A}$: all possible $\{\tilde{S}_i\}$ configurations must be ferromagnetically ordered outside $\mathcal{A}$. If $m_{\mathcal{B}}$ is the number of connected components in $\mathcal{B}$, then the ferromagnetic order holds across each component separately, and from one component to the next the overall sign of the $\tilde{S}$ spins may change. An overall sign change in the spins $\tilde{S}$ is immaterial, as one can see from Eq. (D7), and therefore one needs to introduce a corresponding factor of 1/2 when summing over $\{\tilde{S}_i\}$.

Equation (D1) then becomes

$$\begin{aligned}
\mathcal{Z}^{(P)}(n) &= \left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1})}\right)^n \frac{1}{2^{[\mathcal{N}_{\mathcal{A}}^{(c)}+m_{\mathcal{B}}-1]n}} \frac{1}{2} \sum_{\{\tilde{S}_i\}} \sum_{\substack{\{\{\eta_{ij}^{(\mathcal{A},l)}\}\}_{l=1}^{n} \\ \Pi_{l=1}^{n} \eta_{ij}^{(\mathcal{A},l)} = \tilde{S}_i \tilde{S}_j}} \prod_{l=1}^{n} \left[\sum_{\{S_i^{(l)}\}} \left(\prod_{\langle ij \rangle} \exp(J \eta_{ij}^{(\mathcal{A},l)} S_i^{(l)} S_j^{(l)}) \right) \right] \\
&= \left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1}) 2^{\mathcal{N}_{\mathcal{A}}^{(c)}+m_{\mathcal{B}}-1}}\right)^n \frac{1}{2} \sum_{\{\tilde{S}_i\}} \sum_{\substack{\{\{\eta_{ij}^{(\mathcal{A},l)}\}\}_{l=1}^{n} \\ \Pi_{l=1}^{n} \eta_{ij}^{(\mathcal{A},l)} = \tilde{S}_i \tilde{S}_j}} \sum_{\{\{S_i^{(l)}\}\}_{l=1}^{n}} \left[\prod_{\langle ij \rangle} \exp\left(J \sum_{l=1}^{n} \eta_{ij}^{(\mathcal{A},l)} S_i^{(l)} S_j^{(l)} \right) \right] \\
&= \left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1}) 2^{\mathcal{N}_{\mathcal{A}}^{(c)}+m_{\mathcal{B}}-1}}\right)^n \frac{1}{2} \sum_{\{\tilde{S}_i\}} \sum_{\{\{S_i^{(l)}\}\}_{l=1}^{n}} \prod_{\langle ij \rangle} \left[\sum_{\substack{\{\eta_{ij}^{(\mathcal{A},l)}\}_{l=1}^{n} \\ \Pi_{l=1}^{n} \eta_{ij}^{(\mathcal{A},l)} = \tilde{S}_i \tilde{S}_j}} \exp\left(J \sum_{l=1}^{n} \eta_{ij}^{(\mathcal{A},l)} S_i^{(l)} S_j^{(l)} \right) \right] \\
&= \left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1}) 2^{\mathcal{N}_{\mathcal{A}}^{(c)}+m_{\mathcal{B}}-1}}\right)^n \frac{1}{2} \sum_{\{\tilde{S}_i\}} \sum_{\{\{S_i^{(l)}\}\}_{l=1}^{n}} \sum_{\{\bar{\eta}^{(l)}\}_{l=1}^{n}}^{(\mathrm{even})} \prod_{\langle ij \rangle \in \mathcal{A}} \left[\sum_{\substack{\{\tilde{\eta}_{ij}^{(\mathcal{A},l)}\}_{l=1}^{n} \\ \Pi_{l=1}^{n} \tilde{\eta}_{ij}^{(\mathcal{A},l)} = \tilde{S}_i \tilde{S}_j}} \exp\left(J \sum_{l=1}^{n} \bar{\eta}_{ij}^{(l)} \tilde{\eta}_{ij}^{(\mathcal{A},l)} S_i^{(l)} S_j^{(l)} \right) \right] \\
&\quad \times \prod_{\langle ij \rangle \notin \mathcal{A}} \left[\exp\left(J \sum_{l=1}^{n} \bar{\eta}_{ij}^{(l)} S_i^{(l)} S_j^{(l)} \right) \right],
\end{aligned} \tag{D8}$$

where $\Sigma_{\{\bar{\eta}^{(l)}\}_{l=1}^{n}}^{(\mathrm{even})}$ runs over all $n$tuples, $\{\bar{\eta}^{(l)} \in (\eta^0, \eta^1)\}_{l=1}^{n}$, with an even number of η^1 terms. Notice that the summation

$$\begin{aligned}
&\sum_{\substack{\{\tilde{\eta}_{ij}^{(\mathcal{A},l)}\}_{l=1}^{n} \\ \Pi_{l=1}^{n} \tilde{\eta}_{ij}^{(\mathcal{A},l)} = \tilde{S}_i \tilde{S}_j}} \exp\left(J \sum_{l=1}^{n} \bar{\eta}_{ij}^{(l)} \tilde{\eta}_{ij}^{(\mathcal{A},l)} S_i^{(l)} S_j^{(l)} \right) \\
&\quad = \mathcal{Z}_n(\{J \bar{\eta}_{ij}^{(l)} S_i^{(l)} S_j^{(l)}\}; \tilde{S}_i \tilde{S}_j),
\end{aligned} \tag{D9}$$

where $\mathcal{Z}_n(\{J \bar{\eta}_{ij}^{(l)} S_i^{(l)} S_j^{(l)}\}; \tilde{S}_i \tilde{S}_j)$ can be interpreted as the partition function of an Ising chain of degrees of freedom $\{\tilde{\eta}_{ij}^{(\mathcal{A},l)}\}_{l=1}^{n}$ in a random field of local strength $J \bar{\eta}_{ij}^{(l)} S_i^{(l)} S_j^{(l)}$ and subject to the condition that the product of all Ising spins $\Pi_{l=1}^{n} \tilde{\eta}_{ij}^{(\mathcal{A},l)}$ equals $\tilde{S}_i \tilde{S}_j$. By means of the change of variables $\tilde{\eta}_{ij}^{(\mathcal{A},l)} = m_{ij}^{(\mathcal{A},l)} m_{ij}^{(\mathcal{A},l+1)}$, this becomes the partition function of a nearest-neighbor Ising chain with periodic or antiperiodic BCs depending on the sign of $\tilde{S}_i \tilde{S}_j = \pm 1$ (i.e., $m_{ij}^{(\mathcal{A},n+1)} = m_{ij}^{(\mathcal{A},1)} \tilde{S}_i \tilde{S}_j$),

$$\begin{aligned}
\mathcal{Z}_n &\equiv \mathcal{Z}_n(\{J \bar{\eta}_{ij}^{(l)} S_i^{(l)} S_j^{(l)}\}; \tilde{S}_i \tilde{S}_j) \\
&= \frac{1}{2} \sum_{\substack{\{m_{ij}^{(\mathcal{A},l)}\}_{l=1}^{n} \\ \mathrm{BC} = \tilde{S}_i \tilde{S}_j}} \exp\left(J \sum_{l=1}^{n} \bar{\eta}_{ij}^{(l)} m_{ij}^{(\mathcal{A},l)} m_{ij}^{(\mathcal{A},l+1)} S_i^{(l)} S_j^{(l)} \right).
\end{aligned} \tag{D10}$$

This in turn can be computed exactly,

$$2\mathcal{Z}_n = (2\cosh J)^n + \left[\left(\prod_{l=1}^{n} \bar{\eta}_{ij}^{(l)} S_i^{(l)} S_j^{(l)}\right)\tilde{S}_i\tilde{S}_j\right](2\sinh J)^n$$
$$= (2\cosh J)^n + \left[\left(\prod_{l=1}^{n} S_i^{(l)} S_j^{(l)}\right)\tilde{S}_i\tilde{S}_j\right](2\sinh J)^n. \tag{D11}$$

We also used the fact that $\bar{\eta}_{ij}^{(l)} = +1$ if $\langle ij\rangle \in \mathcal{A}$ by construction. (Notice that this convenient choice does not introduce any limitations. In general, the number of times when a -1 appears in the $l=1,\ldots,n$ sequence of $\bar{\eta}_{ij}^{(l)}$ values must be even, and therefore $\Pi_{l=1}^{n}\bar{\eta}_{ij}^{(l)} = +1$, $\forall i,j$.)

For convenience of notation, let us consider the following change of summation variables:

$$\tilde{S}_i \to \theta_i = \left(\prod_{l=1}^{n} S_i^{(l)}\right)\tilde{S}_i, \tag{D12}$$

so that we can write $\mathcal{Z}_n = \frac{1}{2}e^{A_n}e^{B_n\theta_i\theta_j}$, with A_n and B_n defined as

$$e^{A_n+B_n} = (2\cosh J)^n + (2\sinh J)^n, \tag{D13a}$$

$$e^{A_n-B_n} = (2\cosh J)^n - (2\sinh J)^n. \tag{D13b}$$

Given that $\Pi_{l=1}^{n} S_i^{(l)} = \pm 1$, for all sites i whose adjacent bonds $\langle ij\rangle$ are *solely* in $\mathcal{A}$, the summation over $\{\tilde{S}_i = \pm 1\}$ and the summation over $\{\theta_i = \pm 1\}$ are unconstrained. The case is different for the sites i that have an adjacent bond not in $\mathcal{A}$. The correlation across such bond is, in fact, ferromagnetic by construction, and if $\mathcal{B}$ has only one connected component, the spin $\tilde{S}_i$ has the same sign as all other spins not entirely surrounded by bonds in $\mathcal{A}$. Consequently, all the boundary spins $\tilde{S}_i$ have the same sign, and the values of the associated spins θ_i are determined uniquely by the product $\Pi_{l=1}^{n} S_i^{(l)}$. If $m_{\mathcal{B}}$ is the number of connected components in $\mathcal{B}$, then the ferromagnetic order holds across each component separately, and from one component to the next the overall sign of the $\tilde{S}$ spins may change. This is accounted for by summing over boundary sign variables $q_r = \pm 1$, $r=1,\ldots,m_{\mathcal{B}}$, assigned to each boundary ∂_r defined as the set of sites that have adjacent bonds both in $\mathcal{A}$ and in the rth component of $\overline{\mathcal{A}}$.

In the end, Eq. (D8) becomes

$$\mathcal{Z}^{(P)}(n) = \left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1})2^{\mathcal{N}_{\mathcal{A}}^{(c)}+m_{\mathcal{B}}-1}}\right)^n \frac{1}{2}\sum_{\{\tilde{S}_i\}}\sum_{\{\{S_i^{(l)}\}\}_{l=1}^{n}}\prod_{\langle ij\rangle\in\mathcal{A}}\mathcal{Z}_n(\{JS_i^{(l)}S_j^{(l)}\};\tilde{S}_i\tilde{S}_j)\sum_{\{\bar{\eta}^{(l)}\}_{l=1}^{n}}^{(\mathrm{even})}\prod_{\langle ij\rangle\notin\mathcal{A}}\left[\exp\left(J\sum_{l=1}^{n}\bar{\eta}_{ij}^{(l)}S_i^{(l)}S_j^{(l)}\right)\right]$$
$$= \left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1})2^{\mathcal{N}_{\mathcal{A}}^{(c)}+m_{\mathcal{B}}-1}}\right)^n \frac{1}{2}\sum_{\{\{S_i^{(l)}\}\}_{l=1}^{n}}\sum_{\{\theta_i\}}\prod_{\langle ij\rangle\in\mathcal{A}}\frac{1}{2}e^{A_n}e^{B_n\theta_i\theta_j}\sum_{\{\bar{\eta}^{(l)}\}_{l=1}^{n}}^{(\mathrm{even})}\prod_{\langle ij\rangle\notin\mathcal{A}}\left[\exp\left(J\sum_{l=1}^{n}\bar{\eta}_{ij}^{(l)}S_i^{(l)}S_j^{(l)}\right)\right]$$
$$\times\sum_{\{q_r=\pm 1\}_{r=1}^{m_{\mathcal{B}}}}\prod_{r=1}^{m_{\mathcal{B}}}\prod_{i\in\partial_r}\delta\left(\theta_i\prod_{l=1}^{n}S_i^{(l)}=q_r\right)$$
$$= \left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1})2^{\mathcal{N}_{\mathcal{A}}^{(c)}+m_{\mathcal{B}}-1}}\right)^n \frac{e^{\mathcal{N}_{\mathcal{A}}^{(p)}A_n}}{2^{\mathcal{N}_{\mathcal{A}}^{(p)}}}\frac{1}{2}\sum_{\{\{S_i^{(l)}\}\}_{l=1}^{n}}\sum_{\{\theta_i\}}\prod_{\langle ij\rangle\in\mathcal{A}}e^{B_n\theta_i\theta_j}\sum_{\{\bar{\eta}^{(l)}\}_{l=1}^{n}}^{(\mathrm{even})}\prod_{\langle ij\rangle\notin\mathcal{A}}\left[\exp\left(J\sum_{l=1}^{n}\bar{\eta}_{ij}^{(l)}S_i^{(l)}S_j^{(l)}\right)\right]$$
$$\times\sum_{\{q_r=\pm 1\}_{r=1}^{m_{\mathcal{B}}}}\prod_{r=1}^{m_{\mathcal{B}}}\prod_{i\in\partial_r}\delta\left(\theta_i\prod_{l=1}^{n}S_i^{(l)}=q_r\right). \tag{D14}$$

Notice that $\bar{\eta}_{ij}^{(l)} = +1$ if the plaquette $\langle ij\rangle$ does not belong to the collective operation and that whenever $\langle ij\rangle$ belongs to the collective operation the value of $\bar{\eta}_{ij}^{(l)} = \pm 1$ is the same for all $\langle ij\rangle$. (We restrict here for simplicity to the case where there is at most one collective operation in $\mathcal{A}$. In order to extend to the general case one needs to repeat the derivation for each collective operation separately.)

Notice also that the sum over $S_i^{(l)}$ that are entirely surrounded by bonds in $\mathcal{A}$ is unconstrained, and it contributes a trivial factor $2^{\mathcal{N}_{\mathcal{A}}^{(c)}n}$ to the sum over the remaining spins. In the following, we use this simplification and all summations over $S_i^{(l)}$ are intended as constrained only to the remaining spins (for convenience, we do not increase the already complex notation).

Let us focus on the boundary condition,

$$\sum_{\{q_r=\pm 1\}_{r=1}^{m_{\mathcal{B}}}}\prod_{r=1}^{m_{\mathcal{B}}}\prod_{i\in\partial_r}\delta\left(\theta_i\prod_{l=1}^{n}S_i^{(l)}=q_r\right). \tag{D15}$$

Given that the θ and the S spins can assume only the values ± 1, then the quantity $\theta_i + \Sigma_{l=1}^{n} S_i^{(l)}$ can only assume the values $n+1, n-1, n-3, \ldots, -(n-1), -(n+1)$.[23] In particular, the product $\theta_i\Pi_{l=1}^{n}S_i^{(l)}$ is positive whenever said summation equals $n+1, n-3, n-7, \ldots$, and it is negative otherwise. We

can therefore rewrite the delta function in the above equation as

$$\delta\left(\theta_i\prod_{l=1}^{n}S_i^{(l)}=q_r\right)=\sum_{p=0}^{\lfloor(n+q_r)/2\rfloor}\delta\left(\theta_i+\sum_{l=1}^{n}S_i^{(l)}=n+q_r-4p\right),$$

where $\lfloor\cdot\rfloor$ stands for the integer part of its argument. In other words, the sum $\theta_i+\Sigma_{l=1}^{n}S_i^{(l)}$ must equal $n+q\pmod 4$ or

$$\theta_i+\sum_{l=1}^{n}S_i^{(l)}-(n+q_r)=0\pmod 4. \tag{D16}$$

Using the function

$$f(x)=\frac{1}{4}\sum_{k=0}^{3}\exp\left(i\frac{\pi}{2}kx\right)=\begin{cases}1 & \text{if } x=0\pmod 4\\ 0 & \text{if } x=1,2,3\pmod 4,\end{cases} \tag{D17}$$

we can finally write the delta function as

$$\delta\left(\theta_i\prod_{l=1}^{n}S_i^{(l)}=q_r\right)=\frac{1}{4}\sum_{k_i}\exp\left[i\frac{\pi}{2}k_i\left(\theta_i+\sum_{l=1}^{n}S_i^{(l)}-(n+q_r)\right)\right]. \tag{D18}$$

Substituting into Eq. (D14), we obtain

$$\begin{aligned}\mathcal{Z}^{(P)}(n)=&\left(\frac{1}{Z_J^{\rm tot}(\mathbf{1})2^{\mathcal{N}_{\overline{\mathcal{A}}}^{(c)}+m_{\mathcal{B}}-1}}\right)^n\frac{e^{\mathcal{N}_{\mathcal{A}}^{(p)}A_n}}{2^{\mathcal{N}_{\mathcal{A}}^{(p)}}}2^{\mathcal{N}_{\mathcal{A}}^{(c)}n}\frac{1}{2}\sum_{\{\{S_i^{(l)}\}\}_{l=1}^{n}}\sum_{\{\theta_i\}}\prod_{\langle ij\rangle\in\mathcal{A}}e^{B_n\theta_i\theta_j}\sum_{\{\overline{\eta}^{(l)}\}_{l=1}^{n}}^{\rm (even)}\prod_{\langle ij\rangle\notin\mathcal{A}}\left[\exp\left(J\sum_{l=1}^{n}\overline{\eta}_{ij}^{(l)}S_i^{(l)}S_j^{(l)}\right)\right]\\&\times\sum_{\{q_r=\pm 1\}_{r=1}^{m_{\mathcal{B}}}}\prod_{r=1}^{m_{\mathcal{B}}}\prod_{i\in\partial_r}\frac{1}{4}\sum_{k_i}\exp\left[i\frac{\pi}{2}k_i\left(\theta_i+\sum_{l=1}^{n}S_i^{(l)}-(n+q_r)\right)\right]\\=&\left(\frac{1}{Z_J^{\rm tot}(\mathbf{1})2^{m_{\mathcal{B}}-1}}\right)^n\frac{e^{\mathcal{N}_{\mathcal{A}}^{(p)}A_n}}{2^{\mathcal{N}_{\mathcal{A}}^{(p)}}}\frac{1}{2}\frac{1}{4^{\mathcal{N}_\partial}}\sum_{\{q_r=\pm 1\}_{r=1}^{m_{\mathcal{B}}}}\sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}\sum_{\{\theta_i\}}\prod_{\langle ij\rangle\in\mathcal{A}}e^{B_n\theta_i\theta_j}\\&\times\sum_{\{\overline{\eta}^{(l)}\}_{l=1}^{n}}^{\rm (even)}\sum_{\{S_i^{(1)}\}}\left[\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}\overline{\eta}_{ij}^{(1)}S_i^{(1)}S_j^{(1)}\right)\right]\exp\left[i\frac{\pi}{2}\sum_{r=1}^{m_{\mathcal{B}}}\sum_{i\in\partial_r}k_i(\theta_i+S_i^{(1)}-1-q_r)\right]\\&\times\sum_{\{\{S_i^{(l)}\}\}_{l=2}^{n}}\left[\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}\sum_{l=2}^{n}\overline{\eta}_{ij}^{(l)}S_i^{(l)}S_j^{(l)}\right)\right]\exp\left\{i\frac{\pi}{2}\sum_{i\in\partial}k_i\left[\sum_{l=2}^{n}(S_i^{(l)}-1)\right]\right\},\end{aligned} \tag{D19}$$

where ∂ and $\mathcal{N}_\partial$ are, respectively, the full set and the total number of boundary sites, i.e., sites that have adjacent bonds both in $\mathcal{A}$ and outside $\mathcal{A}$. In the language introduced earlier, $\mathcal{N}^{(c)}=\mathcal{N}_{\mathcal{A}}^{(c)}+\mathcal{N}_{\overline{\mathcal{A}}}^{(c)}=\mathcal{N}_{\mathcal{A}}^{(c)}+\mathcal{N}_{\mathcal{B}}^{(c)}+\mathcal{N}_\partial$ and therefore $\mathcal{N}_{\overline{\mathcal{A}}}^{(c)}=\mathcal{N}_{\mathcal{B}}^{(c)}+\mathcal{N}_\partial$.

Note that the last line in Eq. (D19) does not depend on the $S^{(1)}$ or θ spins. If we introduce the partition functions,

$$Z_{\{k_i\}}^{\overline{\mathcal{A}},+}=\sum_{\{S_i\}}\exp\left[J\sum_{\langle ij\rangle\notin\mathcal{A}}S_iS_j+i\frac{\pi}{2}\sum_{i\in\partial}k_i(S_i-1)\right],$$

$$Z_{\{k_i\}}^{\overline{\mathcal{A}},-}=\sum_{\{S_i\}}\exp\left[J\sum_{\langle ij\rangle\notin\mathcal{A}}\eta_{ij}^{1}S_iS_j+i\frac{\pi}{2}\sum_{i\in\partial}k_i(S_i-1)\right],$$

we can carry out the summation over the even number of collective operations $\{\overline{\eta}^{(l)}\}_{l=1}^{n}$ explicitly and arrive at

$$\begin{aligned}\mathcal{Z}^{(P)}(n)=&\left(\frac{1}{Z_J^{\rm tot}(\mathbf{1})2^{m_{\mathcal{B}}-1}}\right)^n\frac{e^{\mathcal{N}_{\mathcal{A}}^{(p)}A_n}}{2^{\mathcal{N}_{\mathcal{A}}^{(p)}}}\frac{1}{2}\sum_{\{q_r=\pm 1\}_{r=1}^{m_{\mathcal{B}}}}\frac{1}{4^{\mathcal{N}_\partial}}\sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}\sum_{\{\theta_i\}}\exp\left(B_n\sum_{\langle ij\rangle\in\mathcal{A}}\theta_i\theta_j\right)\sum_{\{S_i^{(1)}\}}\exp\left[i\frac{\pi}{2}\sum_{r=1}^{m_{\mathcal{B}}}\sum_{i\in\partial_r}k_i(\theta_i+S_i^{(1)}-1-q_r)\right]\\&\times\frac{1}{2}\left\{\left[\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}S_i^{(1)}S_j^{(1)}\right)+\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}\eta_{ij}^{1}S_i^{(1)}S_j^{(1)}\right)\right](Z_{\{k_i\}}^{\overline{\mathcal{A}},+}+Z_{\{k_i\}}^{\overline{\mathcal{A}},-})^{n-1}\right.\\&\left.+\left[\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}S_i^{(1)}S_j^{(1)}\right)-\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}\eta_{ij}^{1}S_i^{(1)}S_j^{(1)}\right)\right](Z_{\{k_i\}}^{\overline{\mathcal{A}},+}-Z_{\{k_i\}}^{\overline{\mathcal{A}},-})^{n-1}\right\}.\end{aligned} \tag{D20}$$

We are finally in the position to take the derivative with respect to n and to compute the von Neumann entropy of the bipartition,

$$S_{\mathrm{vN}}^{(P)}(\mathcal{A};T/\lambda_B) = -\lim_{n\to 1}\partial_n \mathcal{Z}^{(P)}(n)$$

$$= -\lim_{n\to 1}\partial_n\left\{\left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1})2^{m_{\mathcal{B}}-1}}\right)^n \frac{e^{\mathcal{N}_{\mathcal{A}}^{(p)}A_n}}{2^{\mathcal{N}_{\mathcal{A}}^{(p)}}}\right\}\sum_{\{\theta_i\}}\exp\left(B_1\sum_{\langle ij\rangle\in\mathcal{A}}\theta_i\theta_j\right)$$

$$\times\sum_{\{S_i^{(1)}\}}\left[\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}S_i^{(1)}S_j^{(1)}\right)\right]\frac{1}{2}\sum_{\{q_r\}_{r=1}^{m_{\mathcal{B}}}}\frac{1}{4^{\mathcal{N}_\partial}}\sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}\exp\left[i\frac{\pi}{2}\sum_{r=1}^{m_{\mathcal{B}}}\sum_{i\in\partial_r}k_i(\theta_i+S_i^{(1)}-1-q_r)\right] \tag{D21a}$$

$$-\left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1})2^{m_{\mathcal{B}}-1}}\right)\frac{e^{\mathcal{N}_{\mathcal{A}}^{(p)}A_1}}{2^{\mathcal{N}_{\mathcal{A}}^{(p)}}}\lim_{n\to 1}\partial_n\left[\sum_{\{\theta_i\}}\exp\left(B_n\sum_{\langle ij\rangle\in\mathcal{A}}\theta_i\theta_j\right)\right]$$

$$\times\sum_{\{S_i^{(1)}\}}\left[\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}S_i^{(1)}S_j^{(1)}\right)\right]\frac{1}{2}\sum_{\{q_r\}_{r=1}^{m_{\mathcal{B}}}}\frac{1}{4^{\mathcal{N}_\partial}}\sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}\exp\left[i\frac{\pi}{2}\sum_{r=1}^{m_{\mathcal{B}}}\sum_{i\in\partial_r}k_i(\theta_i+S_i^{(1)}-1-q_r)\right] \tag{D21b}$$

$$-\left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1})2^{m_{\mathcal{B}}-1}}\right)\frac{e^{\mathcal{N}_{\mathcal{A}}^{(p)}A_1}}{2^{\mathcal{N}_{\mathcal{A}}^{(p)}}}\sum_{\{\theta_i\}}\exp\left(B_1\sum_{\langle ij\rangle\in\mathcal{A}}\theta_i\theta_j\right)$$

$$\times\sum_{\{S_i^{(1)}\}}\frac{1}{2}\sum_{\{q_r\}_{r=1}^{m_{\mathcal{B}}}}\frac{1}{4^{\mathcal{N}_\partial}}\sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}\exp\left[i\frac{\pi}{2}\sum_{r=1}^{m_{\mathcal{B}}}\sum_{i\in\partial_r}k_i(\theta_i+S_i^{(1)}-1-q_r)\right]$$

$$\times\frac{1}{2}\left\{\left[\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}S_i^{(1)}S_j^{(1)}\right)+\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}\eta_{ij}^1S_i^{(1)}S_j^{(1)}\right)\right]\ln(Z_{\{k_i\}}^{\overline{\mathcal{A}},+}+Z_{\{k_i\}}^{\overline{\mathcal{A}},-})\right.$$

$$\left.+\left[\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}S_i^{(1)}S_j^{(1)}\right)-\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}\eta_{ij}^1S_i^{(1)}S_j^{(1)}\right)\right]\ln(Z_{\{k_i\}}^{\overline{\mathcal{A}},+}-Z_{\{k_i\}}^{\overline{\mathcal{A}},-})\right\}. \tag{D21c}$$

The summation over $\{k_i\}$ can be carried out explicitly both in contributions (D21a) and (D21b). This leads to a delta function that identifies $\theta_i=q_rS_i^{(1)}$, $i\in\partial_r$, and $r=1,\ldots,m_{\mathcal{B}}$. One can verify that the factor q_r is actually immaterial, and the θ and $S^{(1)}$ terms in the above equation can be gathered into a single partition function,

$$\sum_{\{\theta_i\}}\exp\left(B_1\sum_{\langle ij\rangle\in\mathcal{A}}\theta_i\theta_j\right)\sum_{\{S_i^{(1)}\}}\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}S_i^{(1)}S_j^{(1)}\right)$$

$$=\sum_{\{S_i\}}\exp\left(J\sum_{\langle ij\rangle}S_iS_j\right)\equiv Z_J^{\mathrm{tot}}(\mathbf{1}), \tag{D22}$$

where we used the fact that $B_1=J$ [see Eqs. (D26) below]. The summation over $\{q_r=\pm 1\}_{r=1}^{m_{\mathcal{B}}}$ becomes then trivial, yielding an overall factor $2^{m_{\mathcal{B}}}$.

In contribution (D21c), each summation over $q_r=\pm 1$ yields a factor $2\cos(\pi\Sigma_{i\in\partial_r}k_i/2)$, which vanishes unless Σk_i is even. Thus, we can constrain the summation over $\{k_i=0,\ldots,3\}_{i\in\partial_r}$ to satisfy this condition, and we can drop the terms $\exp[i\frac{\pi}{2}\Sigma_{i\in\partial_r}k_i(1-q_r)]$ since $1-q_r$ is even and the term is identically one. The summation over $\{q_r=\pm 1\}_{r=1}^{m_{\mathcal{B}}}$ becomes again trivial. In particular,

$$\exp\left[i\frac{\pi}{2}\sum_{i\in\partial}k_i(\theta_i+S_i^{(1)})\right]$$

$$=\exp\left\{i\frac{\pi}{2}\sum_{i\in\partial}k_i[(\theta_i-1)+(S_i^{(1)}-1)]\right\} \tag{D23}$$

for the same reasoning, and we can write the θ and $S^{(1)}$ terms in a more compact form using the definition of $Z_{\{k_i\}}^{\overline{\mathcal{A}},\pm}$ and introducing the notation

$$Z_{\{k_i\}}^{\overline{\mathcal{B}},+}=\sum_{\{\theta_i\}}\exp\left[J\sum_{\langle ij\rangle\in\mathcal{A}}\theta_i\theta_j+i\frac{\pi}{2}\sum_{i\in\partial}k_i(\theta_i-1)\right]. \tag{D24}$$

(The labeling $\overline{\mathcal{B}}$ instead of $\mathcal{A}$ is used here as a reminder that the summation over $\{\theta_i\}$ includes both spins surrounded only by bonds in $\mathcal{A}$ and spins on the boundary ∂. Therefore, the total number of θ spins is $\mathcal{N}_{\overline{\mathcal{B}}}^{(c)}=\mathcal{N}_{\mathcal{A}}^{(c)}+\mathcal{N}_\partial$.)

These considerations allow us to simplify Eq. (D21) to

$$S_{\mathrm{vN}}^{(P)}(\mathcal{A};T/\lambda_B) = -\lim_{n\to 1}\partial_n\left\{\left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1})2^{m_{\mathcal{B}}-1}}\right)^{n-1}\frac{e^{\mathcal{N}_{\mathcal{A}}^{(p)}A_n}}{2^{\mathcal{N}_{\mathcal{A}}^{(p)}}}\right\} \tag{D25a}$$

$$-\left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1})}\right)\frac{e^{\mathcal{N}_{\mathcal{A}}^{(p)}A_1}}{2^{\mathcal{N}_{\mathcal{A}}^{(p)}}}\lim_{n\to 1}\partial_n\left[\sum_{\{S_i\}}\exp\left(B_n\sum_{\langle ij\rangle\in\mathcal{A}}S_iS_j + J\sum_{\langle ij\rangle\notin\mathcal{A}}S_iS_j\right)\right] \tag{D25b}$$

$$-\left(\frac{1}{Z_J^{\mathrm{tot}}(\mathbf{1})}\right)\frac{e^{\mathcal{N}_{\mathcal{A}}^{(p)}A_1}}{2^{\mathcal{N}_{\mathcal{A}}^{(p)}}}\frac{1}{4^{\mathcal{N}_\partial}}\sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}^{(\mathrm{even})}\frac{Z_{\{k_i\}}^{\mathcal{B},+}}{2}\{(Z_{\{k_i\}}^{\overline{\mathcal{A}},+}+Z_{\{k_i\}}^{\overline{\mathcal{A}},-})\ln(Z_{\{k_i\}}^{\overline{\mathcal{A}},+}+Z_{\{k_i\}}^{\overline{\mathcal{A}},-}) + (Z_{\{k_i\}}^{\overline{\mathcal{A}},+}-Z_{\{k_i\}}^{\overline{\mathcal{A}},-})\ln(Z_{\{k_i\}}^{\overline{\mathcal{A}},+}-Z_{\{k_i\}}^{\overline{\mathcal{A}},-})\}. \tag{D25c}$$

In order to proceed further, let us first study some of the terms in Eq. (D25) separately. From Eq. (D13) we have that

$$A_n = \frac{1}{2}\ln\{[(2\cosh J)^n + (2\sinh J)^n] \times [(2\cosh J)^n - (2\sinh J)^n]\} = \frac{1}{2}\ln\{(2\cosh J)^{2n} - (2\sinh J)^{2n}\}, \tag{D26a}$$

$$B_n = \frac{1}{2}\ln\frac{(2\cosh J)^n + (2\sinh J)^n}{(2\cosh J)^n - (2\sinh J)^n}, \tag{D26b}$$

$$A_1 = \ln 2, \tag{D26c}$$

$$B_1 = \frac{1}{2}\ln\frac{1+\tanh J}{1-\tanh J} = J, \tag{D26d}$$

$$\left.\frac{d}{dn}A_n\right|_{n=1} = \ln 2 + \cosh^2 J\,\ln(\cosh J) - \sinh^2 J\,\ln(\sinh J), \tag{D26e}$$

$$\left.\frac{d}{dn}B_n\right|_{n=1} = \sinh J\cosh J\,\ln\frac{\sinh J}{\cosh J}. \tag{D26f}$$

Notice that $\frac{d}{dn}A_n|_{n=1}\to\ln 2$ for $J\to 0$, $\frac{d}{dn}A_n|_{n=1}\sim J+1/2+\mathcal{O}(e^{-2J})$ for $J\to\infty$ and that $\frac{d}{dn}B_n|_{n=1}\to 0$ for $J\to 0$, $\frac{d}{dn}B_n|_{n=1}\to -1/2+\mathcal{O}(e^{-2J})$ for $J\to\infty$.

We can also carry out the derivative in Eq. (D25),

$$\lim_{n\to 1}\partial_n\left[\sum_{\{S_i\}}\exp\left(B_n\sum_{\langle ij\rangle\in\mathcal{A}}S_iS_j + J\sum_{\langle ij\rangle\notin\mathcal{A}}S_iS_j\right)\right] = \left.\frac{d}{dn}B_n\right|_{n=1}\sum_{\{S_i\}}\left(\sum_{\langle ij\rangle\in\mathcal{A}}S_iS_j\right)\exp\left(B_1\sum_{\langle ij\rangle\in\mathcal{A}}S_iS_j + J\sum_{\langle ij\rangle\notin\mathcal{A}}S_iS_j\right)$$

$$= \sinh J\cosh J\,\ln\frac{\sinh J}{\cosh J}\sum_{\{S_i\}}\left(\sum_{\langle ij\rangle\in\mathcal{A}}S_iS_j\right)\exp\left(J\sum_{\langle ij\rangle}S_iS_j\right)$$

$$= \sinh J\cosh J\,\ln\frac{\sinh J}{\cosh J}\langle E_{\mathcal{A}}\rangle_{Z_J^{\mathrm{tot}}(\mathbf{1})}Z_J^{\mathrm{tot}}(\mathbf{1}), \tag{D27a}$$

$$\langle E_{\mathcal{A}}\rangle_{Z_J^{\mathrm{tot}}(\mathbf{1})} \equiv \frac{\sum_{\{S_i\}}(\sum_{\langle ij\rangle\in\mathcal{A}}S_iS_j)\exp(J\sum_{\langle ij\rangle}S_iS_j)}{Z_J^{\mathrm{tot}}(\mathbf{1})}, \tag{D27b}$$

where $E_{\mathcal{A}}$ is the extensive energy of the bonds in $\mathcal{A}$ (in units of J) in the Ising model described by the equilibrium partition function $Z_J^{\mathrm{tot}}(\mathbf{1})$. The last calculation we still need is

$$\left.\frac{d}{dn}e^{\mathcal{N}_{\mathcal{A}}^{(p)}A_n}\right|_{n=1} = \mathcal{N}_{\mathcal{A}}^{(p)}2^{\mathcal{N}_{\mathcal{A}}^{(p)}}[\ln 2 + \cosh^2 J\,\ln(\cosh J) - \sinh^2 J\,\ln(\sinh J)]. \tag{D28}$$

Combining all the results in Eqs. (D26)–(D28), Eq. (D25) reduces to

$$S_{\mathrm{vN}}^{(P)}(\mathcal{A};T/\lambda_B) = \ln(2^{m_{\mathcal{B}}-1}) + \ln Z_J^{\mathrm{tot}}(\mathbf{1}) - \mathcal{N}_{\mathcal{A}}^{(p)}[\ln 2 + \cosh^2 J\,\ln(\cosh J) - \sinh^2 J\,\ln(\sinh J)] \tag{D29a}$$

$$-\sinh J\cosh J\,\ln\frac{\sinh J}{\cosh J}\langle E_{\mathcal{A}}\rangle_{Z_J^{\mathrm{tot}}(\mathbf{1})} \tag{D29b}$$

$$-\frac{1}{Z_J^{\rm tot}(\mathbf{1})}\frac{1}{4^{\mathcal{N}_\partial}}\sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}^{(\rm even)}\frac{Z^{\mathcal{B},+}_{\{k_i\}}}{2}[(Z^{\overline{\mathcal{A}},+}_{\{k_i\}}+Z^{\overline{\mathcal{A}},-}_{\{k_i\}})\ln(Z^{\overline{\mathcal{A}},+}_{\{k_i\}}+Z^{\overline{\mathcal{A}},-}_{\{k_i\}})+(Z^{\overline{\mathcal{A}},+}_{\{k_i\}}-Z^{\overline{\mathcal{A}},-}_{\{k_i\}})\ln(Z^{\overline{\mathcal{A}},+}_{\{k_i\}}-Z^{\overline{\mathcal{A}},-}_{\{k_i\}})]. \tag{D29c}$$

Recall that Σk_i is even and therefore $\Sigma k_i S_i$ is also even, irrespective of the values of the spins $\{S_i=\pm 1\}$. In particular,

$$\begin{aligned}&\exp\left[i\frac{\pi}{2}\sum_{i\in\partial}k_i(S_i-1)\right]\\&=\prod_{i\in\partial}\left[e^{-i(\pi/2)k_i}\cos\frac{\pi}{2}k_i+ie^{-i(\pi/2)k_i}\left(\sin\frac{\pi}{2}k_i\right)S_i\right]\\&=\prod_{i\in\partial}[\delta_{k_i\ \rm even}+S_i\delta_{k_i\ \rm odd}],\end{aligned} \tag{D30}$$

and both $Z^{\mathcal{A},+}_{\{k_i\}}$ and $Z^{\overline{\mathcal{A}},+}_{\{k_i\}}$ can be rewritten as

$$\begin{aligned}Z^{\mathcal{A},+}_{\{k_i\}}&=\sum_{\{S_i\}}\left(\prod_{i\in\partial}^{k_i\rm odd}S_i\right)\exp\left(J\sum_{\langle ij\rangle\in\mathcal{A}}S_iS_j\right)\\&=Z^{\mathcal{A},+}\langle\prod_{i\in\partial}^{k_i\rm odd}S_i\rangle,\end{aligned} \tag{D31}$$

$$\begin{aligned}Z^{\overline{\mathcal{A}},+}_{\{k_i\}}&=\sum_{\{S_i\}}\left(\prod_{i\in\partial}^{k_i\rm odd}S_i\right)\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}S_iS_j\right)\\&=Z^{\overline{\mathcal{A}},+}\langle\prod_{i\in\partial}^{k_i\rm odd}S_i\rangle,\end{aligned} \tag{D32}$$

where $Z^{\mathcal{A},+}=\Sigma_{\{S_i\}}\exp(J\Sigma_{\langle ij\rangle\in\mathcal{A}}S_iS_j)$ and $Z^{\overline{\mathcal{A}},+}=\Sigma_{\{S_i\}}\exp(J\Sigma_{\langle ij\rangle\notin\mathcal{A}}S_iS_j)$; similarly for $Z^{\mathcal{A},-}_{\{k_i\}}$ and $Z^{\overline{\mathcal{A}},-}_{\{k_i\}}$. Thus, all these quantities can be interpreted as correlation functions of boundary spins located at the odd entries of the set $\{k_i\}$ times a partition function. Note that the constraint $\Sigma_{i\in\partial_r}k_i$ even, $\forall r$, requires that the number of such odd entries is also even separately on each boundary component $r=1,\ldots,m_{\mathcal{B}}$.

If we are interested in computing the topological entropy of the system, it is convenient to decompose the last term in Eq. (D29) so that

$$S^{(P)}_{\rm vN}(\mathcal{A};T/\lambda_B)=\ln(2^{m_{\mathcal{B}}-1})+\ln Z_J^{\rm tot}(\mathbf{1})-\mathcal{N}^{(p)}_{\mathcal{A}}[\ln 2+\cosh^2 J\ln(\cosh J)-\sinh^2 J\ln(\sinh J)] \tag{D33a}$$

$$-\sinh J\cosh J\ln\frac{\sinh J}{\cosh J}\langle E_{\mathcal{A}}\rangle_{Z_J^{\rm tot}(\mathbf{1})} \tag{D33b}$$

$$-\frac{1}{4^{\mathcal{N}_\partial}}\sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}^{(\rm even)}\frac{Z^{\mathcal{B},+}_{\{k_i\}}Z^{\overline{\mathcal{A}},+}_{\{k_i\}}}{Z_J^{\rm tot}(\mathbf{1})}\ln Z^{\overline{\mathcal{A}},+}_{\{k_i\}} \tag{D33c}$$

$$-\frac{1}{4^{\mathcal{N}_\partial}}\sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}^{(\rm even)}\frac{Z^{\mathcal{B},+}_{\{k_i\}}Z^{\overline{\mathcal{A}},+}_{\{k_i\}}}{Z_J^{\rm tot}(\mathbf{1})}\frac{1}{2}\left[\left(1+\frac{Z^{\overline{\mathcal{A}},-}_{\{k_i\}}}{Z^{\overline{\mathcal{A}},+}_{\{k_i\}}}\right)\ln\left(1+\frac{Z^{\overline{\mathcal{A}},-}_{\{k_i\}}}{Z^{\overline{\mathcal{A}},+}_{\{k_i\}}}\right)+\left(1-\frac{Z^{\overline{\mathcal{A}},-}_{\{k_i\}}}{Z^{\overline{\mathcal{A}},+}_{\{k_i\}}}\right)\ln\left(1-\frac{Z^{\overline{\mathcal{A}},-}_{\{k_i\}}}{Z^{\overline{\mathcal{A}},+}_{\{k_i\}}}\right)\right]. \tag{D33d}$$

The result in Eqs. (D33a) holds for $n_{\mathcal{A}}=1$ (i.e., there is only one collective operation in $\mathcal{A}$). In order to compute the topological entropy of the system with the bipartition scheme in Sec. III, we also need to consider the case where $n_{\mathcal{A}}=0$. Repeating the derivation above, from Eq. (D19) to Eq. (D33), in the absence of collective operations leads rather straightforwardly to the result that

$$\begin{aligned}\mathcal{Z}^{(P)}(n)=&\left(\frac{1}{Z_J^{\rm tot}(\mathbf{1})2^{m_{\mathcal{B}}-1}}\right)^n\frac{e^{\mathcal{N}^{(p)}_{\mathcal{A}}A_n}}{2^{\mathcal{N}^{(p)}_{\mathcal{A}}}}\frac{1}{2}\sum_{\{q_r\}_{r=1}^{m_{\mathcal{B}}}}\frac{1}{4^{\mathcal{N}_\partial}}\sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}\sum_{\{\theta_i\}}\exp\left(B_n\sum_{\langle ij\rangle\in\mathcal{A}}\theta_i\theta_j\right)\\&\times\sum_{\{S_i^{(1)}\}}\exp\left[i\frac{\pi}{2}\sum_{r=1}^{m_{\mathcal{B}}}\sum_{i\in\partial_r}k_i(\theta_i+S_i^{(1)}-1-q_r)\right]\left[\exp\left(J\sum_{\langle ij\rangle\notin\mathcal{A}}S_i^{(1)}S_j^{(1)}\right)\right](Z^{\overline{\mathcal{A}},+}_{\{k_i\}})^{n-1},\end{aligned} \tag{D34}$$

and

$$S_{\mathrm{vN}}^{(P)}(\mathcal{A};T/\lambda_B) = \ln(2^{m_{\mathcal{B}}-1}) + \ln Z_J^{\mathrm{tot}}(\mathbf{1}) - \mathcal{N}_{\mathcal{A}}^{(p)}[\ln 2 + \cosh^2 J \ln(\cosh J) - \sinh^2 J \ln(\sinh J)] \tag{D35a}$$

$$- \sinh J \cosh J \ln\frac{\sinh J}{\cosh J}\langle E_{\mathcal{A}}\rangle_{Z_J^{\mathrm{tot}}(\mathbf{1})} \tag{D35b}$$

$$- \frac{1}{4^{\mathcal{N}_\partial}} \sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}^{(\mathrm{even})} \frac{Z_{\{k_i\}}^{\mathcal{B},+} Z_{\{k_i\}}^{\overline{\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{\overline{\mathcal{A}},+}. \tag{D35c}$$

Notice that Eq. (D35) differs from Eq. (D33) only in that it lacks contribution (D33d).

We can finally compute the plaquette contribution to the topological entropy $S_{\mathrm{topo}}^{(P)}(T/\lambda_B)$ using the full bipartition scheme. All the terms that do not carry a topological contribution cancel. Namely, as discussed in Sec. III,

$$\mathcal{N}_{1\mathcal{A}}^{(p)} + \mathcal{N}_{4\mathcal{A}}^{(p)} = \mathcal{N}_{2\mathcal{A}}^{(p)} + \mathcal{N}_{3\mathcal{A}}^{(p)} \tag{D36a}$$

and on similar grounds

$$\langle E_{1\mathcal{A}}\rangle_{Z_J^{\mathrm{tot}}(\mathbf{1})} + \langle E_{4\mathcal{A}}\rangle_{Z_J^{\mathrm{tot}}(\mathbf{1})} = \langle E_{2\mathcal{A}}\rangle_{Z_J^{\mathrm{tot}}(\mathbf{1})} + \langle E_{3\mathcal{A}}\rangle_{Z_J^{\mathrm{tot}}(\mathbf{1})}, \tag{D36b}$$

likewise for bipartitions 5–8. Recall also that $m_{\mathcal{B}}=1$ and $n_{\mathcal{A}}=0$ for all bipartitions, except bipartitions 4 and 5 (which have $m_{\mathcal{B}}=1$ and $n_{\mathcal{A}}=1$) and bipartition 1 (which has $m_{\mathcal{B}}=2$ and $n_{\mathcal{A}}=0$). Using Eqs. (D33) and (D35) accordingly, we obtain

$$\begin{aligned} S_{\mathrm{topo}}^{(P)}(T/\lambda_B) = {} & \ln(2^{-m_{1\mathcal{B}}+m_{2\mathcal{B}}+m_{3\mathcal{B}}-m_{4\mathcal{B}}}) \\ & + \frac{1}{4^{\mathcal{N}_\partial}} \sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}^{(\mathrm{even})} \left\{ \frac{Z_{\{k_i\}}^{\overline{1\mathcal{B}},+} Z_{\{k_i\}}^{\overline{1\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{1,\mathcal{A},+} - \frac{Z_{\{k_i\}}^{\overline{2\mathcal{B}},+} Z_{\{k_i\}}^{\overline{2\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{2,\mathcal{A},+} - \frac{Z_{\{k_i\}}^{\overline{3\mathcal{B}},+} Z_{\{k_i\}}^{\overline{3\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{3,\mathcal{A},+} + \frac{Z_{\{k_i\}}^{\overline{4\mathcal{B}},+} Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{\overline{4\mathcal{A}},+} \right\} \\ & + \frac{1}{4^{\mathcal{N}_\partial}} \sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}^{(\mathrm{even})} \frac{Z_{\{k_i\}}^{\overline{4\mathcal{B}},+} Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \frac{1}{2} \left[\left(1 + \frac{Z_{\{k_i\}}^{\overline{4\mathcal{A}},-}}{Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}\right) \ln\left(1 + \frac{Z_{\{k_i\}}^{\overline{4\mathcal{A}},-}}{Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}\right) + \left(1 - \frac{Z_{\{k_i\}}^{\overline{4\mathcal{A}},-}}{Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}\right) \ln\left(1 - \frac{Z_{\{k_i\}}^{\overline{4\mathcal{A}},-}}{Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}\right) \right] \\ & + (\text{partitions } 5-8). \end{aligned} \tag{D37}$$

Using the fact that $m_{1\mathcal{B}}-m_{2\mathcal{B}}-m_{3\mathcal{B}}+m_{4\mathcal{B}}=1$, that $m_{5\mathcal{B}}-m_{6\mathcal{B}}-m_{7\mathcal{B}}+m_{8\mathcal{B}}=0$, and that $Z_{\{k_i\}}^{\overline{4\mathcal{A}},\pm}(J) \equiv Z_{\{k_i\}}^{\overline{5\mathcal{A}},\pm}(J)$ since bipartitions 4 and 5 are in fact identical, one arrives to the result

$$\begin{aligned} S_{\mathrm{topo}}^{(P)}(T/\lambda_B) = {} & -\ln 2 \\ & + \frac{1}{4^{\mathcal{N}_\partial}} \sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}^{(\mathrm{even})} \left\{ \frac{Z_{\{k_i\}}^{\overline{1\mathcal{B}},+} Z_{\{k_i\}}^{\overline{1\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{\overline{1\mathcal{A}},+} - \frac{Z_{\{k_i\}}^{\overline{2\mathcal{B}},+} Z_{\{k_i\}}^{\overline{2\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{\overline{2\mathcal{A}},+} - \frac{Z_{\{k_i\}}^{\overline{3\mathcal{B}},+} Z_{\{k_i\}}^{\overline{3\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{\overline{3\mathcal{A}},+} + \frac{Z_{\{k_i\}}^{\overline{4\mathcal{B}},+} Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{\overline{4\mathcal{A}},+} \right\} \\ & + \frac{1}{4^{\mathcal{N}_\partial}} \sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}^{(\mathrm{even})} \left\{ \frac{Z_{\{k_i\}}^{\overline{5\mathcal{B}},+} Z_{\{k_i\}}^{\overline{5\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{\overline{5\mathcal{A}},+} - \frac{Z_{\{k_i\}}^{\overline{6\mathcal{B}},+} Z_{\{k_i\}}^{\overline{6\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{\overline{6\mathcal{A}},+} - \frac{Z_{\{k_i\}}^{\overline{7\mathcal{B}},+} Z_{\{k_i\}}^{\overline{7\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{\overline{7\mathcal{A}},+} + \frac{Z_{\{k_i\}}^{\overline{8\mathcal{B}},+} Z_{\{k_i\}}^{\overline{8\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \ln Z_{\{k_i\}}^{\overline{8\mathcal{A}},+} \right\} \\ & + \frac{1}{4^{\mathcal{N}_\partial}} \sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}^{(\mathrm{even})} \frac{Z_{\{k_i\}}^{\overline{4\mathcal{B}},+} Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}{Z_J^{\mathrm{tot}}(\mathbf{1})} \left[\left(1 + \frac{Z_{\{k_i\}}^{\overline{4\mathcal{A}},-}}{Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}\right) \ln\left(1 + \frac{Z_{\{k_i\}}^{\overline{4\mathcal{A}},-}}{Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}\right) + \left(1 - \frac{Z_{\{k_i\}}^{\overline{4\mathcal{A}},-}}{Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}\right) \ln\left(1 - \frac{Z_{\{k_i\}}^{\overline{4\mathcal{A}},-}}{Z_{\{k_i\}}^{\overline{4\mathcal{A}},+}}\right) \right]. \end{aligned} \tag{D38}$$

This expression can be cast in a more useful way by noticing the following. Factors like

$$\mathcal{P}^p_{\{k_i\}} \equiv \frac{1}{4^{\mathcal{N}_{\partial^p}}} \frac{Z^{\overline{p\mathcal{B}},+}_{\{k_i\}} Z^{\overline{p\mathcal{A}},+}_{\{k_i\}}}{Z^{\mathrm{tot}}_J(\mathbf{1})} = \frac{1}{4^{\mathcal{N}_{\partial^p}}} \frac{Z^{\overline{p\mathcal{B}},+} Z^{\overline{p\mathcal{A}},+}}{Z^{\mathrm{tot}}_J(\mathbf{1})} \left\langle \prod_{i\in\partial^p}^{k_i \text{ odd}} \theta_i \right\rangle \left\langle \prod_{i\in\partial^p}^{k_i \text{ odd}} S_i \right\rangle \quad \text{(D39)}$$

are greater than or equal to zero, for each of the partitions $p=1-8$. This is because the expectation values of the products of spins are always non-negative when the interactions are *ferromagnetic* (this can be shown explicitly in a high-temperature expansion, for example). Recall that the set $\{k_i\}$ contains always an even number of odd k_i's.

Moreover, one can check that

$$\begin{aligned}
\sum_{\{k_i\}_{i=1}^{\mathcal{N}_{\partial^p}}}^{(\text{even})} \mathcal{P}^p_{\{k_i\}} &= \frac{1}{Z^{\mathrm{tot}}_J(\mathbf{1})} \sum_{\{\theta_i\}} \sum_{\{S_i\}} \exp\left(J \sum_{\langle ij\rangle \in p\mathcal{A}} \theta_i\theta_j\right) \exp\left(J \sum_{\langle ij\rangle \notin p\mathcal{A}} S_iS_j\right) \frac{1}{4^{\mathcal{N}_{\partial^p}}} \sum_{\{k_i\}_{i=1}^{\mathcal{N}_{\partial^p}}}^{(\text{even})} \exp\left[i\frac{\pi}{2} \sum_{i\in\partial^p} k_i(\theta_i + S_i - 2)\right] \\
&= \frac{1}{Z^{\mathrm{tot}}_J(\mathbf{1})} \sum_{\{\theta_i\}} \sum_{\{S_i\}} \exp\left(J \sum_{\langle ij\rangle \in p\mathcal{A}} \theta_i\theta_j\right) \exp\left(J \sum_{\langle ij\rangle \notin p\mathcal{A}} S_iS_j\right) \frac{1}{4^{\mathcal{N}_{\partial^p}}} \sum_{\{k_i\}_{i=1}^{\mathcal{N}_{\partial^p}}}^{(\text{even})} \exp\left[i\frac{\pi}{2} \sum_{i\in\partial^p} k_i(\theta_i + S_i - 1)\right] \\
&\quad \times \left[\cos\left(\frac{\pi}{2} \sum_{i\in\partial^p} k_i\right) - i \sin\left(\frac{\pi}{2} \sum_{i\in\partial^p} k_i\right)\right] \\
&= \frac{1}{Z^{\mathrm{tot}}_J(\mathbf{1})} \sum_{\{\theta_i\}} \sum_{\{S_i\}} \exp\left(J \sum_{\langle ij\rangle \in p\mathcal{A}} \theta_i\theta_j\right) \exp\left(J \sum_{\langle ij\rangle \notin p\mathcal{A}} S_iS_j\right) \frac{1}{2} \sum_{q=\pm 1} \frac{1}{4^{\mathcal{N}_{\partial^p}}} \sum_{\{k_i\}_{i=1}^{\mathcal{N}_{\partial^p}}} \exp\left[i\frac{\pi}{2} \sum_{i\in\partial^p} k_i(\theta_i + S_i - 1 - q)\right] \\
&= \frac{1}{Z^{\mathrm{tot}}_J(\mathbf{1})} \sum_{\{\theta_i\}} \sum_{\{S_i\}} \exp\left(J \sum_{\langle ij\rangle \in \mathcal{A}} \theta_i\theta_j\right) \exp\left(J \sum_{\langle ij\rangle \notin \mathcal{A}} S_iS_j\right) \frac{1}{2} \sum_{q=\pm 1} \delta(\theta_i S_i = q) \\
&= \frac{1}{Z^{\mathrm{tot}}_J(\mathbf{1})} \times Z^{\mathrm{tot}}_J(\mathbf{1}) = 1, \qquad \text{(D40)}
\end{aligned}$$

and thus the $\mathcal{P}^p_{\{k_i\}} \geq 0$ are probability weights. Similarly, we can define a probability

$$\mathcal{P}_{\{k_i\}} = (\mathcal{P}^1\mathcal{P}^4\mathcal{P}^5\mathcal{P}^8)_{\{k_i\}} \qquad \text{(D41)}$$

$$= (\mathcal{P}^2\mathcal{P}^3\mathcal{P}^6\mathcal{P}^7)_{\{k_i\}} \geq 0, \qquad \text{(D42)}$$

where the $\{k_i\}$ are defined on the total boundary of the added partitions, and we used the fact that partitions 1, 4, 5, 8 and 2, 3, 6, 7 have exactly the same total boundary. We can then define averages with respect to this measure,

$$\langle \cdots \rangle_{\{k_i\}} \equiv \sum_{\{k_i\}_{i=1}^{\mathcal{N}_\partial}}^{(\text{even})} \mathcal{P}_{\{k_i\}}(\cdots), \qquad \text{(D43)}$$

and Eq. (D38) reduces to

$$S^{(P)}_{\mathrm{topo}}(T/\lambda_B) = -\ln 2 + \left\langle \ln\left(\frac{Z^{\overline{1\mathcal{A}},+}_{\{k_i\}} Z^{\overline{4\mathcal{A}},+}_{\{k_i\}} Z^{\overline{5\mathcal{A}},+}_{\{k_i\}} Z^{\overline{8\mathcal{A}},+}_{\{k_i\}}}{Z^{\overline{2\mathcal{A}},+}_{\{k_i\}} Z^{\overline{3\mathcal{A}},+}_{\{k_i\}} Z^{\overline{6\mathcal{A}},+}_{\{k_i\}} Z^{\overline{7\mathcal{A}},+}_{\{k_i\}}} \right) \right\rangle_{\{k_i\}} \qquad \text{(D44a)}$$

$$+ \left\langle \left(1 + \frac{Z^{\overline{4\mathcal{A}},-}_{\{k_i\}}}{Z^{\overline{4\mathcal{A}},+}_{\{k_i\}}}\right) \ln\left(1 + \frac{Z^{\overline{4\mathcal{A}},-}_{\{k_i\}}}{Z^{\overline{4\mathcal{A}},+}_{\{k_i\}}}\right) + \left(1 - \frac{Z^{\overline{4\mathcal{A}},-}_{\{k_i\}}}{Z^{\overline{4\mathcal{A}},+}_{\{k_i\}}}\right) \ln\left(1 - \frac{Z^{\overline{4\mathcal{A}},-}_{\{k_i\}}}{Z^{\overline{4\mathcal{A}},+}_{\{k_i\}}}\right) \right\rangle_{\{k_i\}}. \qquad \text{(D44b)}$$

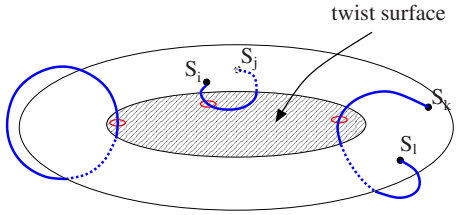

FIG. 8. (Color online) Qualitative examples of terms in the loop expansion that appear with different signs in $Z^{\overline{A},-}_{\{k_i\}}$ and $Z^{\overline{A},+}_{\{k_i\}}$: closed loops that wind around the donut shape and open strings that connect boundary spins S_i (which appear in the high-temperature expansion whenever the corresponding k_i is odd).

We can finally analyze this expression as a function of temperature. Recall that $J=-(1/2)\ln[\tanh(\beta\lambda_B)]$, so that $J\to 0$ when $T\to 0$, and the disordered Ising phase occurs for $T<T_c\simeq 1.313\ 346(3)\lambda_B$. Below the Ising transition at $J=J_c\simeq 0.221\ 654\ 4(3)$, one can use a high-temperature loop expansion to estimate the ratio of $Z^{\overline{A},-}_{\{k_i\}}$ over $Z^{\overline{A},+}_{\{k_i\}}$.

The high-temperature expansion contains either closed loops or open strings that terminate at the boundary because an S_i is inserted for each site i where k_i is odd. The corresponding expansions for $Z^{\overline{A},-}_{\{k_i\}}$ over $Z^{\overline{A},+}_{\{k_i\}}$ differ only by loop terms that intersect the twist surface (generated by the collective operation in Fig. 5, bottom) an odd number of times. These terms appear indeed with opposite sign in the two expansions. This can be achieved only by closed loops that wind around the donut shape and by open strings that connect boundary spins S_i among those identified by the set of odd k_i's (see Fig. 8).

In the high-temperature limit, long loops are exponentially suppressed and we can safely neglect the winding loop contributions when the size of the partition is taken to infinity. Similarly, out of all possible ways of connecting boundary spins in the k_i odd set, only "short" strings between spins "close" to the twist surface need be considered, as illustrated in Fig. 9.

For k_i points near the twist surface, rearranging the way that points are paired does not change the parity of the number of crossings of the twist surface. This is illustrated in Fig. 9, where reconnecting spins 5–8 via the dashed lines instead of the solid lines give 0 instead of 2 crossings, thus not changing the parity. Now, a reconnection that changes the parity involves drawing long strings. Below the Ising transition, the probability $\mathcal{P}_{\{k_i\}}$ keeps the points with odd k_i confined in pairs; thus there are ways to connect them together with short strings. But changing the parity of the intersections requires rematching them in such a way that connections with sites far away are made, and the total length of these strings is of order the system size. This is illustrated in Fig. 9: for example, reconnecting spins 1–4 requires strings whose total length spans the system size.

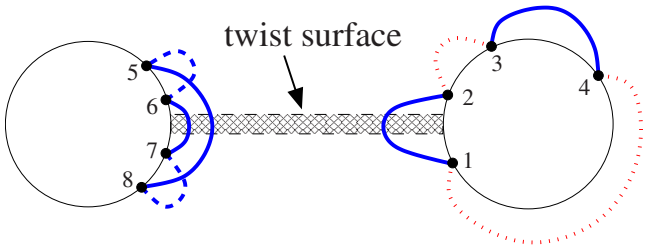

FIG. 9. (Color online) Schematic projected illustration of open strings between boundary spins. The locations of the spins 1–8 are given by the sites where k_i is odd (recall that their total number must be even). One can verify that the parity of the number of intersections with the twist surface is fixed by the choice of the locations 1–8, up to exponentially small corrections such as the red dotted string in the figure, which vanish in the thermodynamic limit of $N_\partial\to\infty$. For example, consider the change upon reconnecting spin 5–8 via the dashed lines instead of the solid lines. (Notice that the case where, say, the points 1–4 are uniformly distributed on the boundary is exponentially suppressed by the probability $\mathcal{P}_{\{k_i\}}$.)

Therefore, one can verify that all the loop terms corresponding to a given choice of k_i's have the same parity in the number of intersections to the twist surface up to corrections that are exponentially small in the size of the bipartition. As a result, the ratio $Z^{\overline{A},-}_{\{k_i\}}/Z^{\overline{A},+}_{\{k_i\}}$ tends to ± 1 in the thermodynamic limit of $N_\partial\to\infty$, and the sign is purely determined by the choice of k_i.

Equation (D44b) is clearly symmetric under the change $Z^{\overline{A},-}_{\{k_i\}}/Z^{\overline{A},+}_{\{k_i\}}\to -Z^{\overline{A},-}_{\{k_i\}}/Z^{\overline{A},+}_{\{k_i\}}$, and we finally arrive at the result that at low temperature $T<T_c$, the term in Eq. (D44b) gives2 ln 2.

In the Ising ordered phase ($T>T_c$ here), on the other hand, the ratio $Z^{\overline{A},-}/Z^{\overline{A},+}\to 0$ in the thermodynamic limit because of the energy cost associated with the twist in boundary condition (domain wall) in the "–" partition. Hence, in this case the term in Eq. (D44b) gives 0.

A similar reasoning gives that the ratios entering Eq. (D44a) are equal to 1 in the thermodynamic limit, and corrections appear only as the correlation length becomes of the order of the size of the bipartitions, i.e., infinite in the thermodynamic limit. Thus, in the low-temperature phase, Eq. (D44a) gives $\ln 1=0$ for $T<T_c$.

On the other hand, for $T>T_c$, the partitions order ferromagnetically, and one must account for the fact that partition $1\overline{\mathcal{A}}$ has two disconnected components; therefore these two components can order in two ways relative to one another, giving a factor of 2 in the ratio appearing in Eq. (D44a), and hence this term gives a contributionln 2.

Putting it all together, we obtain that

$$S^{(P)}_{\text{topo}}(T/\lambda_B)=\begin{cases}\ln 2, & T<T_c\\ 0, & T>T_c,\end{cases} \tag{D45}$$

and $\Delta S^{(P)}_{\text{topo}}(T/\lambda_B)=S^{(P)}_{\text{topo}}(T/\lambda_B)-S^{(P)}_{\text{topo}}(0)$ is given by Eq. (4.32).

[1] F. D. M. Haldane and E. H. Rezayi, Phys. Rev. B **31**, 2529 (1985); X.-G. Wen and Q. Niu, *ibid.* **41**, 9377 (1990); X.-G. Wen, Int. J. Mod. Phys. B **4**, 239 (1990); Adv. Phys. **44**, 405 (1995); Phys. Rev. B **65**, 165113 (2002).
[2] M. Levin and X.-G. Wen, Phys. Rev. Lett. **96**, 110405 (2006).
[3] A. Y. Kitaev and J. Preskill, Phys. Rev. Lett. **96**, 110404 (2006).
[4] C. Castelnovo and C. Chamon, Phys. Rev. B **76**, 184442 (2007).
[5] A. Y. Kitaev, Ann. Phys. (N.Y.) **303**, 2 (2003).
[6] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, J. Math. Phys. **43**, 4452 (2002).
[7] Z. Nussinov and G. Ortiz, arXiv:cond-mat/0605316 (unpublished); arXiv:cond-mat/0702377 (unpublished).
[8] A. Hamma, P. Zanardi, and X.-G. Wen, Phys. Rev. B **72**, 035307 (2005).
[9] Note that this decomposition is visible only at finite temperature, where the full Hamiltonian enters the calculations for the von Neumann entropy of the system via the density matrix. At zero temperature, the two contributions do remain distinct, but they cannot be told apart as there is no explicit dependence on $\lambda_{\mathcal{A}}$ and $\lambda_{\mathcal{B}}$ in the GS wave function.
[10] Our conclusions, from the topological entropy, are in disagreement with the ones obtained by Z. Nussinov and G. Ortiz in Ref. 18. While the authors discuss both phase transitions in the model, at $T=0$ and at finite temperature, they argue that only the former has a topological nature, and they indeed conclude that topological order is fragile at finite temperature. As explained in Sec. IV, this discrepancy is due to the fact that the authors consider winding loop operators as (nonlocal) order parameters, which vanish intrinsically at any finite temperature and cannot be used (at least in a naive way) to investigate the robustness of topological order to thermal fluctuations.
[11] E. Knill, R. Laflamme, H. Barnum, D. Dalvit, J. Dziarmaga, J. Gubernatis, L. Gurvits, G. Ortiz, L. Viola, and W. H. Zurek, Los Alamos Sci. **27**, 2 (2002).
[12] C. Castelnovo and C. Chamon, Phys. Rev. B **76**, 174416 (2007).
[13] The entanglement entropy is invariant under a local spin rotation.
[14] A. Hamma, R. Ionicioiu, and P. Zanardi, Phys. Rev. A **71**, 022315 (2005).
[15] Alternatively, one could replace the von Neumann entropy with its symmetrized version—the mutual information (entropy)—as proposed in Ref. 12.
[16] F. Wegner, J. Math. Phys. **12**, 2259 (1971).
[17] J. B. Kogut, Rev. Mod. Phys. **51**, 659 (1979).
[18] Z. Nussinov and G. Ortiz, Phys. Rev. B **77**, 064302 (2008).
[19] R. Savit, Rev. Mod. Phys. **52**, 453 (1980).
[20] M. Caselle, M. Hasenbusch, and M. Panero, J. High Energy Phys. 2003, 057.
[21] The familiar reader may have noticed that the construction of Z_J^{tot} is based on the well-known duality between the 3D Ising model and the $\mathbb{Z}_2$ Ising gauge theory in three dimensions, discussed, for example, in Refs. 17 and 19.
[22] Note that the classical model at finite T that obtains by setting $\lambda_A=0$ is nothing but a classical $\mathbb{Z}_2$ gauge theory in three dimensions. Therefore, our results show that the topological entropy of this classical system behaves as a proper (nonlocal) order parameter that captures its finite-temperature phase transition.
[23] We thank John Cardy for pointing us in the direction of this replica trick to handle the delta function terms in Eq. (D15).