

Amoeba Formulation of Non-Bloch Band Theory in Arbitrary Dimensions

Hong-Yi Wang¹, Fei Song¹, and Zhong Wang^{1*}

Institute for Advanced Study, Tsinghua University, Beijing 100084, China



(Received 16 July 2023; accepted 5 February 2024; published 16 April 2024)

The non-Hermitian skin effect dramatically reshapes the energy bands of non-Hermitian systems, meaning that the usual Bloch band theory is fundamentally inadequate as their characterization. The non-Bloch band theory, in which the concept of Brillouin zone is generalized, has been widely applied to investigate non-Hermitian systems in one spatial dimension. However, its generalization to higher dimensions has been challenging. Here, we develop a formulation of the non-Hermitian skin effect and non-Bloch band theory in arbitrary spatial dimensions, which is based on a natural geometrical object known as the amoeba. Our theory provides a general framework for studying non-Hermitian bands beyond one dimension. Key quantities of non-Hermitian bands, including the energy spectrum, eigenstates profiles, and the generalized Brillouin zone, can be efficiently obtained from this approach.

DOI: [10.1103/PhysRevX.14.021011](https://doi.org/10.1103/PhysRevX.14.021011)

Subject Areas: Condensed Matter Physics,
Metamaterials, Quantum Physics

I. INTRODUCTION

Non-Hermitian Hamiltonians have wide applications in many branches of physics, ranging from classical wave phenomena to open quantum systems [1,2]. Among various non-Hermitian systems, those that have spatially periodic structures are especially important and extensively studied. Recently, their interplay with topological physics has stimulated the fruitful investigations of non-Hermitian topological states [2,3]. One of the key phenomena uncovered in this direction is the non-Hermitian skin effect (NHSE), which refers to the counterintuitive feature that the nominal “bulk eigenstates” of a non-Hermitian Hamiltonian are exponentially localized at the boundary [4–11]. Unlike the topological boundary modes whose number scales linearly with the boundary area, the number of skin modes is proportional to the system volume. Among other consequences, the NHSE implies that the energy spectra of a non-Hermitian system can be drastically different under the experimentally favored open boundary condition (OBC) and the theoretically efficient periodic boundary condition (PBC). In sharp contrast to the Anderson localization, the NHSE-induced exponential localization can occur in pristine non-Hermitian systems without any disorder.

As a pronounced deviation from the usual Bloch-wave picture, the NHSE implies that the standard Bloch band theory is insufficient to characterize a generic non-Hermitian band. For example, the OBC energy spectrum cannot be calculated within this framework. To address this serious issue, a non-Hermitian generalization of the standard band theory, known as the non-Bloch band theory, has been introduced and applied to various non-Hermitian systems [4,12–19]. A central concept in this theory is the generalized Brillouin zone (GBZ), which is the proper surrogate for the conventional Brillouin zone (BZ) in Hermitian bands. The GBZ allows efficient computation of the continuous OBC energy spectrum of a large system without the need of diagonalizing a large Hamiltonian matrix in real space. Meanwhile, the shape of the GBZ directly tells the information of real-space eigenstates, e.g., the skin localization lengths. Moreover, the topological invariants on the GBZ are able to predict the number of topological boundary modes of non-Hermitian systems, whereas the usual BZ-based topological numbers fail to do so.

So far, the GBZ has been well defined only in one dimension (1D). Finding a general definition and formula for GBZ in two and higher dimensions is challenging, because the well-known 1D approach is not amenable to a straightforward generalization [4,12,20]. In certain special cases, the two-dimensional (2D) GBZ has been approximately defined and calculated [21,22]. However, a general approach without resorting to uncontrolled approximations has been lacking.

In this paper, we present a general formulation of non-Hermitian band theory in arbitrary dimensions. Among other results, it tells how the GBZ and related spectral

*wangzhongemail@tsinghua.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

properties are quantitatively determined beyond 1D. In 1D, our formulation reduces to the well-known GBZ formulation. The formulation is universal in the sense that our main results are applicable to models in arbitrary spatial dimensions, and with arbitrary degrees of freedom in a unit cell.

Our theory is based on a natural geometrical object called the amoeba by mathematicians [23–27]. Inspired by the amoeba and related mathematical tools, we formulate a theory which characterizes the NHSE quantitatively in arbitrary spatial dimensions. In particular, it is possible to directly calculate the energy spectrum and density of states (DOS) in the thermodynamic (i.e., large-size) limit without the troublesome finite-size errors. We show in a theorem that the energy spectrum can be obtained from the shape of the amoeba. We also demonstrate, despite the geometry-dependent NHSE, the existence of a universal spectrum (amoebic spectrum) to which the OBC spectrum under any generic geometry converges. Furthermore, the amoeba inspires a definition and the associated algorithm of the GBZ in arbitrary spatial dimensions. Among other applications, this amoeba-based GBZ provides a starting point for calculating non-Bloch band topology beyond 1D.

The remainder of this article is arranged as follows. In Sec. II, we go through the existing method of determining the GBZ in 1D, and then try to find clues for its higher-dimensional generalization. In Sec. III, we introduce the basic mathematical properties of the amoeba and the associated Ronkin function, which will be useful in calculating the DOS and the GBZ. In Secs. IV–VI, we introduce the amoeba formulation for non-Hermitian systems, and then make use of this formulation and the theory of Toeplitz matrices to establish a universal way to determine the DOS as well as the GBZ. Non-Bloch band topology based on the proposed GBZ theory is studied in Sec. VII. Finally, in Sec. VIII, several useful inequalities on the OBC and PBC spectra are proved from the amoeba approach.

II. MOTIVATION

A. Review of 1D non-Bloch band theory

We start with reviewing the concept of the GBZ in 1D [4, 12, 28, 29], searching for clues of its higher-dimensional generalizations.

A general 1D tight-binding Hamiltonian with OBC can be written as

$$H = \sum_{i,j=1}^L \sum_{a,b} |i, a\rangle (t_{j-i})_{ab} \langle j, b|, \quad (1)$$

where i, j are the position indices, and a, b are the indices for intracell degrees of freedom (band indices). Explicitly, the matrix H looks like

$$H = \begin{pmatrix} t_0 & t_1 & t_2 & \cdots & 0 \\ t_{-1} & t_0 & t_1 & t_2 & \\ t_{-2} & t_{-1} & t_0 & t_1 & \ddots & \vdots \\ & t_{-2} & t_{-1} & t_0 & \ddots & \\ \vdots & & \ddots & \ddots & \ddots & t_2 \\ & & & & t_0 & t_1 \\ 0 & \cdots & & t_{-2} & t_{-1} & t_0 \end{pmatrix}, \quad (2)$$

in which each t_n stands for a square matrix $(t_n)_{ab}$. As the hopping matrix t_{j-i} depends only on the spatial distance $j-i$, we have translational symmetry in the bulk. The Hermiticity condition $t_n = t_{-n}^\dagger$ is not required since non-Hermitian Hamiltonians are our focus. A finite hopping range n_c is assumed so that the hopping matrix $t_n = 0$ when $|n| > n_c$. Given this real-space Hamiltonian, the corresponding Bloch Hamiltonian is the Fourier transform of t_n :

$$h(e^{ik}) = \sum_n t_n (e^{ik})^n. \quad (3)$$

Note that the Bloch Hamiltonian has been written as $h(e^{ik})$ instead of the more frequently used $h(k)$. This simplifies our notations when the real-valued wave vector k is generalized to the complex plane, which amounts to making the substitution $e^{ik} \rightarrow e^{\mu+ik} = \beta$ (k and μ are real-valued). Evidently, $h(\beta) = \sum_n t_n \beta^n$ is a matrix-valued Laurent polynomial of β .

To solve the real-space Schrödinger equation $H|\psi\rangle = E|\psi\rangle$, we first note that the equations have identical form at all spatial coordinates j (except some points near the boundary):

$$E\psi_j = \sum_n t_n \psi_{j+n}. \quad (4)$$

Thanks to the translational invariance in the bulk, the Schrödinger equation is a linear recurrence equation with constant coefficients. The standard approach is to take a trial solution $\psi_j = v\beta^j$ (v is a vector whose dimension is the number of bands), and the bulk equation gives

$$Ev\beta^j = \sum_n t_n v\beta^{j+n}, \quad (5)$$

$$\Leftrightarrow [E - h(\beta)]v = 0. \quad (6)$$

If this trial solution can appear as a component of an eigenstate wave function, it is necessary that

$$\det[E - h(\beta)] = 0, \quad (7)$$

which is called the characteristic equation for this problem. The most general form of a wave function satisfying the bulk equations is then a linear superposition of functions $v_n(\beta_n)^j$, where β_n is a solution of $\det[E - h(\beta)] = 0$, and v_n is the corresponding eigenvector. Thus, an eigenstate of energy E is expressed as

$$|\psi\rangle = \sum_{j=1}^L \sum_a \sum_{n=1}^{M+N} c_n(v_n)_a(\beta_n)^j |j, a\rangle, \quad (8)$$

where the coefficients c_n are to be determined by the boundary conditions. We expand the characteristic polynomial as $\det[E - h(\beta)] = a_{-M}(E)\beta^{-M} + \dots + a_N(E)\beta^N$ with $a_{-M}(E)$, $a_N(E)$ nonzero, and sort its roots as $|\beta_1(E)| \leq |\beta_2(E)| \leq \dots \leq |\beta_{M+N}(E)|$. It has been found that, in the thermodynamic (large-size) limit, the OBC boundary condition results in the following simple equation known as the GBZ equation [4,12,29]:

$$|\beta_M(E)| = |\beta_{M+1}(E)|. \quad (9)$$

It turns out that all the β_M, β_{M+1} solutions form a closed curve called the GBZ on the complex plane of β , which contains key information about the eigenstate profiles, including the conventional wave vector k and the spatial decay rate μ of a skin mode. Once the GBZ is obtained, one can insert $\beta \in \text{GBZ}$ into Eq. (7) to obtain the OBC energy spectrum. Furthermore, the topological boundary modes are dictated by the topological invariants defined in the GBZ rather than in the BZ. This phenomenon is known as non-Bloch bulk-boundary correspondence [4,12]. Thus, in many senses, GBZ plays a similar role as BZ does in Hermitian systems. Band theory based on the GBZ concept is known as the non-Bloch band theory.

The condition $|\beta_M(E)| = |\beta_{M+1}(E)|$ lies at the heart of the 1D non-Bloch band theory. It can be intuitively justified with the simplest example, in which there are only one band and two β 's, β_1 and β_2 , with $|\beta_1| \leq |\beta_2|$. Then, Eq. (8) reads $|\psi\rangle = \sum_{j=1}^L \psi_j |j\rangle$ with $\psi_j = c_1 \beta_1^j + c_2 \beta_2^j$. The open boundary condition amounts to adding fictitious sites $j = 0$ and $j = L + 1$ at the left and right ends, with $\psi_0 = 0$ and $\psi_{L+1} = 0$, respectively. At the left end, we find $\psi_0 = 0 \Rightarrow c_1 = -c_2$. Then, at the right end, it is possible for the two terms to cancel each other only if $|\beta_1| = |\beta_2|$; otherwise, the β_2^{L+1} term would be much larger than the β_1^{L+1} term. In the general case, a rigorous treatment is to write down the system of linear equations for the boundary conditions, and it turns out that the coefficient matrix is an $(M + N) \times (M + N)$ square matrix. Letting its determinant vanish results in $|\beta_M(E)| = |\beta_{M+1}(E)|$ [12,29].

Unfortunately, we will see that in higher dimensions, the above treatment relying on the small rank of the coefficient matrix will become quite intractable. This has been a major obstacle in the attempt to study GBZ in higher dimensions.

B. The way to the amoeba

To demonstrate the applications of non-Bloch band theory, we consider the non-Hermitian Su-Schrieffer-Heeger (SSH) model with the Bloch Hamiltonian [4,5]:

$$h(e^{ik}) = [t_1 + (t_2 + t_3) \cos k] \sigma_x + \left[(t_2 - t_3) \sin k + i \frac{\gamma}{2} \right] \sigma_y, \quad (10)$$

or, in terms of β ,

$$h(\beta) = \left[t_1 + \frac{t_2 + t_3}{2} \left(\beta + \frac{1}{\beta} \right) \right] \sigma_x + \left[\frac{t_2 - t_3}{2i} \left(\beta - \frac{1}{\beta} \right) + i \frac{\gamma}{2} \right] \sigma_y, \quad (11)$$

where $\sigma_{x,y,z}$ are the Pauli matrices. It is known that the eigenstates exhibit NHSE under OBC, and consequently, the OBC and PBC energy spectra are drastically different [4,5]. As an illustration, the OBC and PBC spectra are plotted in Fig. 1(a), for parameter values $t_1 = t_2 = 1$, $t_3 = 0.7$, and $\gamma = 4/3$. This choice of parameters is in the topologically nontrivial regime, and therefore topological edge modes with $E = 0$ are found in the OBC energy spectrum. The characteristic equation Eq. (7) is a quartic equation of β , and its four solutions are shown in Figs. 1(b) and 1(c) for $E = E_1 = -1 + 0.3i$ and $E = E_2 = -1$, respectively. Instead of β itself, we show $\mu = \log |\beta|$, namely, the imaginary part of the wave vector. As stated in Eq. (9), when and only when $|\beta_2(E)| = |\beta_3(E)|$, i.e., $\mu_2(E) = \mu_3(E)$, will E belong to the OBC energy spectrum. This is the case for $E = E_2$ [Fig. 1(c)]. The corresponding $\beta_2(E_2)$ and $\beta_3(E_2)$ belong to the GBZ.

Remarkably, the GBZ equation, Eq. (9), enables us to find the OBC energy spectra and other physical quantities without the need of diagonalizing a large real-space Hamiltonian whose size grows with the system size. The thermodynamic-limit quantities are obtained directly from the GBZ equation. A natural question arises: What is the higher-dimensional counterpart of Eq. (9)?

To be concrete, we consider a single-band model shown in Fig. 2(a). With the notation $\boldsymbol{\beta} = (\beta_x, \beta_y)$, the corresponding Bloch Hamiltonian is

$$h(\boldsymbol{\beta}) = t(\beta_x + \beta_x^{-1} + \beta_y + \beta_y^{-1}) + t'(\beta_x + \beta_x^{-1})(\beta_y + \beta_y^{-1}) + \gamma(\beta_x - \beta_x^{-1} + \beta_y - \beta_y^{-1}), \quad (12)$$

in which the γ terms generate nonreciprocal hoppings [see Fig. 2(a)]. For concreteness, we fix parameters $t = 1$, $t' = 0.5$, and $\gamma = 0.2$. The energy spectra from brute-force numerical diagonalization are shown in Fig. 1(d).

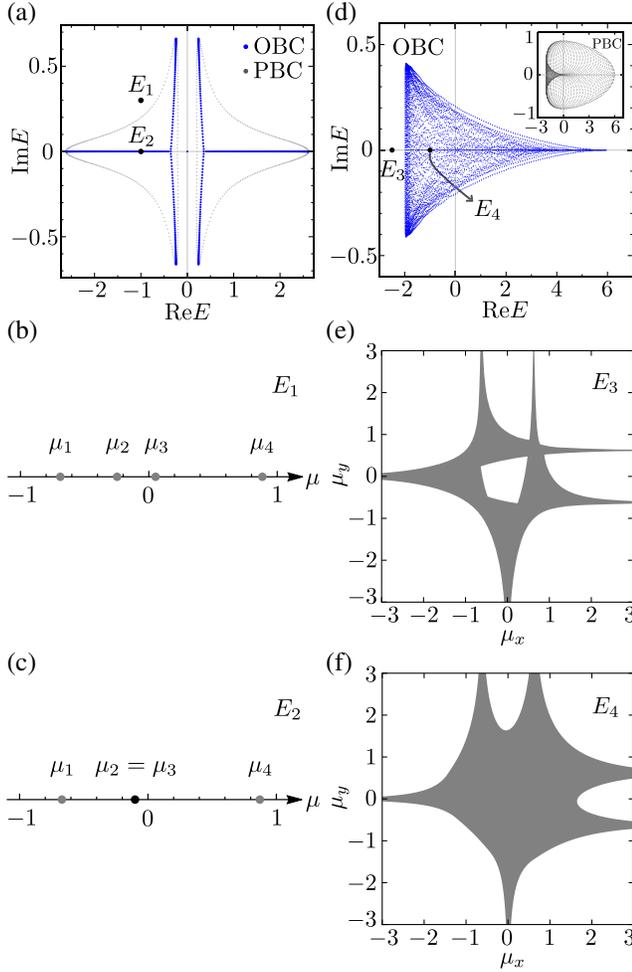


FIG. 1. Energy spectra and amoebae. (a) Energy spectrum of the 1D non-Hermitian SSH model Eq. (11) for a chain with length $L = 300$, under OBC (blue) and PBC (gray), respectively. Parameter values are $t_1 = t_2 = 1$, $t_3 = 0.7$, and $\gamma = 4/3$. (b),(c) Illustrations of $\mu = \log |\beta|$, where β satisfies the 1D characteristic equation $\det[E - h(\beta)] = 0$. E is taken to be $E_1 = -1 + 0.3i$ in (b), and $E_2 = -1$ in (c). For E_2 belonging to the OBC spectrum, we have $\mu_2(E_2) = \mu_3(E_2)$. (d) Energy spectrum of the 2D model Eq. (12). Main figure: disk geometry (OBC) with diameter $L = 140$. Inset: torus geometry (PBC). Parameter values are $t = 1$, $t' = 0.5$, and $\gamma = 0.2$. (e),(f) Illustrations of the 2D amoebae whose points are $(\mu_x, \mu_y) = (\log |\beta_x|, \log |\beta_y|)$, where β satisfies the 2D characteristic equation $\det[E - h(\beta)] = 0$ of the model Eq. (12). $E_3 = -2.5$ in (e), and $E_4 = -1$ in (f). Notably, there is a hole in the amoeba for E_3 outside the OBC spectrum, and no hole for E_4 in the OBC spectrum.

We hope to generalize the non-Bloch band theory to 2D, such that the energy spectrum can be obtained from the GBZ instead of the real space. The characteristic equation takes the same form as in 1D (and the reasoning is also the same):

$$\det[E - h(\beta)] = 0. \quad (13)$$

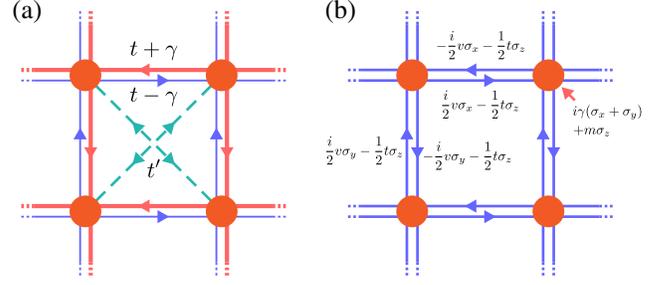


FIG. 2. Illustration of the real-space hopping of the 2D models used in this article. (a) Single-band model Eq. (12). (b) The non-Hermitian Chern-band model Eq. (45). All the parameters t, γ, v, m are real-valued.

For our single-band model, the determinant can be dropped, and therefore the characteristic equation is simply $E - h(\beta) = 0$. Notably, the zero locus of $\det[E - h(\beta)]$, namely, the β -solution space of $\det[E - h(\beta)] = 0$, is (real) two dimensional. In fact, the solution space can be locally parametrized by the complex-valued β_x or β_y , one of which is determined by the other via the characteristic equation. In contrast, the zero locus of $E - h(\beta)$ for a 1D system is zero dimensional, that is, several isolated β points.

The plausible next step, following the approach in the 1D non-Bloch band theory, is to select the legitimate β 's by adding the boundary conditions. For 1D systems with OBC, the number of constraint equations imposed by the boundary condition does not grow with the system size L , which greatly eases the derivation of the GBZ equation, Eq. (9) [4,12]. For 2D OBC systems (e.g., with square or disk geometry), however, the number of constraint equations is proportional to the linear size L . It is therefore challenging to exploit all these boundary-condition equations. Consequently, it is difficult to obtain a 2D counterpart of Eq. (9) from the approach similar to 1D.

Although a straightforward generalization to 2D looks infeasible, we can still find some clues from the 1D GBZ construction. Equation (9) and Figs. 1(b) and 1(c) suggest that the moduli of the solutions to the characteristic equation contain much useful information about the energy spectrum. Correspondingly, we plot the solutions to the 2D characteristic equation, followed by mapping $(\mu_x, \mu_y) = (\log |\beta_x|, \log |\beta_y|)$, in Figs. 1(e) and 1(f). This geometrical object is known as the amoeba in mathematics literature (see Sec. III for an introduction) [23–27]. We notice that a hole exists in the amoeba in Fig. 1(e), for which the energy E_3 does not belong to the OBC spectrum. In contrast, there is no hole in the amoeba in Fig. 1(f), with energy E_4 belonging to the OBC spectrum. Viewed from this amoeba perspective, 1D non-Hermitian energy spectra also exhibit similar behaviors. In 1D, the amoeba consists of discrete points, and the hole is simply the open interval between two adjacent points. For example, the open interval (μ_2, μ_3) in Fig. 1(b) can be viewed as a hole, which closes in Fig. 1(c) with $\mu_2 = \mu_3$.

From the above examples, we observe that the absence (presence) of a hole in the amoeba of the characteristic polynomial could be an indicator of the energy E being (not being) in the OBC energy spectrum. This is a key observation of the present work. To obtain more quantitative results from this observation, it is helpful to know some mathematical properties about the amoeba.

III. MATHEMATICAL PROPERTIES OF THE AMOEBA AND RONKIN FUNCTION

In this section, we introduce the basic concept of amoeba and a closely related analytic tool, the Ronkin function. As a quite recent concept in mathematics, the amoeba was introduced by Gelfand *et al.* in 1994 [23]. Albeit elementary, the notion of amoeba has deep connections with various concepts in algebraic geometry, which has stimulated extensive studies in mathematics [24–27].

Let f be a Laurent polynomial of β_j , $j = 1, 2, \dots, d$, where d will be identified as the spatial dimension in our study. The amoeba of f is defined as the log-moduli of the zero locus of f ,

$$\mathcal{A}_f = \{\log |\boldsymbol{\beta}| : f(\boldsymbol{\beta}) = 0\} \subset \mathbb{R}^d, \quad (14)$$

in which we use the notation $\log |\boldsymbol{\beta}| := (\log |\beta_1|, \dots, \log |\beta_d|)$ to simplify our expressions. Similar notations such as $e^\boldsymbol{\mu} := (e^{\mu_1}, \dots, e^{\mu_d})$ are used hereafter. In our case, the Laurent polynomial in use is $\det[E - h(\boldsymbol{\beta})]$. We can see that the geometric objects in Figs. 1(b) and 1(c) and 1(e) and 1(f) are 1D and 2D amoebae, respectively.

The name amoeba was motivated by its appearance in 2D: It has slim “tentacles” extending to infinity, and sometimes several “vacuoles” (holes) inside its body. Importantly, a particular hole plays an important role in our formulation. It is known that the amoeba in any spatial dimensions is a closed set, and each hole is a convex set [25].

A useful analytic tool in the study of amoeba is the Ronkin function, which is defined as [24,30,31]

$$R_f(\boldsymbol{\mu}) = \int_{T^d} \left(\frac{d\theta}{2\pi}\right)^d \log |f(e^{\boldsymbol{\mu}+i\theta})|, \quad (15)$$

where the domain of integration is the d -dimensional torus $T^d = [0, 2\pi]^d$, and the expression is simplified by the notations $f(e^{\boldsymbol{\mu}+i\theta}) := f(e^{\mu_1+i\theta_1}, \dots, e^{\mu_d+i\theta_d})$, and $(d\theta/2\pi)^d := (d\theta_1/2\pi) \dots (d\theta_d/2\pi)$.

It is beneficial to study the gradient of the Ronkin function [25,30]. To this end, we can express the integrand in R_f as $\log |f| = \text{Re} \log f$. The real part can be taken at the end of the calculation. It turns out that the integral is real-valued before taking the real part, and therefore the “Re” symbol can be discarded. The derivation proceeds as

$$\begin{aligned} \nu_j &= \frac{\partial R_f(\boldsymbol{\mu})}{\partial \mu_j} \\ &= \text{Re} \int_{T^d} \left(\frac{d\theta}{2\pi}\right)^d \partial_{\mu_j} \log f(e^{\boldsymbol{\mu}+i\theta}) \\ &= \text{Re} \int_{T^d} \left(\frac{d\theta}{2\pi}\right)^d \frac{\partial_{\mu_j} f(e^{\boldsymbol{\mu}+i\theta})}{f(e^{\boldsymbol{\mu}+i\theta})} \\ &= \text{Re} \int_{T^d} \left(\frac{d\theta}{2\pi}\right)^d \frac{-i \partial_{\theta_j} f(e^{\boldsymbol{\mu}+i\theta})}{f(e^{\boldsymbol{\mu}+i\theta})}. \end{aligned} \quad (16)$$

In the last line, we use $\partial_{\mu_j} = -i \partial_{\theta_j}$ acting on f . We observe that

$$w_j = \frac{1}{2\pi i} \int_0^{2\pi} d\theta_j \frac{\partial_{\theta_j} f(e^{\boldsymbol{\mu}+i\theta})}{f(e^{\boldsymbol{\mu}+i\theta})} \quad (17)$$

is the winding number of the phase of f along a circle parametrized by θ_j , and therefore it is always real-valued. Thus, the gradient ν_j is the average of the winding number w_j on the $(d-1)$ -dimensional torus parametrized by $(\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d)$:

$$\nu_j = \int_{T^{d-1}} \frac{d\theta_1 \dots d\theta_{j-1} d\theta_{j+1} \dots d\theta_d}{(2\pi)^{d-1}} w_j. \quad (18)$$

For example, in 2D, one has $\nu_1 = \int_0^{2\pi} (d\theta_2/2\pi) w_1$ and $\nu_2 = \int_0^{2\pi} (d\theta_1/2\pi) w_2$.

The next fact we see is that the Ronkin function is exactly linear on each connected component of the complement of the amoeba. We call each of these components (either bounded or unbounded) a hole of the amoeba. In fact, when $\boldsymbol{\mu}$ is not in the amoeba, $f(e^{\boldsymbol{\mu}+i\theta}) \neq 0$ is satisfied in the entire T^d parametrized by $(\theta_1, \dots, \theta_d)$. It follows that w_j is a constant integer as $\{\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d\}$ vary, and therefore the average ν_j is the same integer. Thus, we can assign to each amoeba hole an integer tuple $\boldsymbol{\nu} = (\nu_1, \dots, \nu_d)$ dubbed the order of the hole. The orders of two different holes cannot be the same [25]. Crucially, there exists at most one hole with order $\boldsymbol{\nu} = (0, \dots, 0)$, which we call the central hole. Because the order is zero, the Ronkin function is a constant in this hole. Moreover, the Ronkin function is convex in the entire $\boldsymbol{\mu}$ space; i.e., $R_f(\lambda \boldsymbol{\mu}_1 + (1-\lambda) \boldsymbol{\mu}_2) \leq \lambda R_f(\boldsymbol{\mu}_1) + (1-\lambda) R_f(\boldsymbol{\mu}_2)$ is satisfied for any two points $\boldsymbol{\mu}_{1,2}$ and $0 < \lambda < 1$ [24,30]. An easy corollary of the convexity is that a Ronkin function converges everywhere: If it were $-\infty$ at one point, convexity would imply that it is $-\infty$ everywhere.

In 1D, the Ronkin function is closely related to Jensen’s formula in complex analysis [32], which reads

$$\frac{1}{2\pi} \int_0^{2\pi} d\theta \log |g(Re^{i\theta})| = \log |g(0)| + \sum_{k=1}^l \log \left| \frac{R}{z_k} \right|, \quad (19)$$

where g is a holomorphic function with $g(0) \neq 0$, and z_k ($k = 1, \dots, l$) are the zeros of g enclosed by the circle $|\beta| = R$. Jensen's formula can be readily obtained from Eq. (16) with $d = 1$, in which case the gradient $\nu = w$ (the index $j = 1$ is redundant). In fact, the left-hand side of Eq. (19) is exactly the Ronkin function $R_g(\log |R|)$. To calculate it, we order the zeros of g as $|z_1| \leq |z_2| \leq |z_3| \leq \dots$. According to Eqs. (16) and (17), the gradient $\nu = [\partial R_g / \partial (\log |R|)]$ equals the winding number of g along the circle $|\beta| = R$, which counts the number of enclosed zeros. Therefore, we have $\nu = 0$ for $\log |R| < \log |z_1|$, and $\nu = k$ for $\log |z_k| < \log |R| < \log |z_{k+1}|$. It follows that, when the circle $|\beta| = R$ encloses l zeros,

$$\begin{aligned} R_g(\log |R|) &= \log |g(0)| + \log \frac{|z_2|}{|z_1|} + 2 \log \frac{|z_3|}{|z_2|} + 3 \log \frac{|z_4|}{|z_3|} \\ &\quad + \dots + l \log \frac{|R|}{|z_l|} \\ &= \log |g(0)| + \sum_{k=1}^l \log \left| \frac{R}{z_k} \right|, \end{aligned} \quad (20)$$

which is exactly Jensen's formula Eq. (19).

We now apply the explicit formula of the Ronkin function to $\det[E - h(\beta)] = a_{-M}(E)\beta^{-M} + \dots + a_N(E)\beta^N$, whose roots are $|\beta_1(E)| \leq |\beta_2(E)| \leq \dots \leq |\beta_{M+N}(E)|$. We rewrite it as $\det[E - h(\beta)] = f(\beta)g(\beta)$ with $f(\beta) = a_{-M}(E)\beta^{-M}$, so that $g(\beta)$ has no pole in the complex plane and $g(0) = 1$. Applying Eq. (20) to g , we have

$$\begin{aligned} R_{\det(E-h)}(\mu) &= R_f(\mu) + R_g(\mu) \\ &= \log |a_{-M}| - M\mu + \sum_{k=1}^l (\mu - \log |\beta_k|) \end{aligned} \quad (21)$$

for $\log |\beta_l| \leq \mu \leq \log |\beta_{l+1}|$. Thus, $R_{\det(E-h)}$ is a piecewise linear function of μ in 1D. To relate to the concept of amoeba, we note that each $\log |\beta_k|$ is a component of the amoeba, and each interval $(|\beta_k|, |\beta_{k+1}|)$ is an amoeba hole. Particularly, the open interval $(|\beta_M|, |\beta_{M+1}|)$ is exactly the central hole in which the Ronkin function is flat, because taking a derivative on Eq. (21) we find $\partial R_{\det(E-h)} / \partial \mu = -M + M = 0$. Therefore, Eq. (9) means that an energy E belongs to the OBC bulk spectrum when the central hole shrinks to zero size. Thus, the Ronkin function provides crucial information about the energy spectra. Figures 3(a) and 3(b) are two examples of the 1D Ronkin function. The Ronkin function is flat in the central hole of Fig. 3(a) corresponding to E outside the energy spectrum, and the hole shrinks to a point in Fig. 3(b) corresponding to E in the energy spectrum.

In 2D and higher dimensions, it is challenging to obtain a closed form for the Ronkin function. Nevertheless, regardless of the spatial dimensions, the Ronkin function is always globally convex, and is linear in the amoeba holes,

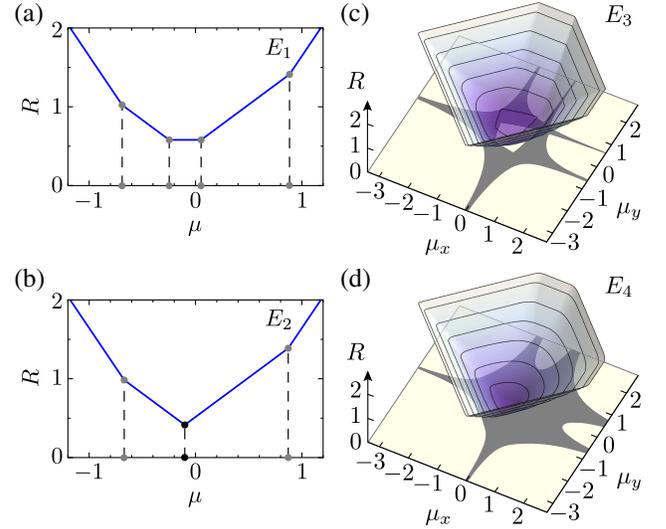


FIG. 3. The Ronkin function in (a),(b) 1D and (c),(d) 2D, taken at energies (a) E_1 , (b) E_2 , (c) E_3 , and (d) E_4 stated in Fig. 1. The Bloch Hamiltonian is Eq. (11) for (a),(b); it is Eq. (12) for (c),(d). The parameter values are the same as stated in Fig. 1. In both 1D and 2D, the Ronkin function is strictly linear on each component (hole) of the complement of the amoeba, where the gradient equals the integer index. The Ronkin function is always convex. Consequently, when the central hole exists, the minimum is reached on the central hole; otherwise, the minimum is reached at a single point in the amoeba.

i.e., connected components of the complement of the amoeba. If there is a central hole, the Ronkin function takes minimum in the entire hole; namely, the function has a flat bottom in this hole [Fig. 3(c)], otherwise the Ronkin minimum is reached at a single point in the amoeba [Fig. 3(d)].

In view of the aforementioned relation between the GBZ and Ronkin function in 1D, it is natural to ask whether there is a deep connection between the non-Hermitian energy spectra and the Ronkin function in higher dimensions. In fact, we have already seen some numerical clues for such a connection. Our observation about 2D non-Hermitian systems at the end of Sec. II can be rephrased in terms of the Ronkin function: If and only if the minimum of the Ronkin function $R_{\det(E-h)}$ is reached at a single point $\mu_{\min}(E)$, instead of in a hole with nonzero size, will E be in the OBC bulk spectrum.

In 1D, the single point where the Ronkin function takes the minimum is $\mu_{\min}(E) = \log |\beta_M(E)| = \log |\beta_{M+1}(E)|$. Note that $|\beta_M(E)| = |\beta_{M+1}(E)|$ is the decay factor of an OBC eigenstate. Thus, the location of the minimum of the Ronkin function precisely determines the decay factor of an OBC eigenstate. In other words, the Ronkin function tells the shape of the GBZ. We propose that this reformulation of GBZ in terms of the Ronkin function is generalizable to non-Hermitian systems in higher dimensions, which is justified in the following sections.

IV. ENERGY SPECTRA AND DENSITY OF STATES

We now introduce the amoeba formulation of non-Hermitian energy band theory in d spatial dimensions. One of our main objectives is to calculate the DOS associated with the energy spectrum, which is defined as the number of states per area on the complex energy plane, divided by the volume of the system, in the thermodynamic limit. Because the DOS has an $O(L^d)$ volume denominator, only the bulk states are relevant in the thermodynamic limit. All contributions from edge states, bound states, etc., vanish in the thermodynamic limit. For example, the number of possible surface states grows with the size as $O(L^{d-1})$, which contributes $O(1/L)$ that vanishes in the thermodynamic limit. Thus, we focus on the bulk spectrum.

A. Statement of the proposal

To describe the DOS of complex energy, it is convenient to use the language of electrostatics. Let us assign electric charge $-1/N$ to each energy eigenvalue ϵ_n , where $N \sim L^d$ is the total number of unit cells (in this convention, the total charge sums up to the number of energy bands, which is independent of the size N). In terms of Dirac's δ function, the DOS of complex energy can be written as $\rho(E) = \lim_{N \rightarrow \infty} \sum_n (1/N) \delta(E - \epsilon_n)$, which is just the absolute value of electric charge density.

Given the electric charge, the corresponding Coulomb potential $\Phi(E)$ is given by

$$\Phi(E) = \frac{1}{N} \sum_{\epsilon_n} \log |E - \epsilon_n| \xrightarrow{N \rightarrow \infty} \int d^2 E' \log |E - E'| \rho(E'). \quad (22)$$

Conversely, the DOS, or the absolute value of the charge density in the electrostatics language, can be readily obtained from the Coulomb potential by taking the Laplacian on the complex energy plane:

$$\rho(E) = \frac{1}{2\pi} \Delta \Phi(E), \quad (23)$$

where $\Delta = [\partial^2 / \partial(\text{Re } E)^2] + [\partial^2 / \partial(\text{Im } E)^2]$, and the $N \rightarrow \infty$ limit is taken for $\Phi(E)$.

One of our main proposals is that the Coulomb potential $\Phi(E)$, in the $N \rightarrow \infty$ limit, can be obtained from the Ronkin function,

$$\Phi(E) = \phi(E), \quad (24)$$

where $\phi(E)$ is the minimum of the Ronkin function of $\det(E - h)$ by varying $\boldsymbol{\mu}$:

$$\phi(E) = \min_{\boldsymbol{\mu}} R_{\det(E-h)}(\boldsymbol{\mu}), \quad (25)$$

or $\phi(E) = R_{\det(E-h)}(\boldsymbol{\mu}_{\min})$, where $\boldsymbol{\mu}_{\min}$ minimizes the function. Therefore, the DOS can be directly obtained from the Ronkin function,

$$\rho(E) = \frac{1}{2\pi} \Delta \phi(E). \quad (26)$$

In the special cases that the spectrum is a 1D object consisting of lines or curves in the complex energy plane, it is more preferable to define the DOS as states per length rather than states per area. In fact, the states per area will diverge in these cases, while states per length is in general finite. This is the case for 1D non-Hermitian systems, and higher-dimensional systems with real spectrum. In a small neighborhood of a segment of the 1D spectrum, suppose that \mathbf{n} is a normal vector to the segment. From the electrostatic analogy and Eq. (26), it is easy to see that the DOS per length denoted by ρ_{1D} is given by

$$\rho_{1D} = \frac{1}{2\pi} \left| \left(\frac{\partial \phi}{\partial \mathbf{n}} \right)_+ - \left(\frac{\partial \phi}{\partial \mathbf{n}} \right)_- \right|, \quad (27)$$

where the two derivatives are taken on opposite sides of the curve segment.

B. Numerical evidence

Before diving into a thorough analytic approach, we provide numerical evidence for the Ronkin-function-based formula, Eq. (26). We take the model Eq. (12) [Fig. 2(a)] as an example. In Fig. 4, we numerically compare the DOS derived from the Ronkin function, via Eq. (26), and that from diagonalizing the OBC Hamiltonian. The Laplacian is implemented by discretizing the complex energy plane into grid points. The DOS obtained from the Ronkin function agrees well with that from real-space Hamiltonians, though there are some differences that can be naturally attributed to the finite-size nature of the real-space calculations. First, the spectra from real-space diagonalization in Figs. 4(b) and 4(c) are slightly narrower in the $\text{Im } E$ direction than the one derived from the Ronkin function in Fig. 4(a). Second, the eigenvalues from real-space diagonalization seem more likely to concentrate on the real axis. These differences can be explained by the non-Bloch PT symmetry, which features a unique size dependence of non-Hermitian spectra in two and higher dimensions [33]. In fact, our real-space Hamiltonian [see Eq. (12) and Fig. 2(a)] has real-valued hoppings and therefore commutes with the complex conjugation operator: $[H, \mathcal{K}] = 0$. Combined with the NHSE, it means that the model has non-Bloch PT symmetry, which implies real energy spectra when the size is small. As the size L grows to infinity, the proportion of real eigenenergies diminishes to zero [33]. Since the length L adopted in Figs. 4(b) and 4(c) is finite, the PT symmetry breaking is incomplete, leaving a nonzero proportion of real eigenenergies.

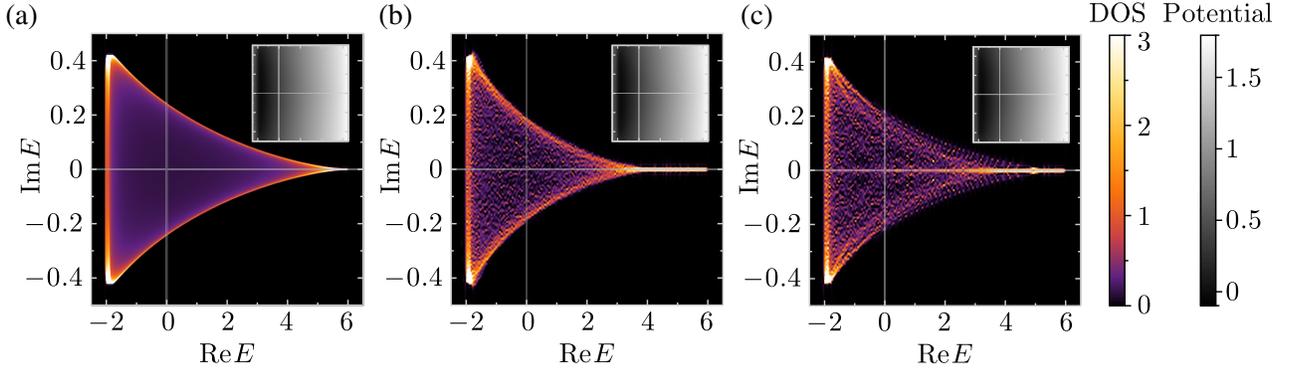


FIG. 4. DOS from the Ronkin function and diagonalization of the real-space Hamiltonian. The Hamiltonian used here is Eq. (12) [Fig. 2(a)], with parameter values $t = 1$, $t' = 0.5$, and $\gamma = 0.2$. (a) DOS from the Ronkin function via Eq. (26). (b) DOS from diagonalizing the real-space Hamiltonian on a square with side length $L = 130$. An on-site random potential distributed uniformly in $[-0.5, 0.5]$ is added at each boundary site. (c) DOS from diagonalizing the real-space Hamiltonian on a disk with diameter $L = 140$ (without random potential at the boundary). The insets show the Coulomb potential, $\phi(E)$ in (a), and $\Phi(E)$ in (b) and (c). To facilitate comparison with (a), the DOS in (b) and (c) is also obtained from the Coulomb potential via Eq. (23), in which $\Phi(E)$ is generated by diagonalizing the real-space Hamiltonian.

We emphasize that the Ronkin function tells the unique universal spectrum and universal DOS of the OBC system, the precise meaning of which we explain below. It is evident that the amoeba and the Ronkin function do not use information about the shape (e.g., square or disk) or boundary details (e.g., clean or locally perturbed) of the OBC system. The Ronkin function yields the same DOS regardless of geometrical details. Therefore, this approach intrinsically assumes a detail-independent (therefore, universal) spectrum with a universal DOS, which is supported by our numerical calculation. Indeed, our numerical results indicate that the DOS of an OBC system with an arbitrary shape converges in the large-size limit to the same universal DOS, at least when certain random local perturbations are added to the boundary [e.g., in Fig. 4(b)]. Note that the DOS thus obtained is independent of the specific forms of boundary randomness [34]. In generic cases, even the boundary randomness is not necessary to ensure the convergence to the universal DOS. For example, the DOS of disk geometry without boundary random potential already resembles the universal DOS [Fig. 4(c)]. In contrast, for a polygon geometry (e.g., a square), boundary randomness significantly helps the DOS to converge to the universal DOS. Without the boundary randomness, the polygon geometry can exhibit the geometry-dependent non-Hermitian skin effect, by which different polygons may have different DOS [35]. Our intuitive understanding of this phenomenon is as follows. The boundary of a polygon consists of straight line segments, which are perfectly reflective in the sense that the wave vector (momentum) parallel to the line segment is conserved during wave reflection. Thus, the boundary fails to fully mix waves with different wave vectors, and therefore the Hamiltonian can be viewed as fine-tuned rather than generic. As such, the spectrum exhibits the fingerprint of specific geometry rather than the universal spectral properties of the Hamiltonian.

Boundary or bulk randomness breaks the wave vector conservation and couples waves with different wave vectors, which generates a more generic energy spectrum. An analogous phenomenon is the critical non-Hermitian skin effect in 1D [36,37]. In the zero-coupling limit of two coupled 1D chains, the straightforward application of the GBZ equation, Eq. (9), does not yield the correct OBC energy spectrum [38]. The zero-coupling limit represents a fine-tuned point, and a small interchain coupling brings the spectrum to that predicted by the GBZ theory [36,37]. In 2D, our numerical results suggest that the Hamiltonian should be viewed as fine-tuned for polygon shapes, and a small randomness restores the universal spectrum characterized by the universal DOS.

To summarize, there exists a geometry-independent universal spectrum that can be calculated from the amoeba and Ronkin function. By nature, it can be called the “amoebic spectrum.” The DOS of an OBC system with a generic shape always approaches the universal DOS in the large-size limit. When the DOS of an OBC system with a certain (nongeneric) shape appears to deviate from the universal DOS, this deviation can be eliminated by adding a small random local perturbation. From an experimental point of view, the universal spectrum is particularly significant because disorders are often unavoidable in realistic systems.

C. Derivation

We establish the proposal Eq. (26) in a few steps. We begin with Szegő’s limit theorem for the determinant of a large Toeplitz matrix. Before doing so, it is appropriate here to introduce the terminology of Toeplitz matrices [39,40]. A matrix A determined by $A_{ij} = a_{j-i}$ is called a Toeplitz matrix; i.e., the matrix element A_{ij} depends on the difference $j - i$ only. It is associated with a symbol, which is a complex-valued function $\sigma(e^{i\theta}) = \sum_n a_n e^{in\theta}$, $\theta \in [0, 2\pi]$.

A Toeplitz matrix is often expressed in terms of its symbol as $A = \mathcal{T}[\sigma]$. By definition, the elements of a Toeplitz matrix are the Fourier components of the symbol:

$$\mathcal{T}[\sigma]_{j_1 j_2} = \int_0^{2\pi} \frac{d\theta}{2\pi} \sigma(e^{i\theta}) e^{i(j_1 - j_2)\theta}. \quad (28)$$

The language of Toeplitz matrices is very useful in addressing tight-binding Hamiltonians. For example, a single-band real-space OBC Hamiltonian H in 1D is a Toeplitz matrix, and the Bloch Hamiltonian is its symbol; conversely, we say H is generated by the Bloch Hamiltonian. In addition, one can generalize the series $\{a_n\}$ from numbers to square matrices. Such a matrix A is called a block Toeplitz matrix, which corresponds to a multiband Hamiltonian. One can also generalize the indices i, j, \dots from integers to integer tuples with several components; for example, in 2D we take $\mathbf{i} = (i_x, i_y)$. The corresponding matrix A is called a multilevel Toeplitz matrix, which can be viewed as the real-space Hamiltonian of a higher-dimensional lattice model. For simplicity, we call them all Toeplitz matrices hereafter.

Szegő's limit theorem was originally established for 1D Hermitian Toeplitz matrices [41], but thereafter generalized by Widom *et al.* to multiband [42,43] and higher-dimensional [44,45] models. To state the theorem, let us consider a subspace Ω of the d -dimensional Euclidean space. Let

$$\sigma(e^{i\theta}) = \sum_{\mathbf{n}} t_{\mathbf{n}} e^{i\mathbf{n}\cdot\theta} = \sum_{n_1, \dots, n_d} t_{n_1, \dots, n_d} e^{i(n_1\theta_1 + \dots + n_d\theta_d)}$$

be a symbol, which generates the Toeplitz matrix $\mathcal{T}[\sigma]$ in Ω . For example, if we take Ω to be the d -dimensional sphere with diameter L (radius $R = L/2$), then each block (or element, in the single-band cases) of $\mathcal{T}[\sigma]$ is $\mathcal{T}[\sigma]_{\mathbf{j}\mathbf{i}} = t_{\mathbf{j}-\mathbf{i}}$, with $\mathbf{i} = (i_1, \dots, i_d)$ and $\mathbf{j} = (j_1, \dots, j_d)$ satisfying $(i_1)^2 + \dots + (i_d)^2 \leq R^2$ and $(j_1)^2 + \dots + (j_d)^2 \leq R^2$. Szegő's limit theorem reveals the asymptotic behavior of the Toeplitz determinant in the $L \rightarrow \infty$ limit:

$$\log \det \mathcal{T}[\sigma] = N \int_{T^d} \left(\frac{d\theta}{2\pi} \right)^d \log \det \sigma(e^{i\theta}) + O(L^{d-1}), \quad (29)$$

where N is the number of lattice points in Ω , which is proportional to L^d (L stands for the linear size), and $T^d = [0, 2\pi]^d$ is the d -dimensional torus. There are two conditions for Eq. (29) to hold: (i) $\sigma(e^{i\theta})$ must be invertible for any θ on the torus T^d , i.e., $\det \sigma(e^{i\theta}) \neq 0$. (ii) The winding number of the phase of $\det \sigma(e^{i\theta})$ along any circle in T^d must be zero, so that a matrix logarithm for $\sigma(e^{i\theta})$ is well defined. A heuristic proof of Szegő's limit theorem is available in Appendix A. Notably, when $\sigma(e^{i\theta})$ and $\mathcal{T}[\sigma]$ are Hermitian, this theorem is consistent with the fact that

the OBC bulk spectrum is asymptotically the same as the PBC spectrum. In fact, the left-hand side of Eq. (29) yields the logarithm of the product of all OBC eigenvalues of $\mathcal{T}[\sigma]$, while the right-hand side yields the PBC counterpart, and these two quantities should be almost equal. Szegő's limit theorem tells us that a similar relation remains valid under the aforementioned conditions even though $\mathcal{T}[\sigma]$ is non-Hermitian.

Now we apply this theorem to our non-Hermitian problem. Specifically, we consider $\sigma(e^{ik}) = E - h(e^{ik})$, then the generated Toeplitz matrix is $\mathcal{T}[E - h(e^{ik})] = E - H$, in which $H = \mathcal{T}[h(e^{ik})]$ is the real-space Hamiltonian corresponding to the Bloch Hamiltonian $h(e^{ik})$. The motivation to consider $E - H$ is the simple identity

$$\Phi(E) = \frac{1}{N} \log |\det(E - H)|, \quad (30)$$

in which $\Phi(E)$ is the Coulomb potential defined in Eq. (22). We observe that this expression bears resemblance to the left-hand side of Eq. (29), which hints useful formulas for the Coulomb potential. Before exploiting Eq. (29), however, we notice that its application relies on the aforementioned two conditions. For example, it would be troublesome if $\det[E - h(e^{ik})] = 0$, and therefore the symbol is not invertible at certain \mathbf{k} points. To use Eq. (29), we perform a similarity transformation $D^{-1}HD$ to the Hamiltonian, with $D_{x,y} = \delta_{x,y} e^{\mu \cdot x}$. It adds an $e^{\mu \cdot (y-x)}$ factor to $H_{x,y}$, the hopping from \mathbf{y} to \mathbf{x} , i.e., $(D^{-1}HD)_{x,y} = H_{x,y} e^{\mu \cdot (y-x)}$. As such, the corresponding Bloch Hamiltonian $h(e^{ik})$ transforms into $h(e^{\mu+ik})$. In the language of Toeplitz matrices, we have $D^{-1}HD = \mathcal{T}[h(e^{\mu+ik})]$. In fact, given $h(e^{ik}) = \sum_{\mathbf{n}} t_{\mathbf{n}} e^{i\mathbf{k}\cdot\mathbf{n}}$, we have $h(e^{\mu+ik}) = \sum_{\mathbf{n}} t_{\mathbf{n}} e^{\mu \cdot \mathbf{n}} e^{i\mathbf{k}\cdot\mathbf{n}}$, from which we can read that $\mathcal{T}[h(e^{\mu+ik})]_{x,y} = t_{y-x} e^{\mu \cdot (y-x)} = H_{x,y} e^{\mu \cdot (y-x)}$. The transformation of $E - H$ can be written as

$$D^{-1} \mathcal{T}[E - h(e^{ik})] D = \mathcal{T}[E - h(e^{\mu+ik})], \quad (31)$$

or $D^{-1} \mathcal{T}[E - h(e^{ik})] D = \mathcal{T}[E - h(\boldsymbol{\beta})]$ with $\boldsymbol{\beta} = e^{\mu+ik}$. It follows from Eq. (31) that the matrix $\mathcal{T}[E - h(\boldsymbol{\beta})]$ has the same spectrum as $\mathcal{T}[E - h(e^{ik})]$ for an arbitrary value of $\boldsymbol{\mu}$. Therefore, we have $\det(E - H) = \det \mathcal{T}[E - h(\boldsymbol{\beta})]$ regardless of the value of $\boldsymbol{\mu}$. We can freely choose $\boldsymbol{\mu}$ such that $\sigma(\boldsymbol{\beta}) = E - h(\boldsymbol{\beta})$ satisfies the conditions required in applying Eq. (29), even if the original symbol $E - h(e^{ik})$ (with $\boldsymbol{\mu} = 0$) does not. In fact, when $\boldsymbol{\mu}$ locates in the central hole of the amoeba of $\det(E - h)$, we have $\det[E - h(\boldsymbol{\beta})] \neq 0$, and the winding number of $\det[E - h(\boldsymbol{\beta})]$ along any k_j circle is 0 (recall the mathematical preparation in Sec. III). Thus, we can apply Eq. (29) to $\sigma(\boldsymbol{\beta}) = E - h(\boldsymbol{\beta})$ and take the real part, resulting in

$$\begin{aligned} & \frac{1}{N} \log |\det \mathcal{T}[E - h(\boldsymbol{\beta})]| \\ &= \int_{T^d} \left(\frac{dk}{2\pi}\right)^d \log |\det[E - h(\boldsymbol{\beta})]| + O(L^{-1}), \end{aligned} \quad (32)$$

where $\boldsymbol{\beta} = e^{\boldsymbol{\mu} + i\mathbf{k}}$ with $\boldsymbol{\mu}$ fixed in the central hole of the amoeba of $E - h(\boldsymbol{\beta})$. As has been explained, the left-hand side of Eq. (32) is $(1/N) \log |\det(E - H)|$. Notably, the integral on the right-hand side is exactly the Ronkin function of $\det(E - h)$. Thus, Eq. (32) can be written as

$$\frac{1}{N} \log |\det(E - H)| = R_{\det(E-h)}(\boldsymbol{\mu}) + O(L^{-1}), \quad (33)$$

in which $\boldsymbol{\mu}$ locates in the central hole of the amoeba. Since the Ronkin function takes its minimum in the central hole, Eq. (33) reduces to Eq. (24) in the $N \rightarrow \infty$ limit.

So far, this proof of Eq. (24) is incomplete because the existence of the central hole of the amoeba is assumed. When the central hole does not exist, Szegő's limit theorem Eq. (29) cannot be applied in its original form. Here, we propose a generalization of Eq. (29) to remove this limitation. The generalization makes essential use of the Ronkin function. The proposed generalization is

$$\log \det \mathcal{T}[\sigma] = N \int_{T^d} \left(\frac{dk}{2\pi}\right)^d \log \det \sigma(\boldsymbol{\beta}) + O(L^{d-1}), \quad (34)$$

in which $\boldsymbol{\beta} = e^{\boldsymbol{\mu}_{\min} + i\mathbf{k}}$ with $\boldsymbol{\mu}_{\min}$ being the minimum location of the Ronkin function; i.e., $R_{\det \sigma}(\boldsymbol{\mu})$ takes its minimum at $\boldsymbol{\mu}_{\min}$. Note that the left-hand side of Eq. (34) can be taken as $\log \det \mathcal{T}[\sigma] = \log \det \mathcal{T}[\sigma(e^{\boldsymbol{\mu} + i\mathbf{k}})]$ with an arbitrary $\boldsymbol{\mu}$. This is because a similarity transformation of the real-space Hamiltonian does not change its determinant [see the discussion below Eq. (31)]. The location $\boldsymbol{\mu}_{\min}$ can be determined by the vanishing of gradient $\partial_{\mu_j} R_{\det \sigma} = 0$ for all j 's (or simply written as $\partial_{\boldsymbol{\mu}} R_{\det \sigma} = \mathbf{0}$).

When the amoeba of $\det \sigma(\boldsymbol{\beta})$ contains a central hole, Eq. (34) can be established by the same approach that leads to Eq. (33). When there is no central hole, we may intuitively view the minimum location $\boldsymbol{\mu}_{\min}$ as an ‘‘infinitesimal central hole,’’ which is consistent with the fact that the Ronkin function takes the minimum in the central hole. Although we do not find a rigorous proof of Eq. (34) in these general cases, our numerical results support its validity (see below). In fact, taking $\sigma(\boldsymbol{\beta}) = E - h(\boldsymbol{\beta})$ in Eq. (34) and extracting the real part lead to

$$\frac{1}{N} \log |\det(E - H)| = R_{\det(E-h)}(\boldsymbol{\mu}_{\min}) + O(L^{-1}) \quad (35)$$

or

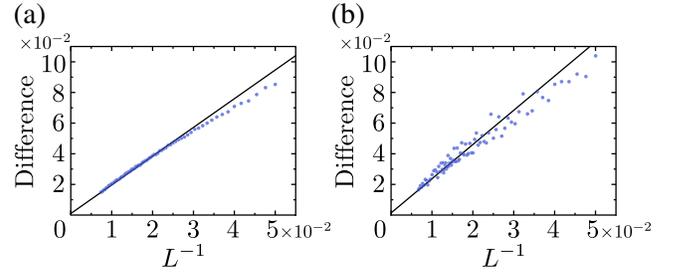


FIG. 5. The maximal difference $\max_E |\phi(E) - \Phi(E)|$ between the OBC-spectrum-based and Ronkin-function-based Coulomb potentials as a function of the linear size L . The adopted Hamiltonian is Eq. (12) [Fig. 2(a)]. The straight line is a linear fitting from data points with $L \geq 55$. (a) The real space is a square of side length L , with disorders added on the boundary [the same as in Fig. 4(b)]. Each data point comes from averaging over six disorder configurations. (b) The real space is a disk of diameter L , without disorder [the same as in Fig. 4(c)].

$$\Phi(E) = \phi(E) + O(L^{-1}), \quad (36)$$

which becomes Eq. (24) in the large-size limit $L \rightarrow \infty$. To confirm this, we numerically calculate the Coulomb potential $\Phi(E)$ for different system sizes, and compare it with the Ronkin minimum $\phi(E)$. In Fig. 5, we plot the maximal difference $\max_E |\phi(E) - \Phi(E)|$ as a function of the system size L . Note that the size dependence comes solely from that of $\Phi(E)$, while $\phi(E)$ is independent of the size. Regardless of the shape of the OBC system, the maximal difference is in good agreement with the L^{-1} behavior, and converges to 0 when extrapolated to the $L \rightarrow \infty$ limit. These behaviors are exactly what Eq. (36) tells.

For Hermitian bands, the DOS of the OBC system determined by our formulation using the Ronkin function is consistent with the familiar fact that the OBC bulk spectrum and the PBC spectrum are asymptotically identical. This can be proved as follows. Since the eigenenergies belong to the real axis \mathbb{R} , one can determine the DOS by Eq. (27), for which knowing $\phi(E)$ for $E \notin \mathbb{R}$ is sufficient. For an arbitrary $E \notin \mathbb{R}$, $\det[E - h(e^{i\mathbf{k}})] \neq 0$ always holds for all $\mathbf{k} \in [0, 2\pi]^d$ because of the Hermiticity of $h(e^{i\mathbf{k}})$. Thus, $\boldsymbol{\mu} = \mathbf{0}$ does not belong to the amoeba of $\det[E - h(\boldsymbol{\beta})]$. In other words, it belongs to one of the amoeba holes. Furthermore, the phase winding number of $\det[E - h(e^{i\mathbf{k}})]$ along every k_j circle is zero. Therefore, the order $\nu_j = 0$ ($j = 1, \dots, d$) at $\boldsymbol{\mu} = \mathbf{0}$, meaning that $\boldsymbol{\mu} = \mathbf{0}$ locates in the central hole. It follows that the Ronkin minimum is reached at $\boldsymbol{\mu} = \mathbf{0}$, i.e., $\phi(E) = R_{\det(E-h)}(\mathbf{0})$. On the other hand, one can readily see that $R_{\det(E-h)}(\mathbf{0})$ by definition is the Coulomb potential of the PBC energy spectrum. Therefore, the OBC DOS obtained from Eq. (27) is the same as the PBC DOS in the Hermitian cases.

V. AMOEBA HOLE CLOSING AND SPECTRAL BOUNDARY

We are now able to prove a powerful theorem about the range of the OBC spectrum in the complex energy plane. It uses only the topology of the amoeba, without having to evaluate the Ronkin function. Hence, it is more efficient when one only wants to know the range of the spectrum.

We denote by Λ the set of E where the amoeba of $\det[E - h(\boldsymbol{\beta})]$ does not possess a central hole. We prove the following theorem:

$$\rho(E) = 0, \quad E \notin \Lambda. \quad (37)$$

Thus, at a certain E , if the amoeba has a central hole, the DOS is zero at this E . In other words, the bulk spectrum is restricted inside Λ .

With the results of the DOS in the previous section, the proof of this theorem is now simple. When an energy E_0 is outside Λ , one must choose $\boldsymbol{\mu}$ in the central hole so that Eq. (33) holds. Because the shape of the central hole varies continuously as E varies, the central hole should still contain this $\boldsymbol{\mu}$ for E sufficiently close to E_0 . Therefore, there exists a neighborhood of E_0 denoted by V , such that for any $E \in V$, the same $\boldsymbol{\mu}$ is in the central hole of the amoeba of $\det(E - h)$. Thus, for any $E \in V$, the Ronkin minimum is $\phi(E) = R_{\det(E-h)}(\boldsymbol{\mu})$. Recalling the definition of amoeba, we see that $\det[E - h(e^{\boldsymbol{\mu}+ik})]$ is nowhere zero in V for any $\mathbf{k} \in T^d = [0, 2\pi]^d$. Furthermore, the Ronkin function can be expressed as

$$R_{\det(E-h)}(\boldsymbol{\mu}) = \int_{T^d} \left(\frac{dk}{2\pi} \right)^d \sum_i \log |E - E_i(\boldsymbol{\beta})| \quad (38)$$

in which $\{E_i(\boldsymbol{\beta})\}$ are the eigenvalues of $h(\boldsymbol{\beta})$ ($\boldsymbol{\beta} = e^{\boldsymbol{\mu}+ik}$ as usual). Because $\det(E - h) \neq 0$ implies $|E - E_i(\boldsymbol{\beta})| \neq 0$, we have $\Delta[\sum_i \ln |E - E_i(\boldsymbol{\beta})|] = 0$, and therefore its integration over T^d vanishes, leading to $\rho(E) = \Delta\phi(E)/2\pi = 0$. This ends our proof of the theorem Eq. (37). Note that this proof makes use only of the original version of Szegő's theorem Eq. (29), without invoking the generalized version Eq. (34).

When $E \in \Lambda$, the DOS is generally nonzero since nothing forces it to vanish. Thus, the boundary of Λ , on which the central hole closes, coincides with the boundary of the energy spectrum. In fact, if we assume the validity of the conjecture Eq. (34) and therefore Eq. (24), we have $\rho(E) = \Delta\phi(E)/2\pi$, which is generally nonzero in Λ . In other words, the support of the DOS is exactly Λ .

In Fig. 6, we illustrate the amoeba-hole closing for the model Eq. (12). Starting from an energy well above the band top (maximum of the real part of the eigenenergies), we decrease the energy along the real axis. For E larger than a certain energy, the central hole of amoeba exists, though its size shrinks as E decreases. At $E = E_t \approx 5.95995$, the central hole closes. According to our proposal, this

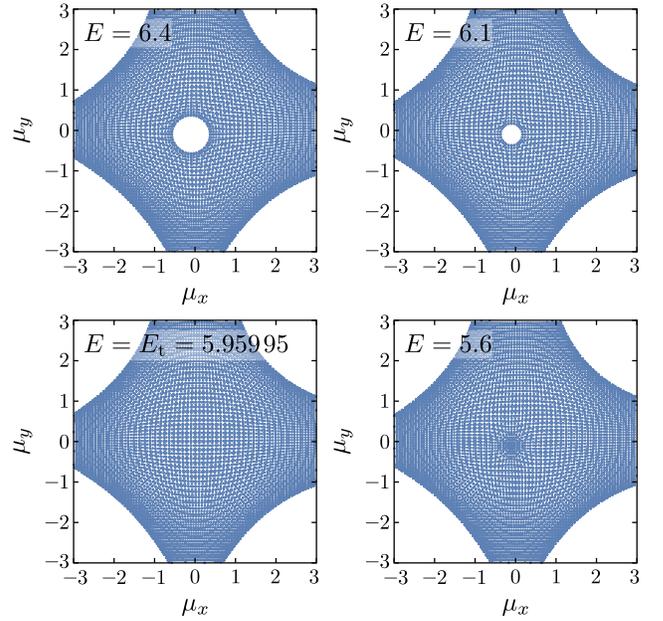


FIG. 6. Amoebae for several energies near the band top of the model Eq. (12). As we decrease the energy along the real axis, the central hole of amoeba closes at $E = E_t \approx 5.95995$. Parameter values are $t = 1$, $t' = 0.5$, and $\gamma = 0.2$, for which the energy spectrum on a disk is shown in Fig. 1(d).

hole-closing point is identified as the band top. Similarly, the entire spectral boundary can be delineated by the hole closing.

In Fig. 7, we compare the results obtained from the amoeba formulation and numerical calculations. We numerically diagonalize the real-space Hamiltonian with increasing sizes and obtain the values of the band top, which are then extrapolated to infinite size [Fig. 7(a)]. The results are strikingly close to the predictions of the amoeba theory [Figs. 7(a) and 7(b)]. We also do a similar comparison for the band bottom (minimum of the real part of the eigenenergies) [Figs. 7(c) and 7(d)]. Note that the band top and band bottom exhibit different scaling behavior when $L \rightarrow \infty$: the finite-size correction $\Delta E_t \propto L^{-2}$ for the former [Fig. 7(a)], and $\Delta E_b \propto L^{-1}$ for the latter [Fig. 7(c)]. It turns out that the former scaling is more accurately obeyed here. Therefore, the error of extrapolation is larger for the latter. Despite larger numerical error, the numerical results are still in good agreement with the amoeba-theoretic prediction.

VI. GENERALIZED BRILLOUIN ZONE

In this section, we establish another proposal mentioned at the end of Sec. III, namely, the location of the Ronkin minimum determines the complex momenta and hence the GBZ. According to Eq. (37), the DOS can be nonzero only when the central hole of the amoeba of $\det(E - h)$ is absent. Thanks to the convexity of the Ronkin function, the

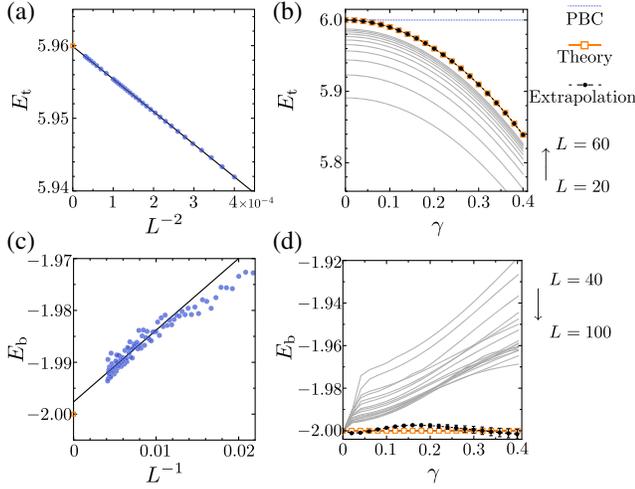


FIG. 7. Band top E_t and bottom E_b obtained from amoeba formulation and numerical calculations. The model is Eq. (12), with $t = 1$, $t' = 0.5$ fixed. In (a),(c), γ is fixed to 0.2. (a) Band top obtained from diagonalizing the real-space Hamiltonian on the disk with diameter L . The extrapolation to $L \rightarrow \infty$ agrees well with the amoeba-hole-closing point marked as the orange square. (b) Band top as a function of γ , for increasing L values. The extrapolation to $L \rightarrow \infty$ is shown as black dots, which are in excellent agreement with the amoeba-hole-closing points marked as orange squares. The PBC result is shown as the dotted line. (c),(d) The counterparts of (a),(b) for the band bottom. The linear fitting in (c),(d) is based on data from $L \in [80, 240]$. Error bars indicate 95% confidence intervals.

minimum location μ_{\min} must be unique in these cases, and therefore the proposal is unambiguous.

The GBZ essentially determines the exponential behavior of the eigenstates of a non-Hermitian lattice Hamiltonian. According to our proposal, an eigenstate with eigenenergy E is expressed asymptotically in the bulk as [46]

$$\psi_E(\mathbf{x}) = \sum_{\mathbf{k}} c_{\mathbf{k}} \exp \{ [i\mathbf{k} + \boldsymbol{\mu}_{\min}(E)] \cdot \mathbf{x} \}, \quad (39)$$

where the sum is over all \mathbf{k} satisfying $\det[E - h(e^{\boldsymbol{\mu}_{\min} + i\mathbf{k}})] = 0$, and $c_{\mathbf{k}}$ are certain E -dependent coefficients. Here, $\boldsymbol{\mu}_{\min}$ plays the role of the imaginary part of the wave vector if we define a complex-valued wave vector

$$\tilde{\mathbf{k}} = \mathbf{k} - i\boldsymbol{\mu}_{\min}. \quad (40)$$

Equivalently, we can use the variable $\boldsymbol{\beta} = e^{i\tilde{\mathbf{k}}}$ (or $\tilde{\mathbf{k}} = -i \log \boldsymbol{\beta}$), so that $\psi_E(\mathbf{x}) = \sum_{\boldsymbol{\beta}} c_{\boldsymbol{\beta}} \prod_{j=1}^d (\beta_j)^{x_j}$. In our theory, the GBZ consists of all points $\boldsymbol{\beta}$ subject to

$$\begin{aligned} \det[E - h(\boldsymbol{\beta})] &= 0, \\ (\log |\beta_1|, \dots, \log |\beta_d|) &= \boldsymbol{\mu}_{\min}(E). \end{aligned} \quad (41)$$

For a d -dimensional non-Hermitian system, the GBZ is a d -dimensional subspace of the $\boldsymbol{\beta}$ space whose real dimension

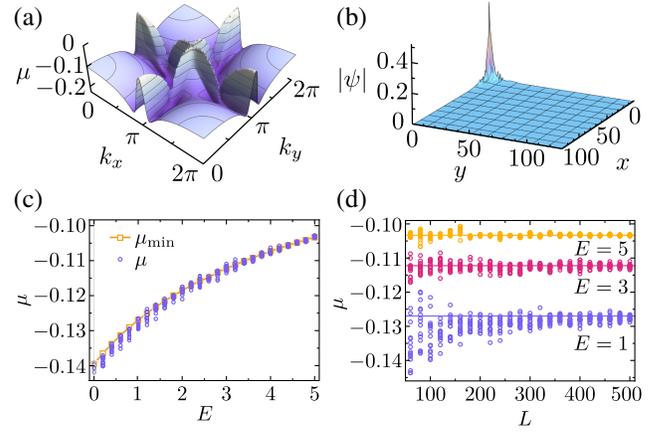


FIG. 8. (a) The generalized Brillouin zone of model Eq. (12). Because of the symmetry of interchanging x and y in this model, we have $(\boldsymbol{\mu}_{\min})_x = (\boldsymbol{\mu}_{\min})_y \equiv \mu$, which is plotted as a function of the real part of the wave vector $\mathbf{k} = (k_x, k_y)$. (b) Typical profile of a bulk eigenstate that exhibits the non-Hermitian skin effect. The eigenstate is taken at $E = -0.003 + 0.056i$. The system is a square of length $L = 130$, with certain random on-site disorders on the boundary. (c),(d) Comparison between μ_{\min} and the exponential decay rate of the Green's function. The latter is defined by linearly fitting μ_x, μ_y from $\log \langle \mathbf{x} | (E - H)^{-1} | \mathbf{0} \rangle \sim \mu_x x + \mu_y y$, in which $\mu_x \approx \mu_y$ for our model, and therefore we present the average $\mu = (\mu_x + \mu_y)/2$. Here, H is a real-space Hamiltonian defined on a disk with diameter L , with random on-site disorder (distributed uniformly in $[-0.5, 0.5]$) on the boundary. Each circle represents the result from a disorder configuration. In the fitting, we discard \mathbf{x} 's that are within distance 20 from the boundary. In (c), we fix $L = 400$. In (d), we vary L at fixed $E = 1, 3, 5$; the corresponding μ_{\min} 's are shown as horizontal lines.

is $2d$. In fact, there are $2d + 2$ real unknowns in Eq. (41), namely, the real and imaginary parts of $(\boldsymbol{\beta}, E)$. Equation (41) then imposes $d + 2$ constraints, meaning that the solution space is d dimensional. Equation (41) provides a general approach to calculate the GBZ for higher-dimensional non-Hermitian systems. In practice, we often parametrize the GBZ by \mathbf{k} , treating $\boldsymbol{\mu}_{\min}$ as its function. This vectorial function $\boldsymbol{\mu}_{\min}(\mathbf{k})$ is a complete representation of the GBZ. As an application of our theory, the GBZ thus obtained for our model Eq. (12) is shown in Fig. 8(a).

Now we provide evidence for Eqs. (39) and (41). First, we show that an eigenstate $|\psi\rangle$ subject to $H|\psi\rangle = E|\psi\rangle$ shares the same exponential behavior with the Green's function $G(E) = (E - H)^{-1}$. To see this, we decompose the OBC Hamiltonian into its eigenstates:

$$H = \sum_n \epsilon_n |nR\rangle \langle nL|, \quad (42)$$

where the sum is over all eigenstates; $|nR\rangle$ and $\langle nL|$ are the right and left eigenstates, respectively, which satisfy

$$H|nR\rangle = \epsilon_n|nR\rangle, \quad \langle nL|H = \epsilon_n\langle nL|. \quad (43)$$

They are orthonormalized as $\langle mL|nR\rangle = \delta_{mn}$. Suppose that \mathbf{x} and $\mathbf{0}$ are real-space locations far from the boundary. The Green's function is related to the eigenstates in the following way:

$$\begin{aligned} & \oint_C \frac{dE'}{2\pi i} \langle \mathbf{x} | (E' - H)^{-1} | \mathbf{0} \rangle \\ &= \sum_n \langle \mathbf{x} | nR \rangle \langle nL | \mathbf{0} \rangle \oint_C \frac{dE'}{2\pi i} \frac{1}{E' - \epsilon_n} \\ &= \sum_{\epsilon_n=E} \langle nL | \mathbf{0} \rangle \langle \mathbf{x} | nR \rangle, \end{aligned} \quad (44)$$

where C is an infinitesimal counterclockwise contour centered at E . We emphasize that E is taken in the OBC energy spectrum, otherwise the integral in Eq. (44) would vanish. The right-hand side of Eq. (44) is a linear combination of all eigenstates with energy E . As our objective is to obtain $\boldsymbol{\mu}$, it suffices to keep both the eigenstates and the Green's functions to the exponential level. Assuming that the Green's function does not change abruptly with respect to E , we conclude that these eigenstates share the same exponential behavior as the Green's function $\langle \mathbf{x} | (E - H)^{-1} | \mathbf{0} \rangle$. We would also like to remark that, while the above argument assumes that the Hamiltonian is diagonalizable, the result is general enough to hold even in the presence of exceptional points [47].

Suppose that the asymptotic behavior is $\langle \mathbf{x} | (E - H)^{-1} | \mathbf{0} \rangle \sim \exp(\sum_{j=1}^d \mu_j x_j)$. To validate Eqs. (39) and (41), we need to show that $\mu_j = (\mu_{\min})_j$. We numerically confirm this relation in Figs. 8(c) and 8(d). Note that in principle one may also validate Eqs. (39) and (41) by the exponential behavior of eigenstates themselves. However, our detour of calculating the Green's function is computationally more accurate and less expensive, enabling calculation for larger L within reasonable time.

VII. NON-BLOCH BAND TOPOLOGY

It has been found recently that the topological numbers defined on the conventional Brillouin zone fail to characterize the topological edge states and bulk-boundary correspondence in non-Hermitian systems. To correctly account for the topological edge modes, the non-Bloch topological invariants defined on the GBZ have been proposed. In practice, most of their applications are restricted to 1D, because a general calculable formulation of GBZ in higher dimension has been lacking. Based on the amoeba formulation, we are now able to address the non-Bloch band topology in higher dimensions. Specifically, the non-Bloch Chern number, which was previously calculated only by continuum approximation, can now be calculated in the entire GBZ as is.

To be concrete, we consider the following non-Hermitian Chern-band model [21]:

$$\begin{aligned} h(e^{i\mathbf{k}}) &= (v \sin k_x + i\gamma)\sigma_x + (v \sin k_y + i\gamma)\sigma_y \\ &+ (m - t \cos k_x - t \cos k_y)\sigma_z. \end{aligned} \quad (45)$$

The real-space Hamiltonian reads

$$\begin{aligned} H &= \sum_{\mathbf{x}} \sum_{j=x,y} |\mathbf{x}\rangle \left(-\frac{i}{2} v \sigma_j - \frac{1}{2} t \sigma_z \right) \langle \mathbf{x} + \mathbf{e}_j | + \text{H.c.} \\ &+ \sum_{\mathbf{x}} |\mathbf{x}\rangle (m \sigma_z + i\gamma \sigma_x + i\gamma \sigma_y) \langle \mathbf{x} |, \end{aligned} \quad (46)$$

where \mathbf{x} are 2D integer coordinates, and \mathbf{e}_j is the unit vector in the j th direction [see Fig. 2(b)]. We set $t = v = 1$ for simplicity.

This model is known to have a Chern insulator phase as well as a trivial insulator phase in the Hermitian limit ($\gamma = 0$) [48]. It has Chern number $C = 1$ when $0 < m < 2$, $C = -1$ when $-2 < m < 0$, and $C = 0$ otherwise. We focus on the phase boundary between $C = 0$ and $C = 1$. When we turn on the non-Hermitian term ($\gamma \neq 0$), the phase boundary extends into a curve in the m - γ plane, which should be predicted by the non-Bloch band theory. In particular, the number of chiral edge modes in each phase is given by the non-Bloch Chern number evaluated on the GBZ, which is a two-dimensional subspace of the four-dimensional $\boldsymbol{\beta}$ space. Alternatively, we may take $\tilde{\mathbf{k}} = \mathbf{k} - i\boldsymbol{\mu} = \log \boldsymbol{\beta}$ as the coordinate system, in which the GBZ is two dimensional because $\boldsymbol{\mu}$ is treated as a function of \mathbf{k} , so that the GBZ can be parametrized by \mathbf{k} . Thus, the non-Bloch Chern number can be viewed as the conventional Chern number of $h(e^{i\tilde{\mathbf{k}}})$, with $\tilde{\mathbf{k}} \in \text{GBZ}$. For a band labeled by α , the non-Bloch Chern number reads

$$C = \frac{1}{2\pi i} \int_{T^2} dk_x dk_y \epsilon^{ij} \partial_i \left\langle \tilde{\mathbf{k}}, \alpha, L \left| \partial_j \tilde{\mathbf{k}}, \alpha, R \right. \right\rangle, \quad (47)$$

where $\epsilon^{xy} = -\epsilon^{yx} = 1$, ∂_j refers to $\partial/\partial k_j$, and $\langle \tilde{\mathbf{k}}, \alpha, L |$ and $|\tilde{\mathbf{k}}, \alpha, R\rangle$ are the left and right eigenstates of the α band, respectively. We focus on the ‘‘valence band’’ ($\text{Re}E < 0$), labeled as $\alpha = -$.

More intuitively, the non-Bloch Chern number can be expressed as the change of Berry phase along circular sections of the GBZ in the y direction, i.e., $C = (1/2\pi) \int_0^{2\pi} [dW_y(k_x)/dk_x] dk_x = [W_y(2\pi) - W_y(0)]/2\pi$, in which the Berry phase

$$\begin{aligned} W_y(k_x) &= -i \int_0^{2\pi} \left\langle \tilde{\mathbf{k}}, -, L \left| \partial_{k_y} \tilde{\mathbf{k}}, -, R \right. \right\rangle dk_y \\ &= - \int_0^{2\pi} \frac{h_x \partial_{k_y} h_y - h_y \partial_{k_y} h_x}{2\hbar(\hbar + h_z)} dk_y, \end{aligned} \quad (48)$$

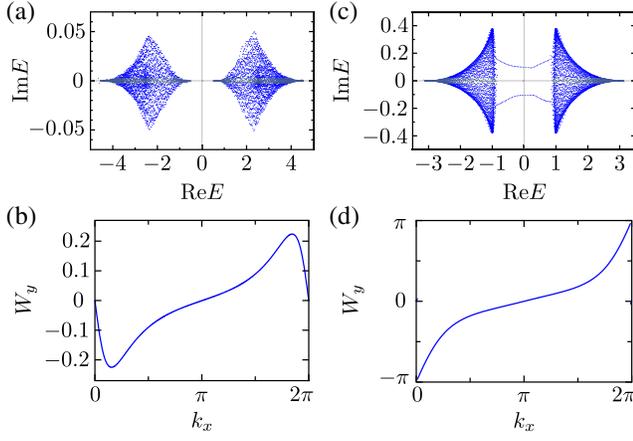


FIG. 9. Energy spectrum and band topology of the non-Hermitian Chern-band model Eq. (45). (a) Energy spectrum by diagonalizing the real-space Hamiltonian on a square of side length $L = 60$, with random on-site disorders on the boundary. Parameters are taken in the topologically trivial regime: $m = 2.6$, $\gamma = 0.4$. (b) The Berry phase $W_y(k_x)$ along a circle in the k_y direction ($k_x = \text{const}$). The non-Bloch Chern number $C = [W_y(2\pi) - W_y(0)]/2\pi = 0$. Parameter values are the same as in (a). (c) Energy spectrum by diagonalization, calculated under the same setting as in (a). Parameters are taken in the topologically nontrivial regime: $m = 1.2$, $\gamma = 0.4$. Topological edge states are observed. (d) The Berry phase as a function of k_x , with $W_y(2\pi) - W_y(0) = 2\pi$ and therefore $C = 1$. Parameter values are the same as in (c).

where we write the Bloch Hamiltonian as $h = h_x\sigma_x + h_y\sigma_y + h_z\sigma_z$, and $\bar{h} = \sqrt{h_x^2 + h_y^2 + h_z^2}$. Here, $h_{x,y,z}$ can be extracted from Eq. (45) by the substitution $\mathbf{k} \rightarrow \tilde{\mathbf{k}}$: $h_x = v \sin \tilde{k}_x + i\gamma$, $h_y = v \sin \tilde{k}_y + i\gamma$, and $h_z = m - t \cos \tilde{k}_x - t \cos \tilde{k}_y$.

In Fig. 9, we show the energy spectrum and the non-Bloch Chern number for the Chern-band model. As an example, we fix $\gamma = 0.4$, and take $m = 2.6$ and $m = 1.2$. From Figs. 9(b) and 9(d), we read for each case that $C = 0$ and $C = 1$, respectively. In Fig. 9(c), topological edge states are indeed seen, being consistent with the nonzero non-Bloch topological invariant $C = 1$.

For the present model, the precise phase boundary can even be analytically determined by the amoeba formulation. In the m - γ plane, the phase boundary between the $C = 0$ and $C = 1$ phases is a curve, which can be viewed as a function $m_c(\gamma)$; i.e., the phase transition point is $m_c(\gamma)$ for a fixed γ . On the phase boundary, the energy gap about $E = 0$ closes, and $E = 0$ belongs to the energy spectrum. Thus, we can find m_c by inspecting when the central hole of the amoeba at $E = 0$ closes. The exact formula for m_c is found to be

$$m_c = 2\sqrt{1 + \gamma^2}. \quad (49)$$

TABLE I. Topological phase boundary m_c as a function of γ . The decay factor μ (equal in both x and y directions) at $E = 0$ and “ m_c amoeba” is determined using the amoeba method, with numerical error $< 1 \times 10^{-6}$. “ m_c numerical” is adopted from Ref. [21], which is obtained by numerical diagonalization of the real-space Hamiltonian with numerical error $< 3 \times 10^{-4}$.

| γ | μ | m_c amoeba | m_c numerical |
|----------|-----------|--------------|-----------------|
| 0.00 | 0.000 000 | 2.000 000 | 2.0000 |
| 0.05 | 0.049 979 | 2.002 498 | 2.0025 |
| 0.10 | 0.099 834 | 2.009 975 | 2.0100 |
| 0.15 | 0.149 443 | 2.022 375 | 2.0225 |
| 0.20 | 0.198 690 | 2.039 608 | 2.0400 |
| 0.25 | 0.247 466 | 2.061 553 | 2.0625 |
| 0.30 | 0.295 673 | 2.088 061 | 2.0885 |
| 0.35 | 0.343 222 | 2.118 962 | 2.1200 |
| 0.40 | 0.390 035 | 2.154 066 | 2.1540 |
| 0.45 | 0.436 050 | 2.193 171 | 2.1940 |
| 0.50 | 0.481 212 | 2.236 068 | 2.2360 |

A perturbative formula for m_c was obtained in Ref. [21], and recently a result to the sixth order of γ was reported in Ref. [49]. Our derivation of Eq. (49) is as follows. We focus on the special solution of the characteristic equation that satisfies $\beta_x = \beta_y \equiv \beta$, as the symmetry of the model allows. The characteristic equation $\det[E - h(\beta_x, \beta_y)] = 0$ becomes

$$(m - \beta - \beta^{-1})^2 - 2\left(\gamma - \frac{\beta - \beta^{-1}}{2}\right)^2 = 0. \quad (50)$$

In a neighborhood of the phase transition, we numerically notice that the four roots of the above quartic equation are all real, and that the second-largest and third-largest (the middle two) roots, under the logarithm, are on the boundary of the central hole of the amoeba. Hence, requiring the central hole to vanish is equivalent to requiring the middle two roots to be equal. Solving this requirement leads to the final result Eq. (49).

Equation (49) is verified up to high precision by numerically locating where the central hole closes, as listed in Table I. From the amoeba-based approach, Fig. 10 illustrates locating the critical value m_c for the case $\gamma = 0.4$. In practice, it is convenient to search for $m_c(\gamma)$ by iteration:

- (1) Plot the amoebae at a consecutive sequence of m 's.
- (2) Find the interval with the smallest central-hole area (closest to hole closing).
- (3) Terminate if the desired accuracy is reached.
- (4) Otherwise, on this new interval, assign a new sequence of m 's and go back to step 1.

On the other hand, m_c can also be determined by numerically diagonalizing the real-space Hamiltonian and finding where the band gap closes [21]. We compare the amoeba-based and numerical-diagonalization results, which agree well with each other.

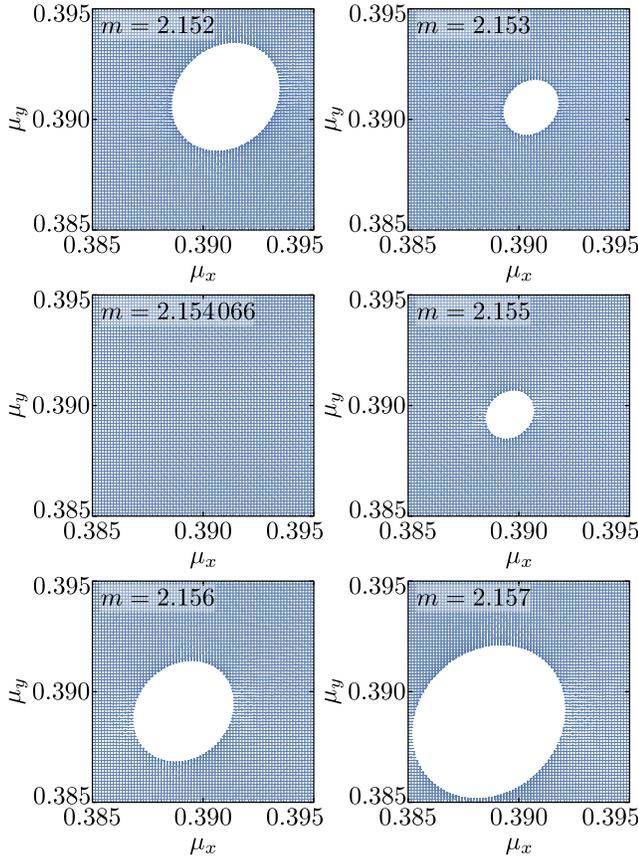


FIG. 10. The amoeba-based procedure of locating the phase boundary in Table I. The figures show the amoebae focused on the central hole, with $\gamma = 0.4$ fixed and m varied. The central hole closes at exactly one point $m = m_c \approx 2.154066$, which is where the topological phase transition occurs.

In the amoeba formulation, an exact expression for the decay factor $\mu_x = \mu_y \equiv \mu$ at the band-closing point $E = 0$ can also be obtained:

$$\mu = \log\left(\sqrt{\gamma^2 + 1} + \gamma\right). \quad (51)$$

This result is very close to the μ values listed in Table I, which are numerically obtained by locating the amoeba-hole-closing points. We remark that a first-order approximation $\mu = \gamma + o(\gamma)$ was reported in Ref. [21].

VIII. SPECTRAL INEQUALITIES

As one of the applications of the amoeba formulation, we prove a general inequality about the spectral radius. For a matrix or operator H , the spectral radius $\rho(H)$ is defined to be the largest absolute value of the eigenvalues. Our spectral inequality states that for a non-Hermitian lattice Hamiltonian with translational symmetry in the bulk, the

OBC spectral radius is less than or equal to its PBC counterpart:

$$\rho(H_{\text{OBC}}) \leq \rho(H_{\text{PBC}}). \quad (52)$$

Note that we are interested in the thermodynamic limit in which all edge-state contribution vanishes. It is evident that Figs. 1(a) and 1(d) satisfy the inequality. The motivation of considering this inequality is a neat inclusion relation in 1D, which states that the OBC spectrum is in the interior of the PBC spectrum [29,50]. As the non-Bloch band theory (which the 1D theorem is based on) aligns with the amoeba formulation in 1D, it is natural to ask if a parallel statement can be established in higher dimensions.

We begin the proof with a simple fact about the electrostatic potential. Suppose that a Coulomb potential Φ is generated by a DOS (charge) distribution ρ in the 2D complex plane, i.e.,

$$\Phi(E) = \int d^2 E' \log |E - E'| \rho(E'). \quad (53)$$

We are interested in the average potential on a circle, say, $|E| = R$, which reads

$$\bar{\Phi} = \frac{1}{2\pi} \int_0^{2\pi} d\theta \Phi(Re^{i\theta}). \quad (54)$$

To calculate this average, one just needs to treat all charges inside the circle as distributed uniformly on the circle, and all charges outside as they are. Because of the additivity of the potential, one needs only to prove this property for a single point charge. The contribution of a point charge q with radial coordinate $r < R$ is $\bar{\Phi} = q \log R$, while that of a point charge with $r > R$ is $\bar{\Phi} = q \log r$. This can be readily proved using Jensen's formula Eq. (19). Taking $g(Re^{i\theta}) = Re^{i\theta} - re^{i\alpha}$ in Eq. (19), the left-hand side of the equation is the average potential generated by a point charge at $E' = re^{i\alpha}$, and Eq. (19) becomes

$$\bar{\Phi} = \begin{cases} \log R, & r < R, \\ \log r, & r > R. \end{cases} \quad (55)$$

Now we proceed with the spectral radius inequality. From Sec. IV, we know that the Coulomb potential generated by the OBC DOS is equal to the minimum of the Ronkin function $\Phi_{\text{OBC}}(E) = \min_{\boldsymbol{\mu}} R_{\det(E-h)}(\boldsymbol{\mu})$. On the other hand, we can also write the Coulomb potential generated by the PBC DOS in terms of the Ronkin function at $\boldsymbol{\mu} = \mathbf{0}$:

$$\Phi_{\text{PBC}}(E) = R_{\det(E-h)}(\mathbf{0}). \quad (56)$$

Hence,

$$\Phi_{\text{OBC}}(E) \leq \Phi_{\text{PBC}}(E). \quad (57)$$

It follows that the average satisfies

$$\bar{\Phi}_{\text{OBC}} \leq \bar{\Phi}_{\text{PBC}}. \quad (58)$$

Taking a circle $|E| = R$ that surrounds the whole PBC spectrum, the average potential on this circle is $\bar{\Phi}_{\text{PBC}} = Q \log R$, where Q is the total charge, which is equal to the number of bands of the specific lattice model. Since the total charge of the OBC DOS is also Q , the aforementioned electrostatic fact implies that $\bar{\Phi}_{\text{OBC}} \geq Q \log R$, with “=” reached when the charge density vanishes outside the circle. Combining this with Eq. (58), we obtain the identity $\bar{\Phi}_{\text{OBC}} = Q \log R$. Consequently, the OBC DOS is zero outside the circle, which means that the OBC spectral radius is no larger than the PBC counterpart.

An alternative proof of the spectral inequality is based on Eq. (37). We denote the PBC spectral (outer) boundary as S , which consists of one or several closed curves (note that the spectrum may also have an inner boundary, which is not included in S ; for example, an annulus-shaped spectrum has an outer and an inner boundary). Consider an energy E outside S . Since E is not in the PBC spectrum, $\boldsymbol{\mu} = \mathbf{0}$ must belong to the complement of the amoeba of $\det(E - h)$. Furthermore, one can see that $\boldsymbol{\mu} = \mathbf{0}$ locates in the central hole. In fact, in the $|E| \rightarrow \infty$ limit (with $\boldsymbol{\mu} = \mathbf{0}$ fixed), the phase of $\det[E - h(\boldsymbol{\beta})]$ is determined solely by E and independent of \mathbf{k} , and therefore, the phase winding number along each k_j circle is zero [cf. Eqs. (16) and (17)]. It follows that, for sufficiently large $|E|$, the order $\nu = \mathbf{0}$ at $\boldsymbol{\mu} = \mathbf{0}$, i.e., $\boldsymbol{\mu} = \mathbf{0}$ locates in the central hole. For any E outside S , one may connect E to infinity by a path that does not intersect S . Since $\boldsymbol{\mu} = \mathbf{0}$ is always in the same amoeba hole as E varies along this path, this amoeba hole must be the central hole. Thus, for any E outside S , $\boldsymbol{\mu} = \mathbf{0}$ locates in the central hole of the amoeba. It follows from Eq. (37) that the OBC DOS $\rho_{\text{OBC}}(E) = 0$ outside S . Therefore, the OBC spectrum is enclosed by the PBC spectral boundary. This statement is slightly stronger than the spectral inequality Eq. (52). Particularly, when the energy spectrum is not convex in the complex plane, this statement implies Eq. (52) but not vice versa.

We can further extend the above result as follows. Taking $\boldsymbol{\mu}$ to be any value other than $\mathbf{0}$, we readily find that the OBC spectrum is enclosed by any TBC($\boldsymbol{\mu}$) spectrum, where TBC($\boldsymbol{\mu}$) refers to a PBC Hamiltonian generated by symbol $h(e^{\boldsymbol{\mu} + i\mathbf{k}})$, i.e., a Hamiltonian under “twisted” periodic boundary condition. We also mention that the development of our theorem is reminiscent of a similar theorem in 1D [50], as they both rely on the winding numbers of the symbol.

A corollary is immediately implied: The spectral range of the real (or imaginary) part of the OBC bulk spectrum is within its PBC counterpart. To be specific, one has the following general inequalities:

$$\max \text{Re}(\epsilon_{n,\text{OBC}}) \leq \max \text{Re}(\epsilon_{n,\text{PBC}}), \quad (59)$$

$$\max \text{Im}(\epsilon_{n,\text{OBC}}) \leq \max \text{Im}(\epsilon_{n,\text{PBC}}), \quad (60)$$

$$\min \text{Re}(\epsilon_{n,\text{OBC}}) \geq \min \text{Re}(\epsilon_{n,\text{PBC}}), \quad (61)$$

$$\min \text{Im}(\epsilon_{n,\text{OBC}}) \geq \min \text{Im}(\epsilon_{n,\text{PBC}}). \quad (62)$$

In view of the significance of the energy spectrum, the spectral inequalities have various implications, some of which have already been exploited. In one dimension, Eq. (60) has played an important role in the directional amplification [51–53]. In the context of open quantum systems, the spectral inequalities are closely related to the boundary sensitivity of Liouvillian gap and relaxation time [54–56]. Our general statement and proof of the spectral inequalities lay the groundwork for their higher-dimensional applications.

IX. CONCLUDING REMARKS

In this work, we have formulated an amoeba theory of the non-Hermitian skin effect and non-Bloch band theory in arbitrary spatial dimensions. It provides a theoretical framework for studying periodic non-Hermitian systems without the serious dimensional limitation. Among other applications, our theory offers a general yet efficient approach to compute the key physical quantities of non-Hermitian systems, such as the energy spectrum, density of states, and generalized Brillouin zone.

Although the initial version of non-Bloch band theory was formulated under the OBC, the concept of GBZ in 1D is also generalizable to other boundary conditions such as the domain-wall cases [57]. Nevertheless, it seems that the amoeba approach, as we now understand, naturally corresponds to the standard OBC systems. Thus, we have focused on the OBC case throughout the present paper. Compared to other boundary conditions, the OBC is especially important in many senses. First, the OBC is experimentally the most relevant because realistic systems often have an OBC. Second, taking the OBC enables the investigation of both bulk and boundary physics, while other boundary conditions including the PBC are blind to boundary phenomena. Third, even certain measurable physical quantities far from the boundary can be naturally expressed in terms of the GBZ from the OBC rather than the conventional BZ associated with the PBC [13,58], though it is in principle free to choose the boundary condition when investigating the physics deep in the bulk. In this sense, the OBC seems to be an advantageous choice even for studying certain bulk physics.

We would also like to remark that not all aspects of this work are mathematically rigorous. Although numerical evidence is supplied whenever a mathematically strict derivation is unavailable, a fully rigorous proof of all our main results is of course desirable.

In view of the ubiquity of periodic structures in both natural and synthetic systems, it is hoped that our theory can find wide applications in the abundant non-Hermitian phenomena. For example, our formulation can be naturally applied to open quantum systems in the free-particle limit, for which the energy spectrum of the non-Hermitian Liouvillian superoperator determines the dynamics and relaxation [54,55]. Our theory immediately enables calculating the relevant quantities beyond 1D. Furthermore, for many-body non-Hermitian systems, our amoeba theory may still be a good starting point for including the interaction effects, which will be left for future work.

ACKNOWLEDGMENTS

This work is supported by NSFC under Grant No. 12125405.

APPENDIX A: A BRIEF PROOF OF SZEGŐ'S LIMIT THEOREM

In this appendix, we sketch a physicist-oriented proof for Szegő's limit theorem Eq. (29), along with an estimation of the inverse of a Toeplitz matrix. We derive concrete conditions for the theorem to hold and explain their intuitions. This proof is inspired by a parallel treatment for translationally invariant differential operators [44,59]. However, the specificity of Toeplitz matrices on lattice requires additional topological constraints, as stated below.

Let us consider a subspace Ω of the d -dimensional Euclidean space. The number of unit cells in Ω is then $O(L^d)$, and the number of unit cells on the boundary is $O(L^{d-1})$, where L is the linear scale of Ω (e.g., the side length of a square, or the diameter of a disk).

Consider two (block-)Toeplitz matrices $\mathcal{T}[\sigma_1]$ and $\mathcal{T}[\sigma_2]$ defined in Ω . Both of them are $N \times N$ block matrices, where N is the number of unit cells (labeled by integer-valued coordinates) in Ω . We first establish the following relation:

$$\|\mathcal{T}[\sigma_1\sigma_2] - \mathcal{T}[\sigma_1]\mathcal{T}[\sigma_2]\|_1 = O(L^{d-1}), \quad (\text{A1})$$

where we use the Schatten norm for operators: $\|T\|_p$ is the p norm of all singular values $\{s_n\}$ of T ,

$$\|T\|_p = \left(\sum_n s_n^p \right)^{1/p}. \quad (\text{A2})$$

For this kind of relation, we use the notation $\mathcal{T}[\sigma_1\sigma_2] \approx \mathcal{T}[\sigma_1]\mathcal{T}[\sigma_2]$. The (\mathbf{x}, \mathbf{y}) matrix element of $\mathcal{T}[\sigma_1\sigma_2] - \mathcal{T}[\sigma_1]\mathcal{T}[\sigma_2]$ reads

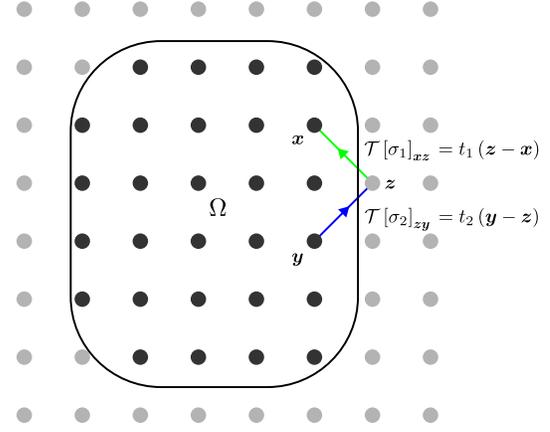


FIG. 11. Illustration of Eq. (A3). The matrix entries of $\mathcal{T}[\sigma_1\sigma_2] - \mathcal{T}[\sigma_1]\mathcal{T}[\sigma_2]$ are nonzero only when x and y are in a neighborhood of the boundary of Ω .

$$\sum_{z \notin \Omega} t_1(z-x)t_2(y-z), \quad (\text{A3})$$

where \mathbf{x} , \mathbf{y} , and \mathbf{z} are integer coordinates; $t_{1,2}$ are the hopping coefficients that appear in $\mathcal{T}[\sigma_{1,2}]$:

$$t_{1,2}(\mathbf{n}) = \int_{T^d} \left(\frac{d\theta}{2\pi} \right)^d \sigma_{1,2}(e^{i\theta}) e^{-in\cdot\theta}. \quad (\text{A4})$$

Note that in Eq. (A3) $\mathbf{x}, \mathbf{y} \in \Omega$ while $\mathbf{z} \in \Omega^c$, where Ω^c is the complement of Ω in the Euclidean space. The terms $\sum_{z \in \Omega} t_1(z-x)t_2(y-z)$ appear in both $\mathcal{T}[\sigma_1\sigma_2]$ and $\mathcal{T}[\sigma_1]\mathcal{T}[\sigma_2]$ and therefore do not contribute to their difference.

It is clear that each term in Eq. (A3) can be nonzero only when both \mathbf{x} and \mathbf{y} are near the boundary (see also Fig. 11). Thus, it is intuitive that the matrix norm of $\mathcal{T}[\sigma_1\sigma_2] - \mathcal{T}[\sigma_1]\mathcal{T}[\sigma_2]$ is of order $O(L^{d-1})$. A more formal proof is based on Hölder's inequality for matrix norms, which states that $\|AB\|_1 \leq \|A\|_p \|B\|_q$ holds for all $p \in [1, \infty]$ and q satisfying $1/p + 1/q = 1$. Defining $(T_1)_{x,z} = t_1(z-x)$ that maps Ω^c to Ω and $(T_2)_{z,y} = t_2(y-z)$ that maps Ω to Ω^c , we see that Eq. (A3) becomes $(T_1 T_2)_{x,y}$. According to Hölder's inequality, we have

$$\|T_1 T_2\|_1 \leq \|T_1\|_2 \|T_2\|_2, \quad (\text{A5})$$

and we are left only with showing $\|T_j\|_2^2 = O(L^{d-1})$ for $j = 1, 2$. Recalling that the singular values of T_1 are the eigenvalues of $\sqrt{T_1 T_1^\dagger}$, $\|T_1\|_2^2$ by definition equals the trace of $T_1 T_1^\dagger$, whose elements read

$$(T_1 T_1^\dagger)_{x,x'} = \sum_{z \in \Omega^c} t_1(z-x)t_1^\dagger(z-x'). \quad (\text{A6})$$

Hence,

$$\begin{aligned}
\|T_1\|_2^2 &= \text{tr } T_1 T_1^\dagger \\
&= \sum_{\mathbf{x} \in \Omega, \mathbf{z} \in \Omega^c} \text{tr } t_1(\mathbf{z} - \mathbf{x}) t_1^\dagger(\mathbf{z} - \mathbf{x}) \\
&= \sum_{\mathbf{x} \in \Omega, \mathbf{z} \in \Omega^c} \|t_1(\mathbf{z} - \mathbf{x})\|_2^2 \\
&= \sum_{\mathbf{u}} \|t_1(\mathbf{u})\|_2^2 \text{vol}((\Omega + \mathbf{u}) \cap \Omega^c). \quad (\text{A7})
\end{aligned}$$

In the last line, we substitute $\mathbf{z} - \mathbf{x}$ by \mathbf{u} . The volume accounts for the multiplicity of the term $\|t_1(\mathbf{u})\|_2^2$, which is the number of lattice points in the overlap of Ω translated by \mathbf{u} and the complement Ω^c . It is roughly $|\mathbf{u}|$ times the surface area of Ω denoted by $|\partial\Omega|$. For short-ranged Toeplitz matrices, we have $t_j(\mathbf{u}) = 0$ when $|\mathbf{u}|$ is beyond the hopping range. Therefore, only a boundary layer contributes to the summation over \mathbf{u} , which means

$$\|T_1\|_2^2 = O(L^{d-1}). \quad (\text{A8})$$

A similar estimation holds for T_2 . These two estimates together prove Eq. (A1). Notably, Eq. (A1) holds even if we add disorders (which is considered in Sec. IV) on the boundary of $\mathcal{T}[\sigma_1]$ and $\mathcal{T}[\sigma_2]$, since the difference on the left-hand side is still localized near the boundary.

A caveat arises when either $\mathcal{T}[\sigma_1]$ or $\mathcal{T}[\sigma_2]$ is not short ranged, meaning, e.g., $t_1(\mathbf{u})$ is nonzero even for large $|\mathbf{u}|$. For a large system size, Eq. (A7) is less than

$$\sum_{\mathbf{u}} \|t_1(\mathbf{u})\|_2^2 |\mathbf{u}| |\partial\Omega|. \quad (\text{A9})$$

Thus, a necessary condition for Eq. (A1) and all the following theorems to hold is that

$$\sum_{\mathbf{u}} |\mathbf{u}| \|t(\mathbf{u})\|_2^2 < \infty \quad (\text{A10})$$

is satisfied for all Toeplitz matrices in consideration. This tells us that our theorems also hold for certain long-ranged Hamiltonians, as long as the hopping decays sufficiently fast at long distance.

Now let $\sigma_1 = \sigma$, $\sigma_2 = \sigma^{-1}$ and assume that σ is invertible, we have

$$\|I - \mathcal{T}[\sigma]\mathcal{T}[\sigma^{-1}]\|_1 = O(L^{d-1}), \quad (\text{A11})$$

where I is the identity matrix. Intuitively, multiplying $\mathcal{T}[\sigma]^{-1}$ on the left, we would expect the following asymptotic inversion formula:

$$\|\mathcal{T}[\sigma]^{-1} - \mathcal{T}[\sigma^{-1}]\|_1 = O(L^{d-1}). \quad (\text{A12})$$

Nevertheless, this multiplication should not be done without discretion. To be precise, we use Hölder's inequality again,

$$\|\mathcal{T}[\sigma]^{-1} - \mathcal{T}[\sigma^{-1}]\|_1 \leq \|\mathcal{T}[\sigma]^{-1}\|_\infty \|I - \mathcal{T}[\sigma]\mathcal{T}[\sigma^{-1}]\|_1. \quad (\text{A13})$$

Therefore, a condition for Eq. (A12) is that $\|\mathcal{T}[\sigma]^{-1}\|_\infty$, which is the largest of its singular values, is finite in the thermodynamic limit $L \rightarrow \infty$. This amounts to asking if $\mathcal{T}[\sigma]$ itself has a zero singular value when $L \rightarrow \infty$. Since most of its singular values form bulk bands and have a finite gap from zero, the zero singular values should be of certain topological origin. They are intimately related to the index theorem, or the bulk-boundary correspondence in free systems. If the symbol has a nontrivial topological index, then the corresponding Toeplitz matrix has either nonempty kernel or nonempty cokernel. Either case implies that $\mathcal{T}[\sigma]\mathcal{T}[\sigma]^\dagger$ is not fully ranked and has a zero eigenvalue. As to 1D Toeplitz matrices, the topological index is well known to be the negative of the winding number of $\det \sigma(e^{i\theta})$ [60]. To ensure that the Toeplitz matrix does not have zero singular values, we impose a condition that the symbol σ is homotopically trivial [61]. This means that there exists a continuous path in the space of invertible symbols subject to Eq. (A10), connecting the symbol $\sigma(e^{i\theta})$ to the constant symbol $\sigma_{\text{const}}(e^{i\theta}) = 1$. This requirement may be too strong for now, but we see in a moment that it arises again when defining the matrix logarithm.

The asymptotic inversion formula Eq. (A12) aims at finding a good surrogate for the inverse of $\mathcal{T}[\sigma]$ in the bulk $\mathcal{T}[\sigma^{-1}]$, which is calculable by integration alone. We remark that Widom also proposed an explicit expression for the subleading term in $\mathcal{T}[\sigma]^{-1} = \mathcal{T}[\sigma^{-1}] + O(L^{d-1}) + o(L^{d-1})$ in one-dimensional case [43,59], and also in higher-dimensional continuous case [44]. It is not used in the current paper and therefore omitted. Szegő's limit theorem also naturally inherits an expression for the subleading term $O(L^{d-1})$.

Next, we move on to Szegő's limit theorem Eq. (29). Since $\log \det A = \text{tr } \log A$, it is equivalent to the following trace formula:

$$\text{tr } \log (\mathcal{T}[\sigma]) = N \int_{T^d} \left(\frac{d\theta}{2\pi} \right)^d \text{tr } \log [\sigma(e^{i\theta})] + O(L^{d-1}), \quad (\text{A14})$$

where N is the number of lattice points in Ω . In order to properly define a logarithm for the symbol, we must rely on the homotopy path imposed earlier. Let $\sigma_\lambda(e^{i\theta})$ be one of these paths, which is a continuous function from $\lambda \in [0, 1]$ to the space of invertible symbols. To ensure the path independence of the integral in Eq. (A16) below, we further

require that $[\sigma'_\lambda, \sigma_\lambda^{-1}] = 0$. For example, a possible construction of such a homotopy path in the context of taking $\sigma(e^{i\theta}) = E - h(e^{\mu+i\theta})$ is as follows. We can find a continuous path E_λ connecting $E_0 = E$ and $E_1 = \infty$ that avoids the spectral range of $h(e^{\mu+i\theta})$ (with $E_\lambda \neq E - 1$), then the path of symbol can be taken as

$$\sigma_\lambda(e^{i\theta}) = \frac{E_\lambda - h(e^{\mu+i\theta})}{E_\lambda - E + 1}. \quad (\text{A15})$$

A consequence of trivial homotopy is that the winding number of σ along any circle on T^d is identically zero. Moreover, it is an interesting question what role is played by point-gap topological invariants other than the winding numbers Eq. (17). Our current conjecture is that the (1D) winding numbers are all that are related to the appearance of NHSE. Mathematically, it means that Szegő's limit theorem holds even if other topological numbers are nonzero. For example, the 3D winding number can protect topological surface states in non-Hermitian band systems, but it does not induce NHSE [62]. Possibly the subleading terms in Szegő's limit theorem will be helpful in quantifying such surface states.

The logarithm in Szegő's limit theorem is now defined as

$$\log \sigma = \int_0^1 \sigma'_\lambda \sigma_\lambda^{-1} d\lambda, \quad (\text{A16})$$

$$\log \mathcal{T}[\sigma] = \int_0^1 \mathcal{T}[\sigma'_\lambda] \mathcal{T}[\sigma_\lambda]^{-1} d\lambda, \quad (\text{A17})$$

where a prime denotes a derivative with respect to λ . Using Eqs. (A12) and (A1), we deduce

$$\begin{aligned} \log \mathcal{T}[\sigma] &\approx \int_0^1 \mathcal{T}[\sigma'_\lambda] \mathcal{T}[\sigma_\lambda^{-1}] d\lambda \\ &\approx \int_0^1 \mathcal{T}[\sigma'_\lambda \sigma_\lambda^{-1}] d\lambda \\ &= \mathcal{T}[\log \sigma]. \end{aligned} \quad (\text{A18})$$

The reason why we focus on the 1-norm becomes clear now, since $\text{tr } A \leq \|A\|_1$ for any matrix or operator A . Taking the trace on both sides, Szegő's limit theorem is proved, because the trace of the matrix on the last line is by definition

$$\begin{aligned} \text{tr } \mathcal{T}[\log \sigma] &= \sum_{x \in \Omega} \int_{T^d} \left(\frac{d\theta}{2\pi} \right)^d \text{tr} \log [\sigma(e^{i\theta})] \\ &= N \int_{T^d} \left(\frac{d\theta}{2\pi} \right)^d \text{tr} \log [\sigma(e^{i\theta})]. \end{aligned} \quad (\text{A19})$$

APPENDIX B: INVARIANCE UNDER CHANGE OF BASIS

An advantage of our method exploiting the amoeba and the Ronkin function is that they are invariant under linear transformations on the basis of the reciprocal space. This fact guarantees that all physical outcomes are independent of a specific choice of basis in the Brillouin zone. We show this invariance in this appendix.

We examine the effect of changing the basis in the reciprocal lattice. Let a basis of the reciprocal lattice be $\{\mathbf{b}_1, \dots, \mathbf{b}_d\}$. An equivalent basis $\{\mathbf{b}'_j\}$ is then related to $\{\mathbf{b}_j\}$ by

$$(\mathbf{b}_1, \dots, \mathbf{b}_d) = (\mathbf{b}'_1, \dots, \mathbf{b}'_d) U. \quad (\text{B1})$$

To ensure that the new basis spans the entire lattice, we require all elements of $U_{d \times d}$ to be integers, and $\det U = \pm 1$. Thus, U is an element drawn from $\text{GL}(d, \mathbb{Z})$.

Let the real-space Hamiltonian be

$$H = \sum_{x,y} |x\rangle t_{y-x} \langle y|, \quad (\text{B2})$$

where we suppress the intracell indices for notational simplicity, but they can be restored without effort. The Bloch Hamiltonian (also known as the symbol) then reads

$$h(\beta_1, \dots, \beta_d) = \sum_{\mathbf{n}} t_{\mathbf{n}} \beta_1^{n \cdot \mathbf{b}_1} \dots \beta_d^{n \cdot \mathbf{b}_d}. \quad (\text{B3})$$

We may write it alternatively as a function of the complex momentum $\tilde{k}_j = -i \log \beta_j$:

$$h(e^{i\tilde{k}_1}, \dots, e^{i\tilde{k}_d}) = \sum_{\mathbf{n}} t_{\mathbf{n}} \exp i(\tilde{k}_1 \mathbf{n} \cdot \mathbf{b}_1 + \dots + \tilde{k}_d \mathbf{n} \cdot \mathbf{b}_d). \quad (\text{B4})$$

After a U transformation, the new Bloch Hamiltonian becomes

$$h'(e^{i\tilde{k}'_1}, \dots, e^{i\tilde{k}'_d}) = \sum_{\mathbf{n}} t_{\mathbf{n}} \exp i(\tilde{k}'_1 \mathbf{n} \cdot \mathbf{b}'_1 + \dots + \tilde{k}'_d \mathbf{n} \cdot \mathbf{b}'_d). \quad (\text{B5})$$

We immediately see that $h(e^{i\tilde{k}_1}, \dots, e^{i\tilde{k}_d}) = h'(e^{i\tilde{k}'_1}, \dots, e^{i\tilde{k}'_d})$ as long as $\tilde{k}_1 \mathbf{b}_1 + \dots + \tilde{k}_d \mathbf{b}_d = \tilde{k}'_1 \mathbf{b}'_1 + \dots + \tilde{k}'_d \mathbf{b}'_d$. In other words, h is an invariant scalar function. It follows that h is invariant if the wave vector transforms as

$$\begin{pmatrix} \tilde{k}'_1 \\ \vdots \\ \tilde{k}'_d \end{pmatrix} = U \begin{pmatrix} \tilde{k}_1 \\ \vdots \\ \tilde{k}_d \end{pmatrix}. \quad (\text{B6})$$

Therefore, the amoeba (coordinates being $\mu_j = -\text{Im } \tilde{k}_j$) undergoes the same linear transformation U when we change the basis. Its properties of physical importance, especially the existence of the central hole, are hence unchanged under a change of basis. Furthermore, for the Ronkin function the following identity holds:

$$R_{\det(E-h)}(\boldsymbol{\mu}) = R_{\det(E-h')}(U\boldsymbol{\mu}), \quad (\text{B7})$$

where $\boldsymbol{\mu}$ stands for the column vector $(\mu_1, \dots, \mu_d)^T$. This is because, by definition,

$$R_{\det(E-h')}(U\boldsymbol{\mu}) = \int_{T^d} \left(\frac{d\theta}{2\pi} \right)^d \log |\det(E - h'(e^{U\boldsymbol{\mu}+i\theta}))|. \quad (\text{B8})$$

Renaming $\boldsymbol{\theta} \rightarrow U\boldsymbol{\theta}$ and noting that the integration measure is unchanged, we find the right-hand side to be equal to $R_{\det(E-h)}(\boldsymbol{\mu})$.

-
- [1] Y. Ashida, Z. Gong, and M. Ueda, *Non-Hermitian physics*, *Adv. Phys.* **69**, 249 (2020).
- [2] E. J. Bergholtz, J. C. Budich, and F. K. Kunst, *Exceptional topology of non-Hermitian systems*, *Rev. Mod. Phys.* **93**, 015005 (2021).
- [3] K. Ding, C. Fang, and G. Ma, *Non-Hermitian topology and exceptional-point geometries*, *Nat. Rev. Phys.* **4**, 745 (2022).
- [4] S. Yao and Z. Wang, *Edge states and topological invariants of non-Hermitian systems*, *Phys. Rev. Lett.* **121**, 086803 (2018).
- [5] F. K. Kunst, E. Edvardsson, J. C. Budich, and E. J. Bergholtz, *Biorthogonal bulk-boundary correspondence in non-Hermitian systems*, *Phys. Rev. Lett.* **121**, 026808 (2018).
- [6] C. H. Lee and R. Thomale, *Anatomy of skin modes and topology in non-Hermitian systems*, *Phys. Rev. B* **99**, 201103(R) (2019).
- [7] V. M. Martinez Alvarez, J. E. Barrios Vargas, and L. E. F. Foa Torres, *Non-Hermitian robust edge states in one dimension: Anomalous localization and eigenspace condensation at exceptional points*, *Phys. Rev. B* **97**, 121401(R) (2018).
- [8] 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 topoelectrical circuits*, *Nat. Phys.* **16**, 747 (2020).
- [9] 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).
- [10] A. Ghatak, M. Brandenbourger, J. van Wezel, and C. Coullais, *Observation of non-Hermitian topology and its bulk-edge correspondence in an active mechanical metamaterial*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 29561 (2020).
- [11] W. Wang, X. Wang, and G. Ma, *Non-Hermitian morphing of topological modes*, *Nature (London)* **608**, 50 (2022).
- [12] K. Yokomizo and S. Murakami, *Non-Bloch band theory of non-Hermitian systems*, *Phys. Rev. Lett.* **123**, 066404 (2019).
- [13] S. Longhi, *Probing non-Hermitian skin effect and non-Bloch phase transitions*, *Phys. Rev. Res.* **1**, 023013 (2019).
- [14] S. Longhi, *Non-Bloch-band collapse and chiral Zener tunneling*, *Phys. Rev. Lett.* **124**, 066602 (2020).
- [15] K. Kawabata, N. Okuma, and M. Sato, *Non-Bloch band theory of non-Hermitian Hamiltonians in the symplectic class*, *Phys. Rev. B* **101**, 195147 (2020).
- [16] Z. Yang, K. Zhang, C. Fang, and J. Hu, *Non-Hermitian bulk-boundary correspondence and auxiliary generalized Brillouin zone theory*, *Phys. Rev. Lett.* **125**, 226402 (2020).
- [17] F. Song, S. Yao, and Z. Wang, *Non-Hermitian topological invariants in real space*, *Phys. Rev. Lett.* **123**, 246801 (2019).
- [18] C. H. Lee, L. Li, R. Thomale, and J. Gong, *Unraveling non-Hermitian pumping: Emergent spectral singularities and anomalous responses*, *Phys. Rev. B* **102**, 085151 (2020).
- [19] Y. Yi and Z. Yang, *Non-Hermitian skin modes induced by on-site dissipations and chiral tunneling effect*, *Phys. Rev. Lett.* **125**, 186802 (2020).
- [20] Z. Wang, *Memorial Volume for Shoucheng Zhang* (World Scientific, Singapore, 2021), Chap. 14, pp. 365–387, 10.1142/9789811231711_0016.
- [21] S. Yao, F. Song, and Z. Wang, *Non-Hermitian Chern bands*, *Phys. Rev. Lett.* **121**, 136802 (2018).
- [22] T. Liu, Y.-R. Zhang, Q. Ai, Z. Gong, K. Kawabata, M. Ueda, and F. Nori, *Second-order topological phases in non-Hermitian systems*, *Phys. Rev. Lett.* **122**, 076801 (2019).
- [23] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, Resultants, and Multidimensional Determinants* (Birkhäuser, Boston, 1994).
- [24] O. Viro, *What is amoeba?*, *Not. Am. Math. Soc.* **49**, 916 (2002), <https://www.ams.org/notices/200208/what-is.pdf>.
- [25] M. Forsberg, M. Passare, and A. Tsikh, *Laurent determinants and arrangements of hyperplane amoebas*, *Adv. Math.* **151**, 45 (2000).
- [26] T. Theobald, *Computing amoebas*, *Exp. Math.* **11**, 513 (2002).
- [27] H. Rullgård, *Polynomial amoebas and convexity*, Ph.D. thesis, Stockholm University, 2001.
- [28] Z. Yang, K. Zhang, C. Fang, and J. Hu, *Non-Hermitian bulk-boundary correspondence and auxiliary generalized Brillouin zone theory*, *Phys. Rev. Lett.* **125**, 226402 (2020).
- [29] K. Zhang, Z. Yang, and C. Fang, *Correspondence between winding numbers and skin modes in non-Hermitian systems*, *Phys. Rev. Lett.* **125**, 126402 (2020).
- [30] M. Passare and H. Rullgård, *Amoebas, Monge-Ampere measures, and triangulations of the Newton polytope*, *Duke Math. J.* **121**, 481 (2004).
- [31] L. Ronkin, *Introduction to the Theory of Entire Functions of Several Variables*, *Translations of Mathematical Monographs* (American Mathematical Society, Providence, 1974).
- [32] E. M. Stein and R. Shakarchi, *Complex Analysis*, Princeton Lectures in Analysis (Princeton University Press, Princeton, NJ, 2010).

- [33] F. Song, H.-Y. Wang, and Z. Wang, *Non-Bloch PT symmetry: Universal threshold and dimensional surprise*, in *A Festschrift in Honor of the CN Yang Centenary: Scientific Papers* (World Scientific, Singapore, 2022), pp. 299–311.
- [34] We also check that adding local randomness in the bulk has a similar effect. Here, the bulk randomness should be sufficiently weak such that the resulting energy spectrum reflects the properties of the pristine Hamiltonian. Specifically, one can add random potential only to N_d unit cells, with $N_d/N \rightarrow 0$ (though $N_d \rightarrow \infty$) when taking the large-size limit $N \rightarrow \infty$ (N is the number of unit cells).
- [35] K. Zhang, Z. Yang, and C. Fang, *Universal non-Hermitian skin effect in two and higher dimensions*, *Nat. Commun.* **13**, 2496 (2022).
- [36] L. Li, C. H. Lee, S. Mu, and J. Gong, *Critical non-Hermitian skin effect*, *Nat. Commun.* **11**, 5491 (2020).
- [37] K. Yokomizo and S. Murakami, *Scaling rule for the critical non-Hermitian skin effect*, *Phys. Rev. B* **104**, 165117 (2021).
- [38] When taking the zero-coupling limit, one sets the chain length to be large and fixed.
- [39] A. Böttcher and B. Silbermann, *Introduction to Large Truncated Toeplitz Matrices*, Universitext (Springer, New York, 1999).
- [40] A. Böttcher and S. M. Grudsky, *Spectral Properties of Banded Toeplitz Matrices* (Society for Industrial and Applied Mathematics, Philadelphia, 2005).
- [41] G. Szegő, *Ein Grenzwertsatz über die toeplitzschen Determinanten einer reellen positiven Funktion*, *Math. Ann.* **76**, 490 (1915).
- [42] H. Widom, *Asymptotic behavior of block Toeplitz matrices and determinants*, *Adv. Math.* **13**, 284 (1974).
- [43] H. Widom, *Asymptotic behavior of block Toeplitz matrices and determinants. II*, *Adv. Math.* **21**, 1 (1976).
- [44] H. Widom, *Szego's limit theorem: The higher-dimensional matrix case*, *J. Funct. Anal.* **39**, 182 (1980).
- [45] R. Y. Doktorskii, *Generalization of G. Szegő's limit theorem to the multidimensional case*, *Sib. Math. J.* **25**, 701 (1984).
- [46] Note that possible boundary terms are omitted, in the same spirit as in 1D.
- [47] In the presence of EPs, we have an additional nilpotent term N in the spectral decomposition: $H = \sum_n \epsilon_n |nR\rangle\langle nL| + N$. The Green's function also acquires an additional term, roughly as $(E - H)^{-1} = \sum_n |nR\rangle\langle nL|/(E - \epsilon_n) + \sum(\text{nilpotent matrix})/(E - \epsilon_n)^p$, where $p \geq 2$ [1]. The final term vanishes after the contour integral in Eq. (44), and therefore, Eq. (44) still holds.
- [48] X. L. Qi, Y. S. Wu, and S. C. Zhang, *Topological quantization of the spin Hall effect in two-dimensional paramagnetic semiconductors*, *Phys. Rev. B* **74**, 085308 (2006).
- [49] Y. Takane, *Bulk–boundary correspondence in a non-Hermitian Chern insulator*, *J. Phys. Soc. Jpn.* **90**, 033704 (2021).
- [50] N. Okuma, K. Kawabata, K. Shiozaki, and M. Sato, *Topological origin of non-Hermitian skin effects*, *Phys. Rev. Lett.* **124**, 086801 (2020).
- [51] A. McDonald, T. Pereg-Barnea, and A. A. Clerk, *Phase-dependent chiral transport and effective non-Hermitian dynamics in a bosonic Kitaev-Majorana chain*, *Phys. Rev. X* **8**, 041031 (2018).
- [52] W.-T. Xue, M.-R. Li, Y.-M. Hu, F. Song, and Z. Wang, *Simple formulas of directional amplification from non-Bloch band theory*, *Phys. Rev. B* **103**, L241408 (2021).
- [53] C. C. Wanjura, M. Brunelli, and A. Nunnenkamp, *Topological framework for directional amplification in driven-dissipative cavity arrays*, *Nat. Commun.* **11**, 3149 (2020).
- [54] F. Song, S. Yao, and Z. Wang, *Non-Hermitian skin effect and chiral damping in open quantum systems*, *Phys. Rev. Lett.* **123**, 170401 (2019).
- [55] T. Haga, M. Nakagawa, R. Hamazaki, and M. Ueda, *Liouvillian skin effect: Slowing down of relaxation processes without gap closing*, *Phys. Rev. Lett.* **127**, 070402 (2021).
- [56] F. Yang, Q.-D. Jiang, and E. J. Bergholtz, *Liouvillian skin effect in an exactly solvable model*, *Phys. Rev. Res.* **4**, 023160 (2022).
- [57] T.-S. Deng and W. Yi, *Non-Bloch topological invariants in a non-Hermitian domain wall system*, *Phys. Rev. B* **100**, 035102 (2019).
- [58] S. Longhi, *Non-Bloch PT symmetry breaking in non-Hermitian photonic quantum walks*, *Opt. Lett.* **44**, 5804 (2019).
- [59] H. Widom, *Asymptotic inversion of convolution operators*, *Publ. Math. IHÉS* **44**, 191 (1974).
- [60] R. G. Douglas, *Toeplitz operators*, in *Banach Algebra Techniques in Operator Theory* (Springer, New York, 1998), pp. 158–184.
- [61] As a heuristic example, consider $\sigma(e^{i\theta}) = \epsilon + e^{i\theta}$. For $\epsilon = 10$, the symbol is homotopic to a constant symbol, and the winding number is zero. One can check that both Eqs. (A11) and (A12) hold. As a comparison, the $\epsilon = 1/10$ case has winding number 1, meaning that the symbol is not homotopic to a constant symbol. Accordingly, Eq. (A11) is satisfied, while Eq. (A12) is not. Instead, it is homotopic to a unitary translation operator, whose symbol is $\sigma(e^{i\theta}) = e^{i\theta}$. It is now obvious that a unitary translation maps the leftmost mode to zero, thus having one zero singular value. Being in the same topological sector, the matrix with $\epsilon = 1/10$ also has a zero singular value (with a correction exponentially small in size).
- [62] M. M. Denner, A. Skurativska, F. Schindler, M. H. Fischer, R. Thomale, T. Bzdušek, and T. Neupert, *Exceptional topological insulators*, *Nat. Commun.* **12**, 5681 (2021).