Topological invariant for multiband non-Hermitian systems with chiral symmetry

Chun-Chi Liu,[1] Liu-Hao Li,[1] and Jin An[1,2,*]

[1]*National Laboratory of Solid State Microstructures, Department of Physics, Nanjing University, Nanjing 210093, China*
[2]*Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing 210093, China*

Topology plays an important role in non-Hermitian systems. How to characterize a non-Hermitian topological system under open-boundary conditions (OBCs) is a challenging problem. A one-dimensional (1D) topological invariant defined on a generalized Brillion zone (GBZ) was recently found to successfully describe the topological property of the two-band Su-Schrieffer-Heeger model. But for a 1D multiband chiral symmetric system under OBCs, it is still controversial how to define the topological invariant. We show in this paper by exact proof and detailed demonstration that to acquire the topological invariant for multiband non-Hermitian models with chiral symmetry, the GBZ as the integral domain should be replaced by a more generalized closed loop. Our work thus establishes the non-Bloch bulk-boundary correspondence for 1D multiband chiral symmetric non-Hermitian systems.

I. INTRODUCTION

Recent studies on non-Hermitian systems have revealed many new concepts and phenomena [1–13], such as non-Hermitian skin effect [14–27], distinct differences between the open-boundary spectra (OBS) and the corresponding periodic-boundary spectra [28–32], and coalescence of states, i.e., exception points [28,31,33–42]. Topology is also found to play an increasingly important role in non-Hermitian systems [11,14,22,31,37,43–75]. The central problem of non-Hermitian topological systems is the non-Bloch bulk-boundary correspondence (BBC). Different from the Hermitian counterparts [76–79], or non-Hermitian systems with half-infinite boundary conditions, where the topological invariants can be defined on the Brillouin zone (BZ) [53,55], we have no fundamental principles to guide us to find a topological invariant to characterize a non-Hermitian topological system under open-boundary conditions (OBCs). Even in one dimension, the non-Bloch BBC has not been well solved yet.

For a one-dimensional (1D) non-Hermitian Su-Schrieffer-Heeger model, a pioneer work recently proposed that the topological invariant should be redefined on a generalized Brillouin zone (GBZ) to capture the correct non-Bloch BBC [14]. The idea of replacement of BZ by GBZ has been further applied to two-dimensional systems and has been proven to be very successful [48,71,80]. This is feasible because in these models the GBZ is a unique closed loop or a unique closed surface. However, for 1D multiband non-Hermitian systems under OBCs, the situation changes completely. Generally, one has multiple sub-GBZs [81] and especially for systems with chiral symmetry, the OBS consists of N pairs of open arcs, and each pair is centrosymmetric about $E = 0$ and corresponds to a closed loop (sub-GBZ). It is found that in some models,

choosing one or some of the sub-GBZs to replace BZ in the definition of the topological invariant could not give the non-Bloch BBC. On the other hand, the number of topologically protected edge states (TPESs) changes even when the system is under a topologically trivial variation [47,82,83]. This means that the number of TPESs cannot be the candidate of the invariant. Thus, how to find out the topological invariant of the 1D multiband non-Hermitian systems with chiral symmetry becomes an open problem.

In this paper we solve this problem by redefining the topological invariant ν on a more generalized closed loop $\mathcal{L}_\beta$ in the complex β plane and then by exactly proving the relevant statement. We show explicitly that the number of TPESs for systems characterized by ν can vary between $|\nu|$ and $2|\nu|$, and clarify the origin of the defectiveness of the TPESs. We discuss the implications of the conditions on the topological phase transitions imposed by our topological invariant.

This paper is organized as follows. In Sec. II, we give a theorem on our topological invariant and then demonstrate by a simple model the consistency between the occurrence of the TPESs and the invariant. In Sec. III, we give the exact proof of our theorem and we also reveal the origin of the defectiveness of the TPESs. In Sec. IV, we discuss the topological phase transitions and their implications imposed by the invariant. In Sec. V, we summarize our results.

II. TOPOLOGICAL INVARIANT AND THE THEOREM

We start from a general non-Hermitian tight-binding model given by

$$H = \sum_j \sum_{m=-M_2}^{M_1} \sum_{\mu,\nu=1}^{2N} |j+m\rangle_\mu (T_m)_{\mu\nu} \langle j|_\nu, \qquad (1)$$

where T_m is the $2N \times 2N$ hopping matrix between the mth nearest-neighbor unit cells, $|j\rangle_\mu$ is the localized μ orbital state

at unit cell j, and M_1 and M_2 denote the hopping ranges along the forward and backward directions, respectively. In terms of $\beta = e^{ik}$, the non-Bloch Hamiltonian of H can be expressed as

$$H(\beta) = \sum_{m=-M_2}^{M_1} T_m \beta^{-m}. \tag{2}$$

Since for a given energy E, $\det[H(\beta) - E]$ can generally be viewed as a polynomial of β and the corresponding secular equation $\det[H(\beta) - E] = 0$ has $(p + q)$ β roots $\beta_i(E)$ with $p = 2NM_1$ and $q = 2NM_2$, we have

$$\det[H(\beta) - E] = \frac{\prod_{i=1}^{p+q} \lceil \beta - \beta_i(E) \rceil}{\beta^p}. \tag{3}$$

These β roots are assumed to be ordered in absolute value as $|\beta_1(E)| \leqslant \ldots \leqslant |\beta_{p+q}(E)|$. Each $\beta_i(E)$ corresponds to an eigenvector ϕ_i, obeying the eigenequation

$$H[\beta_i(E)]\phi_i = E\phi_i, \tag{4}$$

where $i = 1, \ldots, p + q$. A state of energy E belonging to the continuum OBS must obey $|\beta_p(E)| = |\beta_{p+1}(E)|$ [28]. If the system holds chiral symmetry, i.e., there exists a chiral operator $\mathcal{S}$ satisfying $\{\mathcal{S}, H(\beta)\} = 0$ with $\mathcal{S}^2 = 1$, then $H(\beta)$ can always take the following representation ($\mathcal{S} = \sigma_z \otimes I_N$):

$$H(\beta) = \begin{pmatrix} 0 & R_+(\beta) \\ R_-(\beta) & 0 \end{pmatrix}, \tag{5}$$

where $R_\pm$ is a $N \times N$ matrix. When the open system is gapful, the $E = 0$ state would not belong to the continuum OBS, which means $|\beta_p(0)| \neq |\beta_{p+1}(0)|$. When the system is gapless, the $E = 0$ state would be connected to the continuum OBS, with $|\beta_p(0)| = |\beta_{p+1}(0)|$.

In order to study the topological property of this non-Hermitian system, it is very meaningful to consider a class in which all the systems are topologically equivalent to it. Any two systems in the class under OBCs can be deformed continuously to each other without closing the gap. We denote this topologically equivalent class as $\mathcal{C}_\nu$. Non-Hermitian systems $H(\rho\beta)$ with $0 < \rho < \infty$, which can be connected continuously by similarity transformations to $H(\beta)$, obviously belong to $\mathcal{C}_\nu$, since they share exactly the same OBS [32]. However, different systems in the class may have a different number of TPESs. If $\mathcal{C}_\nu$ contains a Hermitian system, then this system under OBCs is expected to have the maximum number of the TPESs in the class. For any closed loop $\mathcal{L}_\beta$ one can define the following winding number:

$$\begin{aligned} \nu &= \frac{1}{2\pi i} \oint_{\mathcal{L}_\beta} \text{Tr}(q^{-1}dq) \\ &= \frac{1}{4\pi i} \oint_{\mathcal{L}_\beta} d\beta \frac{d}{d\beta} \{\ln \det[R_+(\beta)] - \ln \det[R_-(\beta)]\}, \end{aligned} \tag{6}$$

where q comes from the standard occupation projection Q operator [76]. Our theorem is as follows: If the closed loop $\mathcal{L}_\beta$ encircles the first p β roots of $E = 0$ with the other q β roots being kept outside, the winding number ν gives the topological invariant that characterizes the topological feature of the 1D chiral-symmetric non-Hermitian systems under open boundary conditions, i.e., $|\nu|$ is equal to the minimum

number of TPESs of the systems in the topologically equivalent class $\mathcal{C}_\nu$ and the number of TPESs of any system in $\mathcal{C}_\nu$ is no more than $2|\nu|$.

Obviously, $\mathcal{L}_\beta$ is not unique. A useful and convenient selection of $\mathcal{L}_\beta$, particularly in numerical computation of the invariant, is a perfect circle centered at the origin $\beta = 0$ with radius ρ obeying $|\beta_p(0)| < \rho < |\beta_{p+1}(0)|$. This is consistent with Refs. [58,84,85], where the topological invariant is proposed to be defined on a perfect circle with its radius determined self-consistently. On the other hand, if one of the sub-GBZs just contains the first p β roots, it can also serve as a possible $\mathcal{L}_\beta$.

$N = 1$ is a special case which corresponds to the two-band models and has been discussed previously in Ref. [14], where the GBZ C_β is proposed as the closed loop in the integral of the winding number. This is consistent with our theorem, since the GBZ has been proven to be the boundary between the $\beta_p(E)$ and $\beta_{p+1}(E)$ regions for all E [86]. So for any state of E not on the continuum OBS, the GBZ exactly encircles only the first p β roots of E, so the GBZ is actually one of the possible $\mathcal{L}_\beta$ defined here. However, when $N \geqslant 2$, the situation becomes much more complicated since the system generally has $2N$ bands and the GBZ is composed of N closed loops (sub-GBZs). One or several of the sub-GBZs, or even all the sub-GBZs may be chosen as the integral loop(s), but they are all found to be questionable. Therefore, our theorem can be viewed as solving this problem in the case of multiband chiral-symmetric non-Hermitian systems by choosing $\mathcal{L}_\beta$ instead of the GBZ as the integral loop in the definition of the topological invariant.

We take a four-band ($N = 2$) model as an example to illustrate the theorem. The variations of the OBS with parameter λ are shown in Fig. 1(a), while the topological invariant ν calculated by Eq. (6) is shown in Fig. 1(b). The nontrivial regions given by ν agree very well with that predicted by the direct numerical diagonalization of the open chain of the system. For more details, see Appendix D.

III. PROOF OF THE THEOREM

Here we give the proof of our theorem. We will derive the solutions of the isolated $E = 0$ state under OBCs and then connect its degeneracy to the topological invariant ν we defined. For a gapful system, the $E = 0$ state does not belong to the continuum OBS, which means $|\beta_p(0)| \neq |\beta_{p+1}(0)|$. We start from a system with property $|\beta_p(0)| < 1 < |\beta_{p+1}(0)|$. If the non-Hermitian system $H(\beta)$ we study does not have this property, one can always choose a $H(\rho\beta)$ system as the starting one, since its β roots of $E = 0$ are just those of $H(\beta)$, but scaled by a factor ρ^{-1}. Thus for a 1D open chain with L unit cells, the wave function of a state with energy $E = 0$ can be expanded as the superposition of its $p + q$ eigenmodes ϕ_i:

$$\psi(j) = \sum_{i=1}^{p} c_i[\beta_i(0)]^j \phi_i + \sum_{i=1}^{q} c_{i+p}[\beta_{p+i}(0)]^{j-L}\phi_{p+i}, \tag{7}$$

where j represents the jth unit cell with $1 \leqslant j \leqslant L$, and c_i is the superposition coefficient of the eigenmodes ϕ_i.

Now we turn to the boundary conditions. We consider a simplified case where both T_{M_1} and T_{-M_2} are invertible. If this
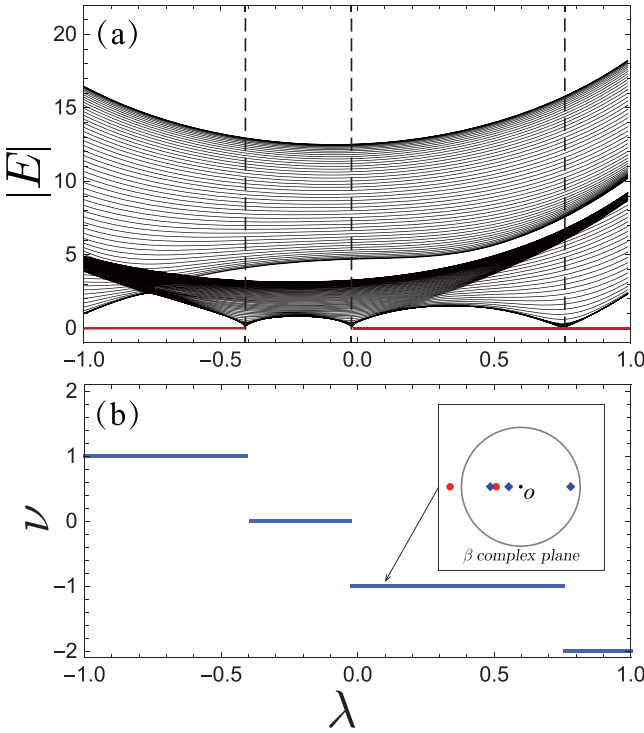
FIG. 1. An example exhibiting the consistency between the occurrence of zero modes and our topological invariant. (a) Absolute value of energy of the states under OBCs as a function of parameter λ, where the thick lines denote the topological zero modes. (b) Topological invariant ν versus λ. Here $R_\pm = t_0^\pm + t_x^\pm \sigma_x + t_y^\pm \sigma_y + t_z^\pm \sigma_z$, where $t_0^+ = 4 + 1.5\beta^+$, $t_x^+ = 4 + 2.5/\beta^+$, $t_y^+ = -i + 0.5i/\beta^+$, $t_z^+ = 3 + 0.5\beta^+$ and $t_0^- = 4.5 + 1.5/\beta^-$, $t_x^- = 3 + 2.5\beta^-$, $t_y^- = -i + 0.5i\beta^-$, $t_z^- = 2.5 + 0.5/\beta^-$ with $\beta^\pm = 4^{\mp\lambda}\beta$. The inset gives the perfect circle in the β plane used to calculate the invariant at a certain λ, with the solid circles (diamonds) being β roots of $E = 0$ from $\det R_+ = 0$ ($\det R_- = 0$).

is not the case in practice, for a gapful system, one can slightly change them to be invertible without closing the gap and so without changing the topological property of the system. By introducing M_1 extra unit cells near the left boundary and M_2 extra unit cells near the right boundary, the range of j in $\psi(j)$ is changed to be $-M_1 + 1 \leqslant j \leqslant L + M_2$, and the OBCs now become (see Appendix A)

$$\psi(0) = \psi(-1) = \ldots = \psi(-M_1 + 1) = 0,$$
$$\psi(L + 1) = \psi(L + 2) = \ldots = \psi(L + M_2) = 0. \quad (8)$$

Below we first study the case in the thermodynamic limit and then we take the effect of a finite and sufficiently large L into account. As $L \to \infty$, the OBCs can be expressed as

$$\sum_{i=1}^{p} c_i [\beta_i(0)]^j \phi_i = 0, \qquad j = 0, -1, \ldots, -M_1 + 1,$$

$$\sum_{i=1}^{q} c_{i+p} [\beta_{i+p}(0)]^j \phi_{i+p} = 0, \qquad j = 1, 2, \ldots, M_2. \quad (9)$$

These two sets of equations can be reexpressed as $\mathcal{M}_\mathcal{L}|\mathcal{L}\rangle = 0$ and $\mathcal{M}_\mathcal{R}|\mathcal{R}\rangle = 0$, where $|\mathcal{L}\rangle = (c_1, c_2, \ldots, c_p)^T$ and $|\mathcal{R}\rangle = (c_{p+1}, c_{p+2}, \ldots, c_{p+q})^T$, respectively, and the $p \times p$ matrix

$\mathcal{M}_\mathcal{L}$ and the $q \times q$ matrix $\mathcal{M}_\mathcal{R}$ are given below:

$$\mathcal{M}_\mathcal{L} = (\Phi_{1,\mathcal{L}}, \Phi_{2,\mathcal{L}}, \ldots, \Phi_{p,\mathcal{L}}),$$
$$\mathcal{M}_\mathcal{R} = (\Phi_{p+1,\mathcal{R}}, \Phi_{p+2,\mathcal{R}}, \ldots, \Phi_{p+q,\mathcal{R}}), \quad (10)$$

with $\Phi_{i,\mathcal{L}} \equiv (\phi_i, \beta_i^{-1}\phi_i, \ldots, \beta_i^{-M_1+1}\phi_i)^T$, $\Phi_{i,\mathcal{R}} \equiv (\beta_i\phi_i, \beta_i^2\phi_i, \ldots, \beta_i^{M_2}\phi_i)^T$. Whether $\mathcal{M}_\mathcal{L}$ or $\mathcal{M}_\mathcal{R}$ is singular determines directly whether there exists an $E = 0$ solution.

The detailed form of $\mathcal{M}_\mathcal{L}$ or $\mathcal{M}_\mathcal{R}$ relies on that of $H(\beta)$ in the presence of chiral symmetry, Eq. (5). Let $\phi_i = (u_i, v_i)^T$. The eigenequation for $E = 0$ becomes

$$R_+(\beta_i(0))v_i = 0,$$
$$R_-(\beta_i(0))u_i = 0, \quad (11)$$

where $i = 1, \ldots, p + q$. If $\beta_i(0)$ comes from the kth root of $\det R_+ = 0$, then $(\beta_i(0), \phi_i)$ is also denoted as $(\beta_k^+(0), \phi_k^+)$ with $\phi_i = (0, v_i)^T$. Otherwise $\beta_i(0)$ may come from the kth root of $\det R_- = 0$, so $(\beta_i(0), \phi_i)$ is also denoted as $(\beta_k^-(0), \phi_k^-)$ with $\phi_i = (u_i, 0)^T$. Therefore, among the $(p + q)$ β roots, half is from $\det R_+ = 0$ and half from $\det R_- = 0$. In the first (last) p (q) roots we assume there are p_1 (q_1) roots from $\det R_+ = 0$ and p_2 (q_2) roots from $\det R_- = 0$. Naturally we have $p_1 + p_2 = p$, $q_1 + q_2 = q$, and $p_1 + q_1 = p_2 + q_2 = (p + q)/2$. Among p_1, p_2, q_1, and q_2, only one is independent and so by introducing an integer w they can be parametrized as follows:

$$p_1 = p/2 + w, \quad p_2 = p/2 - w,$$
$$q_1 = q/2 - w, \quad q_2 = q/2 + w. \quad (12)$$

Our topological invariant ν defined in Eq. (6) can then be given by

$$\nu = (p_1 - p_2)/2 = w. \quad (13)$$

So we have to prove that a state of system denoted by w must possess at least $|w|$ TPESs. Let $\Phi_{i,\mathcal{L}}^\pm \equiv (\phi_i^\pm, (\beta_i^\pm)^{-1}\phi_i^\pm, \ldots, (\beta_i^\pm)^{-M_1+1}\phi_i^\pm)^T$ and $\Phi_{i,\mathcal{R}}^\pm \equiv (\beta_i^\pm\phi_i^\pm, (\beta_i^\pm)^2\phi_i^\pm, \ldots, (\beta_i^\pm)^{M_2}\phi_i^\pm)^T$, and by performing some matrix column exchanges on $\mathcal{M}_\mathcal{L}$ or $\mathcal{M}_\mathcal{R}$, the two matrices would be equivalently changed to be

$$\mathcal{M}_\mathcal{L} \to (\Phi_{1,\mathcal{L}}^+, \ldots, \Phi_{p_1,\mathcal{L}}^+, \Phi_{1,\mathcal{L}}^-, \ldots, \Phi_{p_2,\mathcal{L}}^-) \equiv (\Phi_\mathcal{L}^+, \Phi_\mathcal{L}^-),$$

$$\mathcal{M}_\mathcal{R} \to (\Phi_{p+1,\mathcal{R}}^+, \ldots, \Phi_{p+q_1,\mathcal{R}}^+, \Phi_{p_2+1,\mathcal{R}}^-, \ldots, \Phi_{p_2+q_2,\mathcal{R}}^-)$$
$$\equiv (\Phi_\mathcal{R}^+, \Phi_\mathcal{R}^-) \quad (14)$$

with $|\mathcal{L}\rangle$ also being changed to be $|\mathcal{L}\rangle \to (c_1^+, \ldots, c_{p_1}^+, c_1^-, \ldots, c_{p_2}^-)^T \equiv (C^+, C^-)^T$, where C^+ (C^-) represents the row vector which contains all the components c_i^+ (c_i^-) in $|\mathcal{L}\rangle$. Accordingly, $|\mathcal{R}\rangle$ is also changed to be $|\mathcal{R}\rangle \to (c_{p_1+1}^+, \ldots, c_{p_1+q_1}^+, c_{p_2+1}^-, \ldots, c_{p_2+q_2}^-)^T \equiv (D^+, D^-)^T$.

One particular feature to be noted is that if $\det\mathcal{M}_\mathcal{L} \neq 0$, we must have $p_1 = p_2 = p/2$, namely, $w = 0$. To prove this statement, we assume that, for example, $p_1 > p_2$. Since for any $\Phi_{i,\mathcal{L}}^+$-like vector with p components, one half of its components are zero due to $\phi_i = (0, v_i)^T$, the dimension of the space spanned by this kind of vectors is at most $p/2$. So the rank of $\mathcal{M}_\mathcal{L}$ is at most $p/2 + p_2 < p$, leading to the contradiction that $\det\mathcal{M}_\mathcal{L} = 0$. Similarly, one can prove that if $\det\mathcal{M}_\mathcal{R} \neq 0$, one must have $w = 0$.

We note that the solutions of $\mathcal{M}_{\mathcal{L}}|\mathcal{L}\rangle = 0$ or $\mathcal{M}_{\mathcal{R}}|\mathcal{R}\rangle = 0$ are actually solutions of the $E = 0$ state for the half-infinite system with left-open boundary or right-open boundary. But different systems in the same topologically equivalent class may have a different number of solutions, due to the existence of nontopological edge states. To exclude these edge states not protected by chiral symmetry, the number of solutions must be minimized by continuously deforming the system without closing the gap. In Appendix B we demonstrate another generic important feature that by keeping the open system gapful, one can continuously vary $R_{\pm}(\beta)$ by continuously moving the β roots of $E = 0$, to make $\text{Rank}\mathcal{M}_{\mathcal{L}}$ and $\text{Rank}\mathcal{M}_{\mathcal{R}}$ maximized. Therefore, for any state of system under OBCs, it can always be made to be a topologically equivalent state with $\text{Rank}\mathcal{M}_{\mathcal{L}}$ and $\text{Rank}\mathcal{M}_{\mathcal{R}}$ being maximized. After this rank maximization process, for a state denoted by w, we have $\text{Rank}\mathcal{M}_{\mathcal{L}} = p/2 + (p/2 - |w|) = p - |w|$, and $\text{Rank}\mathcal{M}_{\mathcal{R}} = q/2 + (q/2 - |w|) = q - |w|$. According to linear algebra theory, both $\mathcal{M}_{\mathcal{L}}|\mathcal{L}\rangle = 0$ and $\mathcal{M}_{\mathcal{R}}|\mathcal{R}\rangle = 0$ have $|w|$ independent solutions. These solutions correspond to the $|w|$ left-localized TPESs or the $|w|$ right-localized TPESs when the system is under the left-open or right-open half-infinite boundary conditions. Thus the rank maximization process can be understood as a modification of the system which eliminates the solutions which are not protected by chiral symmetry, since the original state of system denoted by w may have more solutions if the ranks of $\mathcal{M}_{\mathcal{L}}$ and $\mathcal{M}_{\mathcal{R}}$ are not being maximized.

We further show that when the finiteness of L is taken into account, $|w|$ is the minimal number of TPESs for systems in the topological equivalent class. A large but finite L indicates the boundary equations in Eqs. (8) should be replaced by

$$\begin{pmatrix} \mathcal{M}_{\mathcal{L}} & \mathcal{M}_{\mathcal{LR}} \\ \mathcal{M}_{\mathcal{RL}} & \mathcal{M}_{\mathcal{R}} \end{pmatrix}|\psi\rangle = 0,$$
$$|\psi\rangle = (c_1, \ldots, c_p, c_{p+1}, \ldots, c_{p+q})^T, \quad (15)$$

with $\mathcal{M}_{\mathcal{LR}} = (\beta_{p+1}^{-L}\Phi_{p+1,\mathcal{L}}, \ldots, \beta_{p+q}^{-L}\Phi_{p+q,\mathcal{L}})$ and $\mathcal{M}_{\mathcal{RL}} = (\beta_1^{L+1}\Phi_{1,\mathcal{R}}, \ldots, \beta_p^{L+1}\Phi_{p,\mathcal{R}})$, followed by the same column exchanges as $\mathcal{M}_{\mathcal{L}}$ and $\mathcal{M}_{\mathcal{R}}$ in Eq. (14). For concreteness, we assume $w \geqslant 0$. Because $\text{Rank}\Phi_{\mathcal{L}}^+ = p/2 - w$, $\text{Rank}\Phi_{\mathcal{R}}^+ = q/2 - w$, one can easily show that $\mathcal{M}_{\mathcal{L}}|\mathcal{L}\rangle = 0$ ($\mathcal{M}_{\mathcal{R}}|\mathcal{R}\rangle = 0$) would indicate $C^- = 0$ ($D^+ = 0$). This implies that the left-localized states reside on sublattice B, while the right-localized states reside on sublattice A [87]. This is actually required by chiral symmetry which makes the states be the eigenstates of the chiral operator. We denote the solutions as $|\mathcal{L}_m\rangle = (C_m^+, 0)^T$ $[|\mathcal{R}_n\rangle = (0, D_n^-)^T]$, $m(n) = 1, 2, \ldots, w$. Note that each column in $\mathcal{M}_{\mathcal{LR}}$ and $\mathcal{M}_{\mathcal{RL}}$ is vanishingly small for sufficiently large L since $|\beta_p| < 1 < |\beta_{p+1}|$. So one can construct two $w \times w$ perturbation matrices M_{LR} and M_{RL}, whose mn matrix entries are $\langle \mathcal{L}_m|\mathcal{M}_{\mathcal{LR}}|\mathcal{R}_n\rangle$ and $\langle \mathcal{R}_m|\mathcal{M}_{\mathcal{RL}}|\mathcal{L}_n\rangle$, respectively. Because D_n^+ is zero in $|\mathcal{R}_n\rangle$, only D_n^- in $|\mathcal{R}_n\rangle$ has nonzero contribution to the matrix entry $\langle \mathcal{L}_m|\mathcal{M}_{\mathcal{LR}}|\mathcal{R}_n\rangle$, which is then generally proportional to $[\beta_k^-(0)]^{-L}$ with $k \geqslant p_2 + 1 = p/2 - w + 1$. Analogously, the matrix entry $\langle \mathcal{R}_m|\mathcal{M}_{\mathcal{RL}}|\mathcal{L}_n\rangle$ is generally proportional to $[\beta_l^+(0)]^L$ with $l \leqslant p_1 = p/2 + w$. So according to the degenerate perturbation theory, the actual solutions for the TPESs for a finite L are $|\psi_m^\pm\rangle = (\mathcal{A}_m|\mathcal{L}'_m\rangle, \pm\mathcal{B}_m|\mathcal{R}'_m\rangle)^T$, where

$m = 1, 2, \ldots, w$. Here $|\mathcal{L}'_m\rangle$ and $|\mathcal{R}'_m\rangle$ are the linear superpositions of $|\mathcal{L}_m\rangle$ and $|\mathcal{R}_m\rangle$, respectively, and have been normalized. $|\psi_m^+\rangle$ is related to $|\psi_m^-\rangle$ by chiral symmetry: $\mathcal{S}|\psi_m^\pm\rangle = -|\psi_m^\mp\rangle$. $\mathcal{A}_m$ and $\mathcal{B}_m$ are the coefficients, and their ratios are found to be

$$\frac{\mathcal{B}_m}{\mathcal{A}_m} = \frac{\sqrt{\langle \mathcal{R}'_m|\mathcal{M}_{\mathcal{RL}}|\mathcal{L}'_m\rangle}}{\sqrt{\langle \mathcal{L}'_m|\mathcal{M}_{\mathcal{LR}}|\mathcal{R}'_m\rangle}}, \quad (16)$$

which is proportional to $(\beta_l^+(0)\beta_k^-(0))^{L/2}$. See Appendix D for the detailed illustration of these features for a concrete model. So for a relatively small L, among the $2NL$ states of the open chain, $2w$ states belong to the isolated $E = 0$, while the other states belong to the continuum OBS. However, the ratios are very sensitive to the details of the system, and as L is sufficiently large, they can be finite, but can also be vanishingly small or approaching infinity. If one of the ratios in Eq. (16) is not finite, the corresponding pair of TPESs can be viewed as being coalescing into one single state ($|\psi_m^+\rangle \propto |\psi_m^-\rangle$) so the final number of TPESs is actually varying from w to $2w$ for systems in the same topologically equivalent class. This is the origin of the defectiveness of the TPESs. For Hermitian systems, all the ratios are expected to be finite, and the Hermitian open chains always have $2w$ TPESs localized at both ends. In the same topologically equivalent class, there always exist some systems which have the minimum number of TPESs: w. Actually, the systems described by $H(\rho\beta)$ possess w TPESs, as long as ρ is sufficiently small or large. This is because due to the similarity transformation, all superposition coefficients for systems of $H(\rho\beta)$ would be scaled by a factor ρ^{-L}, compared to $H(\beta)$. Therefore, for a sufficiently large (small) ρ, the open systems described by $H(\rho\beta)$ have exactly $|w|$ left-(right-) localized TPESs which all reside on sublattice B (A) if $w > 0$, or on sublattice A (B) if $w < 0$. This means the number of TPESs is not a topological invariant, and $|w|$ or $|\nu|$ is proved to be the minimum number of TPESs and we then complete the proof.

IV. TOPOLOGICAL TRANSITIONS

By varying parameters, the system would undergo topological transitions when the invariant integer ν changes, at which the system becomes gapless, indicating $|\beta_p(0)| = |\beta_{p+1}(0)|$. At this critical point, if the change of the winding number $\Delta\nu = \pm 1$, the state of system denoted by w would be changed to be one denoted by $w \pm 1$, implying $|\beta_{p/2+|w|}^+(0)| = |\beta_{p/2-|w|+1}^-(0)|$ or $|\beta_{p/2+|w|}^-(0)| = |\beta_{p/2-|w|+1}^+(0)|$. So among $\beta_p(0)$ and $\beta_{p+1}(0)$, one is from $\det R_+ = 0$, the other is from $\det R_- = 0$, and their exchange means topological transition. We remark that when both $\beta_p(0)$ and $\beta_{p+1}(0)$ come from $\det R_+ = 0$ or $\det R_- = 0$, their exchanges would not induce topological transitions, but may drive the system to a topologically equivalent state or a gapless state with exceptional points [28]. If $\Delta\nu = \pm 2$, similar discussion would lead to the conclusion that at the transition point, we must have

$$|\beta_{p-1}(0)| = |\beta_p(0)| = |\beta_{p+1}(0)| = |\beta_{p+2}(0)|, \quad (17)$$

and among the four β roots, two must come from $\det R_+ = 0$, with the other two from $\det R_- = 0$. We show schematically the general case of $\Delta\nu = \pm n$ in Fig. 2. Now we make a few
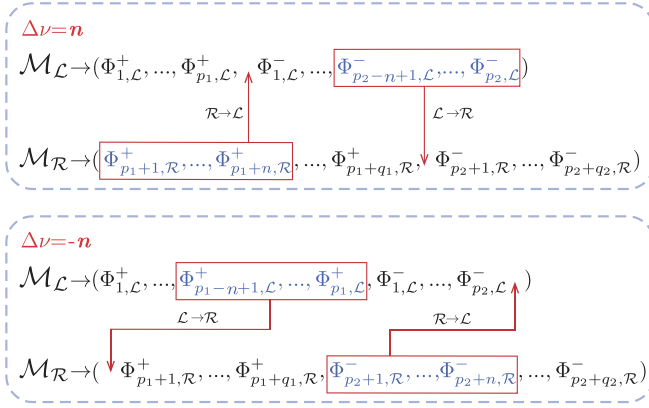
FIG. 2. Schematic demonstration of topological transitions with $\Delta \nu = \pm n$, where n β roots of $\det R_+ = 0$ and n roots of $\det R_- = 0$ are exchanged, obeying $|\beta_{p-n+1}(0)| = \ldots = |\beta_p(0)| = |\beta_{p+1}(0)| = \ldots = |\beta_{p+n}(0)|$.

remarks on the relationship between sub-GBZs and topological transition. At the transition point, the $E = 0$ state is connected to one energy branch of the OBS, which corresponds to a definite sub-GBZ. In Appendix C we prove that near the phase transition, if this sub-GBZ is chosen as the closed loop $\mathcal{L}_\beta$ in the integral in Eq. (6), the change of this newly defined winding number ν' at the transition point with $|\beta_p(0)| = |\beta_{p+1}(0)|$ would be $\Delta \nu' = \pm 1$.

V. FURTHER DISCUSSION AND CONCLUSION

The phase winding of $\det H$ along the closed loop $\mathcal{L}_\beta$ is the sum of those of $\det R_+$ and $\det R_-$, and can be easily checked to be zero. So the topological invariant ν can either be expressed as the phase winding of $\det R_+$ or minus that of $\det R_-$ along $\mathcal{L}_\beta$. This does not mean that either R_+ or R_- can independently determine the invariant because $\mathcal{L}_\beta$ has already contained the information about both of them. For any non-Hermitian chiral-symmetric open system described by $H(\beta)$ with $|\beta_p(0)| \neq |\beta_{p+1}(0)|$, the invariant ν can be computed directly from Eq. (6), but can also be understood physically as follows. Choose its topologically equivalent system $H(\rho\beta)$ obeying $|\beta_p(0)|/\rho < 1 < |\beta_{p+1}(0)|/\rho$. This new system takes the same invariant ν but its $\mathcal{L}_\beta$ can be chosen to the unit circle, implying ν can now be defined on the BZ for this system. Bulk invariants defined on the BZ in non-Hermitian systems means that the invariants are coming from the systems under periodic boundary conditions, and the corresponding BBC connects them to the TPESs of the corresponding half-infinite systems. Our topological invariant characterizes the non-Bloch BBC which connects the invariant to the TPESs under OBCs. So for this system, the topological invariants for the conventional non-Hermitian BBC and non-Bloch BBC become identical. This may shed light on the relationship between the BBC and the non-Bloch BBC in high-dimensional non-Hermitian topological systems.

APPENDIX A: OBCs AND THEIR SIMPLIFICATION

The eigenequation of $H(\beta)$ can be written as

$$\sum_{m=-M_2}^{M_1} T_m \psi(j-m) = E\psi(j), \qquad (A1)$$

where $M_1 < j < L - M_2$. When $1 \leqslant j \leqslant M_1$, the eigenequation becomes the left-boundary conditions, which are given by

$$\sum_{m=-M_2}^{M_1-1} T_m \psi(M_1 - m) = E\psi(M_1),$$

$$\sum_{m=-M_2}^{M_1-2} T_m \psi(M_1 - 1 - m) = E\psi(M_1 - 1),$$

$$\vdots$$

$$\sum_{m=-M_2}^{0} T_m \psi(1 - m) = E\psi(1). \qquad (A2)$$

When $L - M_2 + 1 \leqslant j \leqslant L$, the eigenequation becomes the right-boundary conditions, which are

$$\sum_{m=-M_2+1}^{M_1} T_m \psi(L - M_2 + 1 - m) = E\psi(L - M_2 + 1),$$

$$\sum_{m=-M_2+2}^{M_1} T_m \psi(L - M_2 + 2 - m) = E\psi(L - M_2 + 2),$$

$$\vdots$$

$$\sum_{m=0}^{M_1} T_m \psi(L - m) = E\psi(L). \qquad (A3)$$

The two sets of equations contain M_1 and M_2 equations, respectively. In order to make the eigenequation, Eq. (A1), be obeyed for all unit cells, including those near the left and right boundaries, one can introduce extra unit cells as shown in Fig. 3. The combination of these new equations and the above two sets of equations leads to the following simplified left-boundary conditions:

$$T_{M_1}\psi(0) = 0,$$

$$T_{M_1}\psi(-1) + T_{M_1-1}\psi(0) = 0,$$

$$\vdots$$

$$T_{M_1}\psi(-M_1+1) + T_{M_1-1}\psi(-M_1+2) + \cdots$$
$$+ T_1\psi(0) = 0, \qquad (A4)$$

and right-boundary conditions:

$$T_{-M_2}\psi(L+1) = 0,$$

$$T_{-M_2}\psi(L+2) + T_{-M_2+1}\psi(L+1) = 0,$$

$$\vdots$$

$$T_{-M_2}\psi(L+M_2) + T_{-M_2+1}\psi(L+M_2-1) + \cdots$$
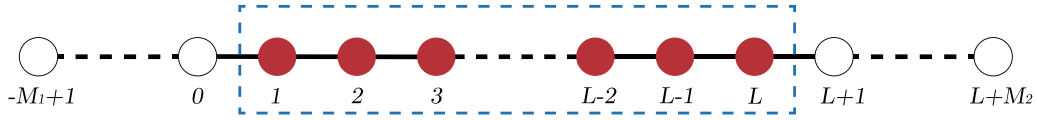$$+ T_{-1}\psi(L+1) = 0. \qquad (A5)$$

FIG. 3. A one-dimensional open chain with L unit cells represented by solid circles, where each unit cell contains $2N$ internal degrees of freedom. Here the hollow circles at both ends of the chain are the extra unit cells added for the discussion of open-boundary conditions.

If both T_{M_1} and T_{-M_2} are invertible, then according to the first equations of the two sets of equations, we have $\psi(0) = \psi(L + 1) = 0$. From the second equations, we further have $\psi(-1) = \psi(L + 2) = 0$. Using a similar argument, we would find the following equivalent boundary equations:

$$\psi(0) = \psi(-1) = \ldots = \psi(-M_1 + 1) = 0,$$
$$\psi(L + 1) = \psi(L + 2) = \ldots = \psi(L + M_2) = 0. \quad (A6)$$

These boundary conditions can be understood as follows: the wave function $\psi(j)$ at the extra unit cells denoted by the hollow circles in Fig. 3 should all be zero. Let us consider the system described by $H(\rho\beta)$ whose wave function is denoted by $\tilde{\psi}(j)$. The real-space Hamiltonian $\mathcal{H}\rho$ of $H(\rho\beta)$ and the real-space Hamiltonian $\mathcal{H}$ of $H(\beta)$ are related by a similarity transformation $\mathcal{H}\rho = S^{-1}\mathcal{H}S$, with $S = \mathrm{diag}(\rho^1, \rho^2, \ldots, \rho^L) \otimes I_{2N}$. Accordingly, we have $\tilde{\psi}(j) = \rho^{-j}\psi(j)$ and the hopping matrices of $H(\rho\beta)$ become $T_m\rho^{-m}$, so for $\tilde{\psi}(j)$ we can obtain similar OBCs to Eq. (A6).

APPENDIX B: RANK MAXIMIZATION OF MATRICES $\mathcal{M}_\mathcal{L}$ AND $\mathcal{M}_\mathcal{R}$

Introduce a continuous parameter Λ, and $R_\pm$ can be viewed as functions of Λ: $R_\pm(\beta, \Lambda)$ obeying $R_\pm(\beta, 0) = R_\pm(\beta)$. Here

we shall prove in this section that for a gapful open system with $|\beta_p(0)| \neq |\beta_{p+1}(0)|$, while keeping the system gapful, one can always vary the parameter Λ to continuously modify the positions of the β roots of $\det R_\pm(\beta, \Lambda) = 0$ to make the rank of matrices of $\mathcal{M}_\mathcal{L}$ and $\mathcal{M}_\mathcal{R}$ in the main text maximized, without breaking the chiral symmetry. This kind of variation of the system is obviously a topologically equivalent one, so the final state should share the same topological invariant ν with the initial state.

By solving the eigenequation of $E = 0$,

$$H(\beta)\phi_i = 0 \rightarrow \begin{cases} R_+(\beta)v_i = 0 \\ R_-(\beta)u_i = 0, \end{cases} \quad (B1)$$

we get the β roots and their corresponding eigenvectors: β_i^+, $\phi_i^+ = (0, v_i)^T$ $[\beta_i^-, \phi_i^- = (u_i, 0)^T]$, $i = 1, 2, \ldots, N(M_1 + M_2)$. From all the eigenvectors v_i of R_+, we construct a $N(M_1 + M_2) \times N(M_1 + M_2)$ matrix $\mathcal{M}_N^+$:

$$\mathcal{M}_N^+ = \left(\Phi_1^+, \Phi_2^+, \ldots, \Phi_{N(M_1+M_2)}^+\right), \quad (B2)$$

with the column vector $\Phi_i^+ = (v_i, (\beta_i^+)^{-1}v_i, \ldots, (\beta_i^+)^{-M_1-M_2+1}v_i)^T = (1, (\beta_i^+)^{-1}, \ldots, (\beta_i^+)^{-M_1-M_2+1})^T \otimes v_i$.

When $N = 1$ or in the two-band case, the matrix $\mathcal{M}_1^+$ becomes

$$\mathcal{M}_1^+ = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ (\beta_1^+)^{-1} & (\beta_2^+)^{-1} & \ldots & (\beta_{M_1+M_2}^+)^{-1} \\ \vdots & \vdots & \vdots & \vdots \\ (\beta_1^+)^{-M_1-M_2+1} & (\beta_2^+)^{-M_1-M_2+1} & \ldots & (\beta_{M_1+M_2}^+)^{-M_1-M_2+1} \end{pmatrix}, \quad (B3)$$

which is a Vandermonde matrix, whose determinant is a well-known polynomial $\mathcal{P}_V$ of $(\beta_i^+)^{-1}$:

$$\det\mathcal{M}_1^+ = \prod_{1 \leqslant i < j \leqslant M_1+M_2} [(\beta_i^+)^{-1} - (\beta_j^+)^{-1}] \equiv \mathcal{P}_V\left((\beta_1^+)^{-1}, (\beta_2^+)^{-1}, \ldots, (\beta_{M_1+M_2}^+)^{-1}\right). \quad (B4)$$

If all β_i^+ are different from each other, then obviously $\det\mathcal{M}_1^+ \neq 0$. Otherwise, there exist repeated β roots, then by fixing the other roots we can easily do a parameter change to make the repeated roots split without closing the gap, since the two roots $\beta_p(0)$ and $\beta_{p+1}(0)$ with unequal magnitude are still left unchanged. So for any state of the system, either this state or its topologically equivalent one would always obey $\det\mathcal{M}_1^+ \neq 0$. For later convenience, we call vectors like $(1, (\beta_i^+)^{-1}, \ldots, (\beta_i^+)^{-M_1-M_2+1})^T \equiv \Phi_{\beta_i^+}$ Vandermonde vectors. The space spanned by n M-dimensional Vandermonde vectors with different β values is n-dimensional if $n < M$, and is M-dimensional otherwise.

When $N \geqslant 2$, each column vector v_i has N components and $R_+(\beta)$ is a $N \times N$ matrix. The (kl) element of $R_+(\beta)$ is a

polynomial of β:

$$R_{kl}(\beta) = \sum_{m=-M_2}^{M_1} t_m^{kl}\beta^{-m}, \quad (B5)$$

where $k, l = 1, 2, \ldots, N$. Design a set of continuous functions of Λ: $\beta_i^+(\Lambda)$, satisfying $\beta_i^+(0) = \beta_i^+$, and

$$\beta_1^+(1) = \beta_2^+(1) = \ldots = \beta_N^+(1) = \tilde{\beta}_1^+,$$
$$\beta_{N+1}^+(1) = \beta_{N+2}^+(1) = \ldots = \beta_{2N}^+(1) = \tilde{\beta}_2^+,$$
$$\vdots$$
$$\beta_{(M_1+M_2-1)N+1}^+(1) = \beta_{(M_1+M_2-1)N+2}^+(1) = \ldots$$
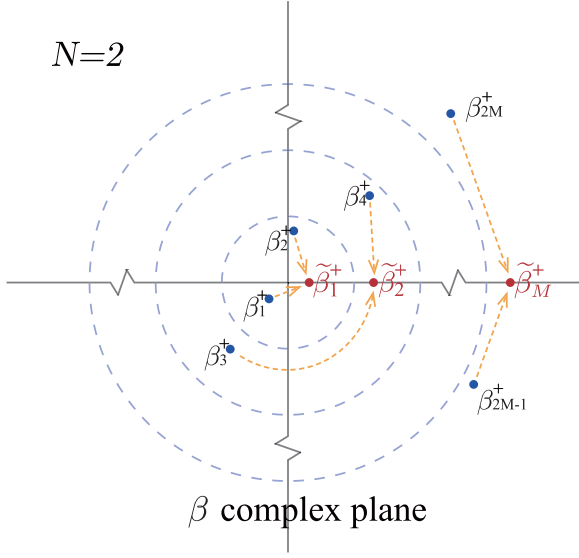$$= \beta_{(M_1+M_2)N}^+(1) = \tilde{\beta}_{M_1+M_2}^+. \quad (B6)$$

$N=2$

FIG. 4. Continuous variation of $R_+(\beta)$, where the β roots of $\det R_+(\beta) = 0$ are varied continuously. The NM roots are equally divided into M groups with the N roots in each group being finally merged into one repeated root, where $M = M_1 + M_2$.

Here $\widetilde{\beta}_i^+$ are the target β roots with different absolute values satisfying $|\beta_{(i-1)N+1}^+| \leqslant |\widetilde{\beta}_i^+| \leqslant |\beta_{iN}^+|$. So all $\beta_i^+(\Lambda)$ are expected to be divided into $M_1 + M_2$ groups, and each group contains N β roots which would merge into one single N-fold root as Λ is varied to be 1. This is shown schematically in Fig. 4. Now we treat all the coefficients t_m^{kl} as unknown variables which can be viewed as functions of Λ: $t_m^{kl}(\Lambda)$. They can be found by solving the equations $\det R_+(\beta_i^+, \Lambda) = 0$, with $i = 1, 2, \ldots, N(M_1 + M_2)$ and imposing the initial conditions, $t_m^{kl}(0) = t_m^{kl}$. The number of nonlinear equations is $N(M_1 + M_2)$, while the number of unknown variables is $N^2(M_1 + M_2 + 1)$. So for any fixed value of Λ, we generally have multiple solutions of $t_m^{kl}(\Lambda)$, but for $\Lambda = 0$, we have only one solution, $t_m^{kl}(0) = t_m^{kl}$. Thus we have found a solution of $R_\pm(\beta, \Lambda)$, which is a continuous function of Λ. When $\Lambda = 1$, each $\widetilde{\beta}_i^+$ is an N-fold root of $\det R_+(\beta_i^+, \Lambda = 1) = 0$, indicating that its eigenspace should be N-dimensional and so $R_+(\beta_i^+, \Lambda = 1) \equiv 0$. Therefore, as varying Λ from 0 to 1, one would continuously transform $R_+(\beta, \Lambda)$ from $R_+(\beta)$ to $R_+(\beta, \Lambda = 1) \propto \beta^{-NM_1} \prod_{1 \leqslant i \leqslant M_1+M_2} (\beta - \widetilde{\beta}_i^+) I_N$. When $\Lambda = 1$, eigenvectors v_i can be chosen to be $(v_1, v_2, \ldots, v_{N(M_1+M_2)}) = (I_N, I_N, \ldots, I_N)$. It is easy to find that $\mathcal{M}_N^+(\Lambda = 1) = \widetilde{\mathcal{M}}_1^+ \otimes I_N$, where $\widetilde{\mathcal{M}}_1^+$ is the Vandermonde matrix of $(\widetilde{\beta}_i^+)^{-1}$, and so $\det \mathcal{M}_N^+(\Lambda = 1) = \mathcal{P}_V^N((\widetilde{\beta}_1^+)^{-1}, (\widetilde{\beta}_2^+)^{-1}, \ldots, (\widetilde{\beta}_{M_1+M_2}^+)^{-1}) \neq 0$.

In a very similar way, from all the eigenvectors u_i of R_- one can construct the $N(M_1 + M_2) \times N(M_1 + M_2)$ matrix $\mathcal{M}_N^-$:

$$\mathcal{M}_N^- = \left(\Phi_1^-, \Phi_2^-, \ldots, \Phi_{N(M_1+M_2)}^- \right), \tag{B7}$$

with the column vectors:

$$\begin{aligned} \Phi_i^- &= \left(u_i, (\beta_i^-)^{-1} u_i, \ldots, (\beta_i^-)^{-M_1-M_2+1} u_i \right)^T \\ &= \left(1, (\beta_i^-)^{-1}, \ldots, (\beta_i^-)^{-M_1-M_2+1} \right)^T \otimes u_i \\ &= \Phi_{\beta_i^-} \otimes u_i. \end{aligned} \tag{B8}$$

We also have $\det \mathcal{M}_N^-(\Lambda = 1) = \mathcal{P}_V^N((\widetilde{\beta}_1^-)^{-1}, (\widetilde{\beta}_2^-)^{-1}, \ldots, (\widetilde{\beta}_{M_1+M_2}^-)^{-1}) \neq 0$, where $\widetilde{\beta}_i^-$ is the target β at $\Lambda = 1$ in this case.

Under this situation, we discuss why the ranks of the matrices $\mathcal{M}_{\mathcal{L}}$ and $\mathcal{M}_{\mathcal{R}}$ in the main text have been maximized. We take $\mathcal{M}_{\mathcal{L}}$ as an example to illustrate this fact. If $w = 0$, then $p_1 = p_2 = p/2 = NM_1$, $q_1 = q_2 = q/2 = NM_2$. From $\mathcal{M}_N^+(\Lambda = 1)$, choose the first NM_1 column vectors $\Phi_i^+ = \Phi_{\beta_i^+} \otimes v_i$. For each vector we remove the last NM_2 components and replace v_i by $\phi_i = (0, v_i)^T$ to obtain $\Phi_{i,\mathcal{L}}^+$. These new vectors are comprised of N sets of M_1-dimensional Vandermonde vectors with each set spanning an independent M_1-dimensional space. From $\mathcal{M}_N^-(\Lambda = 1)$, also choose its first NM_1 column vectors $\Phi_i^- = \Phi_{\beta_i^-} \otimes u_i$. For each vector we also remove the last NM_2 components but replace u_i by $\phi_i = (u_i, 0)^T$ to obtain $\Phi_{i,\mathcal{L}}^-$. These new vectors are also comprised of N sets of M_1-dimensional Vandermonde vectors with each set spanning an independent M_1-dimensional space. Obviously, the two spaces spanned by $\Phi_{i,\mathcal{L}}^+$ or $\Phi_{i,\mathcal{L}}^-$ are orthogonal. Since the matrix $\mathcal{M}_{\mathcal{L}} = (\Phi_{1,\mathcal{L}}^+, \ldots, \Phi_{NM_1,\mathcal{L}}^+, \Phi_{1,\mathcal{L}}^-, \ldots, \Phi_{NM_1,\mathcal{L}}^-)$ is comprised of $\Phi_{i,\mathcal{L}}^+$ and $\Phi_{i,\mathcal{L}}^-$, we have $\text{Rank}\mathcal{M}_{\mathcal{L}} = \text{Rank}(\Phi_{1,\mathcal{L}}^+, \ldots, \Phi_{NM_1,\mathcal{L}}^+) + \text{Rank}(\Phi_{1,\mathcal{L}}^-, \ldots, \Phi_{NM_1,\mathcal{L}}^-) = NM_1 + NM_1 = p$. Thus, when $w = 0$, $\text{Rank}\mathcal{M}_{\mathcal{L}}$ is maximized to be p. Similarly, it can be shown that $\text{Rank}\mathcal{M}_{\mathcal{R}}$ is maximized to be q.

When $w \neq 0$, $p_1 = NM_1 + w$, $p_2 = NM_1 - w$, $q_1 = NM_2 - w$, and $q_2 = NM_2 + w$. We only consider the case of $0 < w < N$ for simplicity. Let us make a slight modification on the continuous variation of Λ. We change the target β roots slightly: only the $(M_1 + 1)$th equation of Eq. (B6) is modified to be

$$\begin{aligned} \beta_{M_1N+1}^+(1) &= \beta_{M_1N+2}^+(1) = \ldots = \beta_{M_1N+w}^+(1) \\ &= \widetilde{\beta}_{M_1+1}^+, \\ \beta_{M_1N+w+1}^+(1) &= \beta_{M_1N+w+2}^+(1) = \ldots = \beta_{(M_1+1)N}^+(1) \\ &= \widetilde{\beta}_{M_1+1}'^+, \end{aligned} \tag{B9}$$

where $|\beta_{M_1N+1}^+| \leqslant |\widetilde{\beta}_{M_1+1}^+| \leqslant |\beta_{M_1N+w}^+| < |\beta_{M_1N+w+1}^+| \leqslant |\widetilde{\beta}'_{M_1+1}| \leqslant |\beta_{(M_1+1)N}^+|$. So here the $(M_1 + 1)$th group of $\beta_i^+(\Lambda)$ roots would merge into two repeated roots $\widetilde{\beta}_{M_1+1}^+$ and $\widetilde{\beta}_{M_1+1}'^+$, where the former is a w-fold root while the latter is a $(N - w)$-fold root. The eigenspace of the two roots formed by their corresponding eigenvectors v_i is w- or $(N - w)$-dimensional, respectively. $R_-(\beta)$ is also varied accordingly in a similar way, where the M_1th group of $\beta_i^-(\Lambda)$ roots would merge into two repeated roots $\widetilde{\beta}_{M_1}^-$ and $\widetilde{\beta}_{M_1}'^-$, with the former being a $(N - w)$-fold root and the latter being a w-fold root. From $\mathcal{M}_N^+(\Lambda = 1)$, choose the first p_1 column vectors $\Phi_i^+ = \Phi_{\beta_i^+} \otimes v_i$. For each vector we remove the last NM_2 components and replace v_i by $\phi_i = (0, v_i)^T$, so we obtain $\Phi_{i,\mathcal{L}}^+$. These new vectors have already contained N sets of M_1-dimensional Vandermonde vectors with each set spanning an independent M_1-dimensional space. So $\text{Rank}(\Phi_{1,\mathcal{L}}^+, \ldots, \Phi_{p_1,\mathcal{L}}^+)$ has been maximized to be $NM_1 = p/2$. From $\mathcal{M}_N^-(\Lambda = 1)$, also choose its first p_2 column vectors $\Phi_i^- = \Phi_{\beta_i^-} \otimes u_i$. For

each vector we also remove the last NM_2 components but replace u_i by $\phi_i = (u_i, 0)^T$, so we obtain $\Phi_{i,\mathcal{L}}^-$. If proper representation of v_i is chosen, these new vectors are found to be comprised of $N - w$ sets of M_1-dimensional Vandermonde vectors with each set spanning an independent M_1-dimensional space, together with w sets of $(M_1 - 1)$-dimensional Vandermonde vectors with each set spanning an independent $(M_1 - 1)$-dimensional space. So $\text{Rank}(\Phi_{1,\mathcal{L}}^-, \ldots, \Phi_{p_2,\mathcal{L}}^-) = NM_1 - w = p/2 - w$. Therefore, we have $\text{Rank}\mathcal{M}_{\mathcal{L}} = \text{Rank}(\Phi_{1,\mathcal{L}}^+, \ldots, \Phi_{p_1,\mathcal{L}}^+) + \text{Rank}(\Phi_{1,\mathcal{L}}^-, \ldots, \Phi_{p_2,\mathcal{L}}^-) = p/2 + (p/2 - w) = p - w$, indicating that it has been maximized. Thus, when $w > 0$, $\text{Rank}\mathcal{M}_{\mathcal{L}}$ is maximized to be $p - w$. Similarly, it can be shown that $\text{Rank}\mathcal{M}_{\mathcal{R}}$ is simultaneously maximized to be $q - w$.

So we have proved that for a state of the open system denoted by w in the main text, by continuously varying $R_{\pm}(\beta)$ without closing the gap, the ranks of $\mathcal{M}_L$ and $\mathcal{M}_R$ in the main text can be maximized to be $p - |w|$ and $q - |w|$, respectively.

APPENDIX C: A THEOREM ON ±1 CHANGE AT TOPOLOGICAL TRANSITIONS OF THE WINDING NUMBER DEFINED ON ONE OF THE SUB-GBZs

In this section we prove that when the non-Hermitian system is sufficiently near the topological transition point, if one replaces the closed loop $\mathcal{L}_{\beta}$ in the integral of the topological invariant ν defined in the main text, by one of the sub-GBZs, the change of the new winding number ν' at the transition point would be $\Delta\nu' = \pm 1$. The sub-GBZ chosen corresponds to the branch of the OBS closest to the $E = 0$ point.

When the system under OBCs is near the topological phase transition, the system must be nearly gapless and then one part of its OBS must be approaching the $E = 0$ point. Under parameter change, the transition process can be summarized into two scenarios, as shown in Fig. 5, where one is the inner-point connection [19] and the other is the end-point connection. At the critical point where the system is gapless, as shown in Figs. 5(a2) and 5(b2), the state of the isolated point $E = 0$ belongs to one pair of the open arcs of the OBS and its $\beta_p(0)$ and $\beta_{p+1}(0)$ would simultaneously fall on the corresponding sub-GBZ. Now consider the situation when the system is sufficiently close to the transition point. The sub-GBZ as a closed loop, is composed of the β_p and β_{p+1} points of the energy branch of the OBS closest to $E = 0$. Choose a special state E_0 on the branch which is nearest to $E = 0$. If the system is sufficiently close to the transition point, then E_0 would be approaching $E = 0$ so that its $\beta_j(E_0)$ is nearly equal to $\beta_j(0)$.

Let us first consider the case where the inner points $\pm E_0$ are going to be touching each other, as shown in Figs. 5(a1) and 5(a2). Expand the polynomial $f(E, \beta) \equiv \det[H(\beta) - E] = 0$ near E_0 and its $\beta_j(E_0)$. We have

$$\left.\frac{\partial f}{\partial E}\right|_{[E_0, \beta_j(E_0)]} \Delta E + \left.\frac{\partial f}{\partial \beta}\right|_{[E_0, \beta_j(E_0)]} \Delta\beta = 0. \quad \text{(C1)}$$

So the small deviations ΔE and $\Delta\beta$ are connected with each other. For any two kinds of deviations $(\Delta E, \Delta\beta)$, and $(\Delta E', \Delta\beta')$, one must have

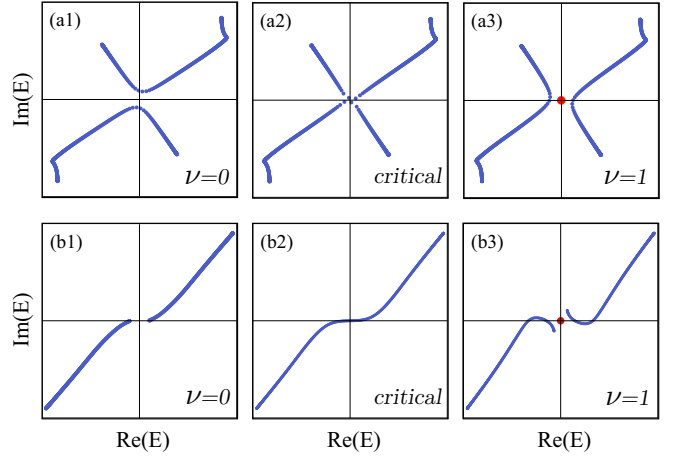$$\Delta E'/\Delta E = \Delta\beta'/\Delta\beta. \quad \text{(C2)}$$

FIG. 5. Two typical topological transition processes under parameter change in a chiral symmetric non-Hermitian system under OBCs, where the gap closes and then opens. Each pair of arcs in each figure represents the corresponding pair of OBSs on the E plane closest to $E = 0$. The isolated solid circle at $E = 0$ in (a3) and (b3) denotes the topological protected zero modes with ν denoting the corresponding topological invariant. (a1)–(a3) Inner-point touching process. (b1)–(b3) End-point touching process.

This is actually the conformal property of analytic functions in complex analysis. Now take ΔE as the deviation along the OBS, and $\Delta E'$ as the deviation along the line segment, as shown in Fig. 6(a1). If j is chosen to be p or $p + 1$, we
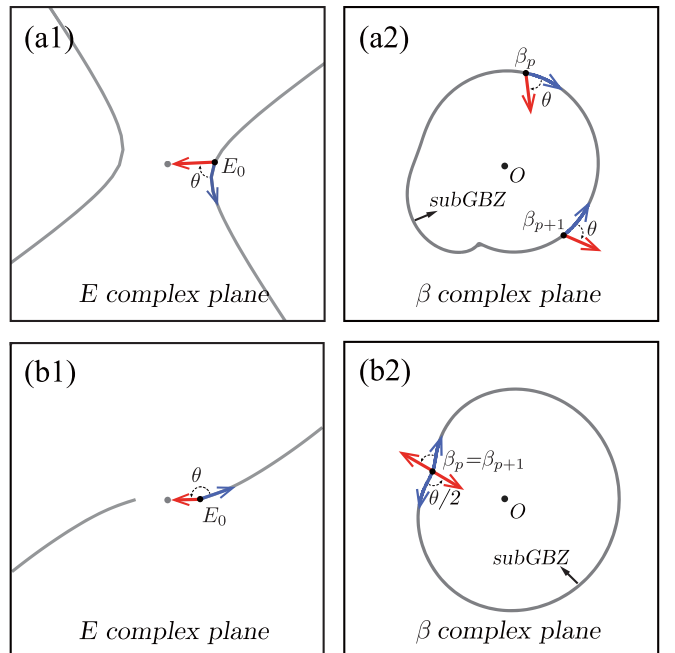

FIG. 6. Two deviation paths in the E plane and the corresponding paths of their β_p and β_{p+1} in the β plane for the two cases when the system is sufficiently near the critical point, as shown in Fig. 5. Here, E_0 is the point on the OBS which is chosen to be the nearest one to $E = 0$. One path on the E plane is the straight line segment connecting E_0 to $E = 0$, and the other path starting from E_0 is a small arc along the same OBS.
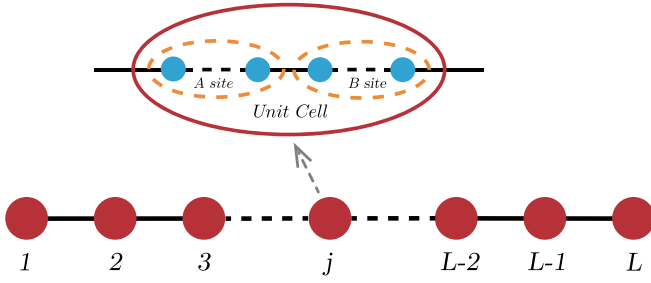
FIG. 7. Open chain and its unit cells which can be divided into two parts: sites belonging to sublattice A and those belonging to sublattice B.

would have two deviations $\Delta\beta$, starting from β_p or β_{p+1}, respectively, as shown in Fig. 6(a2). What is important is that $\Delta\beta$ at β_p and $\Delta\beta$ at β_{p+1} are pointing oppositely along the sub-GBZ [32], namely, one is clockwise while the other is anticlockwise. Therefore, according to the conformal property, Eq. (C2), the deviation $\Delta\beta'$ at β_p (β_{p+1}) would be pointing towards the interior (exterior) of the sub-GBZ. So among the two points $\beta_p(0)$ and $\beta_{p+1}(0)$, the sub-GBZ only encloses $\beta_p(0)$.

For the case of the end points $\pm E_0$ touching each other, as shown in Figs. 6(b1) and 6(b2), since $\beta_p(E_0) = \beta_{p+1}(E_0)$, one has $\partial f/\partial \beta |_{[E_0,\beta_p(E_0)]} = \partial f/\partial \beta |_{[E_0,\beta_{p+1}(E_0)]} = 0$. So the expansion of the polynomial $f(E,\beta)$ near E_0 and its $\beta_p(E_0)$ must be made in the second order of β and we obtain

$$\left.\frac{\partial f}{\partial E}\right|_{[E_0,\beta_p(E_0)]} \Delta E + \frac{1}{2}\left.\frac{\partial^2 f}{\partial \beta^2}\right|_{[E_0,\beta_p(E_0)]} \Delta\beta^2 = 0. \qquad \text{(C3)}$$

So similarly for two kinds of deviations, we must have

$$\Delta E'/\Delta E = (\Delta\beta'/\Delta\beta)^2, \qquad \text{(C4)}$$

where the two kinds of deviations ΔE and $\Delta E'$ are taken similarly as before, as shown in Fig. 6(b1). The two corresponding deviations of $\Delta\beta'$ are found to be pointing oppositely. Thus for $\beta_p(0)$ and $\beta_{p+1}(0)$ points, the sub-GBZ only encircles $\beta_p(0)$.

Therefore for both cases, we have proved that among $\beta_p(0)$ and $\beta_{p+1}(0)$, the sub-GBZ as a closed loop only encircles $\beta_p(0)$, with $\beta_{p+1}(0)$ being kept outside. Consider a topological transition process with $\Delta\nu = \pm1$. Before or after the transition, the system would always be gapful, but is characterized by different topological invariants ν. Let us denote

$\beta_p^i(0)$ $[\beta_p^f(0)]$ and $\beta_{p+1}^i(0)$ $[\beta_{p+1}^f(0)]$ for the system just before (after) the phase transition. According to our theorem, at the transition point, one of the $\beta_p(0)$ and $\beta_{p+1}(0)$ points is from $\det R_+ = 0$, while the other is from $\det R_- = 0$, and they would exchange to induce the topological transition. So during the phase transition, $\beta_p^i(0)$ would continuously change to be $\beta_{p+1}^f(0)$, while $\beta_{p+1}^i(0)$ would continuously change to be $\beta_p^f(0)$. Then the change of the phase winding along the sub-GBZ for $\det R_+$ would be minus that for $\det R_-$. This indicates that the change of the new winding number ν' defined on the sub-GBZ would be ±1. So we complete the proof of the statement at the beginning of this section. We remark that for a two-band system, its GBZ naturally becomes the sub-GBZ discussed here.

APPENDIX D: EXPLICIT MODEL EXAMPLE EXHIBITING RELATIONSHIP BETWEEN THE INVARIANT ν AND THE EDGE STATES

In this section, by studying an explicit model in detail, we will follow the way in the theorem proof in the main text, step by step, to find out the edge states for the model system under OBCs. On the other hand, by direct diagonalization of the Hamiltonian for the corresponding open system, we obtain the numerical results on the edge states of $E = 0$. In the following, we will compare these numerical results with our analytical calculations, to illustrate our conclusions on the TPESs given in the main text.

When a $2N$-band non-Hermitian system with chiral symmetry takes the form Eq. (5) in the main text, it is convenient to introduce the two sublattices A and B, where the first (last) N internal degrees of freedom belong to sublattice A (B), as shown in Fig. 7. We choose the model in the main text, which is a four-band model with only the nearest-neighbor hoppings, implying $N = 2$, $M_1 = M_2 = 1$, and $p = q = 4$.

First, the parameter λ is chosen to be $\lambda = 0.1$, corresponding to a topological nontrivial state with $\nu = -1$. By solving $\det H(\beta) = 0$, we get eight β roots of $E = 0$, which are given by

$$\begin{aligned}
\beta_1 &= \beta_1^- = -0.195, & \beta_2 &= \beta_1^+ = -0.401, \\
\beta_3 &= \beta_2^- = -0.495, & \beta_4 &= \beta_3^- = 0.811, \\
\beta_5 &= \beta_2^+ = -1.149, & \beta_6 &= \beta_3^+ = 2.014, \\
\beta_7 &= \beta_4^- = -2.442, & \beta_8 &= \beta_4^+ = -5.634,
\end{aligned} \qquad \text{(D1)}$$

and accordingly their corresponding eigenmodes are

$$\phi_1 = \phi_1^- = \begin{pmatrix} -0.569 \\ -0.822 \\ 0 \\ 0 \end{pmatrix}, \quad \phi_2 = \phi_1^+ = \begin{pmatrix} 0 \\ 0 \\ -0.664 \\ -0.748 \end{pmatrix}, \quad \phi_3 = \phi_2^- = \begin{pmatrix} -0.084 \\ -0.997 \\ 0 \\ 0 \end{pmatrix}, \quad \phi_4 = \phi_3^- = \begin{pmatrix} -0.464 \\ 0.886 \\ 0 \\ 0 \end{pmatrix},$$

$$\phi_5 = \phi_2^+ = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1.000 \end{pmatrix}, \quad \phi_6 = \phi_3^+ = \begin{pmatrix} 0 \\ 0 \\ -0.409 \\ 0.912 \end{pmatrix}, \quad \phi_7 = \phi_4^- = \begin{pmatrix} -0.714 \\ -0.700 \\ 0 \\ 0 \end{pmatrix}, \quad \phi_8 = \phi_4^+ = \begin{pmatrix} 0 \\ 0 \\ -0.648 \\ -0.761 \end{pmatrix}. \qquad \text{(D2)}$$
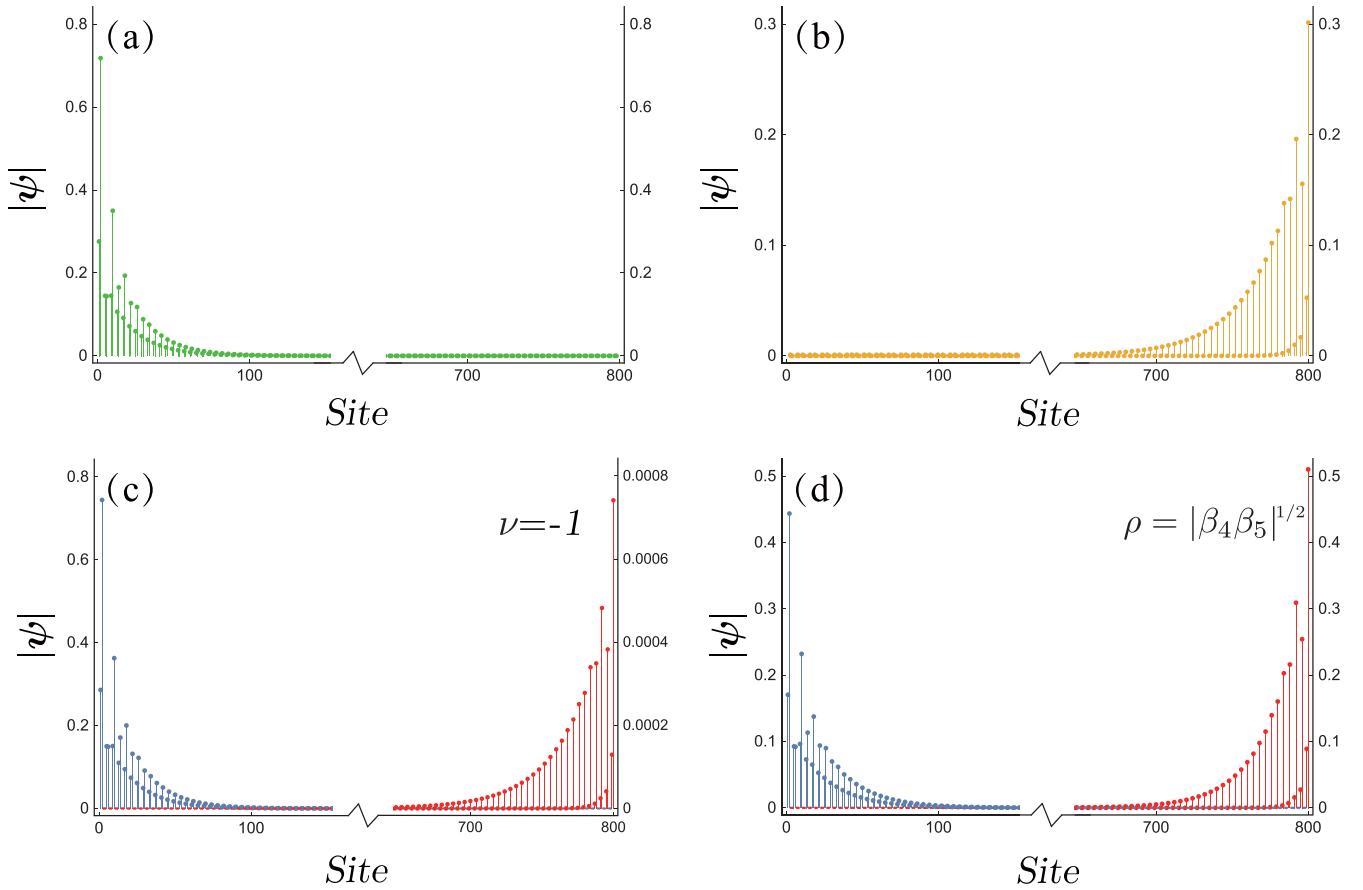
FIG. 8. Space distributions of the edge states with $\nu = -1$ and $\lambda = 0.1$. (a) Edge state $\psi_{\mathcal{L}}$. (b) Edge state $\psi_{\mathcal{R}}$. (c) Numerical matrix diagonalization result of the zero-energy modes, and (d) the corresponding result after a similarity transformation with $\rho = |\beta_4 \beta_5|^{1/2}$, where data on sublattice A (B) are colored in blue (red). We have 800 lattice sites since here $L = 200$ and $N = 2$.

Here the superscript $\pm$ indicates that the corresponding β roots come from $R_{\pm}$. So the matrices $\mathcal{M}_{\mathcal{L}} = (\phi_1^+, \phi_1^-, \phi_2^-, \phi_3^-)$ and $\mathcal{M}_{\mathcal{R}} = (\beta_2^+ \phi_2^+, \beta_3^+ \phi_3^+, \beta_4^+ \phi_4^+, \beta_4^- \phi_4^-)$ can be explicitly expressed as

$$
\mathcal{M}_{\mathcal{L}} = \begin{pmatrix} 0 & -0.569 & -0.084 & -0.464 \\ 0 & -0.822 & 0.996 & 0.886 \\ -0.664 & 0 & 0 & 0 \\ -0.748 & 0 & 0 & 0 \end{pmatrix},
$$

$$
\mathcal{M}_{\mathcal{R}} = \begin{pmatrix} 0 & 0 & 0 & 1.744 \\ 0 & 0 & 0 & 1.710 \\ 0 & -0.824 & 3.649 & 0 \\ -1.149 & 1.840 & 4.292 & 0 \end{pmatrix}. \quad \text{(D3)}
$$

It is easy to check that $\text{Rank}\mathcal{M}_{\mathcal{L}} = \text{Rank}\mathcal{M}_{\mathcal{R}} = 3$, namely, they have been maximized so either $\mathcal{M}_{\mathcal{L}}|\mathcal{L}\rangle = 0$ or $\mathcal{M}_{\mathcal{R}}|\mathcal{R}\rangle = 0$ has one independent solution, which can be expressed as $|\mathcal{L}\rangle = (c_1^+, c_1^-, c_2^-, c_3^-)^T$ and $|\mathcal{R}\rangle = (c_2^+, c_3^+, c_4^+, c_4^-)^T$. The solutions are found to be $|\mathcal{L}\rangle = (0, 0.335, 0.765, -0.550)^T$ and $|\mathcal{R}\rangle = (0.922, 0.377, 0.0852, 0)^T$. Notice that in these expressions the decimal numbers shown here are all retained to a few decimal places, but in the actual calculations they have all been kept in high precision to ensure the accuracy of the calculating results.

From the two solutions one can see that $|\mathcal{L}\rangle$ only contains the coefficients of ϕ^-, while $|\mathcal{R}\rangle$ only contains those of ϕ^+. This indicates that $|\mathcal{L}\rangle$ is a left-localized state residing on sublattice A, while $|\mathcal{R}\rangle$ is a right-localized one residing on sublattice B. By using these coefficients, the corresponding real-space normalized wave functions $\psi_{\mathcal{L}}$ and $\psi_{\mathcal{R}}$ can be written as

$$
\psi_{\mathcal{L}}(j) \propto \sum_{j=1}^{L} c_1^- (\beta_1^-)^j \phi_1^- + c_2^- (\beta_2^-)^j \phi_2^- + c_3^- (\beta_3^-)^j \phi_3^-,
$$

$$
\psi_{\mathcal{R}}(j) \propto \sum_{j=1}^{L} c_2^+ (\beta_2^+)^{j-L} \phi_2^+ + c_3^+ (\beta_3^+)^{j-L} \phi_3^+ + c_4^+
$$

$$
+ (\beta_4^+)^{j-L} \phi_4^+. \quad \text{(D4)}
$$

The space distributions of these two wave functions are shown in Figs. 8(a) and 8(b). For a finite system under OBCs, we have two final real-space wave functions $\psi_{\pm}$ which are the linear combinations of $\psi_{\mathcal{L}}$ and $\psi_{\mathcal{R}}$:

$$
\psi_{\pm} = \mathcal{A}\psi_{\mathcal{L}} \pm \mathcal{B}\psi_{\mathcal{R}}, \quad \text{(D5)}
$$

where ψ_+ and ψ_- are related by chiral symmetry: $\mathcal{S}\psi_{\pm} = \mathcal{A}\psi_{\mathcal{L}} \mp \mathcal{B}\psi_{\mathcal{R}} = \psi_{\mp}$. The coefficients $\mathcal{A}$ and $\mathcal{B}$ are determined by the matrices $\mathcal{M}_{\mathcal{RL}}$ and $\mathcal{M}_{\mathcal{LR}}$, which take the
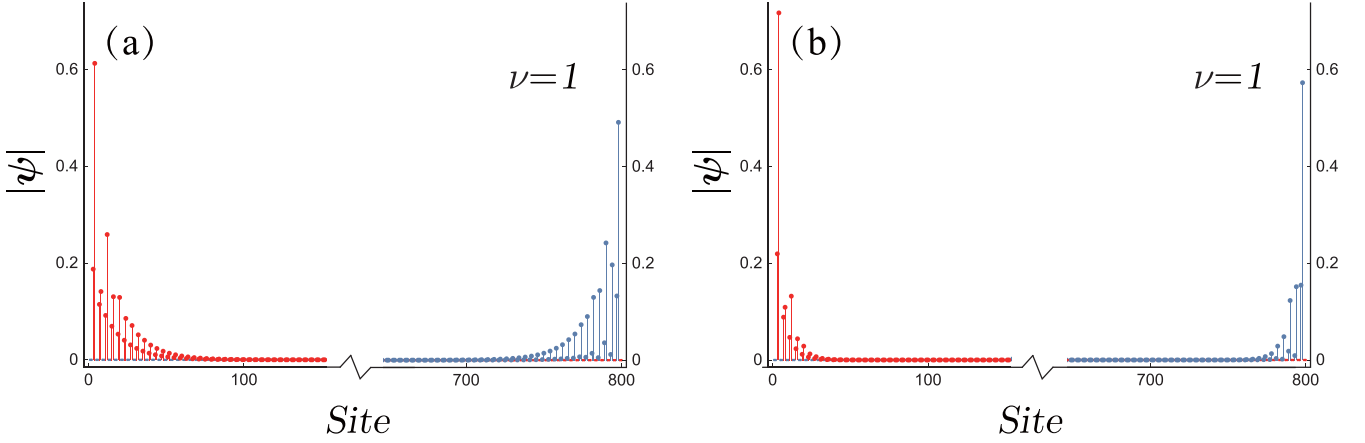
FIG. 9. Space distributions of the zero-energy modes obtained by the numerical matrix diagonalization, with $\nu = 1$, where data on sublattice A (B) are colored in blue (red). (a) $\lambda = -0.6$. (b) $\lambda = -0.9$. Here $L = 200$.

following forms:

$$\mathcal{M}_{\mathcal{RL}} = ((\beta_1^+)^{L+1}\phi_1^+, (\beta_1^-)^{L+1}\phi_1^-, (\beta_2^-)^{L+1}\phi_2^-, (\beta_3^-)^{L+1}\phi_3^-),$$

$$\mathcal{M}_{\mathcal{LR}} = ((\beta_2^+)^{-L}\phi_2^+, (\beta_3^+)^{-L}\phi_3^+, (\beta_4^+)^{-L}\phi_4^+, (\beta_4^-)^{-L}\phi_4^-).$$

$$(D6)$$

By substituting $|\mathcal{L}\rangle = (0, 0.335, 0.765, -0.550)^T$ and $|\mathcal{R}\rangle = (0.922, 0.377, 0.0852, 0)^T$ into the equation

$$\frac{\mathcal{B}}{\mathcal{A}} = \frac{\sqrt{\langle\mathcal{R}|\mathcal{M}_{\mathcal{RL}}|\mathcal{L}\rangle}}{\sqrt{\langle\mathcal{L}|\mathcal{M}_{\mathcal{LR}}|\mathcal{R}\rangle}}, \qquad (D7)$$

one can find $|\mathcal{B}/\mathcal{A}| \approx |\beta_3^-\beta_2^+|^{L/2} = |\beta_4\beta_5|^{L/2}$. Here since $|\beta_4\beta_5| < 1$, when L is sufficiently large, it will be $|\mathcal{B}| \ll |\mathcal{A}|$, indicating the contribution of $\psi_{\mathcal{L}}$ to $\psi_{\pm}$ is much greater than that of $\psi_{\mathcal{R}}$. So for this parameter, it looks like the two TPESs are all localized at the left boundary, and when $L \to \infty$, $|\mathcal{B}|$ will be approaching zero so that ψ_+ and ψ_- linked by chiral symmetry would coalesce into one single state $\psi_{\mathcal{L}}$. This is actually the origin of the defectiveness of the TPESs. In Fig. 8(c) we show the numerical matrix diagonalization result

of the TPESs. The two edge states are nearly overlapping, and are found to be localized at left boundary while at the right boundary the wave function only has a very small weight. This is consistent with our analysis. If we perform a similarity transformation on this model to make $|\beta_4\beta_5/\rho^2| = 1$, i.e., $\rho = |\beta_4\beta_5|^{1/2}$, then this model described by $H(\beta)$ would be transformed to one described by $H(\rho\beta)$ whose wave function $\tilde{\psi}(j) = \rho^{-j}\psi(j)$ and so its ratio $|\mathcal{B}/\mathcal{A}| \approx 1^L = 1$. In our analysis the two TPESs in this situation would be localized at both ends of the open chain, consistent with the numerical results as shown in Fig. 8(d). Meanwhile, notice that a negative $\nu = -1$ means $|\mathcal{L}\rangle$ only has ϕ^- components, while $|\mathcal{R}\rangle$ only has ϕ^+ components, indicating that $|\mathcal{L}\rangle$ ($|\mathcal{R}\rangle$) is a left- (right-)localized A- (B-)sublattice edge state. However, for a positive $\nu = 1$, the $|\mathcal{L}\rangle$ ($|\mathcal{R}\rangle$) state would be a left- (right-)localized B- (A-)sublattice edge state, as mentioned in the main text. This is because when $\nu = +1$, all nonzero components in $|\mathcal{L}\rangle$ ($|\mathcal{R}\rangle$) are ϕ^+ (ϕ^-) instead of ϕ^- (ϕ^+).

In Figs. 9(a) and 9(b), we show the case of $\nu = 1$ with $\lambda = -0.6$ and $\lambda = -0.9$, respectively. The eight β values for the
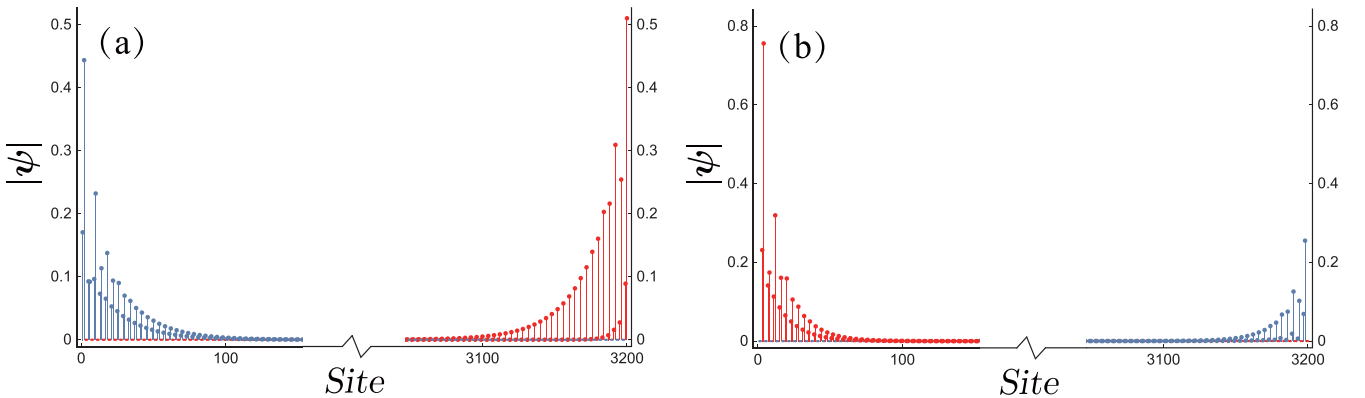

FIG. 10. (a) The same as Fig. 8(d), and (b) the same as Fig. 9(a), except that $L = 800$.
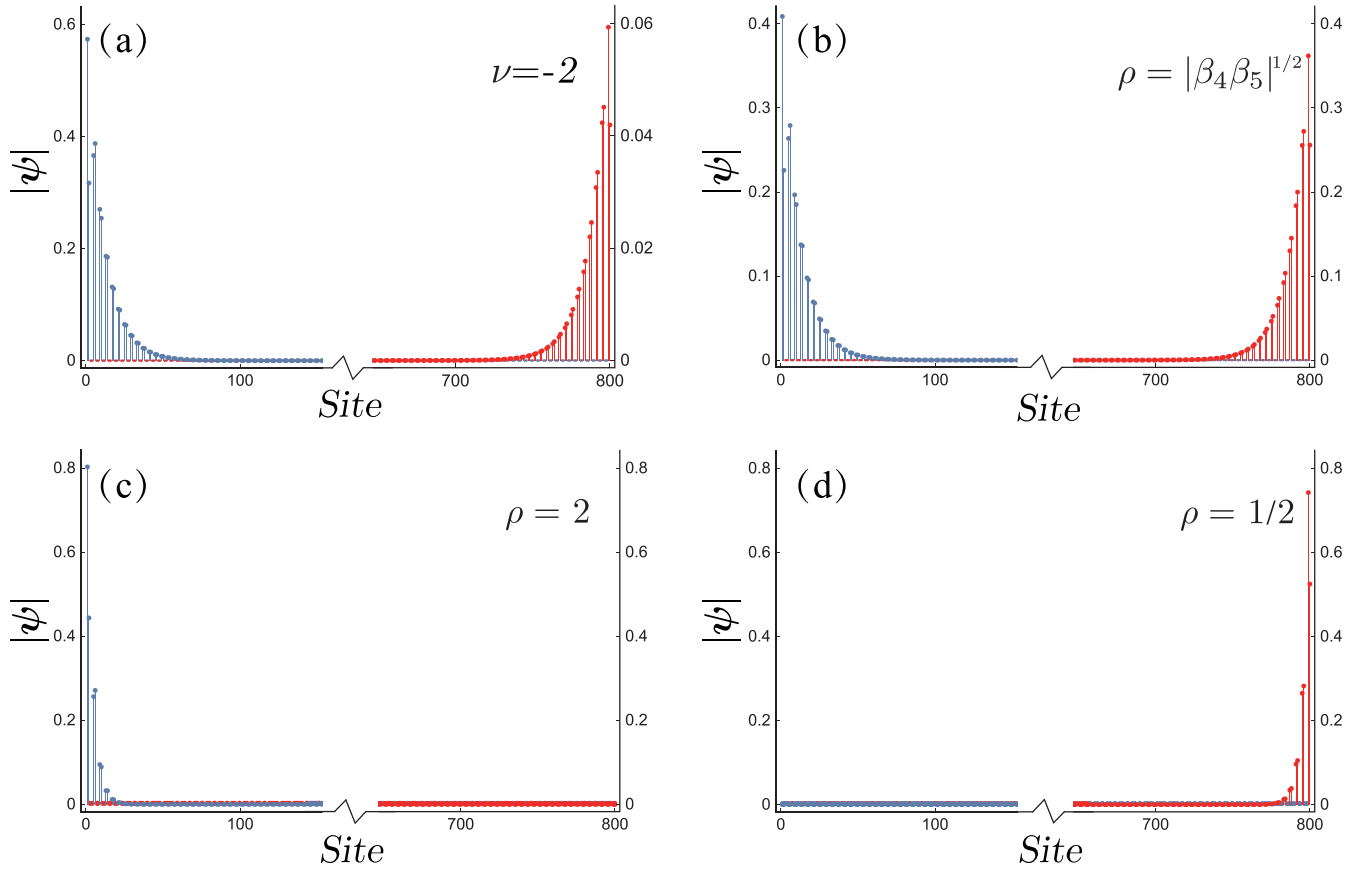
FIG. 11. Space distributions of the edge states with $\nu = -2$ and $\lambda = 1$. (a) One pair of the nearly overlapping zero-energy modes obtained by the numerical matrix diagonalization, and the corresponding result after a similarity transformation with (b) $\rho = |\beta_4\beta_5|^{1/2} = 0.989$, (c) $\rho = 2$, or (d) $\rho = 1/2$, respectively. Here $L = 200$.

two parameters are ordered as usual as follows:

$$|\beta_1| = |\beta_1^+| < |\beta_2| = |\beta_2^+| < |\beta_3| = |\beta_1^-| < |\beta_4| = |\beta_3^+| < 1 < |\beta_5| = |\beta_2^-| < |\beta_6| = |\beta_4^+| < |\beta_7| = |\beta_3^-| < |\beta_8| = |\beta_4^-|,$$

$$|\beta_1| = |\beta_1^+| < |\beta_2| = |\beta_2^+| < |\beta_3| = |\beta_3^+| < |\beta_4| = |\beta_1^-| < 1 < |\beta_5| = |\beta_4^+| < |\beta_6| = |\beta_2^-| < |\beta_7| = |\beta_3^-| < |\beta_8| = |\beta_4^-|,$$

(D8)

where the former corresponds to $\lambda = -0.6$ while the latter corresponds to $\lambda = -0.9$. When $\lambda = -0.6$, $|\mathcal{B}/\mathcal{A}| \approx |\beta_4\beta_5|^{L/2} = |\beta_3^+\beta_2^-|^{L/2}$ and when $\lambda = -0.9$, $|\mathcal{B}/\mathcal{A}| \approx |\beta_3\beta_6|^{L/2} = |\beta_3^+\beta_2^-|^{L/2}$. In both cases, $|\mathcal{B}/\mathcal{A}| \approx 0.997^L$. In our numerical calculation, since $L = 200$, $|\mathcal{B}/\mathcal{A}| \approx 0.74$, which makes $|\mathcal{A}|$ slightly larger than $|\mathcal{B}|$. So each TPES is localized at both boundaries of the open chain. But when $L \to \infty$, only the left-localized components would survive. By setting $L = 800$ we redo the numerical calculation of the edge states in Figs. 8(d) and 9(a), and show the corresponding results in Fig. 10. In Fig. 10(a), the space distribution is nearly unchanged since $|\mathcal{B}/\mathcal{A}| \propto 1$ is also kept unchanged while in Fig. 10(b), the left-localized components are even more dominant than the right-localized ones since the ratio $|\mathcal{B}/\mathcal{A}|$ in this case becomes smaller. These results show that the predicted ratio between the coefficients $\mathcal{A}$ and $\mathcal{B}$ is quite reasonable.

Now we choose $\lambda = 1$ to study the case of $\nu = -2$. The eight β values are ordered as follows:

$$|\beta_1| = |\beta_1^-| < |\beta_2| = |\beta_2^-| < |\beta_3| = |\beta_3^-| < |\beta_4| = |\beta_4^-| < 1 < |\beta_5| = |\beta_1^+| < |\beta_6| = |\beta_2^+| < |\beta_7| = |\beta_3^+| < |\beta_8| = |\beta_4^+|.$$

(D9)

It is found that $\text{Rank}\mathcal{M}_\mathcal{L} = \text{Rank}\mathcal{M}_\mathcal{R} = 2$, namely, they have been maximized so either $\mathcal{M}_\mathcal{L}|\mathcal{L}\rangle = 0$ or $\mathcal{M}_\mathcal{R}|\mathcal{R}\rangle = 0$ has two independent solutions, which can be labeled as $|\mathcal{L}_1\rangle, |\mathcal{L}_2\rangle$ and $|\mathcal{R}_1\rangle, |\mathcal{R}_2\rangle$, respectively. According to degenerate perturbation theory, let us solve $\begin{pmatrix} \mathcal{M}_\mathcal{L} & \mathcal{M}_{\mathcal{L}\mathcal{R}} \\ \mathcal{M}_{\mathcal{R}\mathcal{L}} & \mathcal{M}_\mathcal{R} \end{pmatrix}|\psi\rangle = 0$ which is equivalent to $M|\varphi\rangle = 0$, with M being the $2|\nu| \times 2|\nu|$ matrix given by

$$M = \begin{pmatrix} 0 & M_{LR} \\ M_{RL} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \langle\mathcal{L}_1|\mathcal{M}_{\mathcal{L}\mathcal{R}}|\mathcal{R}_1\rangle & \langle\mathcal{L}_1|\mathcal{M}_{\mathcal{L}\mathcal{R}}|\mathcal{R}_2\rangle \\ 0 & 0 & \langle\mathcal{L}_2|\mathcal{M}_{\mathcal{L}\mathcal{R}}|\mathcal{R}_1\rangle & \langle\mathcal{L}_2|\mathcal{M}_{\mathcal{L}\mathcal{R}}|\mathcal{R}_2\rangle \\ \langle\mathcal{R}_1|\mathcal{M}_{\mathcal{R}\mathcal{L}}|\mathcal{L}_1\rangle & \langle\mathcal{R}_1|\mathcal{M}_{\mathcal{R}\mathcal{L}}|\mathcal{L}_2\rangle & 0 & 0 \\ \langle\mathcal{R}_2|\mathcal{M}_{\mathcal{R}\mathcal{L}}|\mathcal{L}_1\rangle & \langle\mathcal{R}_2|\mathcal{M}_{\mathcal{R}\mathcal{L}}|\mathcal{L}_2\rangle & 0 & 0 \end{pmatrix},$$

(D10)

where $|\varphi\rangle = (l_1, l_2, r_1, r_2)^T$. So $|\psi\rangle = (l_1|\mathcal{L}_1\rangle + l_2|\mathcal{L}_2\rangle, r_1|\mathcal{R}_1\rangle + r_2|\mathcal{R}_2\rangle))^T$. We have four solutions: $|\psi_m^{\pm}\rangle = (\mathcal{A}_m|\mathcal{L}_m'\rangle, \pm\mathcal{B}_m|\mathcal{R}_m'\rangle)^T$, where $m = 1, 2$. Here each normalized $|\mathcal{L}_m'\rangle$ ($|\mathcal{R}_m'\rangle$) is the linear superposition of $|\mathcal{L}_1\rangle$ and $|\mathcal{L}_2\rangle$ ($|\mathcal{R}_1\rangle$ and $|\mathcal{R}_2\rangle$). $|\psi_m^+\rangle$ is also connected with $|\psi_m^-\rangle$ by chiral symmetry: $\mathcal{S}|\psi_m^{\pm}\rangle = |\psi_m^{\mp}\rangle$. From the ratio equation

$$\frac{\mathcal{B}_m}{\mathcal{A}_m} = \frac{\sqrt{\langle\mathcal{R}_m'|\mathcal{M}_{\mathcal{RL}}|\mathcal{L}_m'\rangle}}{\sqrt{\langle\mathcal{L}_m'|\mathcal{M}_{\mathcal{LR}}|\mathcal{R}_m'\rangle}}, \quad m = 1, 2 \tag{D11}$$

we get $|\mathcal{B}_1/\mathcal{A}_1| \approx |\beta_4\beta_5|^{L/2} = |\beta_4^-\beta_1^+|^{L/2} = 0.11567$ and $|\mathcal{B}_2/\mathcal{A}_2| \approx |\beta_3\beta_6|^{L/2} = |\beta_3^-\beta_2^+|^{L/2} = 0.00084$, with $L = 200$. The numerical result for one pair of the TPESs is shown in Fig. 11(a), while the other pair (data not shown) shows analogous behavior in this case. We found that the predicted ratio is consistent with the numerical matrix diagonalization result. As before, if we apply the similarity transformation to this system, i.e., $H(\beta) \to H(\rho\beta)$, the ratios become $|\mathcal{B}_1/\mathcal{A}_1| \approx |\beta_4\beta_5/\rho^2|^{L/2}$

and $|\mathcal{B}_2/\mathcal{A}_2| \approx |\beta_3\beta_6/\rho^2|^{L/2}$. In Fig. 11(b), we set a suitable ρ to make one of the ratios close to 1, so the corresponding pair of the TPESs would be localized at both ends of the open chain. In this situation, the total number of TPESs would be three. In Figs. 11(c) and 11(d), we set $\rho = 2$ and $\rho = 1/2$, respectively. In either case, when L is large, among $\mathcal{B}_m$ and $\mathcal{A}_m$, one would always be much larger than the other, and the smaller one can be viewed as approaching zero, implying that the number of TPESs would be equal to $|\nu| = 2$.

Here we see that we can always choose a sufficiently small or large ρ to perform the similarity transformation so that the number of TPESs can be reduced to the absolute value of the topological invariant ν but is definitely not less than $|\nu|$. If all ratios $|\mathcal{B}_m/\mathcal{A}_m|$ share the same relationship, one can naturally choose a common ρ to make all edge states be localized at both ends, so the number of TPESs reaches its maximum value, $2|\nu|$. However, in most cases of non-Hermitian systems, this is usually unachievable, so for a general non-Hermitian open chain with chiral symmetry, the number of TPESs would vary from $|\nu|$ to $2|\nu|$.

[1] N. Hatano and D. R. Nelson, Phys. Rev. B **56**, 8651 (1997).

[2] N. Hatano and D. R. Nelson, Phys. Rev. B **58**, 8384 (1998).

[3] C. M. Bender, D. C. Brody, and H. F. Jones, Phys. Rev. Lett. **89**, 270401 (2002).

[4] E. M. Graefe, H. J. Korsch, and A. E. Niederle, Phys. Rev. Lett. **101**, 150408 (2008).

[5] I. Rotter, J. Phys. A: Math. Theor. **42**, 153001 (2009).

[6] S. Longhi, Phys. Rev. Lett. **105**, 013903 (2010).

[7] T. E. Lee, F. Reiter, and N. Moiseyev, Phys. Rev. Lett. **113**, 250401 (2014).

[8] T. E. Lee and C.-K. Chan, Phys. Rev. X **4**, 041001 (2014).

[9] R. El-Ganainy, K. G. Makris, M. Khajavikhan, Z. H. Musslimani, S. Rotter, and D. N. Christodoulides, Nat. Phys. **14**, 11 (2018).

[10] V. M. Martinez Alvarez, J. E. Barrios Vargas, M. Berdakin, and L. E. F. Foa Torres, Eur. Phys. J.: Spec. Top. **227**, 1295 (2018).

[11] E. J. Bergholtz, J. C. Budich, and F. K. Kunst, Rev. Mod. Phys. **93**, 015005 (2021).

[12] E. Edvardsson and E. Ardonne, Phys. Rev. B **106**, 115107 (2022).

[13] W.-T. Xue, Y.-M. Hu, F. Song, and Z. Wang, Phys. Rev. Lett. **128**, 120401 (2022).

[14] S. Yao and Z. Wang, Phys. Rev. Lett. **121**, 086803 (2018).

[15] F. Song, S. Yao, and Z. Wang, Phys. Rev. Lett. **123**, 170401 (2019).

[16] Y. Yi and Z. Yang, Phys. Rev. Lett. **125**, 186802 (2020).

[17] C.-H. Liu, K. Zhang, Z. Yang, and S. Chen, Phys. Rev. Res. **2**, 043167 (2020).

[18] N. Okuma, K. Kawabata, K. Shiozaki, and M. Sato, Phys. Rev. Lett. **124**, 086801 (2020).

[19] L. Li, C. H. Lee, S. Mu, and J. Gong, Nat. Commun. **11**, 5491 (2020),

[20] S. Longhi, Phys. Rev. Lett. **124**, 066602 (2020).

[21] C.-X. Guo, C.-H. Liu, X.-M. Zhao, Y. Liu, and S. Chen, Phys. Rev. Lett. **127**, 116801 (2021).

[22] Y. Fu, J. Hu, and S. Wan, Phys. Rev. B **103**, 045420 (2021).

[23] W.-T. Xue, M.-R. Li, Y.-M. Hu, F. Song, and Z. Wang, Phys. Rev. B **103**, L241408 (2021).

[24] J. Claes and T. L. Hughes, Phys. Rev. B **103**, L140201 (2021).

[25] Q.-B. Zeng, Phys. Rev. B **106**, 235411 (2022).

[26] K. Zhang, Z. Yang, and C. Fang, Nat. Commun. **13**, 2496 (2022),

[27] S. Franca, V. Könye, F. Hassler, J. van den Brink, and C. Fulga, Phys. Rev. Lett. **129**, 086601 (2022).

[28] K. Yokomizo and S. Murakami, Phys. Rev. Lett. **123**, 066404 (2019).

[29] K. Kawabata, N. Okuma, and M. Sato, Phys. Rev. B **101**, 195147 (2020).

[30] K. Yokomizo and S. Murakami, Phys. Rev. B **103**, 165123 (2021).

[31] Y. Li, X. Ji, Y. Chen, X. Yan, and X. Yang, Phys. Rev. B **106**, 195425 (2022).

[32] D. Wu, J. Xie, Y. Zhou, and J. An, Phys. Rev. B **105**, 045422 (2022).

[33] C. Dembowski, H.-D. Gräf, H. L. Harney, A. Heine, W. D. Heiss, H. Rehfeld, and A. Richter, Phys. Rev. Lett. **86**, 787 (2001).

[34] C. Dembowski, B. Dietz, H.-D. Gräf, H. L. Harney, A. Heine, W. D. Heiss, and A. Richter, Phys. Rev. E **69**, 056216 (2004).

[35] W. D. Heiss, J. Phys. A: Math. Theor. **45**, 444016 (2012).

[36] B. Zhen, C. W. Hsu, Y. Igarashi, L. Lu, I. Kaminer, A. Pick, S.-L. Chua, J. D. Joannopoulos, and M. Soljačić, Nature (London) **525**, 354 (2015).

[37] H. Shen, B. Zhen, and L. Fu, Phys. Rev. Lett. **120**, 146402 (2018).

[38] H. Zhou, C. Peng, Y. Yoon, C. W. Hsu, K. A. Nelson, L. Fu, J. D. Joannopoulos, M. Soljačić, and B. Zhen, Science **359**, 1009 (2018).

[39] H. Xue, Q. Wang, B. Zhang, and Y. D. Chong, Phys. Rev. Lett. **124**, 236403 (2020).

[40] R. Okugawa, R. Takahashi, and K. Yokomizo, Phys. Rev. B **103**, 205205 (2021).

[41] Z. Yang, A. P. Schnyder, J. Hu, and C.-K. Chiu, Phys. Rev. Lett. **126**, 086401 (2021).

[42] L. Xiao, T. Deng, K. Wang, Z. Wang, W. Yi, and P. Xue, Phys. Rev. Lett. **126**, 230402 (2021).

[43] Y. C. Hu and T. L. Hughes, Phys. Rev. B **84**, 153101 (2011).

[44] K. Esaki, M. Sato, K. Hasebe, and M. Kohmoto, Phys. Rev. B **84**, 205128 (2011).

[45] S.-D. Liang and G.-Y. Huang, Phys. Rev. A **87**, 012118 (2013).

[46] J. M. Zeuner, M. C. Rechtsman, Y. Plotnik, Y. Lumer, S. Nolte, M. S. Rudner, M. Segev, and A. Szameit, Phys. Rev. Lett. **115**, 040402 (2015).

[47] T. E. Lee, Phys. Rev. Lett. **116**, 133903 (2016).

[48] S. Yao, F. Song, and Z. Wang, Phys. Rev. Lett. **121**, 136802 (2018).

[49] F. K. Kunst, E. Edvardsson, J. C. Budich, and E. J. Bergholtz, Phys. Rev. Lett. **121**, 026808 (2018).

[50] H. Jiang, C. Yang, and S. Chen, Phys. Rev. A **98**, 052116 (2018).

[51] V. M. Martinez Alvarez, J. E. Barrios Vargas, and L. E. F. Foa Torres, Phys. Rev. B **97**, 121401(R) (2018).

[52] S. Lieu, Phys. Rev. B **97**, 045106 (2018).

[53] Z. Gong, Y. Ashida, K. Kawabata, K. Takasan, S. Higashikawa, and M. Ueda, Phys. Rev. X **8**, 031079 (2018).

[54] C. Yin, H. Jiang, L. Li, R. Lü, and S. Chen, Phys. Rev. A **97**, 052115 (2018).

[55] K. Kawabata, K. Shiozaki, M. Ueda, and M. Sato, Phys. Rev. X **9**, 041015 (2019).

[56] J. Y. Lee, J. Ahn, H. Zhou, and A. Vishwanath, Phys. Rev. Lett. **123**, 206404 (2019).

[57] L. Jin and Z. Song, Phys. Rev. B **99**, 081103(R) (2019).

[58] K.-I. Imura and Y. Takane, Phys. Rev. B **100**, 165430 (2019).

[59] H. Zhou and J. Y. Lee, Phys. Rev. B **99**, 235112 (2019).

[60] L. Herviou, J. H. Bardarson, and N. Regnault, Phys. Rev. A **99**, 052118 (2019).

[61] F. Song, S. Yao, and Z. Wang, Phys. Rev. Lett. **123**, 246801 (2019).

[62] F. K. Kunst and V. Dwivedi, Phys. Rev. B **99**, 245116 (2019).

[63] T.-S. Deng and W. Yi, Phys. Rev. B **100**, 035102 (2019).

[64] L. Xiao, T. Deng, K. Wang, G. Zhu, Z. Wang, W. Yi, and P. Xue, Nat. Phys. **16**, 761 (2020).

[65] D. S. Borgnia, A. J. Kruchkov, and R.-J. Slager, Phys. Rev. Lett. **124**, 056802 (2020).

[66] C. C. Wojcik, X.-Q. Sun, T. Bzdušek, and S. Fan, Phys. Rev. B **101**, 205417 (2020).

[67] K. Kawabata, K. Shiozaki, and S. Ryu, Phys. Rev. Lett. **126**, 216405 (2021).

[68] H. Hu and E. Zhao, Phys. Rev. Lett. **126**, 010401 (2021).

[69] H.-G. Zirnstein, G. Refael, and B. Rosenow, Phys. Rev. Lett. **126**, 216407 (2021).

[70] Y. Cao, Y. Li, and X. Yang, Phys. Rev. B **103**, 075126 (2021).

[71] Y.-X. Xiao and C. T. Chan, Phys. Rev. B **105**, 075128 (2022).

[72] K. Ding, C. Fang, and G. Ma, Nat. Rev. Phys. **4**, 745 (2022).

[73] J. Cheng, X. Zhang, M.-H. Lu, and Y.-F. Chen, Phys. Rev. B **105**, 094103 (2022).

[74] H. Jiang and C. H. Lee, arXiv:2207.08843.

[75] L. Jezequel and P. Delplace, Phys. Rev. Lett. **130**, 066601 (2023).

[76] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008).

[77] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).

[78] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).

[79] C.-K. Chiu, J. C. Y. Teo, A. P. Schnyder, and S. Ryu, Rev. Mod. Phys. **88**, 035005 (2016).

[80] J. Bartlett and E. Zhao, Phys. Rev. B **107**, 035101 (2023).

[81] Z. Yang, K. Zhang, C. Fang, and J. Hu, Phys. Rev. Lett. **125**, 226402 (2020).

[82] X.-R. Wang, C.-X. Guo, and S.-P. Kou, Phys. Rev. B **101**, 121116(R) (2020).

[83] Y. Fu and S. Wan, Phys. Rev. B **105**, 075420 (2022).

[84] K.-I. Imura and Y. Takane, Prog. Theor. Exp. Phys. **2020**, 12A103 (2020).

[85] Y. Takane, J. Phys. Soc. Jpn. **91**, 054705 (2022).

[86] K. Zhang, Z. Yang, and C. Fang, Phys. Rev. Lett. **125**, 126402 (2020).

[87] Divide the open chain into two parts, sublattice A and sublattice B, which correspond to the u and v components in ϕ, respectively.