Topological edge states in one-dimensional non-Hermitian Su-Schrieffer-Heeger systems of finite lattice size: Analytical solutions and exceptional points

Chong Hou,[1] Lingfang Li,[1] Gangzhou Wu,[1] Yang Ruan,[1] Shihua Chen,[1,2,*] and Fabio Baronio[3,†]

[1]*School of Physics and Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 211189, China*

[2]*Purple Mountain Laboratories, Nanjing 211111, China*

[3]*Consorzio Nazionale Interuniversitario per le Telecomunicazioni and Dipartimento di Ingegneria dell'Informazione, Università di Brescia, Via Branze 38, 25123 Brescia, Italy*

(Received 30 June 2023; revised 7 August 2023; accepted 8 August 2023; published 18 August 2023)

Non-Hermiticity is a common property of a multitude of nonconservative physical systems that exchange energy with the surrounding environment. It brings about altogether new behaviors such as the non-Hermitian skin effect and the appearance of exceptional points, which have no counterparts in Hermitian systems. Despite intensive research activities on non-Hermitian physics, some fundamental issues, such as what the topological zero modes (i.e., the edge states with a degenerate zero eigenvalue that generally occur in the thermodynamic limit) will truly take when the lattice size becomes finite and whether they are definitely attributed to the formation of exceptional points, are still unclear, at least on an analytical level. Here we wish to address these open questions within the framework of the one-dimensional non-Hermitian Su-Schrieffer-Heeger model involving gain and loss. By virtue of this model under the open boundary condition, we obtain analytically the asymptotic solutions of topological edge states along with their explicit energy formulas that are universal for arbitrary lattice size. We also figure out the parameter conditions defining the domain of zero modes as well as the analytical criteria discriminating their localization characteristics, all exhibiting consistency with numerical results. We reveal that the zero modes sitting on either end of the chain are generated from exceptional points, but those sitting on both ends are indeed from diabolic points. In addition, we demonstrate that the skin modes caused by the non-Hermitian skin effect may possess higher-order exceptional points under certain parameter conditions. We anticipate that these analytical results will help illuminate the elusive non-Hermitian topology and may facilitate the experimental studies of some intriguing topological phenomena related to edge states.

DOI: 10.1103/PhysRevB.108.085425

I. INTRODUCTION

In the past decade, non-Hermitian physics [1–4], originally established for quantum field theories [5], has seen tremendous advance on both theoretical and experimental investigations intended for the open classical and quantum systems [6–12], with areas ranging from microwaves [13], mechanics [14], electronics [15,16], acoustics [17], condensed matter physics [18–20], to photonics [21–24]. What makes such non-Hermitian physics so enthralling is the occurrence of the non-Hermitian skin effect under the open boundary condition (OBC) [25–31], which has no Hermitian counterparts. Like the well-known skin effect in classical electrodynamics, where the alternating currents in the bulk conductor tend to move toward its surface, the non-Hermitian skin effect is a quantum phenomenon whereby the non-Hermitian Hamiltonians featured by gain and loss [22,30], or by nonreciprocal couplings [32–34], can host an extensive number of bulk eigenstates that tend to accumulate at the boundaries, now often called skin modes [35]. Due to the presence of skin modes, the conventional bulk-boundary correspondence (BBC) [36–39], which relates the edge states to the bulk topological invariants, encounters some kind of difficulty [40–43], as such bulk modes also collapse into localized edge modes.

Recently, several attempts to understand the non-Hermitian skin effect have been made so as to develop a modified formalism by which the BBC principle holds as well for non-Hermitian systems [25,44–46]. One attempt is the introduction of a concept of the generalized Brillouin zone (GBZ), which is made up of the set of points of $\beta = e^{ik}$ in the complex plane, when the lattice momentum k is no longer real [25,47]. In terms of GBZ, the topological invariants can be well defined and consequently a generalized BBC applicable for non-Hermitian systems is successfully established. The other attempt is to puzzle out the topological origin of the non-Hermitian skin effect [28] and to look for a similar BBC for skin modes. It has turned out that the emergence of skin modes under OBC is closely related to the winding of the energy obtained under the periodic boundary condition (PBC) around the base energy [35]. Most recently, we show further that the BBC for skin modes relies upon a topological winding number that is defined on the Brillouin zone (BZ), distinctly different from that for edge modes whose winding number is defined on GBZ [48]. These developments for non-Hermitian

*cshua@seu.edu.cn
†fabio.baronio@unibs.it

systems constitute what is now known as the non-Bloch band theory [45,49–52].

Another fascinating phenomenon unique to non-Hermitian physics is the appearance of exceptional points (EPs) [53–56], which correspond to the branch point singularity in a parameter space where both the eigenvalue and the associated eigenvectors can coalesce simultaneously [3,4]. Generally, when more than two eigenvalues collapse at one point, then a higher-order EP is created [57]. It should be noted that in Hermitian systems, a similar yet different branch point called the diabolic point (DP) may appear, where only the eigenvalues are degenerate, but the eigenvectors can be made orthogonal to each other [55,58]. In essence, at EPs, the eigenvectors are parallel or antiparallel, and thus the Hamiltonian matrix formed is defective and no longer diagonalizable. From a mathematical standpoint, an EP implies that both the projector and the nilpotent in the Jordan normal form exhibit discontinuous changes [3]. Due to such skewness of the vector space, the non-Hermitian system around an EP behaves as if it loses its dimensionality, resulting in a plethora of counterintuitive phenomena such as loss-induced suppression [59], unidirectional lasing [60], enhanced sensing [61], and formation of the bulk Fermi arc [62].

However, concerning non-Hermitian physics, there are still many intriguing issues that remain to be addressed. Let us take the simple one-dimensional (1D) lattice as an example. It is known that the obtainment of the continuum band, the exact GBZ (eGBZ) under OBC, and the topological edge states with a degenerate zero energy (usually called zero modes) all utilize the thermodynamic limit, i.e., letting the lattice size N be infinite [45,63,64]. Now, when the lattice size is arbitrarily finite, how does one find the asymptotic zero-mode solution along with its explicit universal energy formula? Is it possible that the zero modes in non-Hermitian systems are not attributed to EPs but to DPs in the parameter space? Can skin modes caused by the non-Hermitian skin effect possess both spectral degeneracy and eigenvector coalescence at a parameter point as well, and if so, what is the kind of this EP?

In this article, we wish to answer these questions, using the prototypical 1D non-Hermitian Su-Schrieffer-Heeger (SSH) model, which enables one to explore the intriguing topological phase transitions in such non-Hermitian systems as polyacetylene [65], photonic lattices [66,67], resonator arrays [68–70], electrical circuits [15,16,42], cold atoms [71], and nitrogen-vacancy centers in a diamond [20]. Here, for simplicity, only the non-Hermiticity caused by gain and loss is taken into account, but the results obtained can be extended to other non-Hermitian situations, e.g., the one involving nonreciprocal couplings, which can generally be linked to the former by similarity transformation. We will present the asymptotic zero-mode solutions under OBC that are universal for arbitrary lattice size, the explicit energy formulas and parameter conditions for these zero modes, as well as the valid analytical criteria discriminating their left, right, and double localization characteristics for given system parameters. With analytical results, we find that not all zero modes can be explained within the theory of EPs [63], but rather some of them can indeed be attributed to DPs, although the system in which they were born is non-Hermitian. We also show that the skin modes of the bulk may experience higher-order spectral degeneracy and eigenvector coalescence for special system parameters, and thus can be thought of as higher-order EPs [57] if seen in the parameter space. All analytical predictions are confirmed to be consistent with numerical results. We expect that our results may clarify the above concerns and shed fresh light on non-Hermitian topology, including the topological phase transitions, the non-Hermitian skin effect, and the BBC.

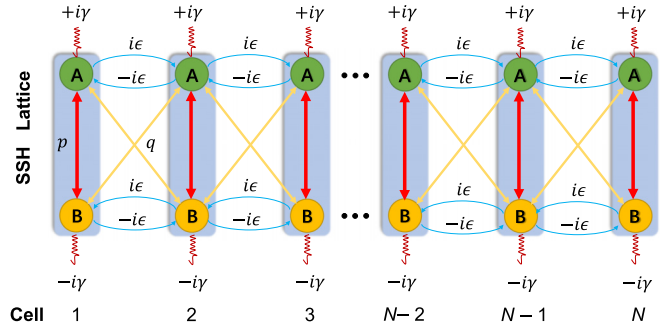

FIG. 1. Geometry of an open 1D tight-binding non-Hermitian SSH model consisting of N cells of sublattices A and B, which admits on-site dissipation characterized by the parameter $i\gamma$ and the intercell dissipation by the parameter $i\epsilon$. The parameters p (red lines) and q (yellow lines) stand for the reciprocal nearest-neighbor and long-range hopping amplitudes, respectively.

The subsequent sections are organized as follows. In Sec. II, we describe a 1D dissipative chain modeled by the non-Hermitian SSH Hamiltonian, and establish a relationship between its solutions and the ones of a simpler Hamiltonian that is obtained by similarity transformation. In Secs. III and IV, we present the asymptotic zero-mode solutions and the relevant parameter conditions for their existence and localization, for the special and general cases, respectively. A thorough discussion of these solutions and their relation to EPs is also given. Finally, we conclude our results in Sec. V.

II. NON-HERMITIAN SSH MODEL FOR OPEN DISSIPATIVE SYSTEMS

Our study begins by considering the following 1D non-Hermitian SSH Hamiltonian with the OBC [40,45]:

$$\begin{aligned}\hat{H} = &\sum_{n=1}^{N}[p(\hat{a}_n^\dagger \hat{b}_n + \hat{b}_n^\dagger \hat{a}_n) + i\gamma(\hat{a}_n^\dagger \hat{a}_n - \hat{b}_n^\dagger \hat{b}_n)] \\ &+ \sum_{n=1}^{N-1}[q(\hat{b}_n^\dagger \hat{a}_{n+1} + \hat{a}_{n+1}^\dagger \hat{b}_n) + q(\hat{b}_{n+1}^\dagger \hat{a}_n + \hat{a}_n^\dagger \hat{b}_{n+1}) \\ &+ i\epsilon(\hat{a}_{n+1}^\dagger \hat{a}_n - \hat{a}_n^\dagger \hat{a}_{n+1}) - i\epsilon(\hat{b}_{n+1}^\dagger \hat{b}_n - \hat{b}_n^\dagger \hat{b}_{n+1})]. \quad (1)\end{aligned}$$

It depicts a dissipative tight-binding chain of finite N cells involving the nearest-neighbor hopping amplitudes p and the long-range hopping amplitudes q (see Fig. 1), with $i\gamma$ denoting the on-site energy exchange with the surrounding environment and $i\epsilon$ the intercell dissipation either due to the intrinsic spin-orbit coupling seen in condensed matter physics [72,73] or introduced by evanescent waves in photonics [74]. $\hat{a}_n$ ($\hat{b}_n$) and $\hat{a}_n^\dagger$ ($\hat{b}_n^\dagger$) are the annihilation and creation operators

of a particle on sublattice A (B) of the unit cell n, respectively. Here, for simplicity, we consider the hoppings to be symmetric or reciprocal so that the non-Hermiticity of Hamiltonian (1) relies solely on the parameter γ; namely, when $\gamma = 0$, the above Hamiltonian is Hermitian; otherwise it is non-Hermitian. All four parameters p, q, γ, and ϵ are assumed to be real. For this model, one can show that it respects a sublattice (or chiral) symmetry, i.e., $\Gamma\hat{H}\Gamma^{-1} = -\hat{H}$, where $\Gamma = \bigoplus_n \sigma_{z,n}$ is the direct sum of z-component Pauli operator $\sigma_{z,n} = \mathrm{diag}(1, -1)$ [2,25,45]. Hence, if $\hat{H}$ has an eigenvector $|\psi\rangle$ with eigenvalue E, then $\Gamma|\psi\rangle$ is also an eigenvector with eigenvalue $-E$. We should point out that this model is a nontrivial generalization of the one adopted in Ref. [40], where $q = \epsilon$ is assumed.

The real-space eigenvalue equation $\hat{H}|\psi\rangle = E|\psi\rangle$ of the Hamiltonian (1) can be analytically solved in a usual way. To be specific, one can write the Hamiltonian as a $2N \times 2N$ matrix, with its matrix elements being defined in terms of the tensor product basis $|n, \alpha\rangle = |n\rangle \otimes |\alpha\rangle \equiv |m\rangle$, where $|n\rangle = \hat{\alpha}_n^\dagger|0\rangle$ $(n = 1, \ldots, N)$ represents the excited state at the cell n, $\alpha \in \{a, b\}$, and $|m\rangle$ $(m = 1, 2, \ldots, 2N)$ are quantum states situated on the lattice sites (1A, 1B, 2A, 2B, ..., NA, NB). In this basis, the eigenvectors $|\psi\rangle$ can now be expressed as $|\psi\rangle = \sum_{m=1}^{2N} \psi_m|m\rangle = (\psi_{1A}, \psi_{1B}, \ldots, \psi_{NA}, \psi_{NB})^{\mathrm{T}}$ (T means transpose). Then, the eigenvalue E and the corresponding eigenvector $|\psi\rangle$ are obtained from this matrix Hamiltonian, within the framework of linear algebra, no matter whether the boundary condition is periodic or open.

Generally, due to the presence of dissipative term $i\gamma$, the Hamiltonian (1) is of the matrix form that involves nonzero diagonal elements. For the current model, one can perform a similarity transformation to reduce it to the one whose diagonal elements become zero, without changing the eigenvalue spectrum. A typical similarity transformation is given below,

$$\hat{H}' = U^{-1}\hat{H}U, \quad U = \mathbf{1}_N \otimes \begin{bmatrix} \sqrt{2}/2 & -i\sqrt{2}/2 \\ -i\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}, \tag{2}$$

which can yield an equivalent Hamiltonian $\hat{H}'$,

$$\hat{H}' = \sum_{n=1}^{N}[(p+\gamma)\hat{a}_n^\dagger\hat{b}_n + (p-\gamma)\hat{b}_n^\dagger\hat{a}_n] + \sum_{n=1}^{N-1}[(q+\epsilon) \times (\hat{b}_n^\dagger\hat{a}_{n+1} + \hat{a}_{n+1}^\dagger\hat{b}_n) + (q-\epsilon)(\hat{a}_n^\dagger\hat{b}_{n+1} + \hat{b}_{n+1}^\dagger\hat{a}_n)]. \tag{3}$$

It follows easily that if E and $|\varphi\rangle$ satisfy the eigenvalue equation $\hat{H}'|\varphi\rangle = E|\varphi\rangle$ of the Hamiltonian (3), then E and $|\psi\rangle = U|\varphi\rangle$ will exactly fulfill $\hat{H}|\psi\rangle = E|\psi\rangle$ of the Hamiltonian (1). An inspection of these two eigenstates reveals that their wave components on each lattice site are related by

$$\psi_{nA} = \frac{\sqrt{2}}{2}(\varphi_{nA} - i\varphi_{nB}), \quad \psi_{nB} = \frac{\sqrt{2}}{2}(-i\varphi_{nA} + \varphi_{nB}). \tag{4}$$

Therefore, below we will derive the analytical solutions of the Hamiltonian (3), which are relatively simple in form, and then give the solutions of Eq. (1) by means of Eqs. (4).

The energy spectra are same for both Hamiltonian Eqs. (1) and (3). By introducing the generalized plane-wave basis $|k\rangle = (1/\sqrt{N})\sum_{n=1}^{N} e^{ink}|n\rangle$, with $k \in \mathbb{C}$, which applies to external degree of freedom of the lattice, the generalized Bloch Hamiltonian $\mathcal{H}'(\beta \equiv e^{ik}) = \langle k|\hat{H}'|k\rangle$ can be obtained, in a matrix form, as [75]

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

where $R_+(\beta) = (q+\epsilon)/\beta + (p+\gamma) + (q-\epsilon)\beta$ and $R_-(\beta) = (q+\epsilon)\beta + (p-\gamma) + (q-\epsilon)/\beta$. Then, in momentum space, the eigenvalue equation under OBC reads as

$$R_+(\beta)R_-(\beta) - E^2 = 0, \tag{6}$$

which will be a quadratic equation of the complex variable β for $q = \epsilon$ but be a quartic one for $q \neq \epsilon$. We should point out that when $k \in [0,\ 2\pi] \in \mathbb{R}$, Eq. (6) can give exactly the spectrum of Hamiltonian (1) or (3) under PBC, which generally encompasses the energy spectrum obtained under OBC.

According to Eq. (6), the energy spectra under OBC would exhibit different characteristics for the $q = \epsilon$ and $q \neq \epsilon$ cases. The former case has been discussed in Ref. [40], but the analytical solutions for topological edge states, especially for those occurring with finite lattice size N, were not well addressed. The latter case involves the four-dimensional parameter space and thus has more complicated behaviors for both bulk and edge states, when the lattice size is finite. To the best of our knowledge, the asymptotic formula for energy E in the latter case was not reported yet in previous studies, not to say of the explicit zero-mode solutions at arbitrary lattice size N and their localization conditions, including their relation to EPs. In this work, we will first derive the asymptotic analytical solutions for zero modes in the simple case $q = \epsilon$, and then extend our investigation to the more general case $q \neq \epsilon$.

III. ANALYTICAL ZERO MODES AND EPs AT $q = \epsilon$

When $q = \epsilon$, Eq. (6) becomes a quadratic equation of β, which we could rewrite explicitly as below,

$$(2\epsilon/\beta + \gamma + p)(2\beta\epsilon - \gamma + p) = E^2. \tag{7}$$

This quadratic equation will always have two roots β_1 and β_2 for a given value of E. As every eigenvalue E has to be associated to one eigenstate $|\varphi\rangle$, the wave components of this eigenstate on lattice sites will also depend on the values β_1 and β_2 only. Therefore, one can assume the state components to take the form

$$\varphi_{nA} = \frac{1}{E}\left[c_1R_+(\beta_1)\beta_1^n + c_2R_+(\beta_2)\beta_2^n\right], \tag{8}$$

$$\varphi_{nB} = c_1\beta_1^n + c_2\beta_2^n, \tag{9}$$

where the function $R_+(\beta)$ is defined in Eq. (5) and c_1 and c_2 are two complex coefficients to be determined. Substituting Eqs. (8) and (9) into the eigenvalue equation $\hat{H}'|\varphi\rangle = E|\varphi\rangle$ yields the system of boundary equations:

$$\begin{bmatrix} 1 & 1 \\ \beta_1^{N+1}R_+(\beta_1) & \beta_2^{N+1}R_+(\beta_2) \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = 0, \tag{10}$$

which gives rise to the following boundary conditions,

$$(\gamma + p)\left(\beta_1^{N+1} - \beta_2^{N+1}\right) + 2\epsilon\left(\beta_1^N - \beta_2^N\right) = 0, \tag{11}$$

when c_1 and c_2 have a nontrivial form

$$c_1 = -c_2 = g. \tag{12}$$

Here the constant g is used to normalize the wave function such that $\langle\varphi|\varphi\rangle = 1$, namely, $\sum_{n=1}^{N}(|\varphi_{nA}|^2 + |\varphi_{nB}|^2) = 1$.

Now, under OBC, by using Eqs. (7) and (11), one can uniquely determine the spectrum E of SSH Hamiltonian (1) or (3) for any given lattice size N. Once E, and therefore β_1 and β_2, are known, the eigenstate $|\varphi\rangle$ that applies for either bulk or edge mode can be found through Eqs. (8) and (9). The resultant eigenvectors $|\psi\rangle$ of Hamiltonian (1) are obtained via Eqs. (4). This can be well understood if one sets

$$\beta_1 = re^{i\phi}, \quad \beta_2 = re^{-i\phi}, \tag{13}$$

which, when inserting back to Eq. (7), yields

$$r = \sqrt{\frac{p-\gamma}{p+\gamma}}, \tag{14}$$

$$E = \pm\sqrt{4r\epsilon(\gamma+p)\cos(\phi) - \gamma^2 + 4\epsilon^2 + p^2}. \tag{15}$$

When substituting Eq. (13) again into Eq. (11), we get

$$\delta\sin[(N+1)\phi] + \sin(N\phi) = 0, \tag{16}$$

where $\delta = (\gamma+p)r/(2\epsilon)$. Compared with Eq. (11), Eq. (16) involves only one single variable ϕ for given system parameters and thus is easier to solve. In fact, Eq. (16) can be transformed to an N-degree polynomial of $\cos\phi$, resulting in N allowed values of ϕ, either real or complex. Each allowed ϕ gives rise to two energy values, as seen in Eq. (15). In this way we obtain the total $2N$ energy points which will be symmetrically distributed with respect to the imaginary axis in the complex plane. In the thermodynamical limit (i.e., $N \to \infty$), these energy points become dense and constitute what we now called an energy spectrum. Generally, if the system Hamiltonian is Hermitian ($\gamma = 0$), the energy spectrum is definitely real. However, if the systems are non-Hermitian ($\gamma \neq 0$), the spectrum may be complex when no additional symmetries (e.g., $\mathcal{PT}$ symmetry [1,6], pseudo-Hermiticity [3]) are imposed. Likewise, in Hermitian systems, the bulk states are extensive on all lattice sites, but in non-Hermitian systems, they may undergo non-Hermitian skin effect and tend to concentrate on the boundaries, as revealed in Refs. [25–31,48].

One of our current concerns is to inspect the parameter conditions under which the zero modes, namely, the edge modes with zero energy, could exist and to figure out what are their asymptotic analytical solutions for finite lattice size N. A key step toward this matter is to find the asymptotic formula of the edge-mode energy E for arbitrary N, which approaches zero when $N \to \infty$. Actually, expressing $\cos\phi$ with E^2 according to Eq. (15) and then inserting it into Eq. (16), one can obtain a $2N$-degree polynomial of E. Using the asymptotic leading-term expansion method [76] and recalling the fact that the energy E would vanish as N approaches infinity, we obtain the edge-mode energy for finite N as

$$E \simeq \pm(1-\delta^2)\delta^{N-1}\sqrt{p^2-\gamma^2}, \tag{17}$$

which holds true for $|\delta| < 1$, or equivalently, for $4\epsilon^2 + \gamma^2 - p^2 > 0$. Obviously, this energy scales with δ^{N-1} and will approach zero when $N \to \infty$.

On the other hand, as E approaches zero at an exponential rate, Eq. (7) has two asymptotic roots $\beta_1' = (\gamma-p)/(2\epsilon)$ and $\beta_2' = -2\epsilon/(\gamma+p)$, which always holds true for a chain of more than five cells long. Hence, from Eqs. (8) and (9), it follows that the edge modes for finite N can be expressed as

$$\varphi_{nA}^{\mathrm{ed}} = g\left(\frac{\gamma-p}{2\epsilon}\right)^n \frac{4\epsilon^2+\gamma^2-p^2}{(\gamma-p)E}, \tag{18}$$

$$\varphi_{nB}^{\mathrm{ed}} = g\left[\left(\frac{\gamma-p}{2\epsilon}\right)^n - \left(-\frac{2\epsilon}{\gamma+p}\right)^n\right], \tag{19}$$

where E is given by Eq. (17) and the constant g takes

$$g = \left\{\sum_{n=1}^{N}\left[\frac{(4\epsilon^2+\gamma^2-p^2)^2}{(\gamma-p)^2}\left|\frac{(\beta_1')^n}{E}\right|^2 + |(\beta_1')^n - (\beta_2')^n|^2\right]\right\}^{-1/2}. \tag{20}$$

The parameter conditions defining the domain of zero modes can be derived from $|\beta_1'| = |\beta_2'|$ [25,48], which can be expressed more explicitly as $4\epsilon^2 + \gamma^2 - p^2 \geqslant 0$ and $4\epsilon^2 + p^2 - \gamma^2 \geqslant 0$. It is worth noting that these two inequalities need to be satisfied simultaneously, and then the domain of zero modes, including its borderlines, can always be determined. Thus, one obtains the asymptotic eigenvalues and eigenvectors of two topological edge modes of Hamiltonian (1) for finite lattice size N by means of Eqs. (4).

Figure 2 shows, from left to right, the numerical OBC energy spectra (red dots) in the complex plane, the intensity distribution of all bulk and edge eigenstates, the numerical intensity profiles (blue circles) of topological edge modes selected from the left contour plots, and the real and imaginary parts of two edge modes for finite lattice size $N = 20$, using three sets of different parameters as indicated in the caption. Here, the symbols $\Re$ and $\Im$ denote the real and imaginary parts of a complex number. In the complex energy plane, we also provide the PBC energy spectra (colored directional lines), which are obtained from Eq. (7) with $\beta = e^{ik}$ and letting the momentum k run from 0 to 2π. It has turned out that these PBC spectra exhibit a topological nature responsible for the non-Hermitian skin effect [28]. As pointed out in our recent work [48], when the PBC spectrum surrounds any OBC energy point in a counterclockwise (clockwise) way, the bulk mode associated to this OBC energy will tend to accumulate at the left (right) boundary, resulting in the formation of skin modes; see Figs. 2(a) and 2(b) for left localization, and Figs. 2(i) and 2(j) for right localization. However, when the PBC spectrum shrinks to a line overlapping the OBC spectrum [see Fig. 2(e)], then all bulk modes are extensively distributed and no skin effects will take place any more [see Fig. 2(f)].

As regards the numerical OBC spectra shown in Figs. 2(a), 2(e), and 2(i), there appears a red point at the origin that would correspond to what we called zero modes. The green insets show a magnified version of this red point, revealing that it actually consists of two distinct points for $N = 20$. As one can check, these two energy points would sit at $\pm 2.6 \times 10^{-12}$, $\pm 9.6 \times 10^{-10}$, and $\pm 4.9 \times 10^{-11}$, respectively, for the above three cases, exhibiting amazing consistence with our analytical formula (17). Hence, only in the thermodynamic limit ($N \to \infty$) can these two energy points merge into the one

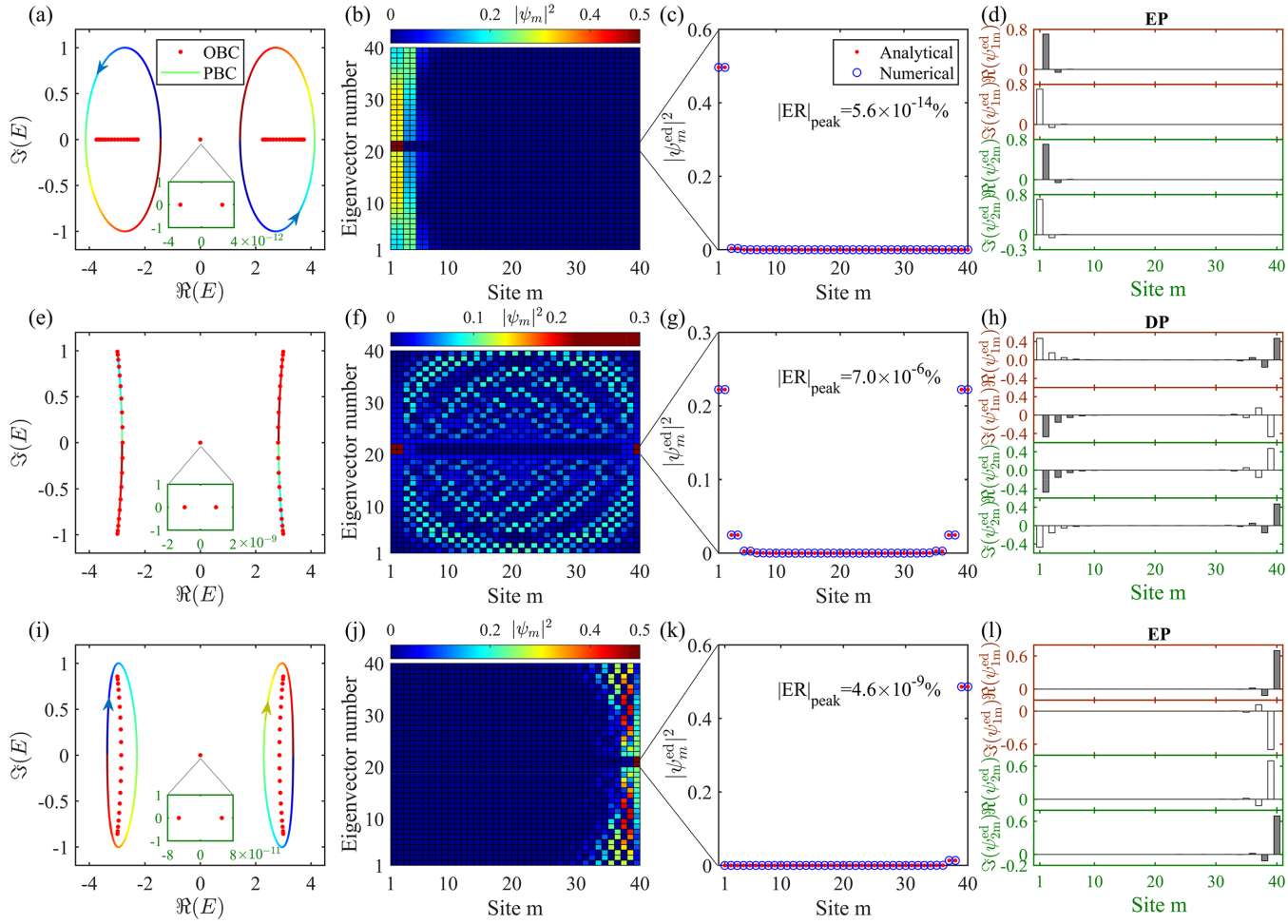

FIG. 2. The numerical OBC energy spectra denoted by red dots (column 1), the contour plots of all bulk and edge eigenstates (column 2), the numerical intensity of topological edge modes compared to their analytical solutions given by Eqs. (4), (18), and (19) (column 3), and the real ($\Re$) and imaginary ($\Im$) parts of the eigenvectors of two zero modes ψ_{jm}^{ed} ($j = 1, 2$) (column 4), all obtained under OBC for a finite lattice size $N = 20$, using three sets of different parameters: (a)–(d) $p = 5/4$, (e)–(h) $p = 0$, and (i)–(l) $p = -1/2$, with $\epsilon = 3/2$ and $\gamma = 1$ being the same. The magnified insets in (a), (e), and (i) show that the red point at the origin consists of two energy points whose coordinates can be well predicted by Eq. (17). The PBC energy spectra (colored directional lines) are also presented therein, to show the topological origin of the non-Hermitian skin effect. In (d), (h), and (l), the white and gray bars denote the wave components on the sublattice site A and B, respectively. Besides, the relative errors $|\mathrm{ER}|_{\mathrm{peak}}$ of our analytical solutions in comparison to numerical ones at the peak of mode intensity have been indicated in (c), (g), and (k).

at the origin (except for some special parameters $p = \pm\gamma$ discussed below) and thus only in such limiting case can the topological edge modes manifest as genuine zero modes whose energy is exactly degenerate. Accordingly, either the EPs or DPs for topological zero modes generally exist only in this thermodynamic limit. However, in our numerical presentations obtained with finite lattice size, we will still call the red point at the origin the zero modes, just for the sake of convenience. In fact, the spectral degeneracy and state behaviors of topological zero modes in the limit of infinite N can be well comprehended by inspecting them with a finite yet relatively large lattice size.

More strikingly, from Figs. 2(c), 2(g), and 2(k), one can see that our analytical solutions expressed by Eqs. (4), (18), and (19) (red dots) coincide perfectly with their numerical ones (blue circles), no matter whether these zero modes are localized at the left end (for short, left localization), in both ends (double localization), or at the right end (right localization). The relative errors at the peak, defined as $|\mathrm{ER}|_{\mathrm{peak}} = |(|\psi_m^{\mathrm{ed}}|^2_{\mathrm{ana}} - |\psi_m^{\mathrm{ed}}|^2_{\mathrm{num}})|/|\psi_m^{\mathrm{ed}}|^2_{\mathrm{num}}$, for these three cases are all unimaginably small. In particular, an inspection of these edge-mode analytical solutions reveals that for $N \to \infty$ where two zero modes are spectrally degenerate, their eigenvectors are only degenerate (i.e., two eigenvectors are parallel or antiparallel) in the left or right localization case [see Figs. 2(d) and 2(l)], whereas in the double localization case, these eigenvectors can be made orthogonal to each other and thus are nondegenerate [see Fig. 2(h)]. This implies that in non-Hermitian systems, not all zero modes can be interpreted as EPs, if seen in the parameter space; they might be DPs in the double localization case, as occurred in Hermitian systems [75]. To better understand it, we demonstrate in

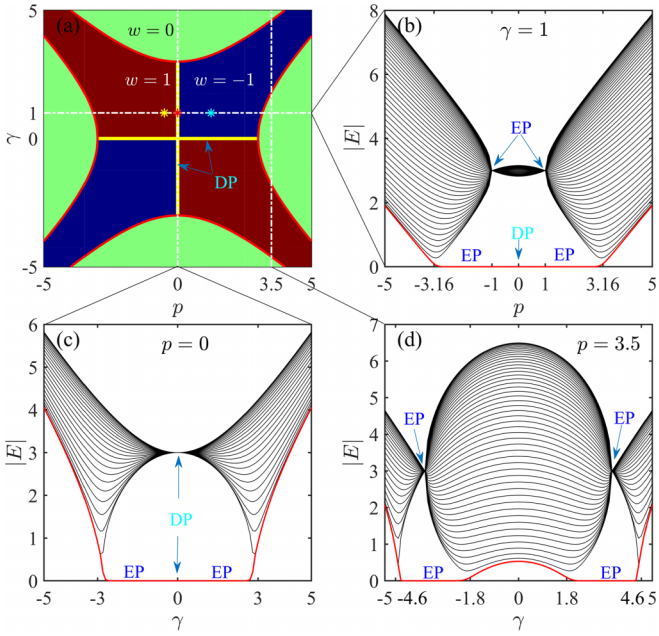

FIG. 3. (a) Map of the winding number in the plane (p, γ) obtained for given $N = 50$ and $\epsilon = 3/2$. The red, blue, and green regions represent $w = 1, -1$, and 0, respectively, with the red demarcation curves being defined by $4\epsilon^2 + \gamma^2 - p^2 = 0$ and $4\epsilon^2 + p^2 - \gamma^2 = 0$. (b), (c), and (d) show the energy bands along the three white dashed lines in (a), where the edge state is highlighted by red curve. Additionally, the parameter points that produce Fig. 2 (from top to bottom) have been indicated in (a) by the cyan, red, and yellow asterisks, respectively, and the yellow lines at $p = 0$ and $\gamma = 0$ denote the set of DPs.

Fig. 3(a) the map of the topological winding number in the plane (p, γ), which is calculated according to the definition (16) in Ref. [48]. As seen, the green region, which features a zero winding number ($w = 0$), prohibits the formation of zero modes, whereas the red (blue) region, which owns the winding number $w = 1$ ($w = -1$), can admit the zero modes, all consistent with the non-Bloch BBC principle developed recently [25]. Meanwhile, we find that these zero modes, once available, tend to accumulate at the left (right) boundary if $|\beta_1'\beta_2'| < 1$ (> 1), but will be localized on both ends when $|\beta_1'\beta_2'| = 1$, where $\beta_{1,2}'$ are the two roots of Eq. (7) when $E = 0$ is taken [48]. Accordingly, one can conclude that both red and blue regions that support topological zero modes can be identified as EPs. However, the yellow solid lines at $\gamma = 0$ (the Hermitian case) and $p = 0$ (the non-Hermitian case), which satisfy $|\beta_1'\beta_2'| = 1$, should fall within the DPs, because the eigenvectors of zero modes are now nondegenerate.

As a matter of fact, EPs not only occur for edge modes, but also for skin modes of the bulk. It is easy to verify that when $p = \pm\gamma$, the skin modes of Hamiltonian (1) with $q = \epsilon$ have also two degenerate energies $E_\pm = \pm 2\epsilon$, each having $(N-1)$-fold degenerate eigenvectors $|\psi_\pm^{\text{skin}}\rangle$, as revealed in Ref. [40]. Specifically, when $p = \gamma$, the degenerate eigenvectors associated to $E_\pm = \pm 2\epsilon$ can be expressed as

$$|\psi_\pm^{\text{skin}}\rangle = \left(\frac{i\gamma}{\epsilon} \pm 1, \pm i + \frac{\gamma}{\epsilon}, i, 1, 0, \ldots, 0\right), \tag{21}$$

which are localized at the left boundary. Contrarily, when $p = -\gamma$, the associated degenerate eigenvectors become

$$|\psi_\pm^{\text{skin}}\rangle = \left(0, \ldots, 0, \frac{\epsilon}{\pm\epsilon - i\gamma}, \frac{-\epsilon}{\pm i\epsilon + \gamma}, -\frac{i\gamma \pm \epsilon}{\gamma \pm i\epsilon}, 1\right), \tag{22}$$

which tend to accumulate at the right boundary. Here we point out that different from the asymptotic edge-mode solutions (18) and (19), Eqs. (21) and (22) are exact analytical solutions of the skin modes that apply for any finite lattice size N. Thus, according to the definition of EP, the special parameter points $p = \pm\gamma$ that lead to the skin modes can be thought of as $(N-1)$-order EPs [57]. To show this, in Figs. 3(b)–3(d), we plot the energy bands along the white dashed lines in Fig. 3(a). It is clear that at $p = \pm\gamma$, the energy bands shrink to a degenerate point $|E| = 2\epsilon$ and the corresponding eigenvectors given by Eqs. (21) and (22) will also be degenerate, indicating the formation of higher-order EPs, as seen by arrows in Figs. 3(b) and 3(d). However, at the special point $p = \gamma = 0$, the system becomes now Hermitian, and thus the energy point $|E| = 2\epsilon$ in Fig. 3(c) is indeed a DP. In addition, for these special system parameters $p = \pm\gamma$, there always appear two degenerate zero modes whose energy is exactly zero, as seen by the red curve in Figs. 3(b)–3(d). The exact analytical solutions for such degenerate zero modes have been first provided by Lee in Ref. [40], see Eqs. (6) and (7) therein.

IV. EXPLICIT ZERO MODES AND EPs FOR $q \neq \epsilon$

Let us now handle the zero-mode problem in the $q \neq \epsilon$ case for the open Hamiltonian system (1). For this end, we rewrite the eigenvalue Eq. (6) in an explicit form

$$\left[\frac{q+\epsilon}{\beta} + p + \gamma + (q-\epsilon)\beta\right]\left[(q+\epsilon)\beta + p - \gamma + \frac{q-\epsilon}{\beta}\right] = E^2, \tag{23}$$

which, for any given E, admits four roots β_j ($j = 1, 2, 3, 4$). Similarly, in terms of these roots, the site components of the eigenstates of Hamiltonian (3) can be expressed as

$$\varphi_{nA} = \frac{1}{E}\sum_{j=1}^{4} c_j R_+(\beta_j)\beta_j^n, \quad \varphi_{nB} = \sum_{j=1}^{4} c_j \beta_j^n, \tag{24}$$

where $R_+(\beta_j) = (q+\epsilon)/\beta_j + p + \gamma + (q-\epsilon)\beta_j$, the same as defined above, and c_j are four complex coefficients that should satisfy the following system of linear equations:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ R_+(\beta_1) & R_+(\beta_2) & R_+(\beta_3) & R_+(\beta_4) \\ \beta_1^{N+1} & \beta_2^{N+1} & \beta_3^{N+1} & \beta_4^{N+1} \\ \beta_1^{N+1}R_+(\beta_1) & \beta_2^{N+1}R_+(\beta_2) & \beta_3^{N+1}R_+(\beta_3) & \beta_4^{N+1}R_+(\beta_4) \end{bmatrix} \times [c_1 \quad c_2 \quad c_3 \quad c_4]^{\mathrm{T}} = 0. \tag{25}$$

Solving this system of equations for nontrivial c_j values, one can obtain the open boundary equation:

$$\beta_{12}\beta_{34}[(\beta_1\beta_2)^{N+1} + (\beta_3\beta_4)^{N+1}] - \beta_{13}\beta_{24}[(\beta_1\beta_3)^{N+1} + (\beta_2\beta_4)^{N+1}] + \beta_{14}\beta_{23}[(\beta_1\beta_4)^{N+1} + (\beta_2\beta_3)^{N+1}] = 0, \tag{26}$$

where $\beta_{ij} = (\beta_i - \beta_j)[\beta_i\beta_j - (q+\epsilon)/(q-\epsilon)]$. Meanwhile, the coefficients c_j are found to be

$$c_1 = g\beta_1\big(\beta_{34}\beta_2^{N+2} + \beta_{23}\beta_4^{N+2} - \beta_{24}\beta_3^{N+2}\big), \tag{27}$$

$$c_2 = -g\beta_2\big(\beta_{34}\beta_1^{N+2} + \beta_{13}\beta_4^{N+2} - \beta_{14}\beta_3^{N+2}\big), \tag{28}$$

$$c_3 = g\beta_3\big(\beta_{24}\beta_1^{N+2} - \beta_{14}\beta_2^{N+2} + \beta_{12}\beta_4^{N+2}\big), \tag{29}$$

$$c_4 = -g\beta_4\big(\beta_{23}\beta_1^{N+2} + \beta_{12}\beta_3^{N+2} - \beta_{13}\beta_2^{N+2}\big), \tag{30}$$

where $\sum_{j=1}^{4} c_j = 0$ and g is a scaling constant used to normalize the eigenvector $|\varphi\rangle$ such that $\langle\varphi|\varphi\rangle = 1$.

In a similar fashion, we set

$$\beta_1 = e^{\frac{i}{2}(\Theta+\Phi-\Psi)}, \quad \beta_2 = e^{\frac{i}{2}(\Theta-\Phi+\Psi)}, \tag{31}$$

$$\beta_3 = e^{\frac{i}{2}(-\Theta+\Phi+\Psi)}, \quad \beta_4 = e^{-\frac{i}{2}(\Theta+\Phi+\Psi)}, \tag{32}$$

where Θ, Φ, and Ψ are complex phase angles. Inserting these formulas into Eq. (23) yields, with the help of Vieta's theorem,

$$E = \pm\sqrt{2(\epsilon^2 - q^2)(X + Y + Z) + 2(\epsilon^2 + q^2) + p^2 - \gamma^2}, \tag{33}$$

$$(X+1)(Y+1)(Z+1) = \frac{2q^2p^2}{(\epsilon^2-q^2)^2}, \tag{34}$$

$$(X-1)(Y-1)(Z-1) = \frac{2\gamma^2\epsilon^2}{(\epsilon^2-q^2)^2}, \tag{35}$$

where $X = \cos\Theta$, $Y = \cos\Phi$, $Z = \cos\Psi$. Inserting them again into Eq. (26), we obtain a more explicit boundary condition:

$$\begin{aligned} &f(X)T_{N+1}(X)(Y-Z) + f(Y)T_{N+1}(Y)(Z-X) \\ &\quad + f(Z)T_{N+1}(Z)(X-Y) = 0, \end{aligned} \tag{36}$$

where $f(x) = (\epsilon^2 + q^2)/(\epsilon^2 - q^2) + x$ and $T_{N+1}(x)$ is a Chebyshev polynomial of the first kind, defined by

$$T_{N+1}(x) = \tfrac{1}{2}(x + i\sqrt{1-x^2})^{N+1} + \tfrac{1}{2}(x - i\sqrt{1-x^2})^{N+1}. \tag{37}$$

Obviously, by using the change of variables, the original Eqs. (23) and (26), which involve five variables E and β_j, are simplified to Eqs. (34)–(36), which involve only three variables X, Y, and Z. Once X, Y, and Z are worked out, the energy spectrum E can be obtained from Eq. (33). Also, the four β_j and then the four c_j are all figured out, using the formulas (27)–(32). Consequently, the whole bulk and edge eigenstates of the Hamiltonian Eq. (1) are obtained from Eqs. (4) and (24).

Basically, by using Eqs. (33)–(35), X, Y, and Z can be expressed in terms of the energy E. Then, the boundary Eq. (36) gives rise to a $2N$-degree polynomial of E, which yields the complete energy spectrum for given system parameters. Here we provide the asymptotic analytical formula of energy for the zero modes applicable for finite lattice size N:

$$E \simeq \pm\sqrt{\frac{2[T_{N+1}(\alpha_1) - T_{N+1}(\alpha_2)][(4\epsilon^2+\gamma^2)q^2 - \epsilon^2p^2]^2}{\eta\epsilon^2q^2\big[T_{N+1}(\alpha_1) + T_{N+1}(\alpha_2) - 2T_{N+1}\big(\frac{q^2+\epsilon^2}{q^2-\epsilon^2}\big)\big]}}, \tag{38}$$

where

$$\alpha_1 = \frac{\gamma^2 - p^2 + \eta}{4(\epsilon^2 - q^2)}, \quad \alpha_2 = \frac{\gamma^2 - p^2 - \eta}{4(\epsilon^2 - q^2)}, \tag{39}$$

$$\eta = \sqrt{[4\epsilon^2 - 4q^2 + (p+\gamma)^2][4\epsilon^2 - 4q^2 + (p-\gamma)^2]}. \tag{40}$$

The parameter conditions for this energy to hold true are

$$(4\epsilon^2 + \gamma^2)q^2 - \epsilon^2p^2 > 0, \quad \text{and} \quad (4\epsilon^2 + p^2)q^2 - \epsilon^2\gamma^2 > 0, \tag{41}$$

which also define the domain of existence of topological zero modes. Of course, when $\eta = 0$, the zero-mode energy formula (38), with the help of L'Hôpital's rule, reduces to

$$E \simeq \pm\sqrt{\frac{T''_{N+1}(\alpha)[(4\epsilon^2+\gamma^2)q^2 - \epsilon^2p^2]^2}{2\epsilon^2q^2(\epsilon^2-q^2)\big[T_{N+1}(\alpha) - T_{N+1}\big(\frac{q^2+\epsilon^2}{q^2-\epsilon^2}\big)\big]}}, \tag{42}$$

where $T''_{N+1}(\alpha) = \frac{d^2T_{N+1}(x)}{dx^2}|_{x=\alpha}$ and $\alpha = \frac{\gamma^2-p^2}{4(\epsilon^2-q^2)}$. As one can check, when $N \to \infty$, the zero-mode energy E, given by either Eq. (38) or (42), could approach zero and thus is degenerate.

The eigenvectors of the topological edge modes under OBC can also be expressed in an explicit analytical form. To do this, let us first write the four roots β'_j of the quartic Eq. (23) with $E = 0$ as follows:

$$\beta'_1 = \frac{\gamma - p + \sqrt{4\epsilon^2 - 4q^2 + (\gamma-p)^2}}{2(\epsilon+q)}, \tag{43}$$

$$\beta'_2 = \frac{\gamma - p - \sqrt{4\epsilon^2 - 4q^2 + (\gamma-p)^2}}{2(\epsilon+q)}, \tag{44}$$

$$\beta'_3 = \frac{\gamma + p - \sqrt{4\epsilon^2 - 4q^2 + (\gamma+p)^2}}{2(\epsilon-q)}, \tag{45}$$

$$\beta'_4 = \frac{\gamma + p + \sqrt{4\epsilon^2 - 4q^2 + (\gamma+p)^2}}{2(\epsilon-q)}. \tag{46}$$

Now, substituting Eqs. (43)–(46) into Eq. (24), we obtain, after some algebra, the state components of the zero modes,

$$\varphi_{nA} = \frac{g}{E}[c_1R_+(\beta'_1)(\beta'_1)^{n+1} - c_2R_+(\beta'_2)(\beta'_2)^{n+1}], \tag{47}$$

$$\begin{aligned} \varphi_{nB} = g\{&c_1\beta'_1[(\beta'_1)^n - (\beta'_4)^n] - c_2\beta'_2[(\beta'_2)^n - (\beta'_4)^n] \\ &+ c_3\beta'_3[(\beta'_3)^n - (\beta'_4)^n]\}, \end{aligned} \tag{48}$$

where $c_1 = \beta'_{23}(\beta'_4)^{N+2} - \beta'_{24}(\beta'_3)^{N+2}$, $c_2 = \beta'_{13}(\beta'_4)^{N+2} - \beta'_{14}(\beta'_3)^{N+2}$, $c_3 = \beta'_{24}(\beta'_1)^{N+2} - \beta'_{14}(\beta'_2)^{N+2} + \beta'_{12}(\beta'_4)^{N+2}$, and g is a free constant enabling $\sum_{n=1}^{N}(|\varphi_{nA}|^2 + |\varphi_{nB}|^2) = 1$. Consequently, substituting Eqs. (47) and (48) into Eqs. (4), one can obtain the resultant zero-mode solutions for the Hamiltonian (1). As such, for any given system parameters, the asymptotic zero-mode solutions along with their energy for a finite lattice size N can be directly determined by Eqs. (38), (41), (47), and (48), without requiring the full knowledge of topological phase transitions of non-Hermitian systems.

Further, as revealed in our recent work [48], one can easily determine which ends of the 1D open chain could host the zero modes, once they are allowed to exist. Specifically, sorting the

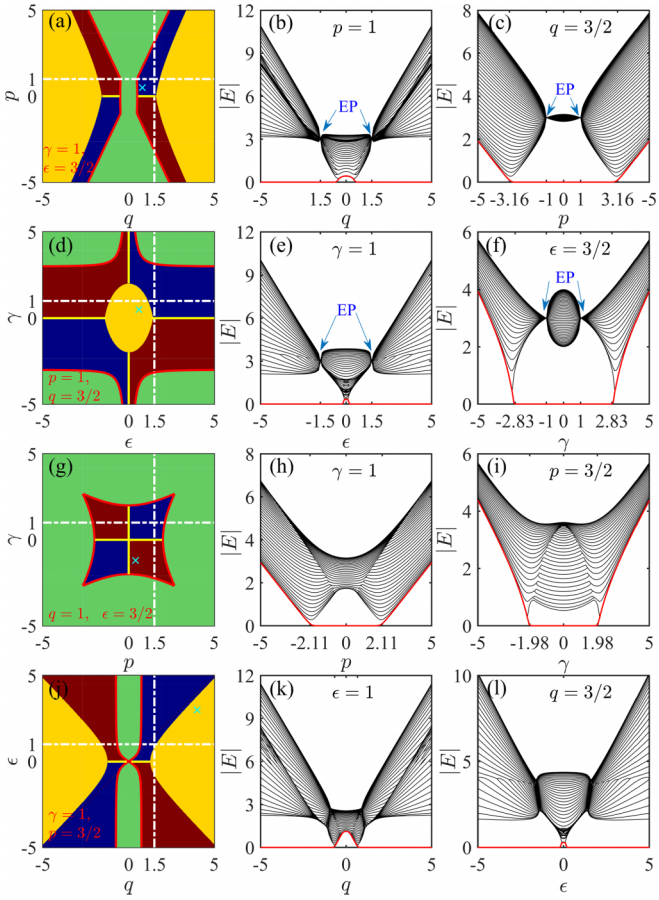

FIG. 4. The phase diagrams defining the domain of zero modes in the parameter plane (left column) and the energy bands calculated with $N = 60$ corresponding to the white dashed lines in each diagram (the middle and right columns), obtained for given system parameters: (a)–(c) $\gamma = 1$, $\epsilon = 3/2$; (d)–(f) $p = 1$, $q = 3/2$; (g)–(i) $q = 1$, $\epsilon = 3/2$; and (j)–(l) $\gamma = 1$, $p = 3/2$. Here, the green zone indicates that the zero modes are forbidden there, but the other red, blue, and yellow regions can admit the zero modes to accumulate at the right boundary, the left boundary, and both boundaries, respectively. In the energy-band plots, all the edge states that can manifest as zero modes have been highlighted red and the branch points that allow both spectral degeneracy and eigenvector coalescence for the skin modes have been labeled as EP. The cyan cross in each phase diagram indicates the parameter point that will be used in Fig. 5.

four β_j' in Eqs. (43)–(46) according to their moduli and renaming them from small to large as β_1^e, β_2^e, β_3^e, and β_4^e, then the zero modes will accumulate on the left end when $|\beta_2^e\beta_3^e| < 1$, on the right end when $|\beta_2^e\beta_3^e| > 1$, and on both ends when $|\beta_2^e\beta_3^e| = 1$, as discussed in the last section. Therefore, by use of these analytical criteria along with the solution conditions (41), one can obtain the phase diagrams in four-dimensional parameter space, which would indicate where the zero modes can exist and which ends are preferable for zero modes to reside on. As typical results, Figs. 4(a), 4(d), 4(g), and 4(j) show four different phase diagrams, which are plotted with respect to (q, p), (ϵ, γ), (p, γ), and (q, ϵ), respectively, while keeping the other two parameters fixed. In each phase diagram, the green region indicates that the zero modes are forbidden there, while the other red, blue, and yellow regions signify that the zero modes are available, but tend to accumulate at the right boundary, the left boundary, and both boundaries, respectively, all exhibiting perfect consistency with the above localization criteria. We show that the double localization case of zero modes in which $|\beta_2^e\beta_3^e| = 1$ is actually not seldom encountered and it may occupy a considerably large parameter regime [see yellow regions in Figs. 4(a), 4(d), and 4(j)]. Hence, with the same reason as used in Figs. 2 and 3, we conclude again that not all zero modes can be attributed to EPs; while those sitting in red and blue regions are generated from the EPs, the zero modes in the yellow region are indeed from the DPs.

We particularly point out that all the above phase diagrams obtained analytically are well consistent with our numerical results based on the Hamiltonian (1). This can be seen first from the energy bands calculated numerically with $N = 60$, which are presented in the middle and right columns in Fig. 4, corresponding to the horizontal and vertical white dashed lines in each phase diagram, respectively. It is clearly seen that the edge states (red lines) of the SSH model can manifest as zero modes in a domain strictly consistent with what the phase diagram reveals, implying that our analytical conditions (41) are accurate for defining the domain of zero modes. The slight discrepancy in Figs. 4(e) and 4(l), which occurs around $\epsilon = 0$, results from the finite lattice size $N = 60$ that we have adopted here. Actually, in the thermodynamical limit $N \to \infty$, this discrepancy would disappear, since the zero modes are now available at $\epsilon = 0$. In addition, as revealed in Eqs. (21) and (22), some of the skin modes can possess both energy degeneracy and eigenvector coalescence as well, provided that $q = \epsilon$ and $p = \pm\gamma$ are simultaneously fulfilled. Naturally, in the parameter space, these singular branch points that lead to such degenerate skin modes belong to EPs as well, as seen in Figs. 4(b), 4(c), 4(e), and 4(f).

Next, we show in Fig. 5 that the localization features of zero modes suggested by color regions in the phase diagrams are also in excellent agreement with our numerical results, where four different sets of system parameters and a finite lattice size $N = 20$ are used (see the four cyan crosses in the phase diagrams of Fig. 4). It is exhibited that for the parameter points in the blue region, the zero modes accumulate at the left boundary [see Figs. 5(b) and 5(c)], whereas for those in the red region, the zero modes will aggregate at the right boundary [see Figs. 5(j) and 5(k)]. However, when the parameters lie just within the yellow region, the zero modes tend to reside at both boundaries [see Figs. 5(f) and 5(g) or Figs. 5(n) and 5(o)]. More strikingly, no matter whether these zero modes are singly localized or doubly localized, our analytical solutions (red dots) agree very well with the numerical results (blue circles), with relative errors $|\mathrm{ER}|_{\mathrm{peak}}$ all far blow 1%, as seen in Figs. 5(c), 5(g), 5(k), and 5(o).

Finally, we emphasize that, due to the non-Hermitian skin effect, the bulk states may also experience strong localization at the boundaries, but they follow different rules. Generally, when the PBC energy spectrum encircles counterclockwise the base energy point in the OBC spectrum [see Figs. 5(a), 5(e), and 5(m)], the skin modes will move toward the left boundary [see Figs. 5(b), 5(f), and 5(n)], whereas if the PBC spectrum rotates clockwise around the OBC energy point

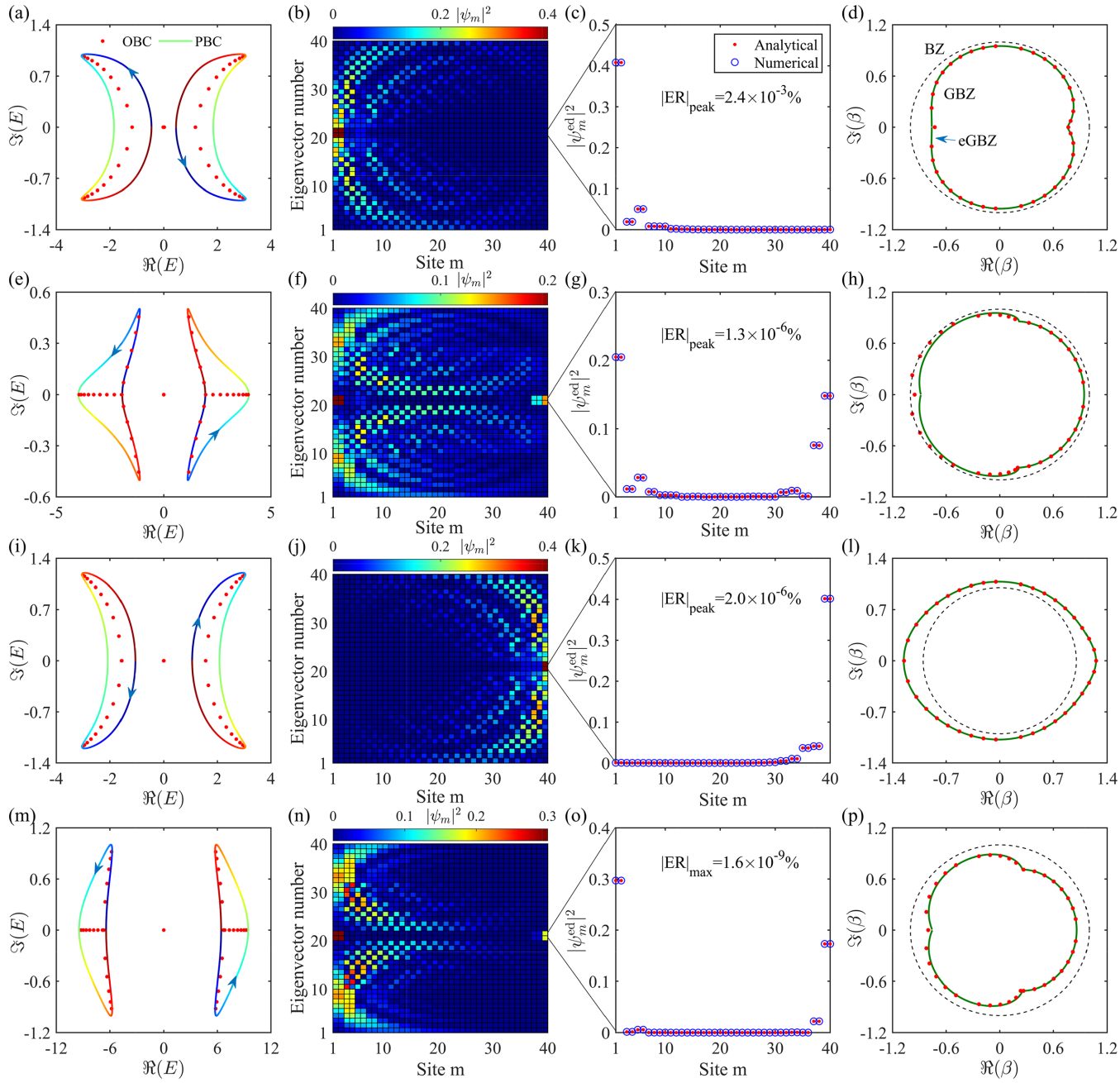

FIG. 5. The OBC and PBC energy spectra in the complex energy plane (the first column), the distribution of all eigenstates (the second column), the analytical zero modes compared with the numerical ones (the third column), and the numerical GBZ compared with the exact GBZ (the fourth column), obtained with $N = 20$ for four different sets of system parameters: (a)–(d) $\gamma = 1$, $\epsilon = 3/2$, $p = 1/2$, $q = 4/5$; (e)–(h) $\gamma = 1/2$, $\epsilon = 3/5$, $p = 1$, $q = 3/2$; (i)–(l) $\gamma = -6/5$, $\epsilon = 3/2$, $p = 2/5$, $q = 1$; and (m)–(p) $\gamma = 1$, $\epsilon = 3$, $p = 3/2$, $q = 4$. These parameters have been indicated by the cyan cross in the phase diagrams in Fig. 4. The corresponding relative errors $|\mathrm{ER}|_{\mathrm{peak}}$ of our analytical solutions compared to numerical ones at the peak of mode intensity have been indicated in (c), (g), (k), and (o). Besides, for the sake of comparison, the BZ (unit circle) has also been given by the thin dashed line in (d), (h), (l), and (p).

[Fig. 5(i)], the skin modes tend to aggregate at the right boundary [Fig. 5(j)]. This is not surprising because the zero modes and skin modes have a different topological origin and thus a different winding number (see Eqs. (15) and (16) in Ref. [48]). Of course, as seen in the rightmost column in Fig. 5, one can also use the GBZ, in comparison with the BZ circle, to judge which sides are preferable for the skin modes to reside in [48].

V. CONCLUSION

In conclusion, we investigated, on an analytical level, the formation of topological edge modes as well as the generation of EPs in an open 1D non-Hermitian SSH model involving dissipative terms, which could offer a prototypical description of topological phase transitions in many physical settings such as photonic lattices [66,67], resonator arrays [68–70],

electrical circuits [15,42], and cold atoms [71]. Using the leading-term perturbation technique [76], we obtained for the first time the asymptotic zero-mode solutions along with universal energy formulas, which are proved to be valid for arbitrary finite lattice size. Meanwhile, we worked out the parameter conditions defining the domain of zero modes and the analytical criteria discriminating their localization characteristics, both of which show perfect consistency with the numerical results. It was revealed that the zero modes sitting on either end of chain are attributed to EPs, but those residing on both ends are indeed from DPs. Besides, we found that the skin modes caused by the non-Hermitian skin effect may possess higher-order EPs under certain parameter conditions.

We demonstrated that our analytical results can provide a direct description of topological zero modes, including their existence domain and localization properties, without requiring a full knowledge of the topological phase transitions in such non-Hermitian systems. This would facilitate greatly the experimental exploration of some topological issues restated to the edge states, for example, the BBC in non-Hermitian systems [12,42]. In view of the universality of the non-Hermitian SSH model [3,71], we expect that our results obtained with this model may find broad applications and open avenues for understanding the elusive non-Hermitian topology in areas ranging from condensed matter physics to photonics.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 11974075) and the Progetti di Ricerca di Interesse Nazionale (PRIN) (Project No. 2020X4T57A).

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

[2] K. Kawabata, K. Shiozaki, M. Ueda, and M. Sato, Symmetry and Topology in Non-Hermitian Physics, Phys. Rev. X **9**, 041015 (2019).

[3] Y. Ashida, Z. Gong, and M. Ueda, Non-Hermitian physics, Adv. Phys. **69**, 249 (2020).

[4] M. Parto, Y. G. N. Liu, B. Bahari, M. Khajavikhan, and D. N. Christodoulides, Non-Hermitian and topological photonics: Optics at an exceptional point, Nanophotonics **10**, 403 (2021).

[5] N. Moiseyev, *Non-Hermitian Quantum Mechanics* (Cambridge University Press, Cambridge, 2011).

[6] C. M. Bender and S. Boettcher, Real Spectra in Non-Hermitian Hamiltonians Having $\mathcal{PT}$ Symmetry, Phys. Rev. Lett. **80**, 5243 (1998).

[7] M. S. Rudner and L. S. Levitov, Topological Transition in a Non-Hermitian Quantum Walk, Phys. Rev. Lett. **102**, 065703 (2009).

[8] H. Cao and J. Wiersig, Dielectric microcavities: Model systems for wave chaos and non-Hermitian physics, Rev. Mod. Phys. **87**, 61 (2015).

[9] V. V. Konotop, J. Yang, and D. A. Zezyulin, Nonlinear waves in $\mathcal{PT}$-symmetric systems, Rev. Mod. Phys. **88**, 035002 (2016).

[10] L. Feng, R. El-Ganainy, and L. Ge, Non-Hermitian photonics based on parity-time symmetry, Nat. Photon. **11**, 752 (2017).

[11] S. Longhi, Parity-time symmetry meets photonics: A new twist in non-Hermitian optics, Europhys. Lett. **120**, 64001 (2017).

[12] L. Xiao, T. Deng, K. Wang, G. Zhu, Z. Wang, W. Yi, and P. Xue, Non-Hermitian bulk-boundary correspondence in quantum dynamics, Nat. Phys. **16**, 761 (2020).

[13] S. Bittner, B. Dietz, U. Günther, H. L. Harney, M. Miski-Oglu, A. Richter, and F. Schäfer, $\mathcal{PT}$ Symmetry and Spontaneous Symmetry Breaking in a Microwave Billiard, Phys. Rev. Lett. **108**, 024101 (2012).

[14] C. M. Bender, B. K. Berntson, D. Parker, and E. Samuel, Observation of $\mathcal{PT}$ phase transition in a simple mechanical system, Am. J. Phys. **81**, 173 (2013).

[15] C. H. Lee, S. Imhof, C. Berger, F. Bayer, J. Brehm, L. W. Molenkamp, T. Kiessling, and R. Thomale, Topolectrical circuits, Commun. Phys. **1**, 39 (2018).

[16] S. Liu, R. Shao, S. Ma, L. Zhang, O. You, H. Wu, Y. J. Xiang, T. J. Cui, and S. Zhang, Non-Hermitian skin effect in a non-Hermitian electrical circuit, Research **2021**, 5608038 (2021).

[17] X. Zhu, H. Ramezani, C. Shi, J. Zhu, and X. Zhang, $\mathcal{PT}$-Symmetric Acoustics, Phys. Rev. X **4**, 031042 (2014).

[18] T. Liu, J. J. He, T. Yoshida, Z.-L. Xiang, and F. Nori, Non-Hermitian topological Mott insulators in one-dimensional fermionic superlattices, Phys. Rev. B **102**, 235151 (2020).

[19] Z. Yang, A. P. Schnyder, J. Hu, and C.-K. Chiu, Fermion Doubling Theorems in Two-Dimensional Non-Hermitian Systems for Fermi Points and Exceptional Points, Phys. Rev. Lett. **126**, 086401 (2021).

[20] W. Zhang, X. Ouyang, X. Huang, X. Wang, H. Zhang, Y. Yu, X. Chang, Y. Liu, D.-L. Deng, and L.-M. Duan, Observation of Non-Hermitian Topology with Nonunitary Dynamics of Solid-State Spins, Phys. Rev. Lett. **127**, 090501 (2021).

[21] S. Weimann, M. Kremer, Y. Plotnik, Y. Lumer, S. Nolte, K. G. Makris, M. Segev, M. C. Rechtsman, and A. Szameit, Topologically protected bound states in photonic parity-time-symmetric crystals, Nat. Mater. **16**, 433 (2017).

[22] K. Takata and M. Notomi, Photonic Topological Insulating Phase Induced Solely by Gain and Loss, Phys. Rev. Lett. **121**, 213902 (2018).

[23] W. Song, W. Sun, C. Chen, Q. Song, S. Xiao, S. Zhu, and T. Li, Breakup and Recovery of Topological Zero Modes in Finite Non-Hermitian Optical Lattices, Phys. Rev. Lett. **123**, 165701 (2019).

[24] H. Wang, X. Zhang, J. Hua, D. Lei, M. Lu, and Y. Chen, Topological physics of non-Hermitian optics and photonics: A review, J. Opt. **23**, 123001 (2021).

[25] S. Yao and Z. Wang, Edge States and Topological Invariants of Non-Hermitian Systems, Phys. Rev. Lett. **121**, 086803 (2018).

[26] F. Song, S. Yao, and Z. Wang, Non-Hermitian Skin Effect and Chiral Damping in Open Quantum Systems, Phys. Rev. Lett. **123**, 170401 (2019).

[27] S. Longhi, Probing non-Hermitian skin effect and non-Bloch phase transitions, Phys. Rev. Res. **1**, 023013 (2019).

[28] N. Okuma, K. Kawabata, K. Shiozaki, and M. Sato, Topological Origin of Non-Hermitian Skin Effects, Phys. Rev. Lett. **124**, 086801 (2020).

[29] L. Li, C. H. Lee, S. Mu, and J. Gong, Critical non-Hermitian skin effect, Nat. Commun. **11**, 5491 (2020).

[30] S. Longhi, Unraveling the non-Hermitian skin effect in dissipative systems, Phys. Rev. B **102**, 201103(R) (2020).

[31] K. Zhang, Z. Yang, and C. Fang, Universal non-Hermitian skin effect in two and higher dimensions, Nat. Commun. **13**, 2496 (2022).

[32] B. Bahari, A. Ndao, F. Vallini, A. E. Amili, Y. Fainman, and B. Kanté, Nonreciprocal lasing in topological cavities of arbitrary geometries, Science **358**, 636 (2017).

[33] M. Li, X. Ni, M. Weiner, A. Alù, and A. B. Khanikaev, Topological phases and nonreciprocal edge states in non-Hermitian Floquet insulators, Phys. Rev. B **100**, 045423 (2019).

[34] X. Wu, L. Wang, S. Chen, X. Chen, and L. Yuan, Transition characteristics of non-Hermitian skin effects in a zigzag lattice without chiral symmetry, Adv. Phys. Res. **2**, 2300007 (2023).

[35] K. Zhang, Z. Yang, and C. Fang, Correspondence between Winding Numbers and Skin Modes in Non-Hermitian Systems, Phys. Rev. Lett. **125**, 126402 (2020).

[36] R. S. K. Mong and V. Shivamoggi, Edge states and the bulk-boundary correspondence in Dirac Hamiltonians, Phys. Rev. B **83**, 125109 (2011).

[37] K. Yatsugi, T. Yoshida, T. Mizoguchi, Y. Kuno, H. Iizuka, Y. Tadokoro, and Y. Hatsugai, Observation of bulk-edge correspondence in topological pumping based on a tunable electric circuit, Commun. Phys. **5**, 180 (2022).

[38] S. Chen, L. Bu, C. Pan, C. Hou, F. Baronio, P. Grelu, and N. Akhmediev, Modulation instability–rogue wave correspondence hidden in integrable systems, Commun. Phys. **5**, 297 (2022).

[39] Y. Hasegawa, Unifying speed limit, thermodynamic uncertainty relation and Heisenberg principle via bulk-boundary correspondence, Nat. Commun. **14**, 2828 (2023).

[40] T. E. Lee, Anomalous Edge State in a Non-Hermitian Lattice, Phys. Rev. Lett. **116**, 133903 (2016).

[41] Y. Xiong, Why does bulk boundary correspondence fail in some non-Hermitian topological models, J. Phys. Commun. **2**, 035043 (2018).

[42] T. Helbig, T. Hofmann, S. Imhof, M. Abdelghany, T. Kiessling, L. W. Molenkamp, C. H. Lee, A. Szameit, M. Greiter, and R. Thomale, Generalized bulk-boundary correspondence in non-Hermitian topolectrical circuits, Nat. Phys. **16**, 747 (2020).

[43] K. Sone, Y. Ashida, and T. Sagawa, Exceptional non-Hermitian topological edge mode and its application to active matter, Nat. Commun. **11**, 5745 (2020).

[44] Z. Gong, Y. Ashida, K. Kawabata, K. Takasan, S. Higashikawa, and M. Ueda, Topological Phases of Non-Hermitian Systems, Phys. Rev. X **8**, 031079 (2018).

[45] K. Yokomizo and S. Murakami, Non-Bloch Band Theory of Non-Hermitian Systems, Phys. Rev. Lett. **123**, 066404 (2019).

[46] C.-X. Guo, C.-H. Liu, X.-M. Zhao, Y. Liu, and S. Chen, Exact Solution of Non-Hermitian Systems with Generalized Boundary Conditions: Size-Dependent Boundary Effect and Fragility of the Skin Effect, Phys. Rev. Lett. **127**, 116801 (2021).

[47] F. Song, S. Yao, and Z. Wang, Non-Hermitian Topological Invariants in Real Space, Phys. Rev. Lett. **123**, 246801 (2019).

[48] C. Hou, L. Li, S. Chen, Y. Liu, L. Yuan, Y. Zhang, and Z. Ni, Deterministic bulk-boundary correspondences for skin and edge modes in a general two-band non-Hermitian system, Phys. Rev. Res. **4**, 043222 (2022).

[49] A. Alase, E. Cobanera, G. Ortiz, and L. Viola, Generalization of Bloch's theorem for arbitrary boundary conditions: Theory, Phys. Rev. B **96**, 195133 (2017).

[50] E. Cobanera, A. Alase, G. Ortiz, and L. Viola, Generalization of Bloch's theorem for arbitrary boundary conditions: Interfaces and topological surface band structure, Phys. Rev. B **98**, 245423 (2018).

[51] H. Shen, B. Zhen, and L. Fu, Topological Band Theory for Non-Hermitian Hamiltonians, Phys. Rev. Lett. **120**, 146402 (2018).

[52] S. Yao, F. Song, and Z. Wang, Non-Hermitian Chern Bands, Phys. Rev. Lett. **121**, 136802 (2018).

[53] W. D. Heiss, Exceptional points of non-Hermitian operators, J. Phys. A: Math. Gen. **37**, 2455 (2004).

[54] M.-A. Miri and A. Alù, Exceptional points in optics and photonics, Science **363**, eaar7709 (2019).

[55] Ş. K. Özdemir, S. Rotter, F. Nori, and L. Yang, Parity-time symmetry and exceptional points in photonics, Nat. Mater. **18**, 783 (2019).

[56] B. Zhen, C. W. Hsu, Y. Igarashi, L. Lu, I. Kaminer, A. Pick, S.-L. Chua, J. D. Joannopoulos, and M. Soljačić, Spawning rings of exceptional points out of Dirac cones, Nature (London) **525**, 354 (2015).

[57] E. M. Graefe, U. Günther, H. J. Korsch, and A. E. Niederle, A non-Hermitian $\mathcal{PT}$ symmetric Bose-Hubbard model: Eigenvalue rings from unfolding higher-order exceptional points, J. Phys. A: Math. Theor. **41**, 255206 (2008).

[58] J. Wiersig, Distance between exceptional points and diabolic points and its implication for the response strength of non-Hermitian systems, Phys. Rev. Res. **4**, 033179 (2022).

[59] B. Peng, Ş. K. Özdemir, S. Rotter, H. Yilmaz, M. Liertzer, F. Monifi, C. M. Bender, F. Nori, and L. Yang, Loss-induced suppression and revival of lasing, Science **346**, 328 (2014).

[60] B. Peng, Ş. K. Özdemir, M. Liertzer, W. Chen, J. Kramer, H. Yılmaz, J. Wiersig, S. Rotter, and L. Yang, Chiral modes and directional lasing at exceptional points, Proc. Natl. Acad. Sci. USA **113**, 6845 (2016).

[61] W. Chen, Ş. K. Özdemir, G. Zhao, J. Wiersig, and L. Yang, Exceptional points enhance sensing in an optical microcavity, Nature (London) **548**, 192 (2017).

[62] H. Zhou, C. Peng, Y. Yoon, C. W. Hsu, K. A. Nelson, L. Fu, J. D. Joannopoulos, M. Soljačić, and B. Zhen, Observation of bulk Fermi arc and polarization half charge from paired exceptional points, Science **359**, 1009 (2018).

[63] K. Yokomizo, *Non-Bloch Band Theory of Non-Hermitian Systems* (Springer, Singapore, 2022).

[64] A. Alase, *Boundary Physics and Bulk-Boundary Correspondence in Topological Phases of Matter* (Springer, Cham, 2019).

[65] W. P. Su, J. R. Schrieffer, and A. J. Heeger, Solitons in Polyacetylene, Phys. Rev. Lett. **42**, 1698 (1979).

[66] L. Li, Z. Xu, and S. Chen, Topological phases of generalized Su-Schrieffer-Heeger models, Phys. Rev. B **89**, 085111 (2014).

[67] M. Bello, G. Platero, J. I. Cirac, and A. González-Tudela, Unconventional quantum optics in topological waveguide QED, Sci. Adv. **5**, eaaw0297 (2019).

[68] C. Poli, M. Bellec, U. Kuhl, F. Mortessagne, and H. Schomerus, Selective enhancement of topologically induced interface states in a dielectric resonator chain, Nat. Commun. **6**, 6710 (2015).

[69] M. Parto, S. Wittek, H. Hodaei, G. Harari, M. A. Bandres, J. Ren, M. C. Rechtsman, M. Segev, D. N. Christodoulides, and M. Khajavikhan, Edge-Mode Lasing in 1D Topological Active Arrays, Phys. Rev. Lett. **120**, 113901 (2018).

[70] S. Longhi, Non-Hermitian gauged topological laser arrays, Ann. Phys. (Berlin) **530**, 1800023 (2018).

[71] D.-W. Zhang, Y.-Q. Zhu, Y. X. Zhao, H. Yan, and S.-L. Zhu, Topological quantum matter with cold atoms, Adv. Phys. **67**, 253 (2018).

[72] Y. Yi and Z. Yang, Non-Hermitian Skin Modes Induced by On-Site Dissipations and Chiral Tunneling Effect, Phys. Rev. Lett. **125**, 186802 (2020).

[73] Z. Ren, D. Liu, E. Zhao, C. He, K. K. Pak, J. Li, and G.-B. Jo, Chiral control of quantum states in non-Hermitian spin-orbit-coupled fermions, Nat. Phys. **18**, 385 (2022).

[74] K. Y. Bliokh, F. J. Rodríguez-Fortuño, F. Nori, and A. V. Zayats, Spin-orbit interactions of light, Nat. Photon. **9**, 796 (2015).

[75] J. K. Asbóth, L. Oroszlány, and A. Pályi, *A Short Course on Topological Insulators* (Springer, Cham, 2016).

[76] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory* (Springer, New York, 1999), Chap. 7.