Editors' Suggestion

Scanning tunneling spectroscopy of Majorana zero modes in a Kitaev spin liquid

Tim Bauer,[1,2] Lucas R. D. Freitas,[2,1] Rodrigo G. Pereira,[2] and Reinhold Egger[1]
[1]*Institut für Theoretische Physik, Heinrich-Heine-Universität, D-40225 Düsseldorf, Germany*
[2]*International Institute of Physics and Departamento de Física Teórica e Experimental, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte 59078-970, Brazil*

(Received 25 September 2022; accepted 8 February 2023; published 22 February 2023)

We describe scanning tunneling spectroscopic signatures of Majorana zero modes (MZMs) in Kitaev spin liquids. The tunnel conductance is determined by the dynamical spin correlations of the spin liquid, which we compute exactly, and by spin-anisotropic cotunneling form factors. Near a $\mathbb{Z}_2$ vortex, the tunnel conductance has a staircase voltage dependence, where conductance steps arise from MZMs and (at higher voltages) from additional vortex configurations. By scanning the probe tip position, one can detect the vortex locations. Our analysis suggests that topological magnon bound states near defects or magnetic impurities generate spectroscopic signatures that are qualitatively different from those of MZMs.

DOI: 10.1103/PhysRevB.107.054432

I. INTRODUCTION

Presently a major goal in condensed matter physics is to realize, detect, and manipulate topologically ordered phases of frustrated quantum magnets, commonly referred to as quantum spin liquids (QSLs). A famous exactly solvable paradigm is given by Kitaev's two-dimensional (2D) honeycomb lattice spin model with bond-dependent anisotropic exchange which, in a magnetic field, describes a gapped non-Abelian chiral QSL [1]. Emergent excitations of the Kitaev spin liquid include Majorana zero modes (MZMs) bound to $\mathbb{Z}_2$ vortices ("visons"), which are Ising anyons of interest for quantum information processing, as well as gapped bulk fermions and a chiral Majorana edge mode at the boundary. Being excitations of an insulating magnet, they are electrically neutral. Sizable Kitaev couplings are expected [2] and have been reported in various material platforms for Mott insulators with strong spin-orbit coupling, e.g., in iridate compounds or in α-$RuCl_3$, where the smallness of interlayer couplings justifies the use of 2D models. For recent reviews, see Refs. [3–12]. Despite the impressive experimental progress achieved over the past decade, however, no consensus has emerged whether α-$RuCl_3$ or any other known material harbors a QSL. In particular, the half-quantized thermal Hall conductivity due to the chiral Majorana edge mode reported in Refs. [13–15] has not been found in other experiments [16,17]. In fact, some spin-liquid predictions can be mimicked by topological magnons in a polarized phase [18–20].

We here show that characteristic signatures of Ising anyons should be seen in scanning tunneling spectroscopy (STS) experiments [21] on a 2D Kitaev layer [22–25] by scanning the probe-tip position in the vicinity of an isolated $\mathbb{Z}_2$ vortex (located far away from all other vortices and from the sample boundary) and/or by changing the applied voltage; see Fig. 1. Below we will also compare our results to an alternative scenario with topological magnon bound states near defects or magnetic impurities, which could also cause low-energy features in the STS tunnel conductance. Such a comparison is important as evidenced by the corresponding topological superconductor case [26], where the tunnel conductance has a zero-bias anomaly with quantized peak conductance $2e^2/h$ due to MZM-mediated resonant Andreev reflection [27–30]. STS experiments have found such zero-bias anomalies near vortex cores in various superconducting materials and attributed them to MZMs [21,31–34]. A major obstacle to this interpretation is that very similar conductance peaks can be caused by conventional disorder-induced Andreev bound states [35]. However, the magnetic QSL case is rather different and warrants a separate investigation. The absence of a Cooper pair condensate implies that the charge of an electron (tunneling in from the tip via the MZM) is much harder to accommodate. For the pure Kitaev model, the infinite charge gap implies a vanishing tunnel conductance, $G(V) = 0$.

To obtain a finite $G(V)$, we start from the Hubbard-Kanamori model for Kitaev materials [2,36–39]. Adding a tunneling Hamiltonian for the QSL couplings to tip and substrate, see Fig. 1(a), and projecting to states with energy below the charge gap, we obtain $H = H_K + H_{\rm cot}$, where H_K describes the Kitaev model [2] and the cotunneling Hamiltonian $H_{\rm cot}$ encodes tip-substrate electron transfer due to virtual excursions to high-energy intermediate states [40–42]. We compute $H_{\rm cot}$ for arbitrary tip position and find that it is anisotropic in spin space. One then obtains $G(V)$ from the dynamical spin correlations of the QSL [43–47], which can be computed exactly [48–53]. However, in the presence of $\mathbb{Z}_2$ vortices, we encounter a technical challenge described and resolved below.

As a function of voltage, we predict a characteristic sequence of conductance steps linked to MZMs. By scanning the tip location at fixed voltage, one can locate MZMs in real space and obtain information about the vortex configurations contributing to the conductance. It stands to reason that experimental tests of our theory will help in identifying QSLs. (For other proposals aimed at the electric detection of QSLs,

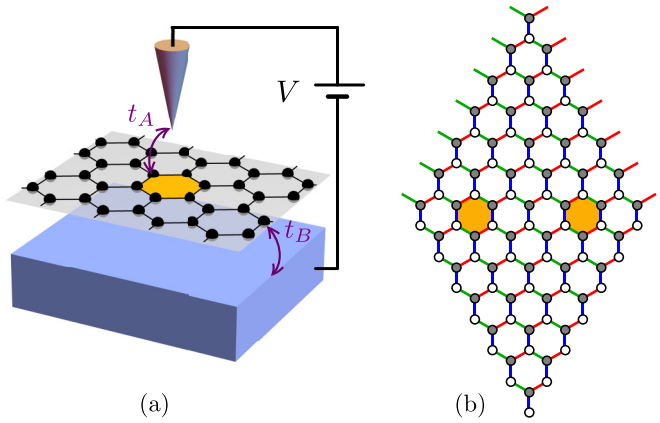


FIG. 1. (a) Schematic STS setup. Tunnel couplings t_A and t_B connect the QSL layer to the tip and the substrate, respectively. The differential conductance $G(V) = \frac{dI}{dV}$ follows by measuring the tunnel current I from tip to substrate as function of the applied voltage V. (b) Finite 2D Kitaev honeycomb lattice with $L \times L$ unit cells and periodic boundary conditions, shown for $L = 7$ and a configuration $\mathcal{G}$ with two $\mathbb{Z}_2$ vortices (shaded). Full and open circles represent the two sublattices. Nearest-neighbor bonds $\langle jl\rangle_\alpha$ of type $\alpha \in \{x, y, z\}$ are distinguished by different colors.

see Refs. [39,54–57].) Our study of an alternative topological magnon scenario suggests that MZM signatures obtained by STS on a Kitaev layer are easier to distinguish from other mechanisms than in the superconducting case.

The structure of the remainder of this article is as follows. In Sec. II, we derive the low-energy theory used for calculating the differential conductance, where technical details have been relegated to the Appendix. We then show in Sec. III how to compute the conductance in terms of an exact evaluation of dynamical spin-spin correlation functions of the Kitaev layer. Our results for the conductance profile are shown in Sec. IV. In Sec. V, we then address a complementary topological magnon scenario. Finally, we offer concluding remarks in Sec. VI.

II. EFFECTIVE LOW-ENERGY THEORY

We consider the setup in Fig. 1(a), where a scanning probe tip at position $\mathbf{r} = (x, y, d)$ is tunnel-coupled to a 2D Kitaev layer at vertical distance d. The layer is also coupled to a metallic substrate. Throughout, we assume weak and spin-independent tunnel amplitudes. Due to the charge gap in the magnetic layer, electron transport at subgap voltages V, applied between the tip [with conduction electron creation operator $\Psi^\dagger_{A\tau}(\mathbf{r})$ for spin projection $\tau = \uparrow, \downarrow$] and the substrate [with $\Psi^\dagger_{B\tau}(\mathbf{R}_j)$ below lattice site $\mathbf{R}_j$], can only take place via cotunneling [41,43,58]. We use the Hubbard-Kanamori model for strongly correlated d^5 electrons in α-$RuCl_3$ or related materials [2,36,37,39], where on-site correlations are captured by a large Coulomb energy U and a Hund coupling J_H. Including a tunneling Hamiltonian for the contacts to tip and substrate, the projection to energies below the charge gap $\sim U$ can be performed by a canonical transformation [2,39]. We show this calculation in some detail in the Appendix.

TABLE I. Kitaev couplings reported from different methods for several materials.

Material	K (meV)	Method
α-$RuCl_3$	5.0	experimental analysis [60]
	6.7	exact diagonalization [38]
	8.0–8.25	ab initio [61,62]
	10.6	density functional theory [63]
Na_2IrO_3	16.8	exact diagonalization [38]
	16.9	quantum chemistry methods [64]
	29.4	perturbation theory [65]
α-Li_2IrO_3	6.3–9.8	exact diagonalization [38]
Li_2RhO_3	2.9–11.7	quantum chemistry methods [66]

The low-energy theory is described by spin-1/2 operators, $\mathbf{S}_j = \frac{1}{2}\boldsymbol{\sigma}_j$, in the QSL layer, where $H = H_K + H_{\rm cot}$ includes the Kitaev model [1,2]

$$H_K = -K\sum_{\langle jl\rangle_\alpha} \sigma^\alpha_j \sigma^\alpha_l - \kappa \sum_{\langle jk\rangle_\alpha, \langle kl\rangle_\beta} \sigma^\alpha_j \sigma^\gamma_k \sigma^\beta_l , \tag{2.1}$$

with $\langle jl\rangle_\alpha$ denoting a nearest-neighbor bond of type $\alpha \in \{x, y, z\}$; see Fig. 1(b). The term $\propto \kappa$ describes a magnetic field [1,52], where $(\alpha\beta\gamma)$ is a cyclic permutation of (xyz) and the sum runs over triangles (jkl) with two adjacent nearest-neighbor bonds. We measure lengths in units of the lattice spacing a_0, where $a_0 \approx 5.9$Å for α-$RuCl_3$ [59]. The projection scheme yields a ferromagnetic (positive) Kitaev coupling $K \propto J_H$ [2], where experimental analysis gives $K \approx 5$ meV for α-$RuCl_3$ [60]. Theoretical estimates for K in different Kitaev materials have been reported in Refs. [38,61–66]; see Table I.

Similarly, summing over all lattice sites, the cotunneling Hamiltonian follows as

$$H_{\rm cot} = \sum_j \Psi^\dagger_A(\mathbf{r})[T_0(\mathbf{r} - \mathbf{R}_j)\mathbb{1}_j + \mathbf{T}(\mathbf{r} - \mathbf{R}_j)\cdot\boldsymbol{\sigma}_j] \times \Psi_B(\mathbf{R}_j) + \text{H.c.}, \tag{2.2}$$

where $\boldsymbol{\sigma}_j$ and $\mathbb{1}_j$ act in Kitaev spin space. The 2×2 matrices T_0 and T^α, with $\mathbf{T} = (T^x, T^y, T^z)$, act in conduction electron spin space. All T-matrix elements scale $\propto t_A t_B/U$, with real-valued tunnel couplings t_A (t_B) from tip (substrate) to a given site. We assume a constant substrate coupling t_B. The tip couplings depend on the overlap between the spherically symmetric tip wave function and the respective t_{2g} orbital (labeled by $\alpha = x, y, z$) for the d^5 electrons. With an energy scale t_0 and a tunneling length $l_0 \lesssim a_0$, we write [41,43]

$$t_{A\alpha}(\mathbf{r}, \mathbf{R}_j) = t_0 e^{-|\mathbf{r}\pm\mathbf{v}_\alpha - \mathbf{R}_j|/l_0}, \tag{2.3}$$

with the overall coupling $t_A \equiv \sqrt{t^2_{Ax} + t^2_{Ay} + t^2_{Az}}$. The vectors $\mathbf{v}_\alpha$ with $|\mathbf{v}_\alpha| \approx 0.1a_0$ encode the orbital overlaps, where the $\pm$ signs in Eq. (2.3) label the sublattice type of site $\mathbf{R}_j$; see the Appendix. The exponential scaling in Eq. (2.3) implies that only a few sites near the tip location $\mathbf{r}$ contribute. Analytical but lengthy expressions for T_0 and $\mathbf{T}$ are given in the Appendix.

Simpler results emerge by approximating $\mathbf{v}_\alpha = 0$, which gives exact results for a tip located on top of a lattice site and

otherwise causes deviations ∼10% in the tunnel couplings. (For the figures shown below, we have used the full expressions.) We then obtain

$$H_{\rm cot} = \sum_j \frac{t_A(\mathbf{r}-\mathbf{R}_j)t_B}{U}\Psi_A^\dagger(\mathbf{r})[\eta_0+\eta_1\boldsymbol{\tau}\cdot\boldsymbol{\sigma}_j + \eta_2(\tau^x+\tau^y+\tau^z)(\sigma^x+\sigma^y+\sigma^z)_j]\Psi_B(\mathbf{R}_j)+\text{H.c.}, \tag{2.4}$$

with $\frac{J_H}{U}$-dependent numbers $\eta_j \sim O(1)$; see the Appendix. The SU(2) spin rotation symmetry assumed in Refs. [43–45] is in fact lowered to a $\mathbb{Z}_3$ symmetry around the [111] axis.

III. DIFFERENTIAL CONDUCTANCE

At this point, it is straightforward to compute the differential conductance, $G(V) = \frac{dI}{dV}$, from Fermi's golden rule [43,44,58]. In the zero-temperature limit, we find

$$G(V) = \sum_{jl,\alpha\beta} C_{jl}^{\alpha\beta}(\mathbf{r})\int_0^{eV} d\omega S_{jl}^{\alpha\beta}(\omega) = \frac{e^2}{\hbar}\int_0^{eV} d\omega S_G(\omega), \tag{3.1}$$

with the dynamical spin correlation function of the QSL,

$$S_{jl}^{\alpha\beta}(\omega) = \int \frac{dt}{2\pi} e^{i\omega t}\langle\Phi|\sigma_j^\alpha(t)\sigma_l^\beta(0)|\Phi\rangle. \tag{3.2}$$

The second step in Eq. (3.1) defines the averaged dynamical spin correlator $S_G(\omega)$, which follows by weighting $S_{jl}^{\alpha\beta}(\omega)$ with its form factor,

$$C_{jl}^{\alpha\beta}(\mathbf{r}) = \frac{2e^2 d_A d_B}{\hbar}\text{Tr}[T^\alpha(\mathbf{r}-\mathbf{R}_j)T^\beta(\mathbf{r}-\mathbf{R}_l)], \tag{3.3}$$

with the tip (substrate) density of states d_A (d_B) and a trace over conduction electron spin space. Note that $\frac{dG}{dV} \propto S_G(V)$. The term $\propto T_0$ in Eq. (2.2) generates a voltage-independent background (including a mixing term of T_0 and $\mathbf{T}$) not contained in Eq. (3.1). However, this term is insensitive to $\mathbb{Z}_2$ vortices and can be disentangled from Eq. (3.1).

The correlation function (3.2) can be computed exactly for H_K by means of a Majorana representation of the spin degrees of freedom [48–51]. By writing $\sigma_j^\alpha = ic_j c_j^\alpha$ in terms of Majorana fermions with a local parity constraint, $D_j = c_j c_j^x c_j^y c_j^z = +1$, one obtains an exactly solvable noninteracting Hamiltonian for "matter" Majorana fermions, $\{c_j\}$, which move in a conserved $\mathbb{Z}_2$ gauge field $u_{\langle jl\rangle_\alpha} = ic_j^\alpha c_l^\alpha = \pm 1$ [1],

$$H_K = iK\sum_{\langle jl\rangle_\alpha} u_{\langle jl\rangle_\alpha} c_j c_l - i\kappa \sum_{\langle jk\rangle_\alpha,\langle kl\rangle_\beta} u_{\langle jk\rangle_\alpha} u_{\langle kl\rangle_\beta} c_j c_l. \tag{3.4}$$

All eigenstates of H_K can be written as a projected tensor product of a matter fermion state, $|\varphi(\mathcal{G})\rangle$, for given static gauge field configuration $|\mathcal{G}\rangle$,

$$|\Phi\rangle = \mathcal{P}|\mathcal{G}\rangle|\varphi(\mathcal{G})\rangle, \tag{3.5}$$

with $H_K|\Phi\rangle = E_\Phi|\Phi\rangle = E_{\varphi(\mathcal{G})}|\Phi\rangle$, where the projection $\mathcal{P} = \prod_j \frac{1+D_j}{2}$ projects onto the physical subspace. Defining gauge-invariant plaquette operators,

$$W_p = \prod_{\langle jl\rangle_\alpha\in p} u_{\langle jl\rangle_\alpha} = \pm 1, \tag{3.6}$$

the ground state has $W_p = +1$ for all hexagonal plaquettes p [1]. Plaquettes with $W_p = -1$ then define $\mathbb{Z}_2$ vortices, which are expected near vacancies or magnetic impurities [67–69] and harbor MZMs. In order to study the case shown in Fig. 1(a), we will then consider $|\Phi\rangle$ as the matter ground state $|\varphi_0(\mathcal{G})\rangle$ for a gauge configuration $\mathcal{G}$ with two well-separated $\mathbb{Z}_2$ vortices. We note that $\mathcal{G}$ can be constructed from a zero-vortex configuration $\mathcal{G}_0$ (with all bond variables $u_{\langle jl\rangle_\alpha} = +1$ for j in sublattice $\mathcal{A}$ and l in sublattice $\mathcal{B}$) by reversing the bond variables along an arbitrary string connecting both vortices.

For explicit calculations, we consider a finite honeycomb lattice with $L\times L$ unit cells and periodic boundary conditions. The $2N = 2L^2$ matter Majoranas are written as $c_j = c_\lambda(m,n)$, where $\lambda \in (\mathcal{A},\mathcal{B})$ labels the sublattice and $m,n = 1,\ldots,L$ the unit cell at $\mathbf{R}_j = m\hat{\mathbf{e}}_1 + n\hat{\mathbf{e}}_2$, with the primitive lattice vectors $\hat{\mathbf{e}}_1 = \frac{1}{2}\hat{\mathbf{x}} + \frac{\sqrt{3}}{2}\hat{\mathbf{y}}$ and $\hat{\mathbf{e}}_2 = -\frac{1}{2}\hat{\mathbf{x}} + \frac{\sqrt{3}}{2}\hat{\mathbf{y}}$. We next define the $2N$-dimensional Majorana vector $c = (c_\mathcal{A}, c_\mathcal{B})^T$, with the ordering convention $c_\lambda = (c_\lambda(1,1),\ldots,c_\lambda(L,1),c_\lambda(1,2),\ldots,c_\lambda(L,L))^T$, and a complex fermion for each unit cell, $f(m,n) = \frac{1}{2}[c_\mathcal{A}(m,n) - ic_\mathcal{B}(m,n)]$. With an N-dimensional vector f formed in analogy to c_λ, the linear transformation between both representations is given by

$$c = T\begin{pmatrix} f \\ f^\dagger\end{pmatrix}, \quad T = \begin{pmatrix} \mathbb{1}_N & \mathbb{1}_N \\ i\mathbb{1}_N & -i\mathbb{1}_N\end{pmatrix}, \tag{3.7}$$

with the $N\times N$ identity $\mathbb{1}_N$ and $T^{-1} = \frac{1}{2}T^\dagger$. The projection $\mathcal{P}$ here implies a parity constraint for the total number N_f of f fermions and the total number N_χ of bond fermions $\chi_{\langle jl\rangle_\alpha} = \frac{1}{2}(c_j^\alpha - ic_l^\alpha)$ [49–51],

$$(-1)^{N_f+N_\chi} = 1, \tag{3.8}$$

where we assume a vanishing boundary condition twist parameter in Ref. [51]. We note that N_χ is uniquely determined by the bond variables $\{u_{\langle jl\rangle_\alpha}\}$ defining the gauge configuration $\mathcal{G}$. Using the f fermions, we obtain

$$H_K = \frac{1}{2}(f^\dagger\ f)\,T^\dagger\begin{pmatrix}\mathcal{H}^\mathcal{G}_{\mathcal{AA}} & \mathcal{H}^\mathcal{G}_{\mathcal{AB}} \\ \mathcal{H}^\mathcal{G}_{\mathcal{BA}} & \mathcal{H}^\mathcal{G}_{\mathcal{BB}}\end{pmatrix}T\begin{pmatrix} f \\ f^\dagger\end{pmatrix}, \tag{3.9}$$

where the $N\times N$ matrices $\mathcal{H}^\mathcal{G}_{\lambda\lambda'}$ for given $\mathcal{G}$ can be read off from Eq. (3.4); see Ref. [39] for explicit expressions.

We next apply a unitary Bogoliubov transformation,

$$\begin{pmatrix} f \\ f^\dagger\end{pmatrix} = U_\mathcal{G}\begin{pmatrix} a \\ a^\dagger\end{pmatrix}, \tag{3.10}$$

in order to diagonalize Eq. (3.9) in terms of new (complex) matter fermions a_μ,

$$H_K = \frac{1}{2}\sum_{\mu=1}^N \varepsilon_\mu\left(2a_\mu^\dagger a_\mu - 1\right), \tag{3.11}$$

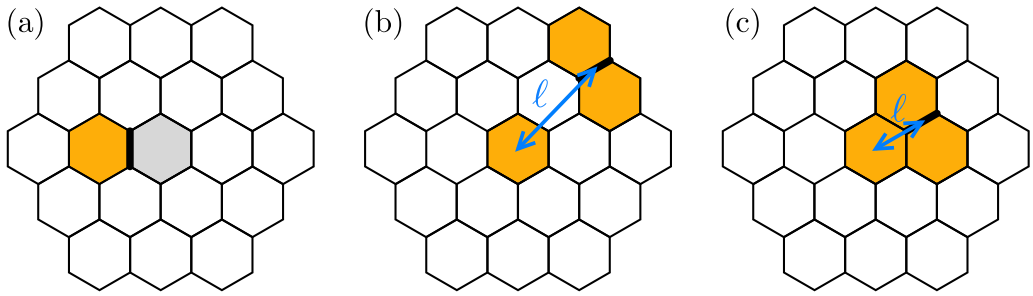


FIG. 2. Illustration of several gauge configurations $\mathcal{G}_j^\alpha$ contributing to the dynamical spin correlation functions determining the tunnel conductance. The central plaquette always refers to one of the two well-separated $\mathbb{Z}_2$ vortices (the other one is not shown) present in the reference configuration $\mathcal{G}$. Thick black bonds indicate a flip of the corresponding bond variable $u_{\langle jl\rangle_\alpha} \to -u_{\langle jl\rangle_\alpha}$. (a) The vortex is translated by one plaquette. (b) An additional pair of adjacent vortices at relatively large distance ℓ (blue double arrow) is created by the bond flip. (c) Same as (b) but for small distance ℓ.

where ε_μ are the non-negative eigenenergies ordered as

$$0 \leqslant \varepsilon_1 \leqslant \varepsilon_2 \leqslant \cdots \leqslant \varepsilon_N. \tag{3.12}$$

We often use the additional index $\mathcal{G}$, i.e., $a_\mu \to a_{\mathcal{G},\mu}$ and $\varepsilon_\mu \to \varepsilon_{\mathcal{G},\mu}$, to emphasize that those operators and energies refer to the corresponding gauge configuration. The matter ground state, $|\varphi_0(\mathcal{G})\rangle$, is determined by the conditions $a_\mu|\varphi_0(\mathcal{G})\rangle = 0$ (for all μ) and has the energy

$$E_{\mathcal{G},0} = -\frac{1}{2}\sum_{\mu=1}^{N} \varepsilon_{\mathcal{G},\mu}. \tag{3.13}$$

However, we still have to check that this state respects the parity constraint (3.8). To that end, we first note that the parity of the a fermions, $(-1)^{N_a}$ with $N_a = \sum_\mu a_\mu^\dagger a_\mu$, satisfies the relation

$$(-1)^{N_f} = (-1)^{N_a} \det U_{\mathcal{G}}, \tag{3.14}$$

where we have verified that the proof for Eq. (3.14) given in Ref. [51] for $\kappa = 0$ can be extended to $\kappa \neq 0$. Equation (3.8) can therefore be written as

$$(-1)^{N_a}\pi_{\mathcal{G}} = 1, \quad \pi_{\mathcal{G}} = (-1)^{N_\chi} \det U_{\mathcal{G}}, \tag{3.15}$$

where the ground-state parity operator, $\pi_{\mathcal{G}} = \pm 1$, is gauge invariant. For configurations with $\pi_{\mathcal{G}} = -1$, the matter ground state $|\varphi_0(\mathcal{G})\rangle$ is not in the physical subspace. One then has to add a single fermion to the ε_1 level for satisfying the parity constraint (3.15). The corresponding changes,

$$|\varphi_0(\mathcal{G})\rangle \to a_{\mu=1}^\dagger|\varphi_0(\mathcal{G})\rangle, \quad E_{\mathcal{G},0} \to E_{\mathcal{G},0} + \varepsilon_1, \tag{3.16}$$

are implicitly understood below.

We now turn to the dynamical spin correlator, where a Fourier transformation gives the Lehmann representation (with j in sublattice $\mathcal{A}$)

$$S_{jl}^{\alpha\beta}(\omega) = \sum_{\Phi'} \langle\Phi|\sigma_j^\alpha|\Phi'\rangle\langle\Phi'|\sigma_l^\beta|\Phi\rangle\, \delta(\omega + E_\Phi - E_{\Phi'}). \tag{3.17}$$

We consider $|\Phi\rangle$ as the matter ground state $|\varphi_0(\mathcal{G})\rangle$ for a given gauge configuration $\mathcal{G}$ (which we will later choose to contain two vortices), with energy $E_0 = E_{\mathcal{G},0}$ in Eq. (3.13). Inserting the Majorana decomposition into Eq. (3.17), we next observe that c_j^α commutes with all terms in H_K that do not contain $u_{\langle jl\rangle_\alpha}$, but anticommutes with all terms that do. Starting from $\mathcal{G} = \{u_{\langle j'l'\rangle_{\alpha'}}\}$, we then define a new gauge configuration $\mathcal{G}_j^\alpha = \{\tilde{u}_{\langle j'l'\rangle_{\alpha'}}\}$, see Fig. 2, with the bond variables

$$\tilde{u}_{\langle j'l'\rangle_{\alpha'}} = \begin{cases} -u_{\langle j'l'\rangle_{\alpha'}}, & \text{if } \langle j'l'\rangle_{\alpha'} = \langle jl\rangle_\alpha\,, \\ u_{\langle j'l'\rangle_{\alpha'}}, & \text{otherwise.} \end{cases} \tag{3.18}$$

With this definition, Eq. (3.17) yields [49–51]

$$S_{jl}^{\alpha\beta}(\omega) = \sum_{\varphi(\mathcal{G}_j^\alpha)} \big\langle\varphi_0(\mathcal{G})|c_j|\varphi\big(\mathcal{G}_j^\alpha\big)\big\rangle\big\langle\varphi\big(\mathcal{G}_j^\alpha\big)|c_l|\varphi_0(\mathcal{G})\big\rangle \times \delta\big(\omega + E_0 - E_{\varphi(\mathcal{G}_j^\alpha)}\big)\big(\delta_{jl} - iu_{\langle jl\rangle_\alpha}\delta_{\langle jl\rangle_\alpha}\big)\delta_{\alpha\beta}. \tag{3.19}$$

Here $\delta_{\langle jl\rangle_\alpha} = 1$ if (jl) form a nearest-neighbor bond of type $\langle jl\rangle_\alpha$, and zero otherwise. Hence $S_{jl}^{\alpha\beta}(\omega) \neq 0$ is possible only for equal spin indices ($\alpha = \beta$) and on-site terms or nearest-neighbor bonds. As sketched in Fig. 2, $\mathcal{G}$ and $\mathcal{G}_j^\alpha$ are connected by either moving a vortex by one plaquette, or by creating two additional vortices. We note that the zero-frequency peak in $S_G(\omega)$ is connected to the configurations in Fig. 2(a). Since we expect this peak to move to a finite but very small frequency ω_0 in practice, see Sec. IV, we have taken it into account with the full weight of the delta function peak in the tunnel conductance (3.1), even though the integral in Eq. (3.1) runs over positive frequencies only.

Since matter states for two different gauge configurations are needed in Eq. (3.19), it is convenient to use the notations

$$a_\mu = a_{\mathcal{G},\mu}, \quad b_\mu = a_{\mathcal{G}_j^\alpha,\mu}, \qquad |0_a\rangle = |\varphi_0(\mathcal{G})\rangle, \quad |0_b\rangle = \big|\varphi_0\big(\mathcal{G}_j^\alpha\big)\big\rangle, \tag{3.20}$$

with the N-component spinors $a = (a_1, \ldots, a_N)^T$ and $b = (b_1, \ldots, b_N)^T$. The a matter fermions with ground state $|0_a\rangle$ thus refer to the gauge configuration $\mathcal{G}$, while the b fermions with ground state $|0_b\rangle$ refer to $\mathcal{G}_j^\alpha$. The corresponding ground-state energies are denoted by $E_{|0_a\rangle}$ and $E_{|0_b\rangle}$, respectively. From Eq. (3.10), the a and b fermions must be connected by a unitary Bogoliubov transformation [50,51,70],

$$\begin{pmatrix} b \\ b^\dagger \end{pmatrix} = \mathcal{W}\begin{pmatrix} a \\ a^\dagger \end{pmatrix}, \quad \mathcal{W} = U_{\mathcal{G}_j^\alpha}^\dagger U_{\mathcal{G}} = \begin{pmatrix} X^* & Y^* \\ Y & X \end{pmatrix}, \tag{3.21}$$

where the $N \times N$ matrices X and Y satisfy the relations

$$XX^\dagger + YY^\dagger = 1, \quad X^\dagger X + Y^T Y^* = 1,$$
$$XY^T + YX^T = 0, \quad X^T Y^* + Y^\dagger X = 0.$$

For $\det \mathcal{W} = +1$, we next observe that $|0_b\rangle$ can be obtained from $|0_a\rangle$ by means of the Thouless theorem [72]. As a result, one finds [51,73]

$$|0_b\rangle = [\det(X^\dagger X)]^{1/4} \exp\left(-\tfrac{1}{2} a^\dagger X^{*-1} Y^* a^\dagger\right) |0_a\rangle \,. \quad (3.22)$$

The matrix elements needed in Eq. (3.19) are of the form

$$\langle \varphi_0(\mathcal{G}) | c_j | \varphi(\mathcal{G}_j^\alpha)\rangle = \langle 0_a | c_j b^\dagger_{\mu_1} \ldots b^\dagger_{\mu_n} | 0_b\rangle \,, \quad (3.23)$$

where $\mu_1 \leqslant \cdots \leqslant \mu_n$ and n is constrained by $(-1)^n = \pi_{\mathcal{G}_j^\alpha}$. One can understand this constraint by noting that Eq. (3.23), which is a matrix element of the single fermion operator c_j, must vanish if $|\varphi_0(\mathcal{G})\rangle$ and $|\varphi(\mathcal{G}_j^\alpha)\rangle$ have the same fermion parity. We note that for $\det \mathcal{W} = 1$, exactly one of the two fermionic vacua $|0_a\rangle$ and $|0_b\rangle$ will not be in the physical subspace since the $\pi_{\mathcal{G}}$ operator will change sign when flipping a bond. As discussed above, we therefore have to add a single fermion to one of the two states. Using Eq. (3.22) and the relation $c = T U_{\mathcal{G}} \, (a, a^\dagger)^T$, which follows from Eqs. (3.7) and (3.10), we can finally express all matrix elements (3.23) exclusively in terms of a and $a^\dagger$ operators, facilitating their practical computation.

For a numerical implementation, we restrict the number n of excitations in Eq. (3.23) by imposing $0 \leqslant n \leqslant n_{\max}$. Under this truncation, exactness of the computed dynamical spin correlations is ensured only for frequencies

$$\omega < \omega_{\max} = E_{|0_b\rangle} - E_{|0_a\rangle} + \sum_{\mu=1}^{n_{\max}+2} \varepsilon_{\mathcal{G}_j^\alpha,\mu} . \quad (3.24)$$

However, already for $n_{\max} = 2$, accurate results can be obtained even for $\omega > \omega_{\max}$ in the vortex-free configuration $\mathcal{G}_0$ [51]. For the two-vortex configuration $\mathcal{G}$, rapid convergence of the numerical results upon increasing $n_{\max}$ was observed. Since the characteristic MZM features stem from the low-frequency part of $S^{\alpha\beta}_{jl}(\omega)$, in all cases shown here, a truncation with $n_{\max} = 2$ was sufficient to reach convergence for $\omega < \omega_{\max}$.

However, for selected bonds $\langle jl\rangle_\alpha$ in the two-vortex configuration $\mathcal{G}$, we find that $\det \mathcal{W} = -1$. In such cases, the Thouless theorem breaks down and X in Eq. (3.21) is a singular $N \times N$ matrix. As a result, Eq. (3.22) does not apply anymore. For computing the STS tunnel conductance near a single $\mathbb{Z}_2$ vortex, it is essential to resolve this issue. For closely related problems, Refs. [71,72] have obtained a solution by interchanging the ground-state occupancies of a single particle and its hole partner. We follow their approach and define the matrices $X^{(\mu)}$ and $Y^{(\mu)}$, see Eq. (3.21), according to

$$X^{(\mu)}_{kl} = \begin{cases} X_{kl}, & l \neq \mu, \\ Y_{kl}, & l = \mu, \end{cases} \quad Y^{(\mu)}_{kl} = \begin{cases} Y_{kl}, & l \neq \mu, \\ X_{kl}, & l = \mu, \end{cases} \quad (3.25)$$

where μ refers to the index of the interchanged particle and hole. This interchange of columns renders $X^{(\mu)}$ nonsingular as it corresponds to a Bogoliubov transformation with positive determinant. We can then use the Thouless theorem again, such that after the operation (3.25), we can effectively use Eq. (3.22). The thereby obtained state, $|0'_b\rangle$, has the energy $E_{|0'_b\rangle} = E_{\mathcal{G}_j^\alpha,0} + \varepsilon_{\mathcal{G}_j^\alpha,\mu}$, and the chosen index μ should minimize $\varepsilon_{\mathcal{G}_j^\alpha,\mu}$. For instance, if it corresponds to a zero mode, $\varepsilon_{\mathcal{G}_j^\alpha,\mu} = 0$, the interchange (3.25) introduces no approximation, the energy ordering in Eq. (3.12) remains unaffected, and $|0'_b\rangle$ captures the ground state for the b fermions. For the configurations studied in this work, we can always find a low-energy fermion level that approaches a zero mode in the thermodynamic limit for $\kappa \neq 0$. These low-energy modes are well separated from the fermion continuum which has a finite gap $\propto |\kappa|$.

It is worth mentioning that two consistency checks are passed successfully by our numerical calculations. First, $\lim_{V\to\infty} \int_0^{eV} d\omega \, S^{\alpha\beta}_{jl}(\omega)$ recovers the static equal-time spin correlator [39]. Second, dynamical spin correlations are radially isotropic around an isolated $\mathbb{Z}_2$ vortex despite of the presence of a gauge string.

IV. CONDUCTANCE SIGNATURES OF MZMs

Figure 3 shows numerical results for $S_G(\omega)$ and $G(V)$ for three different tip positions near an isolated $\mathbb{Z}_2$ vortex. The different peaks in each $S_G(\omega)$ curve have a clear physical meaning. First, the $\omega = 0$ peak is directly connected to MZMs and stems from configurations $\mathcal{G}_j^\alpha$ with the vortex translated by one step. (For nonuniform Kitaev couplings, the peak can shift to a small frequency ω_0; see below.) The support for this peak comes only from on-site terms and nearest-neighbor bonds directly enclosing the vortex. Indeed, Fig. 3(a) shows that the peak weight decreases rapidly with the tip-vortex distance. Second, the peaks at $\omega = \Delta E_{2v}(\ell)$ ($\approx 0.1K$ in Fig. 3) correspond to the energy cost for creating a configuration $\mathcal{G}_j^\alpha$ with an additional pair of adjacent vortices by flipping a bond at distance ℓ from the original vortex, with the fermion bound state built from the new overlapping MZM pair unoccupied. This peak may contain several subpeaks since various configurations $\mathcal{G}_j^\alpha$ with different ℓ, and hence different $\Delta E_{2v}(\ell)$, may contribute to $S_G(\omega)$ in this frequency range. Third, the peak structure at $\omega = \Delta E_{2v}(\ell) + \varepsilon_f(\ell) \approx 0.25K$ includes the energy cost $\varepsilon_f(\ell)$ for occupying the fermion bound state. Finally, the onset of the gapped two-fermion continuum is marked by a (small) peak at $\omega = \Delta E_{2f} = \frac{3\sqrt{3}}{2}|\kappa|$ ($\approx 0.5K$ in Fig. 3).

The conductance $G(V)$ in Fig. 3(b) follows by integrating $S_G(\omega)$ and therefore shows steps at the voltages matching a peak in $S_G(\omega)$. One can thus measure the important energy scales ΔE_{2v}, ε_f, and ΔE_{2f} by STS. However, the respective step sizes are not universal because the peak weights in $S_G(\omega)$ depend on the tip position and on the form factors. It is instructive to compare to the vortex-free configuration $\mathcal{G}_0$, see Fig. 3(b), where $G(V)$ is strongly suppressed for $eV < \Delta E_{2v}(\infty) + \varepsilon_f(\infty)$. Indeed, here the lowest-energy excitation probed by $G(V)$ corresponds to adding a vortex pair and filling the fermion bound state in order to respect the parity constraint. In this low-voltage regime, the conductance for the two-vortex configuration $\mathcal{G}$ is instead dominated by MZMs and will be finite at small V, with a step at $eV = \Delta E_{2v}(\ell)$. We also observe from Fig. 3(b) that the "bulk" behavior of $G(V)$, found for arbitrary tip position in configuration $\mathcal{G}_0$, is approached by moving the probe tip far away from the vortex center. We note that the zero-voltage step is particular to the integrable Kitaev model with uniform couplings (assumed in Fig. 3), where the eigenstates are degenerate with respect to the vortex position. In a generic nonintegrable case, vortices

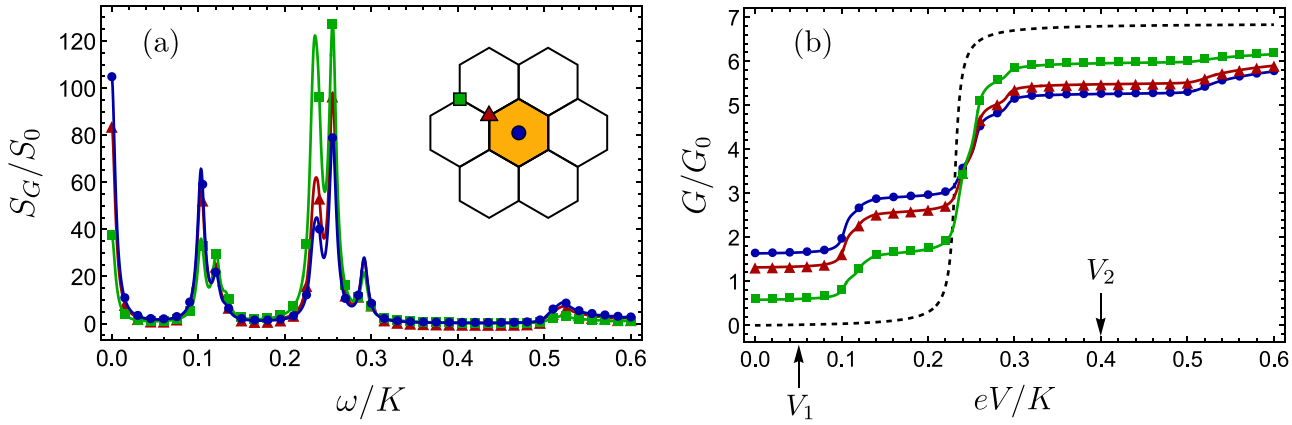


FIG. 3. STS for a Kitaev QSL in a two-vortex configuration $\mathcal{G}$, see Fig. 1(b), for $\kappa = 0.2K$, $L = 37$, $J_H = 0.05U$, $l_0 = 0.75a_0$, and $d = l_0$. For α-$RuCl_3$, one expects $K \approx 5$ meV [60]. (a) Weighted spin correlation function S_G vs ω, see Eq. (3.1), for three tip positions (inset). We plot $S_G(\omega)$ in units of $S_0 = d_A d_B (t_0 t_B/U)^2$, with delta function peaks replaced by Lorentzians of width $\Gamma_L = 0.005K$ due to higher-order tunneling processes. (b) Conductance G (in units of $G_0 = S_0 \frac{e^2}{\hbar}$) vs V, see Eq. (3.1), for the tip positions in (a). The black dashed curve is for the vortex-free configuration $\mathcal{G}_0$. The voltages $V_{1,2}$ are used in Fig. 4.

are mobile but can be trapped by bond disorder, vacancies, magnetic impurities, or by an external electrostatic potential. The $V = 0$ step may then shift to a small finite voltage $eV = \omega_0$, where ω_0 describes the difference in vortex creation energies on different plaquettes. Such shifts may be useful to distinguish MZM-induced conductance steps from the background conductance due to T_0 in Eq. (2.2).

For the voltages $V_{1,2}$ marked in Fig. 3(b), we show the tip-position dependence of the conductance in Fig. 4. For $V = V_1$, see Fig. 4(a), the physics is dominated by the zero-frequency MZM peak in $S_G(\omega)$, and the spatial profile in Fig. 4(a) encodes a convolution of the (squared) MZM wave function [53] with the form factor (3.3). However, in contrast to the standard situation in STS [21], it is not possible to map out the MZM wave function beyond the immediate vicinity of the vortex because only terms from sites or bonds encircling the vortex contribute for $eV < \Delta E_{2v}(\ell)$. The conductance profile for $V = V_2$ in Fig. 4(b) reveals a dip in the center, which arises because for a tip away from the vortex, the form factors enhance the peak contribution for $eV > \Delta E_{2v}(\ell)$. However, this voltage regime involves many vortex configurations $\mathcal{G}_j^\alpha$, rendering it difficult to extract the MZM wave function. Nonetheless, the conductance profile allows us to detect the MZM at the vortex location. Finally, the angular isotropy of the spatial profile approximately found at low voltage is reduced to a C_6 symmetry at higher voltages. While this effect is hardly visible for the tip distance $d = l_0$ in Fig. 4, it becomes more prominent for smaller d.

V. TOPOLOGICAL MAGNONS

In this section, we explore a different mechanism that could in principle generate similar tunnel conductance features to those reported above for MZMs in the spin-liquid phase. To that end, we consider topological magnons in the polarized phase of the Kitaev model in a magnetic field [18,19,43]. Such models have been proposed as an alternative scenario for explaining the observed half-quantized thermal Hall conductivity [18,19]. Below we clarify whether local defects or magnetic impurities are able to generate topological magnon bound states below the magnon gap. If present, such bound states may produce tunnel conductance steps at voltages

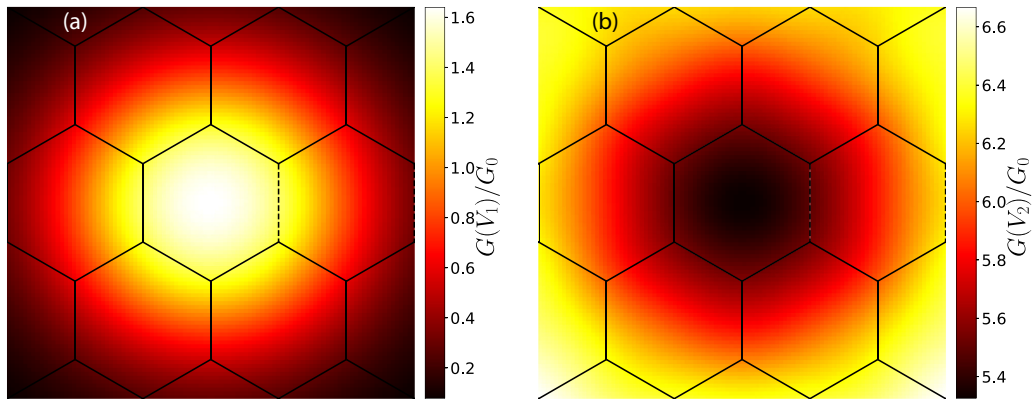


FIG. 4. Spatial conductance profile near a vortex (central plaquette) in the xy plane, for the parameters in Fig. 3 with (a) $V = V_1$ and (b) $V = V_2$; see Fig. 3(b). Note the different color scales.

matching the respective bound-state energies. In analogy to the topological superconductor case, magnon-induced conductance steps could then be difficult to distinguish from those caused by MZMs in a Kitaev spin liquid.

We consider spin-S operators S_i^γ on the 2D honeycomb lattice with Kitaev couplings. The Hamiltonian is given by

$$H_m = -\sum_{\langle ij\rangle_\gamma} K_{ij} S_i^\gamma S_j^\gamma - \sum_j \mathbf{h}_j \cdot \mathbf{S}_j, \tag{5.1}$$

where $\gamma \in \{x, y, z\} \equiv \{1, 2, 3\}$ denotes the spin components as well as the bond directions; see Sec. II. For simplicity, we assume that the local magnetic fields are oriented along the [111] direction, $\mathbf{h}_j = h_j \mathbf{c}$, with the unit vector $\mathbf{c}$ in Eq. (A14); see the Appendix. In the homogeneous case, the Kitaev couplings and local fields are given by $K_{ij} = K$ and $\mathbf{h}_j = \mathbf{h}$, respectively. In order to model a *defect*, we study inhomogeneous Kitaev couplings K_{ij} near a single plaquette corresponding to the defect, similar to models for bond disorder and vacancies [74–76]. Recalling that a large-spin *magnetic impurity* is equivalent to a local change of the magnetic field at a single site [77], we model a magnetic impurity by a local change of the field $h_i \neq h$ at this site relative to the bulk field h. We follow Refs. [18,19] and derive the linear spin wave theory which becomes exact in the large-S limit.

We first rotate the local basis to have the magnetization axis along the c direction. With the orthogonal matrix $R = (\mathbf{a}\,\mathbf{b}\,\mathbf{c})$, see Eq. (A14), we have the rotated spin operators $\tilde{S}_i^\alpha = R_{\alpha\beta} S_i^\beta$. Next, we employ a Holstein-Primakoff transformation to expand around the polarized state,

$$\tilde{S}_i^z = S - b_i^\dagger b_i, \quad \tilde{S}_i^x \approx \sqrt{\frac{S}{2}}(b_i + b_i^\dagger),$$
$$\tilde{S}_i^y \approx -i\sqrt{\frac{S}{2}}(b_i - b_i^\dagger), \tag{5.2}$$

with bosonic magnon operators b_i. Expanding H_m in Eq. (5.1) in powers of $1/S$, we obtain $H_m = E_{\rm cl} + H_1 + H_2 + O(S^{1/2})$. The first term describes the classical ground state energy, $E_{\rm cl} = -\frac{S^2}{3}\sum_{\langle ij\rangle} K_{ij} - S\sum_j h_j$. The second term is linear in the bosons,

$$H_1 = \frac{S^{3/2}}{3}\sum_i \left(\sum_\gamma e^{-i2\pi\gamma/3} K_{i,i+\boldsymbol{\delta}_\gamma}\right) b_i + \text{H.c.}, \tag{5.3}$$

with the in-plane nearest-neighbor vectors

$$\boldsymbol{\delta}_1 = \frac{1}{2}\hat{\mathbf{x}} + \frac{1}{2\sqrt{3}}\hat{\mathbf{y}}, \quad \boldsymbol{\delta}_2 = -\frac{1}{2}\hat{\mathbf{x}} + \frac{1}{2\sqrt{3}}\hat{\mathbf{y}}, \quad \boldsymbol{\delta}_3 = -\frac{1}{\sqrt{3}}\hat{\mathbf{y}}. \tag{5.4}$$

One finds $H_1 = 0$ for $K_{ij} = K$, but in the presence of defects, $H_1 \neq 0$ indicates that we have expanded around the wrong classical state. Due to the anisotropy of the Kitaev interactions, the spins do not align with the [111] direction anymore if the $\mathbb{Z}_3$ symmetry is broken by defect bonds. To correct for this problem, one has to find the correct classical state with an inhomogeneous magnetization and then apply position-dependent R matrices in order to rotate the spins to their local magnetization axis. While such refinements could give quantitative corrections, we here focus on the quadratic term,

$$H_2 = -\frac{S}{3}\sum_{\langle ij\rangle_\gamma} K_{ij}(b_i^\dagger b_j + b_j^\dagger b_i + e^{i2\pi\gamma/3} b_i b_j + e^{-i2\pi\gamma/3} b_j^\dagger b_i^\dagger)$$
$$+ S\sum_i \left(h_i + \frac{1}{3}\sum_j K_{ij}\right) b_i^\dagger b_i. \tag{5.5}$$

Indeed, in general terms, the linear spin wave theory resulting from Kitaev (or other) interactions on the 2D honeycomb lattice must be of the form

$$H_2 = S\sum_{\langle ij\rangle_\gamma}(t_{ij} b_i^\dagger b_j + t_{ij}^* b_j^\dagger b_i + \Delta_{ij} b_i b_j + \Delta_{ij}^* b_j^\dagger b_i^\dagger)$$
$$+ S\sum_i B_i\, b_i^\dagger b_i, \tag{5.6}$$

where B_i is an effective magnetic field including the Weiss field. The misalignment of spins around defects here should give rise to an additional position dependence in the parameters t_{ij}, Δ_{ij}, and B_i in Eq. (5.6), on top of the immediate effects of K_{ij} anisotropy in Eq. (5.5). In what follows, we consider $H_m \simeq H_2$ as given by Eq. (5.5).

We first address the homogeneous case, where Fourier transformation gives $H_2 = S\sum_{\mathbf{k}\in\frac{1}{2}\text{BZ}} \Psi_\mathbf{k}^\dagger M_\mathbf{k} \Psi_\mathbf{k}$. Here $\mathbf{k}$ runs over half the Brillouin zone, $\Psi_\mathbf{k}^\dagger = (b_{\mathbf{k},\mathcal{A}}^\dagger \quad b_{\mathbf{k},\mathcal{B}}^\dagger \quad b_{-\mathbf{k},\mathcal{A}} \quad b_{-\mathbf{k},\mathcal{B}})$ is a four-component spinor (including the sublattice index), and

$$M_\mathbf{k} = \begin{pmatrix} A_\mathbf{k} & B_\mathbf{k} \\ B_{-\mathbf{k}}^* & A_{-\mathbf{k}}^{\rm T} \end{pmatrix}. \tag{5.7}$$

Using the notation $\Gamma_{\mathbf{k},n} = \sum_\gamma e^{-i2\pi n\gamma/3} e^{i\mathbf{k}\cdot\boldsymbol{\delta}_\gamma}$ with $n \in \{0, 1\}$, we have defined the matrices

$$A_\mathbf{k} = \begin{pmatrix} h + K & -\frac{1}{3}K\Gamma_{\mathbf{k},0} \\ -\frac{1}{3}K\Gamma_{-\mathbf{k},0} & h + K \end{pmatrix},$$
$$B_\mathbf{k} = \begin{pmatrix} 0 & -\frac{1}{3}K\Gamma_{\mathbf{k},1} \\ -\frac{1}{3}K\Gamma_{-\mathbf{k},1} & 0 \end{pmatrix}. \tag{5.8}$$

This Hamiltonian can be diagonalized by a Bogoliubov transformation. With $\Sigma = \text{diag}(1, 1, -1, -1)$, we obtain the magnon band dispersion from the positive eigenvalues of $\Sigma M_\mathbf{k}$. The result is illustrated in Fig. 5. We find two bands $\omega_1(\mathbf{k})$ and $\omega_2(\mathbf{k})$, where analytical but lengthy expressions are available. These topological magnon bands cover the energy range

$$h < \omega_1(\mathbf{k}) < \sqrt{h(h + 2K)}, \quad h + K < \omega_2(\mathbf{k}) < h + 2K. \tag{5.9}$$

The magnon band gap is thus given by $\Delta E_m = h$. For $h \to 0$, the lower magnon band becomes a zero-energy flat band, signaling the degeneracy of the classical Kitaev model at zero field.

A. Defect from bond disorder

Next we turn to inhomogeneous Kitaev interactions, where we model a defect by modifying the bonds $K_{ij} \to \xi K$ around

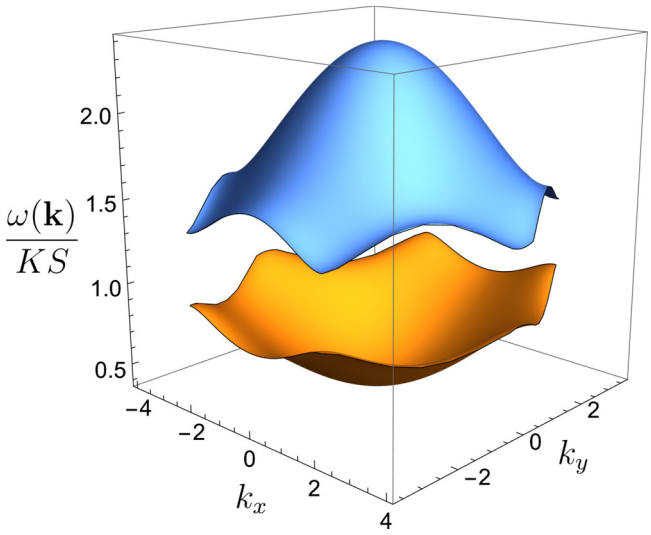


FIG. 5. Topological magnon bands for $h/K = 0.4$ (with momentum unit a_0^{-1}) from linear spin wave theory for the homogeneous model (5.1).

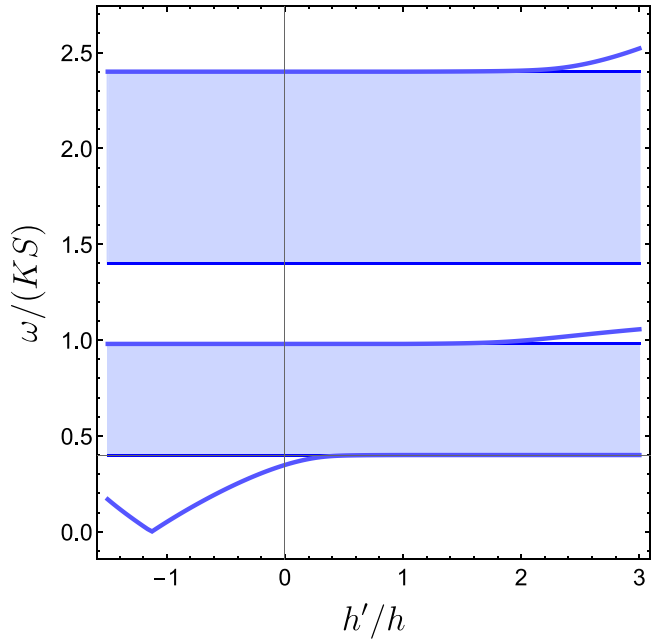


FIG. 6. Magnon spectrum ω vs h'/h for a local magnetic field $h' \neq h$ at a single site. The bulk field is $h = 0.4K$. Shaded regions describe continuum states; see Eq. (5.9). A single subgap bound state can exist for $h' < h$. A high-energy antibound state is visible for $h' > h$, and another bound state exists in the minigap between both bands.

a given plaquette representing the defect by a positive factor $\xi \neq 1$. We have studied two different radially symmetric bond defect patterns. In the first case, we modify only the six bonds directly surrounding the defect plaquette. In the second case, we instead change only the six adjacent bonds pointing radially outward from this plaquette. The conclusions described below are identical for both cases. We have studied the spectrum of H_2 in Eq. (5.5) by numerical diagonalization on a finite $L \times L$ honeycomb lattice as described in Sec. III. We observe that making the bonds stronger ($\xi > 1$) creates a repulsive potential for magnons, which generates antibound states above the top of the upper band, $\varepsilon_m' > h + 2K$. There are also bound states in the gap between both bands. However, even if we make the bonds significantly weaker, $\xi < 1$, we never observe bound states below the lower band, $\varepsilon_m < \Delta E_m$. We conclude that bond defects are unlikely to produce magnon bound states at subgap energies. At the same time, we cannot rule out that a more complex bond defect pattern could cause subgap features that can mimic the Majorana features described in Sec. IV. Future work should investigate this issue in more detail.

B. Magnetic impurity

Another limiting case is to locally modify only the magnetic field h_i in Eq. (5.5), keeping homogeneous Kitaev couplings $K_{ij} = K$. For a radially symmetric inhomogeneous magnetic field profile, $\mathbb{Z}_3$ symmetry remains intact and the linear-boson term H_1 in Eq. (5.3) vanishes. If we change the field only at a single site, $h_i = h' \neq h$, with the bulk field h acting at all other sites, we can find a single subgap bound state for $h' < h$ as shown in Fig. 6. The bound-state energy $\varepsilon_m < \Delta E_m$ vanishes for $h' \approx -1.1h$ for $h = 0.4K$. For smaller h, the vanishing of ε_m occurs at lower values of $h'/h < 0$. For generic values of h'/h, we find that ε_m is positive. The dynamical spin correlation function then will have a peak at $\omega = \varepsilon_m$, and Eq. (3.1) yields a single steplike feature in $G(V)$ at $eV = \varepsilon_m$. Except for the fine-tuned case with $\varepsilon_m = 0$, this step does not occur at zero voltage as expected for the MZM case.

For a wider field profile, with the field change extending over several sites, we typically find several subgap bound states. This case can be realized if the impurity is coupled to several sites. In such cases, from the $G(V)$ curve alone, it can be difficult to disentangle the effects of magnon bound states from those due to MZMs. However, a collection of several nearby magnetic impurities causing such a field profile should be identifiable by concomitant STM surface topography scans.

VI. CONCLUSIONS

Based on the above analysis, we expect that the tunnel conductance features due to MZMs in a spin liquid will be quite robust. For the topological magnon scenario in Sec. V, we find that defects modeled by locally inhomogeneous Kitaev couplings do not bind subgap magnon bound states. On the other hand, a large-spin magnetic impurity can induce a single subgap bound state centered at the corresponding site. One then expects a single conductance step, where the spatial distribution of the STS tunnel conductance peaks at this site. For the MZM case, we instead predict a characteristic sequence of steps and the spatial distribution should peak at the center of the hexagon defining the vortex.

We conclude that the perspectives for STS detection of MZMs in spin liquids appear promising. In fact, tunneling experiments on monolayers of α-$RuCl_3$ have recently observed interesting low-energy excitations [62]. Given the rapid progress in encapsulating and probing atomically thin

materials [78], detailed experimental tests of our predictions will likely soon be available.

ACKNOWLEDGMENTS

We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation), Projektnummer 277101999–TRR 183 (Project No. B04), Normalverfahren Projektnummer EG 96-13/1, and under Germany's Excellence Strategy–Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1–390534769, by the Brazilian ministries MEC and MCTI, by the Brazilian agency CNPq, and by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES)–Finance Code 001.

APPENDIX: DERIVATION OF LOW-ENERGY THEORY

This Appendix provides a derivation of the cotunneling Hamiltonian (2.2) with the corresponding transition matrix elements. As starting point, we take the general Hamiltonian $H_{\rm tot} = H_M + V_{\rm at} + H_{\rm tun}$, where H_M describes noninteracting metallic leads representing the scanning probe tip and the substrate,

$$H_M = \sum_{\nu\in\{A,B\}} \sum_{\tau\in\{\uparrow,\downarrow\}} \sum_{\mathbf{k}} \varepsilon_{\nu\tau}(\mathbf{k}) c^\dagger_{\nu\tau}(\mathbf{k}) c_{\nu\tau}(\mathbf{k}). \tag{A1}$$

The fermion annihilation operators $c_{\nu\tau}(\mathbf{k})$ with $\nu = A, B$ refer to tip and substrate electrons, respectively, where τ is the spin projection and $\varepsilon_{\nu\tau}(\mathbf{k})$ the energy with respect to the Fermi energy. The Pauli matrices $\boldsymbol{\tau}$ used below act in the spin space of the conduction electrons.

For the 2D Kitaev layer, we start from a Hubbard-Kanamori model for the d^5 electrons in an edge-sharing octahedral environment, e.g., those of the Ru^{3+} ions in α-$RuCl_3$. For lowest-order perturbation theory in the tunnel Hamiltonian $H_{\rm tun}$ connecting the layer to the STM tip and to the substrate, only the single-site atomic Hamiltonian $V_{\rm at}$ in the Hubbard-Kanamori model is needed (see, for instance, Ref. [39]),

$$V_{\rm at} = \frac{U-3J_H}{2}(\bar N - 1)^2 - 2J_H \bar{\mathbf{S}}^2 - \frac{J_H}{2}\bar{\mathbf{L}}^2 + \lambda_{\rm so}\bar{\mathbf{L}}\cdot\bar{\mathbf{S}}, \tag{A2}$$

with the on-site Coulomb energy U, the Hund coupling J_H, and the spin-orbit coupling $\lambda_{\rm so}$. The respective couplings can be renormalized by screening processes resulting from the presence of the tip and the substrate, but one expects $U \approx 2$ eV and $\lambda_{\rm so} \ll J_H, U$. For definiteness, we assume $J_H \ll U$. To lowest order in $H_{\rm tun}$, contributions from different lattice sites simply add up. The operators $\bar N$, $\bar{\mathbf{S}}$, and $\bar{\mathbf{L}}$ in Eq. (A2) refer to hole number, spin, and angular momentum, respectively. In terms of the hole annihilation operators h_s with the combined spin-orbital index $s = (\alpha, \sigma)$, they are expressed as

$$\bar N = h^\dagger h, \quad \bar{\mathbf{S}} = \tfrac{1}{2} h^\dagger (\bar{\boldsymbol{\sigma}} \otimes \mathbb{1}_3) h, \quad \bar{\mathbf{L}} = h^\dagger (\mathbb{1}_2 \otimes \bar{\mathbf{l}}) h, \tag{A3}$$

with $h^\dagger = (h^\dagger_{x\uparrow}, h^\dagger_{y\uparrow}, h^\dagger_{z\uparrow}, h^\dagger_{x\downarrow}, h^\dagger_{y\downarrow}, h^\dagger_{z\downarrow})$. The five d electrons in a cubic crystal field occupy three t_{2g} orbitals (xy, yz, zx), denoted here by the complementary index $\alpha = (z, x, y)$. The Pauli matrices $\bar{\boldsymbol{\sigma}}$ act in the spin space of the magnetic layer site, and $\bar{\mathbf{l}} = (\bar l^x, \bar l^y, \bar l^z)$ represents the $l_{\rm eff} = 1$ orbital angular momentum of the corresponding t_{2g} states, with explicit matrix representations specified in Ref. [39]. Following standard practice, the spin-orbit coupling $\lambda_{\rm so}$ will be taken into account later through a projection to the lowest-lying hole states with total angular momentum $j_{\rm eff} = 1/2$.

Electron transfer between tip (or substrate) and the Mott insulating site is described by a tunneling Hamiltonian $H_{\rm tun} = \mathcal{T}_1 + \mathcal{T}_{-1}$, where $\mathcal{T}_{\pm 1}$ refers to changes of the hole number by $\Delta\bar N = \pm 1$, respectively. With the complex-valued tunnel amplitude $t_{\nu\tau s}(\mathbf{k})$ connecting a conduction electron in lead $\nu = A, B$ with spin τ and momentum $\mathbf{k}$ to the spin-orbital hole state $s = (\alpha, -\sigma)$ on the magnetic site,

$$\mathcal{T}_1 = \sum_{\nu,\tau,s} \sum_{\mathbf{k}} t_{\nu\tau s}(\mathbf{k}) c^\dagger_{\nu\tau}(\mathbf{k}) h^\dagger_s, \quad \mathcal{T}_{-1} = \mathcal{T}_1^\dagger. \tag{A4}$$

We then employ $H_0 = H_M + V_{\rm at}$ as the unperturbed Hamiltonian. The ground-state sector has a single hole at the spin-liquid site, and the intermediate states have either $\bar N = 0$ or $\bar N = 2$ holes, depending on whether $\mathcal{T}_{-1}$ or $\mathcal{T}_1$ is applied to a single-hole state. In the latter case, we have to distinguish between angular momentum channels with $L = 0, 1, 2$. Following Ref. [39], we use the notation $\mathcal{P}^{(n)}_L$ for the projection operators to states with angular momentum L and hole number $n = 0, 1, 2$. We omit the lower index for $n = 0, 1$ because in those cases there is only a single angular momentum channel. The projector to two-hole states is $\mathcal{P}^{(2)} = \sum_L \mathcal{P}^{(2)}_L$. For a lowest-order expansion in $H_{\rm tun}$, the Hilbert space can be truncated to have at most two holes at the magnetic layer site, $\mathbb{1} \simeq \mathcal{P}^{(0)} + \mathcal{P}^{(1)} + \mathcal{P}^{(2)}$.

Next we employ a canonical transformation to perform the projection to the low-energy sector, which is equivalent to a Schrieffer-Wolff transformation. Writing $\tilde H = e^S H e^{-S} = H + [S, H] + \cdots$, the first-order generator $S = S_1$ must then obey $[H_0, S_1] = H_{\rm tun}$. Using the commutators

$$\begin{aligned}
[H_0, \mathcal{P}^{(2)}_L \mathcal{T}_1 \mathcal{P}^{(1)}] &= \sum_{\nu,\tau,s} \sum_{\mathbf{k}} t_{\nu\tau s}(\mathbf{k}) [\Delta E_L + \varepsilon_{\nu\tau}(\mathbf{k})]\, c^\dagger_{\nu\tau}(\mathbf{k}) \mathcal{P}^{(2)}_L h^\dagger_s \mathcal{P}^{(1)}, \\
[H_0, \mathcal{P}^{(0)} \mathcal{T}_{-1} \mathcal{P}^{(1)}] &= \sum_{\nu,\tau,s} \sum_{\mathbf{k}} t^*_{\nu\tau s}(\mathbf{k}) [\Delta E_0 - \varepsilon_{\nu\tau}(\mathbf{k})]\, \mathcal{P}^{(0)} h_s \mathcal{P}^{(1)} c_{\nu\tau}(\mathbf{k}),
\end{aligned} \tag{A5}$$

and writing $S_1 = S_1^{(+)} - S_1^{(-)}$ with $S_1^{(-)} = S_1^{(+)\dagger}$, the part increasing the hole number at the magnetic site is

$$S_1^{(+)} = \sum_{\nu,\tau,s} \sum_{\mathbf{k}} c^\dagger_{\nu\tau}(\mathbf{k}) \left(-\frac{t_{\nu\tau s}(\mathbf{k})}{\Delta E_0 - \varepsilon_{\nu\tau}(\mathbf{k})} \mathcal{P}^{(1)} h^\dagger_s \mathcal{P}^{(0)} + \sum_L \frac{t_{\nu\tau s}(\mathbf{k})}{\Delta E_L + \varepsilon_{\nu\tau}(\mathbf{k})} \mathcal{P}^{(2)}_L h^\dagger_s \mathcal{P}^{(1)} \right). \tag{A6}$$

The excitation energies ΔE_L are given by

$$\Delta E_0 = \frac{U}{2} + J_H, \quad \Delta E_1 = \frac{U}{2} - 4J_H, \quad \Delta E_2 = \frac{U}{2} - 2J_H, \tag{A7}$$

where the energy for the transition to a state with zero holes is the same as for the transition to two holes with $L = 0$. The charge gap is set by the smallest of those energies, $E_g = \Delta E_1$.

The canonical transformation then results in the cotunneling Hamiltonian

$$H_{\rm cot} = -\tfrac{1}{2}\mathcal{P}^{(1)}(\mathcal{T}_{-1}S_1^{(+)} - \mathcal{T}_1 S_1^{(-)})\mathcal{P}^{(1)} + \text{H.c.}, \tag{A8}$$

which accurately describes the low-energy subspace with energy scales below E_g. Inserting the above expressions, we find the explicit representation

$$\begin{aligned} H_{\rm cot} = &-\frac{1}{2}\sum_{\nu_1,\tau_1,s_1}\sum_{\nu_2,\tau_2,s_2}\sum_{\mathbf{k}_1,\mathbf{k}_2}\frac{t_{\nu_2\tau_2 s_2}(\mathbf{k}_2)t^*_{\nu_1\tau_1 s_1}(\mathbf{k}_1)}{\Delta E_0 - \varepsilon_{\nu_1\tau_1}(\mathbf{k}_1)}\mathcal{P}^{(1)}h^\dagger_{s_2}h_{s_1}\mathcal{P}^{(1)}c^\dagger_{\nu_2\tau_2}(\mathbf{k}_2)c_{\nu_1\tau_1}(\mathbf{k}_1) \\ &-\frac{1}{2}\sum_{\nu_1,\tau_1,s_1}\sum_{\nu_2,\tau_2,s_2}\sum_{\mathbf{k}_1,\mathbf{k}_2}\sum_{L}\frac{t^*_{\nu_2\tau_2 s_2}(\mathbf{k}_2)t_{\nu_1\tau_1 s_1}(\mathbf{k}_1)}{\Delta E_L + \varepsilon_{\nu_1\tau_1}(\mathbf{k}_1)}\mathcal{P}^{(1)}h_{s_2}\mathcal{P}^{(2)}_L h^\dagger_{s_1}\mathcal{P}^{(1)}c_{\nu_2\tau_2}(\mathbf{k}_2)c^\dagger_{\nu_1\tau_1}(\mathbf{k}_1) + \text{H.c.} \end{aligned} \tag{A9}$$

We next compute the required matrix elements between spin-orbital states (where $\bar\sigma = -\sigma$ for $\sigma = \uparrow, \downarrow = +1, -1$),

$$\begin{aligned} &\langle s'|h^\dagger_{s_2}h_{s_1}|s\rangle = \delta_{s's_2}\delta_{ss_1}, \quad \langle s'|h_{s_2}\mathcal{P}^{(2)}_{L=0}h^\dagger_{s_1}|s\rangle = \tfrac{1}{3}\sigma_2\sigma_1\delta_{\alpha'\alpha_2}\delta_{\alpha\alpha_1}\delta_{\sigma_2\bar\sigma'}\delta_{\sigma_1\bar\sigma}, \\ &\langle s'|h_{s_2}\mathcal{P}^{(2)}_{L=1}h^\dagger_{s_1}|s\rangle = \tfrac{1}{2}(\delta_{\alpha_2\alpha_1}\delta_{\alpha'\alpha} - \delta_{\alpha_2\alpha}\delta_{\alpha'\alpha_1})(\delta_{\sigma_2\sigma_1}\delta_{\sigma\sigma'} + \delta_{\sigma_2\sigma}\delta_{\sigma_1\sigma'}), \\ &\langle s'|h_{s_2}\mathcal{P}^{(2)}_{L=2}h^\dagger_{s_1}|s\rangle = \delta_{s_2s_1}\delta_{ss'} - \delta_{s_2s}\delta_{s_1s'} - \langle s'|h_{s_2}\mathcal{P}^{(2)}_{L=0}h^\dagger_{s_1}|s\rangle - \langle s'|h_{s_2}\mathcal{P}^{(2)}_{L=1}h^\dagger_{s_1}|s\rangle. \end{aligned} \tag{A10}$$

We then obtain the matrix elements of $H_{\rm cot}$ in spin-orbital space as

$$\begin{aligned} (H_{\rm cot})_{s's} = &-\frac{1}{2}\sum_{\mathbf{k}_1\nu_1\tau_1}\sum_{\mathbf{k}_2\nu_2\tau_2}F_{s's}(\mathbf{k}_2,\nu_2,\tau_2;\mathbf{k}_1,\nu_1,\tau_1)c^\dagger_{\nu_2\tau_2}(\mathbf{k}_2)c_{\nu_1\tau_1}(\mathbf{k}_1) \\ &-\frac{1}{2}\sum_{\mathbf{k}_1\nu_1\tau_1}\sum_{\mathbf{k}_2\nu_2\tau_2}\sum_{L=0}^{2}G^L_{s's}(\mathbf{k}_2,\nu_2,\tau_2;\mathbf{k}_1,\nu_1,\tau_1)c_{\nu_2\tau_2}(\mathbf{k}_2)c^\dagger_{\nu_1\tau_1}(\mathbf{k}_1) + \text{H.c.}, \end{aligned} \tag{A11}$$

with the definitions

$$\begin{aligned} F_{s's}(\mathbf{k}_2,\nu_2,\tau_2;\mathbf{k}_1,\nu_1,\tau_1) &= \frac{t_{\nu_2\tau_2 s'}(\mathbf{k}_2)t^*_{\nu_1\tau_1 s}(\mathbf{k}_1)}{\Delta E_0 - \varepsilon_{\nu_1\tau_1}(\mathbf{k}_1)}, \\ G^L_{s's}(\mathbf{k}_2,\nu_2,\tau_2;\mathbf{k}_1,\nu_1,\tau_1) &= \sum_{s_1,s_2}\frac{t^*_{\nu_2\tau_2 s_2}(\mathbf{k}_2)t_{\nu_1\tau_1 s_1}(\mathbf{k}_1)}{\Delta E_L + \varepsilon_{\nu_1\tau_1}(\mathbf{k}_1)}\langle s'|h_{s_2}\mathcal{P}^{(2)}_L h^\dagger_{s_1}|s\rangle. \end{aligned} \tag{A12}$$

In a low-energy approach, we can now assume low energies, $|\varepsilon_{\nu\tau}(\mathbf{k})| \ll E_g$, for all conduction electron states involved in virtual processes. For simplicity, we also consider effectively $\mathbf{k}$-independent, spin-conserving, and spin-independent tunneling amplitudes,

$$t_{\nu\tau s}(\mathbf{k}) = t_{\nu\alpha}\delta_{\tau\sigma}, \tag{A13}$$

with $s = (\alpha, -\sigma)$. Tunneling between the substrate ($\nu = B$) and the magnetic layer is modeled by a featureless isotropic coupling, $t_{B\alpha} = t_B$. However, the tunnel couplings connecting the tip ($\nu = A$) to a magnetic site depend on the t_{2g} orbital (α) as well as on the relative position between tip and site. For definiteness, we model the t_{2g} orbitals by real wave functions with the proper symmetry. For instance, for the xy orbital centered at $\mathbf{R}_j = 0$, we take $\Phi_{xy}(\mathbf{r}') \propto x'y'e^{-|\mathbf{r}'|/l_d}$, where l_d sets the size of the orbital. Here the components of $\mathbf{r}'$ refer to the axes fixed by the octahedral environment of the magnetic ion; see Fig. 7(a). In these coordinates, the unit vectors for the conventional crystallographic directions are given by

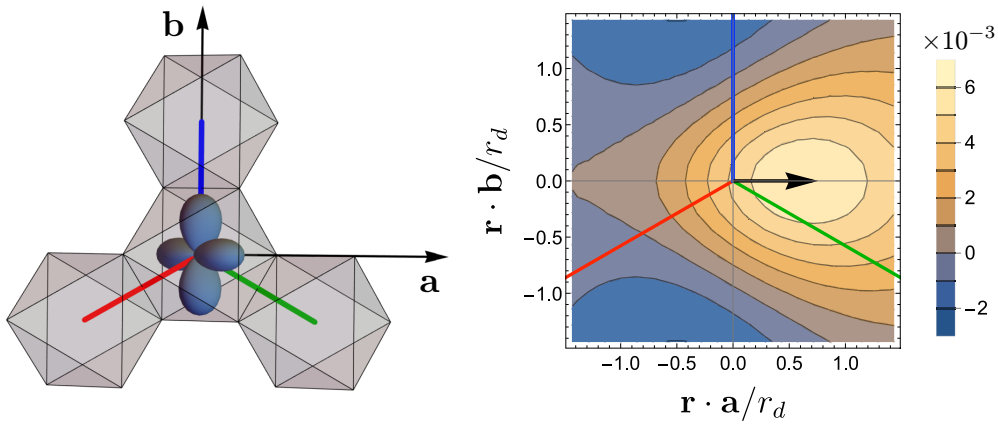
$$\mathbf{a} = \frac{1}{\sqrt{6}}\begin{pmatrix}1\\1\\-2\end{pmatrix}, \quad \mathbf{b} = \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\\0\end{pmatrix}, \quad \mathbf{c} = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}, \tag{A14}$$

where $\mathbf{c}$ is perpendicular to the honeycomb plane. As the wave function for the tip at position $\mathbf{r}$, we consider

$$\Phi_s(\mathbf{r}') \propto e^{-|\mathbf{r}'-\mathbf{r}|/l_s}, \tag{A15}$$

with characteristic length l_s.

In Fig. 7(b) we show the overlap between Φ_{xy} and Φ_s as a function of the tip position, keeping the tip height $\mathbf{r}\cdot\mathbf{c} > 0$ constant and varying the coordinates parallel to the honeycomb plane. The coordinates are scaled by the effective radius of the t_{2g} orbitals, $r_d = \int d^3r'\, r'|\Phi_\alpha(\mathbf{r}')|^2 = 7l_d/2$. We denote by $\mathbf{v}_\alpha$ the in-plane vector that corresponds to the relative position of maximum overlap between the tip and the α orbital. Note that $\mathbf{v}_\alpha$ lies in the direction perpendicular to the α bond. This shift in the position of maximum overlap can be



(a) (b)

FIG. 7. Orbital and spatial dependence of tunnel couplings. (a) xy orbital in the edge-sharing octahedra geometry of α-$RuCl_3$. The red, green, and blue lines represent the directions of x, y, and z bonds in the honeycomb plane, respectively. (b) Overlap between the xy orbital at $\mathbf{R}_j = 0$ and the wave function for electrons in the tip, modeled as an s orbital centered at position $\mathbf{r}$; see Eq. (A15). The arrow indicates the point of maximum overlap, corresponding to the vector $\mathbf{v}_z$. Here we set $l_s = 4l_d$ and $\mathbf{r}\cdot\mathbf{c} = 3l_d$.

interpreted in terms of the direction in which the α orbital points above the plane; see Fig. 7(a). Comparing the ionic radius of Ru^{3+} with the lattice spacing of α-$RuCl_3$, we estimate $|\mathbf{v}_\alpha| \approx 0.1a_0$. To capture the orbital and position dependence in the tunnel couplings within a simple analytical expression, we parametrize $t_{A\alpha}(\mathbf{r}, \mathbf{R}_j)$ as given in Eq. (2.3), with tunneling length $l_0 \sim l_s \lesssim a_0$.

For given $\mathbf{r}$ and $\mathbf{R}_j$, it is convenient to express the tunnel couplings $t_{A\alpha}$ in terms of spherical angles $\varphi \in [0, 2\pi)$ and $\theta \in [-\pi, \pi]$,

$$\begin{pmatrix} t_{Ax} \\ t_{Ay} \\ t_{Az} \end{pmatrix} = t_A \begin{pmatrix} \cos\varphi \sin\theta \\ \sin\varphi \sin\theta \\ \cos\theta \end{pmatrix}. \tag{A16}$$

Inserting the above expressions into Eq. (A12) and using Eq. (A10), we finally perform the projection to the $j_{\text{eff}} = 1/2$ subspace selected by the spin-orbit coupling. The corresponding basis states are [39]

$$|+\rangle = \frac{1}{\sqrt{3}}(-|z,\uparrow\rangle - i|y,\downarrow\rangle - |x,\downarrow\rangle),$$
$$|-\rangle = \frac{1}{\sqrt{3}}(|z,\downarrow\rangle + i|y,\uparrow\rangle - |x,\uparrow\rangle). \tag{A17}$$

The spin operator appearing in the Kitaev model for this site, $\mathbf{S} = \frac{1}{2}\boldsymbol{\sigma}$, acts in the space spanned by Eq. (A17). The cotunneling Hamiltonian follows as

$$H_{\text{cot}} = -\sum_{\mathbf{k}_1\nu_1}\sum_{\mathbf{k}_2\nu_2}\frac{t_{\nu_1}t_{\nu_2}}{2\Delta E_0}c^\dagger_{\nu_2}(\mathbf{k}_2)(f_0\mathbb{1} + \mathbf{f}\cdot\boldsymbol{\sigma})c_{\nu_1}(\mathbf{k}_1) - \sum_{\mathbf{k}_1\nu_1\tau_1}\sum_{\mathbf{k}_2\nu_2\tau_2}\sum_L \frac{t_{\nu_1}t_{\nu_2}}{2\Delta E_L}c_{\nu_2}(\mathbf{k}_2)\big(g^L_0\mathbb{1} + \mathbf{g}^L\cdot\boldsymbol{\sigma}\big)c^\dagger_{\nu_1}(\mathbf{k}_1) + \text{H.c.}, \tag{A18}$$

with f_0 and $\mathbf{f} = (f_x, f_y, f_z)$ given by

$$f_0 = \frac{F_{\uparrow\uparrow} + F_{\downarrow\downarrow}}{2}, \quad f_x = \frac{F_{\uparrow\downarrow} + F_{\downarrow\uparrow}}{2}, \quad f_y = i\frac{F_{\uparrow\downarrow} - F_{\downarrow\uparrow}}{2}, \quad f_z = \frac{F_{\uparrow\uparrow} - F_{\downarrow\downarrow}}{2}, \tag{A19}$$

and likewise for g^L_0 and $\mathbf{g}^L$. For given (σ, σ') indices, the 2×2 matrices $F_{\sigma\sigma'}$ and $G^L_{\sigma\sigma'}$ act in conduction electron spin space. We find

$$F_{\uparrow\uparrow} = \frac{1}{3}\begin{pmatrix} \cos\theta & (1+i)\cos\theta \\ e^{-i\varphi}\sin\theta & (1+i)e^{-i\varphi}\sin\theta \end{pmatrix}, \quad F_{\uparrow\downarrow} = \frac{1}{3}\begin{pmatrix} (1-i)\cos\theta & -\cos\theta \\ (1-i)e^{-i\varphi}\sin\theta & -e^{-i\varphi}\sin\theta \end{pmatrix}, \quad G^0_{\uparrow\uparrow} = \frac{F_{\downarrow\downarrow}}{3}, \quad G^0_{\uparrow\downarrow} = -\frac{F_{\uparrow\downarrow}}{3},$$

$$G^1_{\uparrow\uparrow} = \frac{1}{6}\begin{pmatrix} (1-i)e^{i\varphi}\sin\theta + 2\cos\theta & (1+i)\cos\theta \\ e^{-i\varphi}\sin\theta & (\sin\varphi+\cos\varphi)\sin\theta \end{pmatrix}, \quad G^1_{\uparrow\downarrow} = \frac{1}{6}\begin{pmatrix} e^{-i\varphi}\sin\theta & (\sin\varphi+\cos\varphi)\sin\theta \\ (1-i)e^{-i\varphi}\sin\theta & -(1-i)\cos\theta \end{pmatrix},$$

$$G^2_{\uparrow\uparrow} = [(\cos\varphi+\sin\varphi)\sin\theta + \cos\theta]\mathbb{1} - F_{\uparrow\uparrow} - G^0_{\uparrow\uparrow} - G^1_{\uparrow\uparrow}, \quad G^2_{\uparrow\downarrow} = -F_{\uparrow\downarrow} - G^0_{\uparrow\downarrow} - G^1_{\uparrow\downarrow}. \tag{A20}$$

The remaining matrices are obtained by using a time-reversal operation,

$$F_{\downarrow\downarrow} = \tau_y F^*_{\uparrow\uparrow}\tau_y, \quad F_{\downarrow\uparrow} = -\tau_y F^*_{\uparrow\downarrow}\tau_y, \quad G^L_{\downarrow\downarrow} = \tau_y (G^L)^*_{\uparrow\uparrow}\tau_y, \quad G^L_{\downarrow\uparrow} = -\tau_y (G^L)^*_{\uparrow\downarrow}\tau_y, \tag{A21}$$

with Pauli matrices $\boldsymbol{\tau}$ in conduction electron spin space. In the second term of Eq. (A18), we now use

$$c_{\nu_2\tau_2}(\mathbf{k}_2)c^\dagger_{\nu_1\tau_1}(\mathbf{k}_1) = -c^\dagger_{\nu_1\tau_1}(\mathbf{k}_1)c_{\nu_2\tau_2}(\mathbf{k}_2) + \delta_{\nu_1\nu_2}\delta_{\tau_1\tau_2}\delta_{\mathbf{k}_1\mathbf{k}_2}.$$

The factor $\delta_{\tau_1\tau_2}$ in the last term implies a trace over the 2×2 matrices for conduction electrons. As a result, only the identity can contribute. We thereby obtain the cotunneling Hamiltonian (2.2), where $\Psi_A(\mathbf{r}) = \sum_{\mathbf{k}} c_A(\mathbf{k})$ is a real-space two-component spinor field describing conduction electrons on the tip at position $\mathbf{r}$. Likewise, $\Psi_B(\mathbf{R})$ refers to the substrate spinor field below the site with position $\mathbf{R}$. Cotunneling processes are then characterized by the transition matrices T_0 and T^α, with $\mathbf{T} = (T^x, T^y, T^z)$, which act in conduction electron spin space and are given by

$$T_0 = -\frac{t_A t_B}{\Delta E_0} f_0 + \sum_{L=0}^{2} \frac{t_A t_B}{\Delta E_L} g_0^L,$$

$$\mathbf{T} = -\frac{t_A t_B}{\Delta E_0}\mathbf{f} + \sum_{L=0}^{2} \frac{t_A t_B}{\Delta E_L}\mathbf{g}^L. \tag{A22}$$

All matrix elements scale $\propto t_A t_B/U$, where individual contributions carry $\frac{J_H}{U}$-dependent factors. We emphasize that T_0 and $\mathbf{T}$ depend on $\mathbf{r}-\mathbf{R}_j$, with the tip (site) position $\mathbf{r}$ ($\mathbf{R}_j$).

The above expressions can be simplified considerably when neglecting the orbital-dependent shifts $\mathbf{v}_\alpha$ in Eq. (2.3). This approximation becomes exact for a tip placed right on top of a magnetic site, and otherwise causes quantitative ($\approx$10%) deviations in the tunnel couplings. We then obtain

$$f_0 = \frac{1}{2\sqrt{3}}\mathbb{1}, \quad f_\alpha = \frac{1}{3\sqrt{3}}(\tau^x+\tau^y+\tau^z) - \frac{1}{2\sqrt{3}}\tau^\alpha,$$
$$g_0^0 = \frac{1}{6\sqrt{3}}\mathbb{1}, \quad g_\alpha^0 = -\frac{29}{2\sqrt{3}}(\tau^x+\tau^y+\tau^z) + \frac{1}{6\sqrt{3}}\tau^\alpha,$$
$$g_0^1 = \frac{1}{2\sqrt{3}}\mathbb{1}, \quad g_\alpha^1 = \frac{13}{2\sqrt{3}}\tau^\alpha,$$
$$g_0^2 = \frac{23}{3\sqrt{3}}\mathbb{1}, \quad g_\alpha^2 = -\frac{55}{9\sqrt{3}}(\tau^x+\tau^y+\tau^z) + \frac{26}{3\sqrt{3}}\tau^\alpha,$$

and $H_{\rm cot}$ takes the form (2.4), where we define the $\frac{J_H}{U}$-dependent coefficients ($j = 0, 1, 2$)

$$\eta_j = \frac{U}{2\sqrt{3}\Delta E_0}\zeta_j + \sum_{L=0}^{2}\frac{U}{2\sqrt{3}\Delta E_L}\zeta_j^L \tag{A23}$$

with ΔE_L in Eq. (A7) and the numbers

$$\zeta_0 = 1, \quad \zeta_0^1 = \frac{1}{3}, \quad \zeta_0^2 = 1, \quad \zeta_0^3 = \frac{46}{3},$$
$$\zeta_1 = -\frac{1}{2}, \quad \zeta_2 = \frac{1}{3}, \quad \zeta_1^0 = \frac{1}{6}, \quad \zeta_2^0 = -\frac{1}{9}, \quad \zeta_1^1 = 0,$$
$$\zeta_2^1 = \frac{1}{6}, \quad \zeta_1^2 = \frac{1}{3}, \quad \zeta_2^2 = -\frac{7}{18}.$$

[1] A. Kitaev, Ann. Phys. **321**, 2 (2006).
[2] G. Jackeli and G. Khaliullin, Phys. Rev. Lett. **102**, 017205 (2009).
[3] L. Savary and L. Balents, Rep. Prog. Phys. **80**, 016502 (2017).
[4] Y. Zhou, K. Kanoda, and T.-K. Ng, Rev. Mod. Phys. **89**, 025003 (2017).
[5] X. G. Wen, Rev. Mod. Phys. **89**, 041004 (2017).
[6] S. M. Winter, A. A. Tsirlin, M. Daghofer, J. van den Brink, Y. Singh, P. Gegenwart, and R. Valentí, J. Phys.: Condens. Matter **29**, 493002 (2017).
[7] M. Hermanns, I. Kimchi, and J. Knolle, Annu. Rev. Condens. Matter Phys. **9**, 17 (2018).
[8] J. Knolle and R. Moessner, Annu. Rev. Condens. Matter Phys. **10**, 451 (2019).
[9] H. Takagi, T. Takayama, G. Jackeli, G. Khaliullin, and S. E. Nagler, Nat. Rev. Phys. **1**, 264 (2019).
[10] Y. Motome and J. Nasu, J. Phys. Soc. Jpn. **89**, 012002 (2020).
[11] C. Broholm, R. J. Cava, S. A. Kivelson, D. G. Nocera, M. R. Norman, and T. Senthil, Science **367**, eaay0668 (2020).
[12] S. Trebst and C. Hickey, Phys. Rep. **950**, 1 (2022).
[13] Y. Kasahara, T. Ohnishi, Y. Mizukami, O. Tanaka, S. Ma, K. Sugii, N. Kurita, H. Tanaka, J. Nasu, Y. Motome, T. Shibauchi, and Y. Matsuda, Nature (London) **559**, 227 (2018).
[14] T. Yokoi, S. Ma, Y. Kasahara, S. Kasahara, T. Shibauchi, N. Kurita, H. Tanaka, J. Nasu, Y. Motome, C. Hickey, S. Trebst, and Y. Matsuda, Science **373**, 568 (2021).
[15] J. Bruin, R. Claus, Y. Matsumoto, N. Kurita, H. Tanaka, and H. Takagi, Nat. Phys. **18**, 401 (2022).
[16] P. Czajka, T. Gao, M. Hirschberger, P. Lampen-Kelley, A. Banerjee, J. Yan, D. G. Mandrus, S. E. Nagler, and N. Ong, Nat. Phys. **17**, 915 (2021).
[17] P. Czajka, T. Gao, M. Hirschberger, Paula Lampen-Kelley, A. Banerjee, N. Quirk, D. G. Mandrus, S. E. Nagler, and N. P. Ong, Nat. Mater. **22**, 36 (2023).
[18] L. E. Chern, E. Z. Zhang, and Y. B. Kim, Phys. Rev. Lett. **126**, 147201 (2021).
[19] E. Z. Zhang, L. E. Chern, and Y. B. Kim, Phys. Rev. B **103**, 174402 (2021).
[20] D. Wulferding, Y. Choi, S. H. Do, C. H. Lee, P. Lemmens, C. Faugeras, Y. Gallais, and K. Y. Choi, Nat. Commun. **11**, 1603 (2020).
[21] J.-X. Yin, S. H. Pan, and M. Zahid Hasan, Nat. Rev. Phys. **3**, 249 (2021).
[22] M. Ziatdinov, A. Banerjee, A. Maksov, T. Berlijn, W. Zhou, H. B. Cao, J.-Q. Yan, C. A. Bridges, D. G. Mandrus, S. E. Nagler, A. P. Baddorf, and S. V. Kalinin, Nat. Commun. **7**, 13774 (2016).
[23] D. Weber, L. M. Schoop, V. Duppel, J. M. Lippmann, J. Nuss, and B. V. Lotsch, Nano Lett. **16**, 3578 (2016).
[24] L. Du, Y. Huang, Y. Wang, Q. Wang, R. Yang, J. Tang, M. Liao, D. Shi, Y. Shi, and X. Zhou, 2D Mater. **6**, 015014 (2018).
[25] W. Ruan, Y. Chen, S. Tang, J. Hwang, H.-Z. Tsai, R. L. Lee, M. Wu, H. Ryu, S. Kahn, F. Liou, C. Jia, A. Aikawa, C. Hwang, F. Wang, Y. Choi, S. G. Louie, P. A. Lee,

Z.-X. Shen, S.-K. Mo, and M. F. Crommie, Nat. Phys. **17**, 1154 (2021).
[26] J. Alicea, Rep. Prog. Phys. **75**, 076501 (2012).
[27] K. Sengupta, I. Žutić, H. J. Kwon, V. M. Yakovenko, and S. Das Sarma, Phys. Rev. B **63**, 144531 (2001).
[28] K. T. Law, P. A. Lee, and T. K. Ng, Phys. Rev. Lett. **103**, 237001 (2009).
[29] K. Flensberg, Phys. Rev. B **82**, 180516(R) (2010).
[30] A. Zazunov, R. Egger, and A. Levy Yeyati, Phys. Rev. B **94**, 014502 (2016).
[31] T. Machida, Y. Sun, S. Pyon, S. Takeda, Y. Kohsaka, T. Hanaguri, T. Sasagawa, and T. Tamegai, Nat. Mater. **18**, 811 (2019).
[32] Q. Liu, C. Chen, T. Zhang, R. Peng, Y. J. Yan, C.-H.-P. Wen, X. Lou, Y.-L. Huang, J.-P. Tian, X.-L. Dong, G.-W. Wang, W.-C. Bao, Q.-H. Wang, Z.-P. Yin, Z.-X. Zhao, and D.-L. Feng, Phys. Rev. X **8**, 041056 (2018).
[33] L. Kong, S. Zhu, M. Papaj, H. Chen, L. Cao, H. Isobe, Y. Xing, W. Liu, D. Wang, P. Fan, Y. Sun, S. Du, J. Schneeloch, R. Zhong, G. Gu, L. Fu, H. J. Gao, and H. Ding, Nat. Phys. **15**, 1181 (2019).
[34] S. Zhu, L. Kong, L. Cao, H. Chen, M. Papaj, S. Du, Y. Xing, W. Liu, D. Wang, C. Shen, F. Yang, J. Schneeloch, R. Zhong, G. Gu, L. Fu, Y. Y. Zhang, H. Ding, and H. J. Gao, Science **367**, 189 (2020).
[35] E. Prada, P. San-Jose, M. W. A. de Moor, A. Geresdi, E. J. H. Lee, J. Klinovaja, D. Loss, J. Nygård, R. Aguado, and L. P. Kouwenhoven, Nat. Rev. Phys. **2**, 575 (2020).
[36] J. G. Rau, Eric Kin-Ho Lee, and H.-Y. Kee, Phys. Rev. Lett. **112**, 077204 (2014).
[37] J. G. Rau, E. K.-H. Lee, and H.-Y. Kee, Annu. Rev. Condens. Matter Phys. **7**, 195 (2016).
[38] S. M. Winter, Y. Li, H. O. Jeschke, and R. Valentí, Phys. Rev. B **93**, 214431 (2016).
[39] R. G. Pereira and R. Egger, Phys. Rev. Lett. **125**, 227202 (2020).
[40] J. Fernández-Rossier, Phys. Rev. Lett. **102**, 256802 (2009).
[41] J. Fransson, O. Eriksson, and A. V. Balatsky, Phys. Rev. B **81**, 115454 (2010).
[42] F. Delgado and J. Fernández-Rossier, Phys. Rev. B **84**, 045439 (2011).
[43] J. Feldmeier, W. Natori, M. Knap, and J. Knolle, Phys. Rev. B **102**, 134423 (2020).
[44] E. J. König, M. T. Randeria, and B. Jäck, Phys. Rev. Lett. **125**, 267206 (2020).
[45] M. Carrega, I. J. Vera-Marun, and A. Principi, Phys. Rev. B **102**, 085412 (2020).
[46] G. Chen and J. L. Lado, Phys. Rev. Res. **2**, 033466 (2020).
[47] M. Udagawa, S. Takayoshi, and T. Oka, Phys. Rev. Lett. **126**, 127201 (2021).
[48] G. Baskaran, S. Mandal, and R. Shankar, Phys. Rev. Lett. **98**, 247201 (2007).
[49] F. L. Pedrocchi, S. Chesi, and D. Loss, Phys. Rev. B **84**, 165414 (2011).
[50] J. Knolle, D. L. Kovrizhin, J. T. Chalker, and R. Moessner, Phys. Rev. Lett. **112**, 207203 (2014).
[51] F. Zschocke and M. Vojta, Phys. Rev. B **92**, 014403 (2015).
[52] X.-Y. Song, Y.-Z. You, and L. Balents, Phys. Rev. Lett. **117**, 037209 (2016).
[53] D. Otten, A. Roy, and F. Hassler, Phys. Rev. B **99**, 035137 (2019); however, one can easily take into account such excitations.
[54] D. Aasen, R. S. K. Mong, B. M. Hunt, D. Mandrus, and J. Alicea, Phys. Rev. X **10**, 031014 (2020).
[55] M. G. Yamada and S. Fujimoto, Phys. Rev. Lett. **127**, 047201 (2021).
[56] R. Chari, R. Moessner, and J. G. Rau, Phys. Rev. B **103**, 134444 (2021).
[57] S. Banerjee and S. Z. Lin, arXiv:2208.06887.
[58] F. Elste and C. Timm, Phys. Rev. B **75**, 195341 (2007).
[59] D. A. S. Kaib, S. Biswas, K. Riedl, S. M. Winter, and R. Valentí, Phys. Rev. B **103**, L140402 (2021).
[60] S. M. Winter, K. Riedl, P. A. Maksimov, A. L. Chernyshev, A. Honecker, and R. Valentí, Nat. Commun. **8**, 1152 (2017).
[61] Y. Sugita, Y. Kato, and Y. Motome, Phys. Rev. B **101**, 100410(R) (2020).
[62] B. Yang, Y. M. Goh, S. H. Sung, G. Ye, S. Biswas, D. A. S. Kaib, R. Dhakal, S. Yan, C. Li, S. Jiang, F. Chen, H. Lei, R. He, R. Valentí, S. M. Winter, R. Hovden, and A. W. Tsen, Nat. Mater. **22**, 50 (2023).
[63] Y. S. Hou, H. J. Xiang, and X. G. Gong, Phys. Rev. B **96**, 054410 (2017).
[64] V. M. Katukuri, S. Nishimoto, V. Yushankhai, A. Stoyanova, H. Kandpal, S. Choi, R. Coldea, I. Rousochatzakis, L. Hozoi, and J. van den Brink, New J. Phys. **16**, 013056 (2014).
[65] Y. Yamaji, Y. Nomura, M. Kurita, R. Arita, and M. Imada, Phys. Rev. Lett. **113**, 107201 (2014).
[66] V. M. Katukuri, S. Nishimoto, I. Rousochatzakis, H. Stoll, J. van den Brink, and L. Hozoi, Sci. Rep. **5**, 14718 (2015).
[67] K. Dhochak, R. Shankar, and V. Tripathi, Phys. Rev. Lett. **105**, 117201 (2010).
[68] A. J. Willans, J. T. Chalker, and R. Moessner, Phys. Rev. B **84**, 115146 (2011).
[69] M. Vojta, A. K. Mitchell, and F. Zschocke, Phys. Rev. Lett. **117**, 037202 (2016).
[70] J.-P. Blaizot and G. Ripka, *Quantum Theory of Finite Systems*, Vol. 3 (MIT Press, Cambridge, MA, 1986).
[71] M. Cozzini, P. Giorda, and P. Zanardi, Phys. Rev. B **75**, 014439 (2007).
[72] G. Bertsch, J. Dobaczewski, W. Nazarewicz, and J. Pei, Phys. Rev. A **79**, 043602 (2009).
[73] Some of the relations in Sec. III, in particular, Eqs. (3.22) and (3.23), assume that $|0_a\rangle$ and $|0_b\rangle$ are the respective vacuum states without extra excitations to satisfy the fermion parity constraint. The corresponding expressions in the presence of such excitations follow accordingly.
[74] J. Knolle, R. Moessner, and N. B. Perkins, Phys. Rev. Lett. **122**, 047202 (2019).
[75] W.-H. Kao, J. Knolle, G. B. Hálasz, R. Moessner, and N. B. Perkins, Phys. Rev. X **11**, 011034 (2021).
[76] V. Dantas and E. C. Andrade, Phys. Rev. Lett. **129**, 037204 (2022).
[77] Y. Imry and S. Ma, Phys. Rev. Lett. **35**, 1399 (1975).
[78] D. Rhodes, S. H. Chae, R. Ribeiro-Palau, and J. Hone, Nat. Mater. **18**, 541 (2019).