Rapid Exploration of Topological Band Structures Using Deep Learning

Vittorio Peano

Max Planck Institute for the Science of Light, Erlangen, Germany

Florian Sapper and Florian Marquardt

Max Planck Institute for the Science of Light, Erlangen, Germany
and Department of Physics, Friedrich-Alexander Universität Erlangen-Nürnberg, Germany

The design of periodic nanostructures allows to tailor the transport of photons, phonons, and matter waves for specific applications. Recent years have seen a further expansion of this field by engineering topological properties. However, what is missing currently are efficient ways to rapidly explore and optimize band structures and to classify their topological characteristics for arbitrary unit-cell geometries. In this work, we show how deep learning can address this challenge. We introduce an approach where a neural network first maps the geometry to a tight-binding model. The tight-binding model encodes not only the band structure but also the symmetry properties of the Bloch waves. This allows us to rapidly categorize a large set of geometries in terms of their band representations, identifying designs for fragile topologies. We demonstrate that our method is also suitable to calculate strong topological invariants, even when (like the Chern number) they are not symmetry indicated. Engineering of domain walls and optimization are accelerated by orders of magnitude. Our method directly applies to any passive linear material, irrespective of the symmetry class and space group. It is general enough to be extended to active and nonlinear metamaterials.

I. INTRODUCTION

Wave propagation in a periodic medium is governed by a band structure that substantially modifies the transport of those waves. While these effects were first explored for electrons inside crystals, with the atomic arrangement dictated by chemistry, band structures are also encountered in many other areas across physics where modern advances make it possible to engineer the periodic medium: photonic [1] and phononic [2] crystals as well as optical lattices [3] are well-known examples. This approach the opportunity to explore freely the space of possible designs and search for band structures with peculiar desired properties.

One particularly exciting target for such explorations is the topological features that have become a centerpiece of modern band-structure theory [3–6]. Recent theoretical breakthroughs [6,7] have allowed the exploration of large databases of natural materials to uncover thousands of topological materials [8]. For engineered materials, on the other hand, the configuration space is even infinite dimensional. There, an efficient method to rapidly extract the band structure and topology for any given unit-cell geometry would be a crucial tool which could pave the way to discoveries that would otherwise not be feasible. Ideally, such a method should (i) provide answers for completely arbitrary geometries, (ii) be easily transferrable to different underlying wave equations, (iii) allow a substantial speed-up compared to state-of-the art methods, and (iv) predict topological properties.

We believe that deep learning approaches are uniquely suited to address these challenges. Up to now, the first applications of neural networks to band structures have focused on learning the mapping of a few selected model parameters (describing the geometry of the periodic medium) to the bands [9–12], band gaps [13,14], or topological invariants [15–17]. However, neural networks (NNs) can clearly be designed to make predictions for arbitrary unit-cell geometries, enabling the exploration of a much wider design space. This task is closely related to the well-developed domains of image recognition and image-to-image mapping. While that would already be an important step on its own, such a NN would still be oblivious of any property imprinted in the Bloch waves, including any topological property.
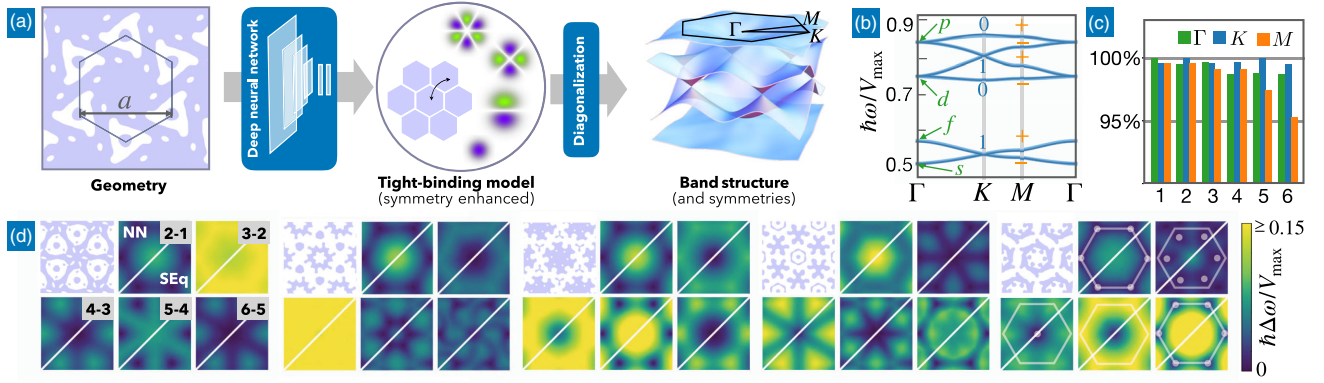
FIG. 1. Neural-network-based prediction of band structures from unit-cell geometries or potentials. (a) The geometry is fed into a multilayer convolutional network (Appendix A) producing the coefficients of a symmetry-enhanced tight-binding model, which is then diagonalized to obtain the band structure. (b) Cut along the **k** path indicated in (a), for the same band structure (now including also the first two bands), comparing the Schrödinger equation (dark thin lines) and NN (thick light lines). The two types of predictions are difficult to distinguish with the bare eye. The symmetry labels are also indicated ($0/1$ indicate the quasiangular momentum; $+$ and $-$ label the even and odd states, respectively). (c) Fraction of correctly predicted symmetry labels at each high-symmetry point (Γ, K, and M) and for each band on validation geometries unknown to the NN. (d) Comparing the **k**-resolved band gap $\omega_{n+1}(\mathbf{k}) - \omega_n(\mathbf{k})$ for five validation geometries (NN vs Schrödinger). For the last potential, the band-gap zeros are marked by white dots. The crossings of the second and third band define a sextuplet of Dirac cones that are not pinned to any high-symmetry point. In all plots, the height of the potential is $V_{\max} = (16\hbar)^2/(2ma^2)$.

The solution we advocate here is to have the NN turn an arbitrary unit-cell geometry into the parameters of a tight-binding (TB) model (see Fig. 1). In a subsequent step, this small TB model is then efficiently diagonalized to yield the full band structure as well as the topologically relevant features of the Bloch waves. Essential constraints imposed by the symmetries of the underlying geometry can be straightforwardly implemented in such a TB model. The whole approach is an example of "known-operator learning" [18], where one embeds into a NN a function that implements a complex (but known) operation that is useful in the given context.

We show that the rapid exploration made possible by our NN is a powerful tool to aid in physical discovery. It addresses challenges in design and optimization, answering questions like the following: Is it possible to implement, under given physical constraints, a band structure of interest, e.g., as produced by a simpler toy model? If yes, which combinations of model parameters are accessible? How abundant are topological bands for an arbitrary distribution of designs? What is the distribution of a topological invariant like the Chern number?

II. SCOPE AND CASE STUDIES

Our approach has a broad scope, as it directly applies to any linear metamaterial supporting band structures encoded in Hermitian Hamiltonians. As such, it is applicable to electronic, photonic, phononic systems, and beyond. For each case study, the symmetry group can be arbitrarily chosen but it remains fixed. In other words, in our method a NN is trained to predict a distribution of band structures

which share one of the 230 space groups (or in 2D, one of the 17 wallpaper groups) and one of the ten symmetry classes. The latter classification accounts for so-called generalized symmetries: the time-reversal, the particle-hole, and the chiral symmetries [19].

Below, we present the essential elements of our method in a general framework. For the sake of concreteness and to prove its practical value, we also discuss in detail a few interesting case studies. The case study which we use to demonstrate most applications is the 2D Schrödinger equation $[-\hbar^2\triangle/2m + V(\mathbf{x})]\psi_n = \hbar\omega_n\psi_n$ with $\mathcal{C}_6$ symmetric (translationally invariant) potentials. This equation has wallpaper group $p6$ and belongs to the symmetry class AI (conserved time-reversal symmetry whose square is the identity and no particle-hole or chiral symmetry). Inspired by the situation that is encountered in photonic or phononic crystals, where the geometry of two materials (solid or air) defines the unit cell, we focus on steplike potentials. This case study relates specifically to the propagation of light in photonic-crystal-type optical waveguides in the paraxial approximation, which generally has been an important playground for photonic topological physics recently [20–24]. We note that the symmetry class AI does not support any strong topological invariant in 2D but still allows for fragile topological phases [25]. Below we show that our method is especially well suited to identify this type of topological phase. Furthermore, we demonstrate that its realm of applications extends to systems supporting strong topological phases. For this purpose, we consider as a case study the 2D Dirac equation with a position-dependent mass; see Sec. VIII. This equation has particle-hole symmetry (squaring to the identity) but broken

time-reversal symmetry and, thus, belongs to the symmetry class D, which supports topological phases with nontrivial Chern numbers. This model is of interest on its own, as it captures the large-wavelength spin-polarized (or mirror-symmetry-polarized) physics of a HgTe/CdTe quantum well [26] whose geometrical parameters are varied periodically to realize a pattern of alternating trivial and topological insulator domains.

Finally, we demonstrate another aspect of the flexibility of our method, implementing it for the 3D Schrödinger equation in the presence of potentials with a nonsymmorphic space group (the space group $p4_222$).

In the context of applications of deep neural networks for topology, our band-structure-based approach, with direct predictions based on the underlying geometry, is of a different nature from other approaches where the network tries to identify (topological) phases of matter based on observing, e.g., simulated snapshots of system configurations or correlators [27–31].

III. TIGHT-BINDING NEURAL NETWORK

In the standard setting of band-structure theory, a wave equation is solved on a periodic lattice, giving rise to a set of bands $\omega_n(\mathbf{k})$, where n is the band index and $\mathbf{k} \in \mathrm{BZ}$ the wave vector inside the Brillouin zone (BZ). The waves are subject to a periodic modulation of a potential (in the case of the Schrödinger equation), a dielectric index (for the Maxwell equations), or material density and elastic moduli (for phononic crystals). To keep our description general, we simply refer to "the unit-cell geometry" in either case.

In our case, we propose to use the NN to generate a TB Hamiltonian: $\hat{H} = \hat{H}(F_\theta[V(\cdot)])$. Here, $V(\cdot)$ represents the network's input (the unit-cell geometry, i.e., a potential or a material distribution), θ is a vector collecting all the network's parameters (weights and biases), and F_θ is the network's output: a vector that contains the energies and hopping matrix elements of the TB model.

The band structure, in turn, results from writing this Hamiltonian in k space, and diagonalizing the resulting $N \times N$ matrix $\hat{H}_\mathbf{k}(F_\theta[V(\cdot)]) = \langle \mathbf{k}|\hat{H}(F_\theta[V(\cdot)])|\mathbf{k}\rangle$. The number N of TB orbitals is chosen depending on how many bands we would like to predict, more on this later. Overall, for any given wave vector $\mathbf{k}$, we generate a vector $\omega = (\omega_1, \omega_2, \ldots, \omega_N)$ of eigenfrequencies,

$$\hbar\omega(\mathbf{k}) = \mathbf{Diag}(\hat{H}_\mathbf{k}(F_\theta[V(\cdot)])).$$

As we indicate above, it is important for network training that the diagonalization operation **Diag** is differentiable with respect to the entries of the Hamiltonian matrix. Indeed, from first-order Rayleigh-Schrödinger perturbation theory, one finds

$$\frac{\partial\omega_n(\mathbf{k})}{\partial\theta} = \sum_l \left\langle \phi_n(\mathbf{k}) \left| \frac{\partial\hat{H}_\mathbf{k}(F_\theta[V(\cdot)])}{\partial F_\theta^{(l)}} \right| \phi_n(\mathbf{k}) \right\rangle \frac{\partial F_\theta^{(l)}[V(\cdot)]}{\partial\theta}. \tag{1}$$

Here, $|\phi_n(\mathbf{k})\rangle$ is the eigenvector in the basis of TB orbitals, $\hat{H}_\mathbf{k}|\phi_n(\mathbf{k})\rangle = \omega_n(\mathbf{k})|\phi_n(\mathbf{k})\rangle$, and $F_\theta^{(l)}$ are the parameters inside the tight-binding Hamiltonian that have been predicted by the NN.

The cost function during training is prescribed as the quadratic deviation between the true band structure and the predictions obtained from the network, averaged over all training samples $V(\cdot)$, the bands n, and the quasimomentum $\mathbf{k}$:

$$C(\theta) = \langle [\omega_n^{\mathrm{NN}}(\mathbf{k}) - \omega_n^{\mathrm{true}}(\mathbf{k})]^2 \rangle_{V(\cdot),n,\mathbf{k}}. \tag{2}$$

The set of $\mathbf{k}$ points is a grid covering the full BZ. See Appendix B for details on the implementation of the resulting gradient descent (using TENSORFLOW).

A. Symmetry-enhanced tight-binding model

One of the important advantages of this approach is the ability to take care of the space group and other symmetries in an elegant and efficient way by imposing them on the TB model. This is particularly important for topological band structures whose topological features are well known to be constrained (and in some cases, even determined) by the underlying symmetry properties.

We call such a TB model "symmetry enhanced." This TB model shares the same space group and symmetry class as the training samples. In order to define its Hilbert space, we select a basis of localized Wannier orbitals. The choice of a suitable set of orbitals depends not only on the space group and symmetry class but also on the potential distribution and the number of bands we would like to predict. The space group and the generalized symmetries impose constraints on the hopping and on-site energies of our TB model (that depends on localization position and point symmetry of the orbitals). Each output neuron of our NN encodes an independent parameter of the underlying Hamiltonian $\hat{H}$; see Appendix C for more details.

During training, we require that the Bloch-wave symmetries at a discrete set of so-called maximal $\mathbf{k}$ points (e.g., Γ, K, and M for the $p6$ group) are reproduced correctly. This also ensures the correct behavior at all other high-symmetry points or lines [6] that may occur in general for arbitrary space groups. For these $\mathbf{k}$ points, the Hamiltonian decomposes into blocks corresponding to an irreducible representation (irrep) of the proper symmetry group of $\mathbf{k}$ (for nonsymmorphic groups, the little group, see Sec. VII). In practice, we enforce the right behavior by applying the cost function (2) separately to each block at the maximal $\mathbf{k}$

points, demanding a match to the training data for each symmetry sector separately.

IV. TRAINING

An important challenge in NN training is the choice of training data. If data are generated by simulation (as is the case here), one can train on random input with a distribution close to the envisaged applications. Our approach is to generate Gaussian random fields in the unit cell (with independent Fourier components, here $\langle |A_{\mathbf{k}}|^2 \rangle \sim |\mathbf{k}|^{-1}$; see Appendix D). The Fourier components are enforced to be of the appropriate symmetry. When required, steplike potentials can be implemented by digitizing the initially continuous random field to two values, $V(\mathbf{x}) = 0$ or $V(\mathbf{x}) = V_{\max}$.

For the 2D Schrödinger equation case study, we train the network on the six lowest-energy bands using a $\mathbf{k}$ grid inside a triangular region covering uniformly one sixth of the Brillouin zone (sufficient for $\mathcal{C}_6$ symmetry); see Appendix E for more details. As we discuss above, afterward our symmetry-enhanced TB model still allows us to predict the band structure with arbitrary $\mathbf{k}$-space resolution. The results produced using our NN (Fig. 1) are essentially indistinguishable from the true bands: Our NN can predict the band structure with about 2% accuracy (relative to typical band gaps; see Appendix F) after training on 50 000 samples, and it is about 1000 times faster than Lanczos-type diagonalization. For a more detailed discussion of the performance gain allowed by our NN, see Appendix G.

In addition, our NN also predicts the underlying proper-group irreps for the Bloch waves at the maximal $\mathbf{k}$ points. Throughout the paper, we refer to the labels identifying such irreps as symmetry labels. We take the example of the $p6$ group to illustrate how symmetries automatically give rise to robust features of the band structure that would be difficult to predict otherwise. For $p6$, the proper group for each maximal $\mathbf{k}$ point is a rotational group $\mathcal{C}_n$, with $n = 6$ for Γ, $n = 3$ for K, and $n = 2$ for M; cf. Fig. 1. The combination of time-reversal symmetry and $\mathcal{C}_n$ rotations gives rise to robust features: (i) At the Γ point, p and d Bloch waves come in pairs with opposite quasiangular momentum and lead to parabolic band touching [Fig. 1(b)]. (ii) Likewise, at the K points, essential degeneracies arise from pairs of states with opposite quasiangular momentum $m_K = \pm 1$, leading to Dirac cones; cf. Figs. 1(b) and 1(d).

We emphasize that such features are automatically enforced by our symmetry-enhanced TB model. This is one of its main advantages over a naive approach. A statistical analysis of a set of validation samples shows that the fraction of correctly predicted symmetry labels is about 99%; cf. Fig. 1(c). This figure of merit is limited only by the rms band-structure deviation: the NN is likely to exchange the ordering of two levels with different symmetry labels only if their splitting happens be so small that the NN is not able to resolve it; see discussion in Sec. VII.

The central focus of modern band-structure theory is the study of topological properties. These properties cannot be deduced from the band structure $\omega_n(\mathbf{k})$ itself but only from the behavior of Bloch waves. We show that, remarkably, our NN learns to predict correctly such properties despite having only very limited implicit information regarding the Bloch waves (via the symmetries). This is a crucial advantage because training for the full eigenstates throughout the BZ would drastically increase the size of the NN and slow down training.

V. DESIGN OF BAND INVERSIONS

The bulk-boundary correspondence provides a link from the bulk topology to the existence of robust gapless excitations at a physical boundary or domain wall. This connection paves the way to using a NN that has been trained on the bulk band structure and Bloch-wave symmetries as a tool to design topological edge states.

For topological insulators, a generic mechanism leading to a nontrivial topology and helical edge states is the so-called band inversion in which the usual ordering of a pair of bands is exchanged. For photonic and phononic crystals, a band inversion of p and d orbitals can be engineered by
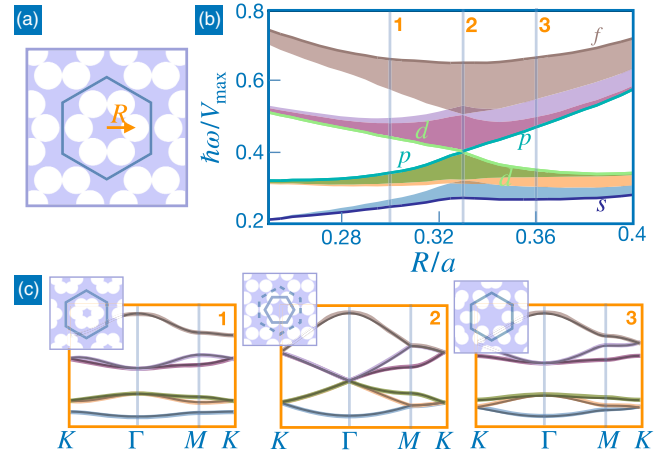

FIG. 2. Designing a band inversion for topological transport using rapid band-structure evaluation and symmetry predictions provided by a neural network. (a) Geometry of the potential: Six circular holes of fixed radius are placed at a varying distance R from the $\mathcal{C}_6$ center. (b) Energy spectrum at the Γ point as a function of R. The energies of the p and d bands cross for $R = a/3$. (c) Band structure for three different values of R [marked in panel (b) by the horizontal lines] before, at, and after the band-inversion transition (thick lines, NN; thin lines, SEq). The corresponding potentials are shown as insets. At the band-inversion transition, the Wigner-Seitz primitive cell becomes smaller. The resulting folded band structure supports a pair of degenerate Dirac cones at the Γ point. Moreover, two pairs of bands become degenerate along the $\mathbf{k}$ path from M to K. These are essential degeneracies enforced by the rotational symmetry and the smaller unit cell. The NN is able to reproduce these features, although it has not been trained on potentials with a smaller unit cell.
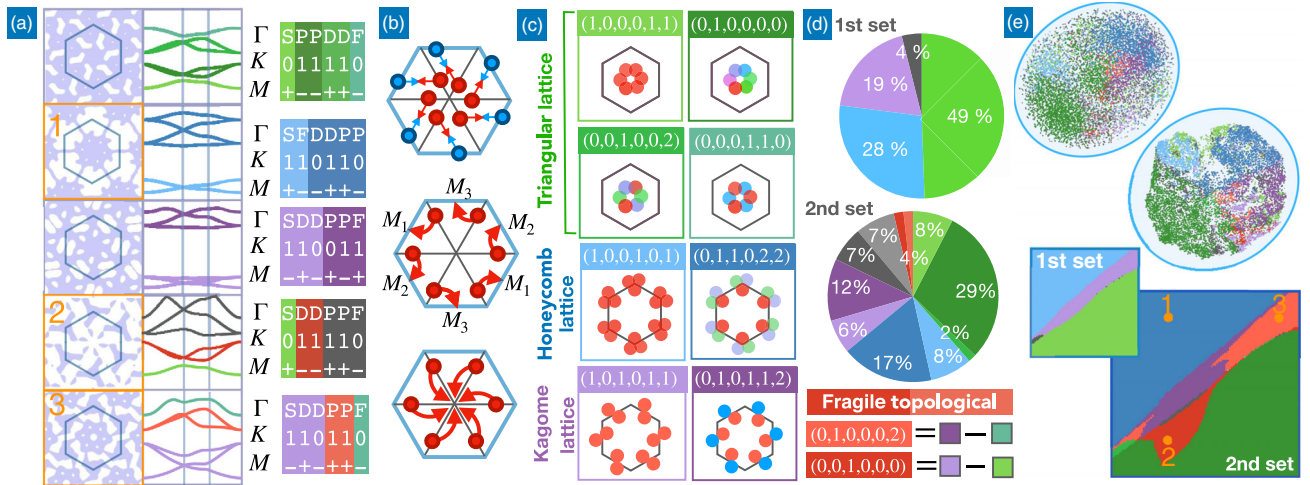
FIG. 3. Using the NN to explore the topological properties of large sets of potentials. (a) Examples of randomly generated potentials and the band structure predicted by the NN. The tables indicate the irreps predicted by the NN for each of the first six bands. (b) Possible pathway leading to the annihilation of two (one) sextuplets of Dirac cones. For one sextuplet, this process is possible only at Γ or M. (c) The EBRs for the wallpaper group p_6. The sketches depict the "Wyckoff" positions and point symmetry of the underlying orbitals. The "occupations" n_i in the "symmetry fingerprint" $(n_s, n_p, n_d, n_f, n_0, n_+)$ count the number of (pairs of) Bloch waves for the corresponding irreps; e.g., $n_p = 1$ means that there is one pair of p waves at Γ. (d) Distribution of (quasi-) BRs for the first (upper chart) and second (lower chart) set of bands. The light gray slice represents sets of bands that cannot be classified based on the first six bands only but are likely to be composite BRs; see Appendix I. All other composite BRs are grouped in the dark gray slice. (e) t-SNE visualization [36], with each point representing a random potential (left) or the corresponding output (TB coefficients) of the NN (right); the latter visualization allows for improved clustering (colors indicate EBRs). (f) Phase map indicating EBR and fragile topology for the first (inset) and the second sets of bands. The parameter space interpolates between the potentials "1," "2," and "3" marked in panel (a), in terms of their underlying Fourier coefficients.

purely geometrical means [32]. Based on this concept, our NN helps efficiently design domain walls of this type. In Fig. 2, the geometry is tuned to decrease the energy of a d orbital while increasing the energy of a p orbital until their order is inverted. The very close agreement between the network predictions and the true spectrum is remarkable, given that the potential designs adopted here look very different from the random training potentials.

VI. EXPLORING BAND REPRESENTATIONS AND FRAGILE TOPOLOGICAL PHASES

Topological band-structure theory originally relied entirely on momentum-space properties defining topological invariants based on the behavior of Bloch waves across the Brillouin zone. Only relatively recently, it was realized that important additional information can be extracted by analyzing the tension between momentum-space and real-space descriptions. The resulting mathematical theories [6–8,33] (sometimes known as "topological quantum chemistry") build onto the theory of band representations 34]] to offer a very general theoretical framework to classify all natural materials according to their topological properties. This theoretical formalism has been so far mostly used to investigate electronic properties of natural materials. However, its range of potential applications extends to any periodic medium; see Ref. [35] for a

pioneering application to photonics. Here, we demonstrate how our NN-based approach combined with topological quantum chemistry allows the rapid exploration and statistical analysis of the topological properties of large sets of band structures.

Band representation (BR) theory tries to understand isolated sets of bands (separated from the remaining bands everywhere by local gaps) in terms of their underlying Wannier orbitals. Mathematically, a BR is a (time-reversal-symmetric) space-group representation that is defined on a basis of Wannier states in the so-called atomic limit [33]. Intuitively, this is the limit where all Wannier states have a localization length that is much shorter than the lattice length scale. A group of bands corresponds to a BR if it is possible to reach the atomic limit by continuously modifying the Hamiltonian without closing the relevant band gaps. For topological bands, it is not possible to reach the atomic limit under continuous deformations. Topological quantum chemistry aims to identify materials hosting such bands. Remarkably, in most cases this is possible based solely on the band structure and the irreps at the maximal $\mathbf{k}$ points, exploiting the fact that all BRs can be decomposed in terms of building blocks known as elementary band representations (EBRs) [34]. Crucially, this information is also made available by our NN [Fig. 3(a)].

We demonstrate the power of the NN by analyzing randomly generated potentials. They are sampled from a

distribution which, in practical applications, might be dictated by experimental design constraints. Here, we illustrate it for $\langle |A_{\mathbf{k}}|^2 \rangle \sim |\mathbf{k}|^{-1/2}$. Even though this distribution is different from the training distribution, the network performs very well.

In a first step, one needs to identify isolated sets of connected bands, which in topological quantum chemistry is commonly done by checking for connections only at high-symmetry points. Our approach allows us to go beyond that by efficiently searching for connections *away* from these points—looking for π defects in the Berry flux on a fine $\mathbf{k}$ grid (much finer than the training grid) evaluated rapidly, thanks to the small Hilbert space of the NN-generated tight-binding model. In this way, we can easily scan large (approximately 10^4) sets of potentials using this method that would be otherwise computationally expensive. Our numerical results show that any clustering of bands based only on connections at high-symmetry points would be incorrect for a substantial fraction of the potentials (approximately 10% for the second set of connected bands). In most cases, this error translates into a wrong topological classification of the bands; see below.

Inspired by these observations, we set out to investigate how robust the connections are away from high-symmetry points. More precisely, we wonder whether—as is often assumed, e.g., Ref. [7]—it is possible to eliminate them without rearranging the order of bands at those points. Band touchings are protected by the $C_2\mathcal{T}$ antiunitary symmetry [37] and can, thus, be eliminated only by pairwise cone annihilation. This property leads us to distinguish two scenarios: (i) If an *odd* number of cones is present in $1/6$ of the BZ, the cones can be annihilated only at the Γ point or at the M points; cf. Fig. 3(c). This implies a rearrangement of the band order at the high-symmetry points. (ii) Otherwise (for an even number), the cones can be annihilated anywhere [Fig. 3(c)] without rearrangement. Using the NN, we discover that the first scenario occurs in the overwhelming majority of cases (approximately 95% for our potential distribution). The presence of robust connections in this scenario seems to point to a missing compatibility relation. Indeed, such a relation can be identified as a consequence of a previous finding in the literature [38]. In our time-reversal invariant system, the sum of the Chern numbers for a set of connected bands is always zero. As shown in Ref. [38], the overall parity of the C_2 eigenvalues at the C_2-symmetric $\mathbf{k}$ points (the parity of the number of odd states for a set of connected bands) is equal to the parity of the Chern number, which therefore means that every connected set of bands must have overall even parity in our system.

As a final step toward identifying topological sets of bands, we enumerate all EBRs, assigning to each a unique symmetry fingerprint ($\mathbb{N}^n$ array) that lists the number of (degenerate) orbitals for each irrep at each symmetry point [Fig. 3(d)] [7,39]. For the group $p6$, the eight possible

irreps at the Γ, K, M points result in $n = 6$ by noting the constraints imposed by the appropriate compatibility relations; see Appendix H. If the fingerprint computed for an isolated set of bands cannot be written as a sum of such EBR fingerprints, the set must be topological (sometimes labeled "quasi-BR").

We use our NN to determine (quasi-) BRs for 10^4 potentials [Fig. 3(e)]. For 4% of the samples in this distribution, the second set of bands is topological. Strictly speaking, this figure depends on the statistical distribution of potentials, but we expect qualitatively similar behavior for other distributions; see Appendix I. The standard analysis without taking into account connections away from the high-symmetry points would overestimate this figure significantly, predicting 14% of topological samples. On the other hand, it turns out that this discrepancy is eliminated once the connections predicted by our C_2 compatibility relation are taken into account. In this case, one recovers with high statistical precision the results already obtained using the much more numerically expensive Berry flux method. This method also gives a way to check our results solving directly the Schrödinger equation; see Appendix I for more details. Besides providing statistical insights, our study also represents an efficient random search, uncovering hundreds of topological samples. Moreover, we obtain important qualitative information: All quasi-BRs discovered here belong to one of two cases [Fig. 3(e)], where the set of bands is obtained by splitting a BR into a topological band and another BR. This is the defining feature of the recently discovered fragile topological phases [25,35,40,41].

A further task rendered feasible by the NN is the creation of high-resolution multidimensional maps that explore the topological and hybridization phase transitions encountered while interpolating between potentials [Fig. 3(f)].

VII. APPLICATION TO A NONSYMMORPHIC EXAMPLE IN 3D

In this section, we aim to demonstrate the flexibility of our method by applying it to 3D band structures. In doing so, we switch our focus from wallpaper groups to space groups. Space groups, unlike wallpaper groups, cannot always be decomposed into a direct sum of lattice translations and the point group. When this is not possible, the space group is said to be nonsymmorphic. For nonsymmorphic groups, the irreps classification—an important step of our method—requires us to take into account transformations that combine point symmetries with translations by a fraction of a lattice vector, i.e., screw rotations and glide mirrors. This is done generalizing the concept of the proper group by introducing the so-called little group. This is an infinite-dimensional subgroup of the space group that leaves invariant a particular quasimomentum and, in contrast to the proper group, can include also some translations. It is also well known that TB models with
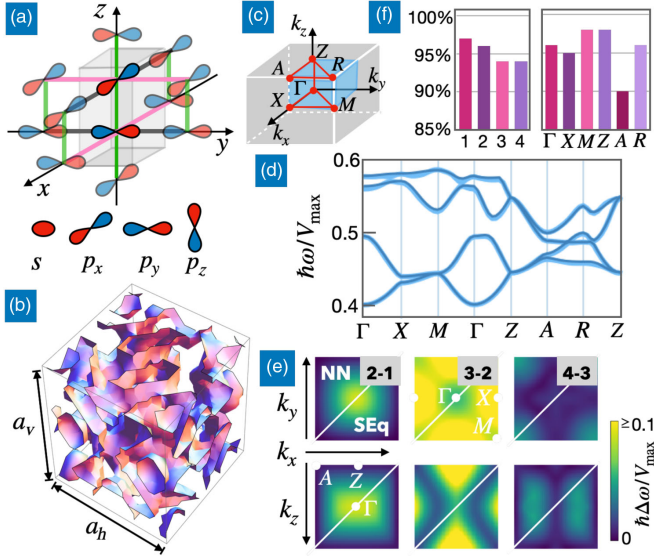
FIG. 4. Neural-network-based prediction of 3D band structures for potential distributions in the $p4_222$ nonsymmorphic space group. (a) Sketch of the symmetry-enhanced TB model. Pairs of orbitals localized about different sublattices are mapped onto each other via screw rotations. Symmetry-related nearest-neighbors transitions are highlighted in the same color. The different orbital types (bottom) transform according to the different representations of the point group $\mathcal{D}_2$ (twofold rotations about the axes x, y, and z); see the Appendix. (b) Example of steplike potential with $p4_222$ space group. (c) BZ and high-symmetry path for the primitive tetragonal lattice. (d) Band structure along the high-symmetry path. (e) Comparison of NN and Schrödinger equation predictions for the first three band gaps on two different 2D cuts of the BZ. [The results in (d) and (e) refer to the validation potential shown in (b).] (f) Fraction of correctly predicted symmetry labels ordered by band and high-symmetry point.

nonsymmorphic space groups can be implemented only on lattices with a basis [42]. Thus, a lattice with a basis will be required for our symmetry-enhanced TB model; cf. Fig. 4(a). We remark that the majority of space groups in 3D are nonsymmorphic (157 out of 230) and that nonsymmorphic space groups can even arise for quasi-2D systems; see, e.g., Ref. [43]. Motivated by the added complexity and the ubiquity of nonsymmorphic space groups, we demonstrate our method for a nonsymmorphic example.

We consider the 3D Schrödinger equation for steplike potentials with space group $p4_222$ (which has point group $\mathcal{D}_4$ and a primitive-tetragonal lattice); cf. Fig. 4(b) for an example of such a potential. In addition to the lattice translations, this group is generated by a screw rotation about the z axis (a rotation by $\pi/2$ accompanied by a translation by half a lattice vector), and a twofold rotation about the x axis; cf. Figs. 4(a) and 4(b). As usual, we train our NN using a coarse grid, here covering 1/8 of the BZ; cf. blue region in Fig. 4(c). As we explain in Sec. III A and

Appendix J, the training cost function includes a contribution for each of the irreps of the little groups at the maximal $\mathbf{k}$ points. Here, we have six maximal $\mathbf{k}$ points (Γ, M, X Z, A, and R) and a total of 28 different irreps (substantially larger than our previous $p6$ example). After taking the steplike ansatz for the potentials and rescaling the energy, we are left with only two free parameters: the potential height V_{max} and the vertical lattice constant a_v (in appropriate units set by the horizontal lattice constant a_h). We train our NN to predict band structures for $V_{max} = (10\hbar)^2/(ma_h^2)$ and $a_v = a_h$. In addition, we discretize the unit cell as a $20 \times 20 \times 20$ grid (comparatively low resolution, with the goal of saving computational resources in producing the training data for this illustrative example).

The results are summarized in Figs. 4(d)–4(f). The band structures for validation potentials evaluated along a high-symmetry path or high-resolution 2D cuts are in good agreement with exact results; cf. Figs. 4(d) and 4(e). More quantitatively, the rms deviation of the band-structure prediction averaged over 2000 validation samples is $2 \times 10^{-3} V_{max}$. The fraction of correctly predicted symmetry labels is above 95%. The error decreases to levels of around 1‰ if one takes into account only the levels and symmetry-protected doublets that are separated by the neighboring levels by more than 3 times the rms deviation. This reduction indicates that the fraction of correctly predicted symmetry labels is limited only by the precision of the band-structure predictions.

Overall, these results show that our method represents a powerful tool to predict 3D band structures that extends to systems with very complex symmetry constraints including nonsymmorphic systems.

VIII. STRONG TOPOLOGICAL PHASES OF TWO-COMPONENT TOPOLOGICAL METAMATERIALS

In our approach to the prediction of band structures and topological properties, the symmetry labels are the only information about the underlying Bloch waves provided to the NN during training. This approach has the advantage of being very efficient, and it is clearly suitable to predict symmetry-indicated topological features. These also cover some (but not all) strong topological invariants. It remains an open question whether our method can be adopted to investigate non-symmetry-indicated topological features. The aim of this section is to address this question. In particular, we focus on the most prominent and well-known example of a non-symmetry-indicated strong topological invariant, the Chern number for 2D systems.

A straightforward approach to predict Chern numbers (or any other non-symmetry-indicated topological invariant) using a NN consists of providing them to the NN during training and one-hot encode the NN's predictions in *ad hoc* output neurons [15–17]. While it would be possible to extend our method based on this approach, we instead use

our standard method. However, this time, we train many NNs on the same training data but with different random initial conditions for the NN. We then use the information regarding the Chern numbers to postselect a NN that performs well on the Chern numbers. This approach is far more elegant, as the Chern numbers are then directly encoded in the symmetry-enhanced TB model.

At first sight, one may think that our approach is doomed to fail: Since we aim to predict the Chern numbers for an infinite number of geometries, one might naively expect that the number of NNs that would be required to be trained should also be infinite. However, one should keep in mind that the predictions of Chern numbers for different geometries are highly correlated: Since a Chern number can only change whenever the corresponding band gap closes, a NN that is trained to accurately predict (via the symmetry-enhanced TB model) the band structure will also automatically find the right topological phase transition hypersurfaces in the space of all geometries. Hence, if the NN is successfully postselected to predict the Chern numbers for even a single geometry for each topological phase, it will automatically provide the right Chern numbers for all other geometries as well. In practice, the Chern numbers might still be wrong in those cases where the band gaps are smaller than the precision of the NN. Only the Chern numbers corresponding to band gaps much larger than the precision will be guaranteed to be correct. However, since, as we see, the NN precision is usually very high, this limitation does not present a significant problem in practice. The remainder of this section is devoted to demonstrating that our expectations are fully confirmed for an interesting case study.

Even after restricting ourselves to 2D systems, there is ample choice of possible case studies where the topological phases are described by nontrivial Chern numbers. Formally, the only precondition is that the relevant Hamiltonian has broken time-reversal symmetry. This is, for example, the case for charged particles in the presence of a magnetic field or for certain systems under a suitable time-dependent drive. Since the spin-orbit interaction is akin to a magnetic field when acting on spin-polarized electrons, the Chern numbers are also suitable to describe the topological phases of this type of excitation. In this framework, the Hamiltonian of interest is "half" (one of each spin-polarized block) of the full time-reversal-symmetric spin-conserving Hamiltonian. Thus, the nontrivial Chern numbers refer to only one spin sector and are known as "spin" Chern numbers. This is the situation in several well-known models [44,45] and for our case study of choice.

Just like before, we want to apply our NN to predict band structures involving nontrivial unit-cell geometries. Starting from any homogeneous (bulk) topological model, we can vary the underlying parameters spatially in a periodic fashion, giving rise to a topological metamaterial. To keep things simple and comparable to other examples, we choose
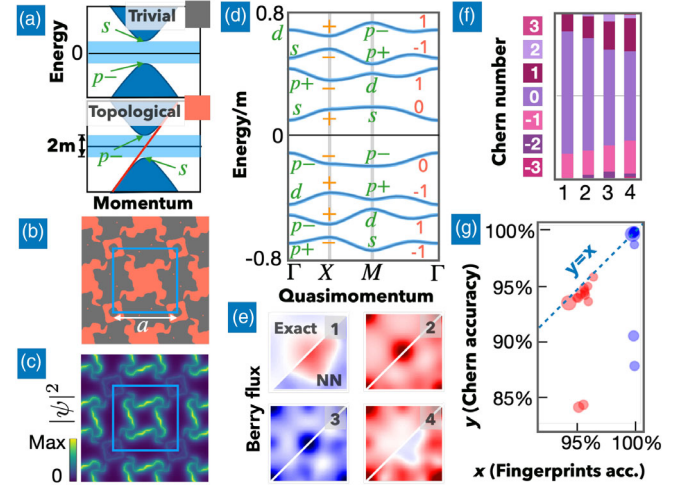

FIG. 5. Using the NN to investigate the spin Chern numbers in a family of 2D topological insulator metamaterials. (a) Sketch of the spin-polarized excitation spectrum in a semi-infinite-plane geometry for a trivial (top) and a topological (bottom) material close to the Γ point, close to the Fermi energy $E_F = 0$. The topological material supports gapless edge states (red line). (b) Sketch of a metamaterial geometry. Its unit cell (blue) is subdivided into trivial (gray) and topological (red) domains according to a randomly drawn fourfold rotationally symmetric pattern. (c) Probability field for a metamaterial's Bloch wave with energy within the single-domain band gap. (d) Metamaterial band structure. Shown are the first (last) four positive (negative) bands together with the symmetry labels and the Chern numbers (in red). Exact results and NN predictions are not distinguishable with bare eyes. (e) Exact and NN-predicted Berry fluxes for the first four bands. [(c)–(e) refer to the validation geometry shown in (b).] (f) Stacked bar chart of the Chern number distribution for the first four positive bands. (g) Scatter plot of the Chern and fingerprint accuracies for 13 trained NNs. Each NN is represented by a red and a blue circle. For the red circles, the rms band-structure deviation (radius of the circles), the fraction of correct fingerprints (x coordinate), and Chern numbers (y coordinate) are obtained, averaging over 2500 validation samples and the eight central bands. The blue circles take into account only bands separated from neighboring bands by a minimal splitting larger than $0.01m$.

only two different values for the parameter in question (similar to the potential landscape discussed so far).

Specifically, we consider a minimal model of a two-component topological metamaterial—a model with many intriguing features that are interesting in their own right, not only as a benchmark for our approach. In this metamaterial, we combine spatial regions of trivial (spin Chern number $\mathcal{C} = 0$) and topological (spin Chern number $\mathcal{C} \neq 0$) bulk material. Both regions are assumed to share the same band gap of width $2m$; cf. Fig. 5(a). The two materials are arranged according to a random unit-cell geometry with fourfold rotational symmetry and lattice constant a; cf. Fig. 5(b). The mismatch in Chern numbers at the closed domain walls separating the two types of regions will give

rise to topological excitations inside the homogeneous bulk band gap; cf. Fig. 5(a). The edge channels propagating along the domain walls themselves form a lattice of closed loops. An isolated loop would have a discrete eigenspectrum. These excitations will propagate chirally about any given closed domain wall but also tunnel to adjacent domain walls; cf. Fig. 5(c). These processes will give rise to the band structure that we set out to investigate.

We model each of the two-component materials using the Bernevig-Hughes-Zhang model [45]. This model is known to capture well the physics of HgTe/CdTe semiconductor quantum wells [26]. For these materials, the valence and the conduction bands have minimal splitting at the Γ point; cf. Fig. 5(a). At this high-symmetry point, the Bloch waves are eigenstates of the quasiangular momentum and are, thus, orbital polarized; cf. the labels in Fig. 5(a). This feature allows us to define the mass or band-gap parameter M as half of the energy difference between the two Bloch waves, $M = \pm m$. The sign of M is determined by the ordering of the bands and a change of ordering is accompanied by a change of Chern number (by one unit). In our conventions (see Appendix K), the Chern number is $\mathcal{C} = 0$, 1 for positive and negative mass, respectively.

We model the composite metamaterial using the Bernevig-Hughes-Zhang model but allowing a lattice-site-dependent mass $M(\mathbf{x})$. Each random configuration $M(\mathbf{x})$ belongs to the p_4 wallpaper group [lattice, square; point group, $\mathcal{C}_4$]. Here, the position $\mathbf{x}$ is defined on a "microscopic" square lattice of lattice constant a/N. The metamaterial unit cell contains $N \times N$ unit cells of the microscopic lattice and has lattice constant a. This leads to a folding of the BZ giving rise to N^2 bands for each band of the Bernevig-Hughes-Zhang model. Subsequent bands are separated by local band gaps, and we can assign to each band an integer Chern number.

As we discuss above, we are primarily interested in the band structure formed by the topological excitations in the bandwidth of the single-domain band gap; cf. Figs. 5(c) and 5(d). The distribution of the Chern numbers for the first four bands above the Fermi energy (here located in the middle of the whole band structure) is shown in the stacked bar chart Fig. 5(f). We consider the mesoscopic regime where the typical localization length ξ, i.e., the transverse extent of the edge channels at the domain walls, is larger than the microscopic lattice constant but smaller than the macroscopic one: $a/N \ll \xi \ll a$. Since a sets the scale for the typical distance between adjacent domain walls, we expect the interdomain hopping to be exponentially suppressed and, thus, the bands to be well separated. In this regime, the band structure is well approximated within a large-wavelength description encapsulated in the Dirac equation

$$H_D = M(\hat{\mathbf{x}})\hat{\sigma}_z + v(\hat{k}_x\hat{\sigma}_x + \hat{k}_y\hat{\sigma}_y). \qquad (3)$$

Here, the energy is counted off from the Fermi energy and v is the speed of the excitations. Moreover, $\hat{\sigma}_{i=x,y,z}$ is a set of Pauli matrices whose basis states are an s and a p_- orbital for $\sigma_z = 1$ and $\sigma_z = -1$, respectively. Importantly, the antiunitary transformation $\Xi = \mathcal{K}\sigma_x$ (where $\mathcal{K}$ is the complex conjugation) is a particle-hole symmetry, $\Xi^{-1}H_D\Xi = -H_D$. Since $\Xi^2 = \mathbb{1}$, our family of metamaterials is in the symmetry class D.

We note that the assumption that the Pauli matrices are defined on a specific basis of atomic orbitals affects only the symmetry labels but not the band structure or the Chern numbers and, thus, does not imply any loss of generality. In this sense, Eq. (3) and, thus, all results presented here go beyond our specific microscopic model. This includes (but it is not limited to) scenarios in which the spin is not conserved (but the mirror out-of-plane transformation M_z is a symmetry) and/or the particle-hole symmetry is an emergent symmetry not present in the microscopic model; see Appendix K.

We train 13 NNs to predict the eight bands around the Fermi energy and the corresponding Chern numbers for $\xi = v/M = a/10$ (after a trivial rescaling of the energy, this length is the only free parameter in the large-wavelength description). We use the same method described in Sec. IV. However, since we aim to predict the band structure in the middle of the spectrum (instead of starting from the minimal energy), we now face an additional challenge: Given that the symmetry-enhanced TB model and the original lattice model have a different number of bands, it is not clear which band should correspond to which (see Appendix K).

All of the NNs trained with this approach perform well for most of the geometries, even without postselection, both in the prediction of the symmetry labels and of the Chern numbers. This includes even cases where the training grid (a 9×9 k-space grid covering one fourth of the BZ) would be too coarse to calculate the Chern numbers as the sum of the Berry fluxes across the BZ [46]. For the purpose of validating the Chern numbers after the training, we use a fine 62×62 k-space grid to obtain reliable results. We note that the somewhat surprisingly good performance in the Chern number predictions, even before the postselection, is favored by the especially stringent constraints posed by the symmetry labels on the Chern numbers for our p_4 wallpaper group. For this group, the symmetry labels of a band determine the Chern number modulo 4 [38]. In other words, $\mathcal{R} = \mathcal{C} \bmod (4)$ can be viewed as a symmetry-indicated topological invariant. Even after taking this constraint into account, there is still an infinite number of possible Chern numbers for each value of $\mathcal{R}$. Thus, the task of predicting the Chern numbers remains nontrivial. We further note that due to the natural correlations between band structure and Berry curvature, even the Berry curvature predicted by the NN is in most cases qualitatively correct; cf. Fig. 5(f). This is remarkable because the NNs

have not received any information regarding the Bloch waves away from the high-symmetry points.

We compare the predictions by different NNs by plotting in a scatter plot the rms band-structure deviation (radius of the circles) and the fraction of correctly predicted Chern numbers (y axis) and symmetry fingerprints (x axis); cf. Fig. 5(g). Below, we refer to the latter two quantities as Chern and fingerprints accuracy, respectively. Remarkably, all NNs perform at a similar level for the rms band-structure deviation and the fingerprint accuracy, while they can be roughly divided into two groups when the Chern accuracy is also taken into account. For the first group, the Chern accuracy is significantly lower than the fingerprints accuracy. This reflects that, as it should be expected for a non-symmetry-indicated topological invariant, for a significant number of cases, the Chern number is wrong even though the symmetry fingerprints are correct. On the other hand, for the other much larger group (11 out of 13 NNs), the two figures almost perfectly coincide and assume a value larger than 95% [cf. red circles in Fig. 5(g)]. Thus, by postselecting any of the NNs of the second group, we obtain a high Chern accuracy.

As anticipated above, the remaining errors are mostly due to band gaps that are too small to be reliably resolved by the NN. This is confirmed by recalculating the Chern and fingerprint accuracy, now taking into account only the bands separated from neighboring bands by a minimal splitting larger than a small threshold ($0.01m$, or roughly 3 standard deviations of the band-structure rms deviation). In this case, both accuracies are very close to 100% (the residual error is at the level of around $1‰$); cf. the blue circle in Fig. 5(g). This proves that the accuracy of our Chern number predictions after NN postselection is limited only by the band-structure precision (a similar conclusion then holds also for the symmetry labels).

IX. OPTIMIZATION

Gradient-based optimization search for a geometry that maximizes some reward is a powerful but numerically intensive design tool for photonic devices [47–50]. The numerical effort involved in calculating a large number of finite element method simulations represents a substantial bottleneck for explorative designs. NNs offer a natural way out of this challenge as it has been demonstrated in a handful of pioneering works [10,51–53]. In contrast to these works, our approach allows us to search for an arbitrary input geometry. As we explain above, this geometry is parametrized via the Fourier coefficients of a smooth field that is then discretized via a sigmoid function (see Appendix L).

An important goal consists of solving the inverse problem, where we try to reach a given target band structure. This approach might be used, for example, to find a physical implementation of some TB model of

interest (sharing the goal of Ref. [54]) under the given experimental constraints.

In Fig. 6, we illustrate the procedure for a TB model [55] that underlies fruitful applications in topological photonics [56,57] and phononics [58,59]. The presence of local minima in the optimization landscape is easily addressed by running multiple trials and postselecting outcomes, thanks to the 1000-fold acceleration produced by the NN.

We observe that the optimal geometry is not defined uniquely [Fig. 6(a)] since we demand only a match in the first few bands. This feature could be exploited to select for structures that are easy to fabricate. Conversely, however, it is not generally possible to reach arbitrary band structures, due to physical constraints like the allowed values of the potential (the refractive-index contrast in the photonic case) and the unit-cell size. To delineate the accessible regions of the TB model parameters, a scan with repeated optimization runs is required. Performing such scan for a 3D parameter space [Fig. 6(b)] even on a coarse grid, the number of evaluations runs in the millions (Appendix L), which does not present a problem for the NN but would be very impractical otherwise. The resulting map can be used
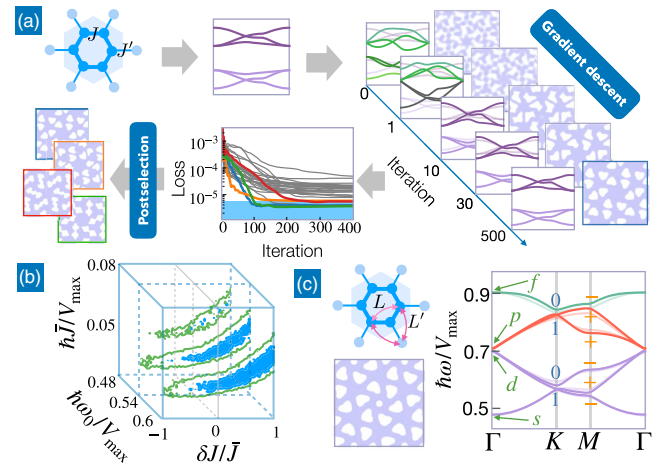

FIG. 6. Using the NN for gradient-based optimization. (a) Pipeline for finding a physical implementation of a given TB model (unit cell shown). The band structure of the model is provided as a target. A randomly initialized geometry is evolved until the NN-predicted band structure (dark lines) approaches the target (bright lines); see also the animation in the Supplemental Video [60]. We select those random trials that end up at a loss comparable to the NN accuracy itself (blue shaded region). (b) Physically accessible regions of the TB model parameter space: Contours show the minimal loss achieved after 50 trials as a function the parameters (on-site energy ω_0, average hopping $\bar{J} = (J + J')/2$, hopping difference $\delta J = J' - J$). In the blue region, the minimal loss is lower than the network accuracy. (c) Solutions for a more general model, including also next-nearest-neighbor hopping supporting fragile topological phases; see Appendix M for more details. The parameters are $\hbar\omega_0 = 0.6V_{max}$, $\hbar J' = 0.038V_{max}$, $\hbar J = 0.035V_{max}$, $\hbar L' = 0.002V_{max}$, and $\hbar L = -0.002V_{max}$.

as a starting point for realizing extended TB models, e.g., implementing fragile topological phases with next-nearest-neighbor hopping [Fig. 6(c)].

Other reward functions can be used to optimize only for specific feature combinations (like band gaps, group velocities, selected band representations, etc.). More generally, one might even optimize potential landscapes where smooth geometry deformations in real space lead to some band-structure evolution that, e.g., produces edge states with desired properties. One important point in optimization is that the network should give reliable robust predictions even away from training examples. Empirically, this seems to be the case here, in our observations. Nevertheless, this could be the domain of further study, possibly exploiting the concept of adversarial approaches (where one tries to slightly change the input in a deliberately disadvantageous way to maximize the deviation from the correct output; see, e.g., Ref. [61]).

X. OUTLOOK

The tight-binding-network approach introduced here can be directly applied to many other situations. These include, without any alterations in the NN, finite-element calculations for electromagnetic and elastic waves (where the execution speed advantage of the NN is enhanced by further orders of magnitude). Moreover, direct extensions allow us to address band structures for metamaterials with inhomogeneous dissipation and amplification (with complex eigenfrequencies and exceptional point physics in reciprocal space) and driven nonlinear photonic crystals or optomechanical arrays (where excitation pair creation leads to a symplectic Hamiltonian structure and novel topological features). Interactions on the mean-field level can be addressed as well, e.g., using solutions of the Gross-Pitaevskii equation for matter waves in optical lattices or using density-functional-theory results for real materials (where the input could be atomic positions instead of geometries, using the ideas of SCHNET [62]). We expect approaches like the one exemplified here to become a standard part of the toolbox for metamaterial design.

ACKNOWLEDGMENTS

APPENDIX A: NETWORK LAYOUT

Here, we describe the layout of our NN. The network maps a two-dimensional $L \times L$ image $V(\cdot)$ representing the potential inside a (parallelogram-shaped) unit cell onto a finite set of coefficients of the symmetry-enhanced TB model; see Appendix C. In our explorations, we find that a
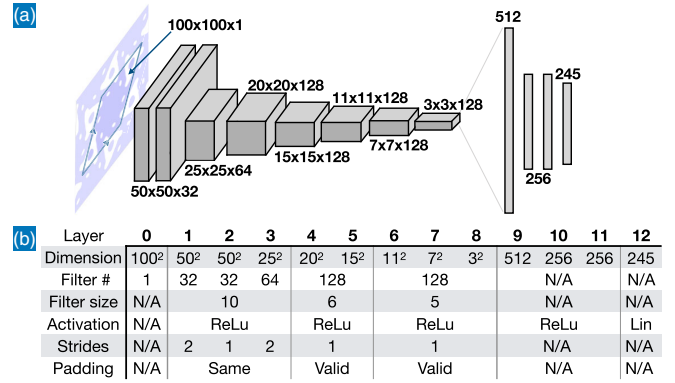

FIG. 7. Details of the NN layout. (a) Sketch of the layout of the neural network. The input image is the unit cell of the potential (the region within the blue contour). The numbers indicate the dimensions of the layers, where for the convolutional the last value is the number of filters. The image is flattened before it is processed by the dense layers. The final dense layer has 245 output neurons encoding the independent coefficients of the TB model. (b) Detailed list of the parameters for the network. "Layer 0" labels the input image.

Layer	0	1	2	3	4	5	6	7	8	9	10	11	12
Dimension	100^2	50^2	50^2	25^2	20^2	15^2	11^2	7^2	3^2	512	256	256	245
Filter #	1	32	32	64	128		128			N/A			N/A
Filter size	N/A	10			6		5			N/A			N/A
Activation	N/A	ReLu			ReLu		ReLu			ReLu			Lin
Strides	N/A	2	1	2	1		1			N/A			N/A
Padding	N/A	Same			Valid		Valid			N/A			N/A

good choice for the number of convolutional layers as well as the kernel sizes is important for robust and successful training, even though the fine details do not matter. The proper structure depends mainly on the size of the input potential. For the case of potentials with size 100×100, we find the layout described below to give good results. The layout is sketched in Fig. 7(a). A detailed list of parameters is given in the table in Fig. 7(b). The entire implementation uses the TENSORFLOW framework.

(i) Multilayer convolutional network: The first eight layers of the neural network are 2D convolutional (conv2D) layers with ReLu (rectified linear unit) as the activation function. One goal of applying successive convolutional layers is the reduction of the image size, which is usually done using pooling layers. In our case, we instead reduce the image size mainly by using stride = 2 in some layers. The combination of the options stride = 2 and padding = same leads to the reduction of the pixels by a factor of 4 (a factor of 2 in each direction). The combination stride = 1 and padding = valid leads also to a reduction of the pixels by eliminating grid points close to the boundaries (when the kernel is not completely within the image).

(ii) Multilayer fully connected network part: After flattening the result of the last conv2D layer, four dense layers are applied with dropout (0.15) between each pair of dense layers. The first three dense layers use also ReLu activations, while the last dense layer uses linear activation to allow rescaling. The number of neurons is 512/256/256/245. The number of neurons in the last layer (245) corresponds to the

number of independent coefficients of our symmetry-enhanced TB model; see Appendix C.

The resulting output of the neural network is then interpreted as the coefficients of the (symmetry-enhanced) tight-binding model (see section above), which is diagonalized numerically. We run everything except the eigenvalue calculations on the graphics processing unit (GPU). However, due to the implementation of the diagonalization in TENSORFLOW, it is important to run tf.linalg.eigh on the CPU instead of the GPU. Otherwise, the diagonalization takes about 2 orders of magnitude longer due to parallelization overhead.

Future improvements might include implementing conv2D layers with periodic boundary conditions and implementing the convolution operations on the actual mesh inside the parallelogram-shaped unit cell (for a triangular lattice, this would be a triangular mesh, instead of the square mesh assumed in the TENSORFLOW implementation). However, as far as we observe, these details do not prevent the network from reaching a very good performance.

APPENDIX B: GRADIENT DESCENT FOR THE COMBINATION OF NEURAL-NETWORK AND TIGHT-BINDING MODEL

One of the unconventional parts of our ansatz is the use of known-operator learning, i.e., having numerical diagonalization be part of the overall pipeline leading from geometry to band structure. To be able to perform gradient descent on this combination, we implement a suitable modification of the cost function introduced above. The main idea is to exploit perturbation theory to obtain the derivative of the eigenvalues of a matrix with respect to its coefficients and to feed this analytical expression into the TENSORFLOW backpropagation pipeline.

The expression for the modified cost function can be derived by starting with the derivative of the original cost function:

$$\frac{\partial C(\theta)}{\partial \theta} = \left\langle \sum_l 2[\omega_{n,\mathbf{F}_\theta[V(\cdot)]}^{\mathrm{NN}}(\mathbf{k}) - \omega_n^{\mathrm{true}}(\mathbf{k})] \right.$$
$$\left. \times \frac{\partial \omega_{n,\mathbf{F}_\theta[V(\cdot)]}^{NN}(\mathbf{k})}{\partial F_\theta^{(l)}[V(\cdot)]} \frac{\partial F_\theta^{(l)}[V(\cdot)]}{\partial \theta} \right\rangle_{V(\cdot),\mathbf{k},n},$$

where $F_\theta^{(l)}[V(\cdot)]$ is determined by a neuron of the output layer. Here we are very careful in spelling out all the dependences; in particular, the band structure depends on the neural-network parameters θ via the tight-binding coefficients.

We can use the Rayleigh-Schrödinger perturbation-theory relation

$$\frac{\partial \omega_{n,\mathbf{F}_\theta[V(\cdot)]}(\mathbf{k})}{\partial F_\theta^{(l)}[V(\cdot)]} = \frac{1}{\hbar}\langle \phi_n(\mathbf{k})| \frac{\partial \hat{H}_\mathbf{k}(F_\theta[V(\cdot)])}{\partial F_\theta^{(l)}} |\phi_n(\mathbf{k})\rangle$$

to calculate the derivative of the eigenvalues. Since $\hat{H}_\mathbf{k}(F_\theta[V(\cdot)])$ is linear in every coefficient $F_\theta^{(l)}[V(\cdot)]$, $\partial \hat{H}_\mathbf{k}(F_\theta[V(\cdot)])/\partial F_\theta^{(l)}$ is a numerical constant. Therefore, we must calculate this expression only once before training, and we can then use it for all training steps. $F_\theta^{(l)}[V(\cdot)]$ is also independent of the sum over the $\mathbf{k}$ and n, which is why we can rewrite the derivative of the cost function as

$$\frac{\partial C(\theta)}{\partial \theta} = \left\langle \sum_l v_l \frac{\partial F_\theta^{(l)}[V(\cdot)]}{\partial \theta} \right\rangle_{V(\cdot),\mathbf{k},n},$$

with

$$v_l[V(\cdot),\mathbf{k},n]$$
$$= \frac{2}{\hbar}[\omega_{n,\mathbf{F}_\theta[V(\cdot)]}^{\mathrm{NN}}(\mathbf{k}) - \omega_n^{\mathrm{true}}(\mathbf{k})]$$
$$\times \left\langle \phi_{n,\mathbf{F}_\theta[V(\cdot)]}(\mathbf{k}) \left| \frac{\partial \hat{H}_\mathbf{k}(F_\theta[V(\cdot)])}{\partial F_\theta^{(l)}} \right| \phi_{n,\mathbf{F}_\theta[V(\cdot)]}(\mathbf{k}) \right\rangle.$$

Hence, we arrive at the conclusion that we can use

$$C^{\mathrm{NN}}(\theta) = \left\langle \sum_l v_l[V(\cdot),\mathbf{k},n] F_\theta^{(l)}[V(\cdot)] \right\rangle_{V(\cdot),\mathbf{k},n}$$
$$= \langle \mathbf{v}[V(\cdot),\mathbf{k},n] \cdot \mathbf{F}_\theta[V(\cdot)] \rangle_{V(\cdot),\mathbf{k},n} \qquad (\mathrm{B1})$$

as the cost function for the neural network.

Indeed, the θ gradient of this cost function is the same as for the original one, as long as we postulate that $\mathbf{v}$ is to be treated as independent of θ. Since $\mathbf{v}$ is a vector with the number of coefficients as the number of entries, this cost function is realized in TENSORFLOW as a simple scalar product between the output layer of the NN and the vector $\mathbf{v}$. Each training step of the neural network consists of the calculation of $\mathbf{v}$ (for a batch of training samples) and the usual gradient descent applied to the cost function C^{NN}.

APPENDIX C: SYMMETRY-ENHANCED TIGHT-BINDING HAMILTONIAN

Here, we give more details on the "symmetry-enhanced" tight-binding models whose parameters are predicted by our NN and subsequently used to calculate the band structures. [There is one such TB model for each symmetry group considered (space group plus time-reversal symmetry when applicable).]

The challenge in defining such TB models is that they should be able to reproduce the low-energy bands of a broad distribution of potentials. Moreover, the number of underlying orbitals and parameters should remain as small as possible to keep the diagonalization of the TB Hamiltonian numerically inexpensive.

1. Wallpaper group p_6 with time-reversal symmetry

In order to estimate how large the Hilbert space of our TB model should be, we define the occupation $n_\xi^{[V(\cdot)]}$ in the lowest seven bands for the potential $V(\cdot)$ and the irreps ξ, $\xi = s, p, d, f, 0, 1, +, -$. The number of orbitals required for our TB model will then depend on the maximal occupations $n_\xi^{(\max)}$ over all training samples, $n_\xi^{(\max)} = \max_{V(\cdot)} n_\xi^{[V(\cdot)]}$.

We decide (somewhat arbitrarily) to build our TB model using only orbitals localized about the $\mathcal{C}_6$ rotocenters. We denote by $\tilde{n}_l^{(\mathrm{TB})}$ the number of l orbitals, $l = s, p, d, f$. The number of different orbitals in real space then determines the number $n_\xi^{(\mathrm{TB})}$ of Bloch waves that are available for each irrep at the high-symmetry points, e.g., $n_0^{(\mathrm{TB})} = \tilde{n}_s^{(\mathrm{TB})} + \tilde{n}_f^{(\mathrm{TB})}$ and so on. Requiring $n_\xi^{(\mathrm{TB})} \geq n_\xi^{(\max)}$ for all irreps (such that for all samples, enough Bloch waves with the right symmetry are available) results in a lower bound on $\tilde{n}_l^{(\mathrm{TB})}$. For the 50 000 training samples used to train our NN, we find $n_\xi^{(\max)} = 3$, for $\xi = s$ and $n_\xi^{(\max)} = 2$ otherwise. Accordingly, we choose to have $\tilde{n}_s^{(\mathrm{TB})} = 4$ s orbitals, $\tilde{n}_p^{(\mathrm{TB})} = 4$ p orbitals, $\tilde{n}_d^{(\mathrm{TB})} = 3$ d orbitals, $\tilde{n}_f^{(\mathrm{TB})} = 3$ f orbitals (well above the lower bound set by $n_{\max}^{(\xi)}$.)

We note that while all unperturbed orbitals for our TB model are localized about the same Wyckoff position, the Wannier states for an isolated set of bands can still be hybridized orbitals localized about different Wyckoff positions. This may happen because the hoppings between different TB orbitals can be larger compared to the typical on-site energy differences. Thus, our choice of the Wyckoff position for the unperturbed orbitals is akin to a choice of basis. We also restrict the hopping to nearest-neighbor orbitals. This choice reduces the number of output neurons and, thus, the overall complexity of the NN while still turning out to be adequate to obtain a well-trained NN in the examples considered in this work.

Next, we derive the explicit form of the TB model described above in terms of the appropriate set of independent on-site energies and hoppings amplitudes. The constraints imposed by the $\mathcal{C}_6$ symmetry that connect hopping rates in different directions are most easily taken into account using a basis of $\mathcal{C}_6$-symmetric Wannier orbitals $\{|\tilde{W}_{n,m}\rangle\}$ where n is the principal quantum number and m is the quasiangular momentum, $\hat{R}_{\pi/3}|\tilde{W}_{n,m}\rangle = e^{-im\pi/3}|\tilde{W}_{n,m}\rangle$ with $m = 0, \pm 1, \pm 2, 3$. In the corresponding basis of Bloch waves, one can then easily add the contributions from all hopping directions to find

$$\hat{H}_{\mathbf{k};n,m;n'm'}(\mathbf{k})/\hbar = \delta_{m,m'}\delta_{n,n'}\omega_{n,|m|} + \tilde{J}_{n,m;n',m'}f_{m-m'}(\mathbf{k}),$$

where $\omega_{n,|m|}$ are the on-site energies, $\tilde{J}_{n,m;n',m'}$ are the hopping amplitudes in the direction of the lattice vector $\mathbf{a}_1 = a(1,0)$, and the functions $f_{\Delta m}(\mathbf{k})$ are independent of the potential,

$$f_0(\mathbf{k}) = \cos(\mathbf{k} \cdot \mathbf{a}_1) + \cos(\mathbf{k} \cdot \mathbf{a}_2) + \cos(\mathbf{k} \cdot \mathbf{a}_3),$$

$$\begin{aligned}
f_1(\mathbf{k}) &= -f_{-1}^*(\mathbf{k}) \\
&= -i[\sin(\mathbf{k} \cdot \mathbf{a}_1) + e^{-i\pi/3}\sin(\mathbf{k} \cdot \mathbf{a}_2) \\
&\quad + e^{-i2\pi/3}\sin(\mathbf{k} \cdot \mathbf{a}_3)],
\end{aligned}$$

$$\begin{aligned}
f_2(\mathbf{k}) &= f_{-2}^*(\mathbf{k}) \\
&= \cos(\mathbf{k} \cdot \mathbf{a}_1) + e^{i4\pi/3}\cos(\mathbf{k} \cdot \mathbf{a}_2) + e^{i2\pi/3}\cos(\mathbf{k} \cdot \mathbf{a}_3),
\end{aligned}$$

$$f_3(\mathbf{k}) = -i[\sin(\mathbf{k} \cdot \mathbf{a}_1) - \sin(\mathbf{k} \cdot \mathbf{a}_2) + \sin(\mathbf{k} \cdot \mathbf{a}_3)],$$

with $\mathbf{a}_2 = a(1, \sqrt{3})/2$, $\mathbf{a}_3 = a(-1, \sqrt{3})/2$. Because of the time-reversal symmetry, the on-site energy is the same for states with equal principal quantum number n and opposite quasiangular momentum m. This results in $\tilde{n}_s^{(\mathrm{TB})} + \tilde{n}_f^{(\mathrm{TB})} + \tilde{n}_p^{(\mathrm{TB})} + \tilde{n}_d^{(\mathrm{TB})}$ real independent on-site energies which are represented by an equal number of output neurons $F^{(l)}$; cf. Eq. (B1). The hopping amplitudes $\tilde{J}_{n,m;n',m'}$ are also constrained by the symmetries of the problem. The relevant constraints are most easily expressed by switching to a time-reversal invariant basis of Wannier states $\{|W_{n,l}\rangle\}$, $l = s, p_1, p_2, d_1, d_2, f$, where $|W_{n,s(f)}\rangle = |W_{n,0(3)}\rangle$, and

$$|W_{n,p(d)_1}\rangle = \frac{1}{\sqrt{2}}(|\tilde{W}_{n,1(2)}\rangle + |\tilde{W}_{n,-1(2)}\rangle),$$

$$|W_{n,p(d)_2}\rangle = \frac{-i}{\sqrt{2}}(|\tilde{W}_{n,1(2)}\rangle - |\tilde{W}_{n,-1(2)}\rangle). \tag{C1}$$

[Here, we also implicitly fix the sum of the phases of states with equal n and opposite quasiangular momentum m by assuming $\mathcal{T}|\tilde{W}_{n,m}\rangle = |\tilde{W}_{n,-m}\rangle$.] Because of the time-reversal symmetry, the hopping amplitudes $J_{n,l;n',l'}$ in the time-symmetric basis are real. Moreover, using the $\mathcal{C}_2$ symmetry and that the Hamiltonian should be Hermitian, one finds the additional constraint

$$J_{n,l;n',l'} = \pm J_{n',l';n,l},$$

where the positive sign applies when both orbitals have the same behavior (odd or even) under the $\mathcal{C}_2$ symmetry and the negative sign applies otherwise, e.g., $+$ when $l = s$ and $l' = d_{1(2)}$ (both orbitals are even) and $-$ for $l = s$ and $l' = f$ (s is even while f is odd). Taking into account these additional constraints, there are $(N_{\mathcal{H}}^2 + N_{\mathcal{H}})/2$ real independent hopping amplitudes, $N_{\mathcal{H}} = \tilde{n}_s^{(\mathrm{TB})} + \tilde{n}_f^{(\mathrm{TB})} + 2(\tilde{n}_p^{(\mathrm{TB})} + \tilde{n}_d^{(\mathrm{TB})})$. These are represented by the same number of output neurons $F^{(l)}$; cf. Eq. (B1).

APPENDIX D: GENERATING TRAINING SAMPLES

When training a network on simulation results, an arbitrary random distribution of training samples can, in principle, be chosen. However, for best accuracy, it is beneficial to have these samples be as close as possible (statistically) to typical use cases encountered in later applications.

We start by generating a periodic smooth 2D random Gaussian field,

$$\phi(\mathbf{x}) = \sum_{\mathbf{k}} A(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{D1}$$

where the wave vectors $\mathbf{k}$ lie on the reciprocal lattice,

$$\mathbf{k} = n_1 \boldsymbol{b}_1 + n_2 \boldsymbol{b}_2. \tag{D2}$$

Here, n_1 and n_2 are integers, and $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are reciprocal lattice vectors.

The Fourier coefficients $A(\mathbf{k})$ respect the underlying symmetry (again, in our chosen example, they are symmetric under 60° rotations). Otherwise, they are complex Gaussian-distributed random numbers (of zero mean), with variance

$$\langle |A(\mathbf{k})|^2 \rangle = \frac{C}{|\mathbf{k}a|^\alpha} f(\mathbf{k}). \tag{D3}$$

The function $f(\mathbf{k})$ is 1 for small $|k|$ and implements a cutoff for larger k. In our case, we choose $C = 2$ and set $f(\mathbf{k}) = 0$ for $n_{1(2)} > 6$; cf. Eq. (D2). The exponent α determines how smooth the field appears (in our case, $\alpha = 1$). After training, we check that our NN still performs well for random validation potentials drawn from a similar distribution but with a substantially larger cutoff. This shows that the band structure for the low-energy bands is insensitive to fine details of the potential (on a length scale smaller than the one set by our cutoff). This makes sense because the underlying Bloch waves should remain smooth to reduce the kinetic energy.

The examples treated in the main text are inspired by photonic or phononic crystals, where only two materials are involved. This means we want to provide a "digitized" potential starting from the smooth field ϕ. That is achieved by the help of the rounded step function, the sigmoid $\sigma(x) = 1/(1 + e^{-x})$:

$$V(\mathbf{x}) = V_{\max} \sigma[\beta\phi(\mathbf{x})], \tag{D4}$$

where smaller β imply a more gradual step. For the training, we use sharp step functions corresponding to the limit $\beta \to \infty$.

APPENDIX E: TRAINING OF THE NEURAL NETWORK

To train the neural network, we use 50 000 samples of random potentials, with the correct band structure evaluated at 79 points. These points are evenly distributed within one-sixth of the Brillouin zone.

Out of these 50 000 samples, 1024 are reserved for calculating the validation loss and the remaining 48 976 samples are used for training. We use the widespread ADAM optimizer, with parameters $\mathrm{Adam}(\mathrm{lr} = 0.0001, \mathrm{epsilon} = 10e - 8)$. The dropout rate between the dense layers is chosen to be 0.15. The training spans many epochs (approximately 1000). Before every training epoch, we reshuffle the training samples.

We recall that (as we mention in the main text) in order to train the NN also on the symmetry of the Bloch waves at the high-symmetry points, we add to the *global* (comprising an average over the BZ) cost function Eq. (2) other *local* terms for each block of the Hamiltonian at each high-symmetry point. Each block corresponds to an irrep of the proper group of the relevant high-symmetry point, and the corresponding *local* cost function term has (except for the average over the BZ) the same form as Eq. (2). The overall cost function will then be a weighted sum of the *global* and the *local* cost functions with the weight ratio r between the *local* and the *global* contributions playing an important role during training.

Since the *global* cost function is oblivious of the symmetry labels, it tends to prevent a change in the ordering of the bands. Thus, the *local* cost functions should dominate the global cost function at least until the fraction of correct symmetry labels is high enough (indicating that, for the overwhelming majority of the geometries, the bands are correctly ordered). On the other hand, a too large weight ratio r might imply an excessive emphasis on the high-symmetry points and, thus, can be detrimental to the overall quality of the band-structure predictions. We note in passing that $r \sim 1$ is to be regarded as comparatively large because, assuming (to fix the ideas) that the band-structure deviations are of the same order across the BZ, it implies that the contribution to the overall training gradient from a single high-symmetry point (via the *local* cost functions) would be similar to the combined contributions from all over the BZ (via the *global* cost function). Indeed, we empirically observe that using a constant r of the order $r \sim 1$ during the whole training run produces accurate NNs and that one can improve even further the NNs by decreasing r toward the end of a training run. As it is often the case for supervised learning, the training is limited by the onset of overfitting. Good results are already obtained on a timescale of approximately 100 epochs, but the onset of overfitting occurs only on a timescale of approximately 1000 epochs. [Thus, it is worth training for approximately 1000 epochs to obtain optimal accuracy.] The onset of overfitting proves that our ansatz for the

symmetry-enhanced TB model (with all orbitals in the Wyckoff position $1a$ and nearest-neighbor coupling; cf. Appendix C) does not represent a bottleneck for the achievable rms band-structure deviation.

We observe that the training allows a large degree of flexibility in the choice of the hyperparameters, such as optimizer, learning rate, batch size, network layout, time dependence of the relative weight ratio r, etc. A good choice of hyperparameters consistently (for all random initial conditions of the network parameters) leads to a low rms band-structure deviation (of the order of approximately $0.001V_{\text{max}}$) and a high fraction of correctly predicted symmetry labels. Nevertheless, the predictions for different training runs might still be qualitatively different; e.g., unlucky initial conditions may lead to spurious Dirac cones or (for the two-component topological metamaterial case study, cf. Sec. VIII) to the wrong Chern numbers. These wrong predictions are strongly correlated from sample to sample. This means that they are either present for a significant portion of validation samples (and are, thus, easy to detect) or are not present at all. Thus, they can be easily eliminated by training a few NNs and discarding the unreliable NNs; cf. Sec. VIII.

Before the training of the NN, we pick a "target" number N_{target} of bands that we aim to predict. In this work, we show results for different N_{target}, $N_{\text{target}} = 4, 6, 8$. In addition, we perform numerical experiments with $N_{\text{target}} = 10$ for the 2D Schrödinger equation, obtaining accurate predictions ($0.004V_{\text{max}}$ rms deviation and 99% of correctly predicted labels). An interesting question is how large can we increase N_{target}? We expect to be able to increase N_{target} somewhat above $N_{\text{target}} = 10$ but that we will encounter a bottleneck in the required number of trainable parameters for the NN. This number scales as the square of the number of independent parameters in our symmetry-enhanced TB which itself scales as N_{target}^2 (thus, overall we have a quartic dependence). One could get around this problem by training different NNs on different band numbers, e.g., one NN for the first ten bands and a second for the next ten bands. If this approach works, the number of NN trainable parameters as a function of the overall number of bands to be predicted will scale as $N_{\text{target}}^{d+1/d}$ where d is the dimension, e.g., $d = 2$ in 2D. This scaling is governed by the number of potential grid points required to converge to the continuum limit ($\propto N_{\text{target}}$). This quantity determines both the number of neurons in each convolutional layer ($\propto N_{\text{target}}$) and the number of convolutional layers ($\propto N_{\text{target}}^{1/d}$).

Since the eigenvalue calculation is performed on the CPU, the duration of one epoch depends strongly on the CPU. The workload on the CPU depends on the number of **k** points in the global cost function. With batch size 16 and 79 reciprocal points, on an NVIDIA RTX 6000 and a
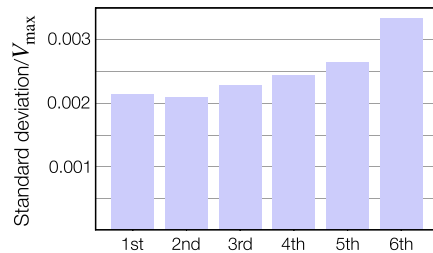

FIG. 8. rms deviation of the individual bands for the NN used to produce Figs. 1–3 and 6.

Xeon Gold 6130 with 16 cores, one epoch takes about 100 s.

APPENDIX F: ACCURACY FOR THE NETWORK

The NN used to produce the results in Figs. 1–3 and 6 has a rms band-structure deviation of approximately $0.0025V_{\text{max}}$. This figure is calculated on a grid with 821 grid points equally distributed within one-sixth of the unit cell (much finer than the training grid which had only 79 grid points). As we note above, the rms deviation will be slightly different for different training runs with equal training hyperparameters (but different initialization of the NN trainable parameters). Nevertheless, it remains of a few ‰ of the overall scale of the band structure (here V_{max}) for a wide range of training hyperparameters in all case studies investigated in this work.

We empirically observe a slight trend for increasing rms deviation for higher bands; cf. Fig. 8. We attribute this to a higher sensitivity of higher-energy bands to the fine details of the potential. We check that this effect can be compensated by increasing the number of training samples.

If one wants to compare these deviations to the band gaps (as a natural scale), we can, e.g., obtain the sample average of the minimal band gaps for disconnected bands, in which case, the deviations represent 2.5% of the band-gap value obtained in that way. On the other hand, the sample average of the **k**-averaged (not minimal) band gaps for disconnected bands is slightly larger, yielding a relative deviation of 2.1%.

APPENDIX G: PERFORMANCE GAIN

1. Speed advantage in predicting band structures

One of several advantages of using a neural network for predicting symmetries and band structures is the dramatically increased speed of calculations vs direct evaluations.

The performance gain offered by the neural network depends on the algorithm which it replaces, as well as on the number of points in reciprocal space, the structure of the neural network, and the number of bands which one is interested in.

For the results of this paper, the neural network should predict the same band structure one would obtain by using

the Schrödinger equation on a periodic potential with 100×100 grid points. In the absence of a trained neural network, this task would be performed for one **k** point in the reciprocal space by calculating the eigenvalues of a sparse $10\,000 \times 10\,000$ matrix, where the diagonal elements correspond to values of the potential on the grid points.

For the numerical calculation, we use for diagonalization scipy.sparse.linalg.eighs, which uses the "implicitly restarted Lanczos method" to find the eigenvalues and eigenvectors for the first six bands in our case. In our comparison, we count only the time needed to calculate the eigenvalues for the case of the numerical method (which works in favor of the numerical method). As reference hardware, both for the direct numerical calculation and the neural network, an i5-6267U (two cores, four threads, 2.9 Ghz) is used, which is as a typical mobile CPU. On this hardware, the diagonalization takes about 2–3 s for three points and 80 s for the 79 point, which we also use for training. On the other hand, the same task takes for the neural network 0.067 s for 79 points (and 0.23 s for a much finer grid of 821 points).

This analysis shows that the neural network performs much

faster than the direct Lanczos-based diagonalization of the Schrödinger equation, even for very few points. The advantage of the NN grows with the number of points: Note that the calculation time in the case of the neural network can be split into the calculation of the coefficients for the tight-binding model and the subsequent calculation of its band structure (by diagonalization of a small matrix), where the former step is independent of the number of **k** points. Since the creation of the tight-binding model is written in PYTHON, one could accomplish further speed-up for the neural network.

2. Overall performance gain

As in all neural-network applications, there are two scenarios to evaluate the cost benefit and overall performance gain of this approach.

(i) The goal is to deploy the network for obtaining speed-up on whatever hardware is available (including, e.g., the type of cluster used for training). In that case, the training effort needs to be accounted for. Breakeven will be reached when the network has been used to accelerate band-structure evaluations on a number of potentials that is at least larger than the initial number of training samples. For our approach, this is easily the case for the optimization of band structures (as well as for large-scale statistical exploration and random discovery).

(ii) The cost-benefit analysis turns out to be even more advantageous when the explicit goal has been to deploy the network on modest computing hardware (e.g., laptops operated by the end users). In that

case, the cost of generating the training samples and performing the training (on a cluster) need not be taken into account, since that hardware by definition would not have been available to the end user.

APPENDIX H: SYMMETRY FINGERPRINTS

Here, we give more details regarding the symmetry fingerprints used to identify EBRs and topological bands in the main text. Equivalent concepts are also presented in Refs. [7,39]. The symmetry fingerprint of an isolated set of bandsgroups in a single $\mathbb{N}^n$ array all the information about the symmetry of the Bloch waves at the maximal **k** points.

At each maximal **k** point, we define the occupation n_ξ as the number of (degenerate) orbitals belonging to each irrep ξ. The occupation numbers n_ξ are subject to linear constraints known as compatibility relations [7,63]. They reduce the number of independent occupations to

$$n = \text{number of irreps} - \text{number of linear constraints.}$$

The simplest compatibility relation is that the number of bands is the same at all maximal **k** points. Another important example of a compatibility relation is realized in crystals with mirror symmetry. For each high-symmetry line that is invariant under a mirror symmetry of the crystal and connects two maximal **k** points, a compatibility relation fixesthe numbers of states with a given parity to be equal at the two maximal **k** points. Such compatibility relations derived from mirror symmetry allow us to predict connections between bands that lie ona high-symmetry line, based only on the spectrum and irreps at the maximal **k** points [6,7].

For the p_6 group, the maximal **k** points are the high-symmetry points Γ, K, and M and the respective proper groups are the rotational groups $\mathcal{C}_n$, with $n = 6$ for Γ, $n = 3$ for K, and $n = 2$ for M. In this case, the time-reversal-symmetric irreps are identified by the absolute value of the quasiangular momentum $|m| \le n/2$. To avoid confusion, here and in the main text we use $\xi = |m| = 0, 1$ to label the irreps of $\mathcal{C}_3$, while using the atomic-physics-inspired labels $\xi = s$, p, d, and f in place of $|m| = 0, 1, 2, 3$ for the irreps of $\mathcal{C}_6$, and the labels $\xi = +$ and $\xi = -$ in place of $m = 0$ and $m = 1$ for the irreps of $\mathcal{C}_2$ (the normal modes are either odd or even). We are, thus, left with eight occupation numbers: $n^{(s)}$, n_p, n_d, and n_f for the Γ point, n_0 and n_1 for the K point, and n_+, n_- for the M point. However, taking into account that the overall number of bands should be the same at all high-symmetrypoints, we find the compatibility relations,

$$n_- = n_s + 2(n_p + n_d) + n_f - n_+,$$
$$n_1 = [n_s + 2(n_p + n_d) + n_f - n_0]/2. \qquad \text{(H1)}$$

Thus, all the information regarding the symmetry labels is grouped in the six-dimensional array $(n_s, n_p, n_d, n_f, n_0, n_+)$.

APPENDIX I: DETAILS OF THE TOPOLOGICAL EXPLORATION

For the topological exploration, the Fourier coefficients $A(\mathbf{k})$ of the random potentials are extracted from the distribution in Eq. (D3) with $\alpha = 1/2$ and $C = 2$. For each potential, we calculate the symmetry fingerprint for the first two sets of bands and use it to identify the set with one of the eight EBRs, a composite BR (cBR), or a topological set of bands if the first two options are excluded. All topological sets of bands discovered correspond to only two different symmetry fingerprints. To determine the fingerprints, two ingredients are required [both represented in the symmetry tables in Fig 3(a) of the main text]: (i) the irreps for each band at each high-symmetry point (this information is provided directly from the NN, (ii) the connectivity of the bands, i.e., which pairs of bands are connected somewhere in the BZ (this information is not provided directly by the NN but has to be inferred by looking at π defects in the Berry flux on a fine $\mathbf{k}$ grid, more on this later, or by using our conjectured compatibility relation).

To calculate the Berry flux, we divide the BZ in small rectangular plaquettes $\mathbf{j}$. The Berry flux $\Phi_{\mathbf{j}}$ is just the Berry phase acquired while encircling each plaquette. It can be easily calculated numerically using the formula

$$\Phi_{\mathbf{j}} = i \ln \left(\frac{o^{(l)}_{\mathbf{j},\mathbf{e}_x} o^{(l)}_{\mathbf{j}+\mathbf{e}_x,\mathbf{e}_y} o^{(l)}_{\mathbf{j}+\mathbf{e}_x+\mathbf{e}_y,-\mathbf{e}_x} o^{(l)}_{\mathbf{j}+\mathbf{e}_y,-\mathbf{e}_y}}{|o^{(l)}_{\mathbf{j},\mathbf{e}_x} o^{(l)}_{\mathbf{j}+\mathbf{e}_x,\mathbf{e}_y} o^{(l)}_{\mathbf{j}+\mathbf{e}_x+\mathbf{e}_y,-\mathbf{e}_x} o^{(l)}_{\mathbf{j}+\mathbf{e}_y,-\mathbf{e}_y}|} \right), \quad \text{(I1)}$$

where $o^{(l)}_{\mathbf{j},\Delta\mathbf{j}} = \langle \mathbf{k}^{(l)}_{\mathbf{j}} | \mathbf{k}^{(l)}_{\mathbf{j}+\Delta\mathbf{j}} \rangle$, $\mathbf{e}_x = (1,0)$, and $\mathbf{e}_y = (0,1)$. This method allows us to find connections efficiently because $|\Phi_{\mathbf{j}}| \approx \pi$ whenever a Dirac cone is inside the plaquette and the plaquette is so small that the band dispersion can be approximated as linear (this property can be proven by approximating the Hamiltonian with a Dirac Hamiltonian). The requirement that the band dispersion should be linear inside a plaquette containing a Dirac cone determines how fine the grid should be to obtain reliable results. Whether this requirement is satisfied for a given grid depends on the specific potential. For this reason, even though a coarse grid would be already enough to obtain reliable results for the majority of the potentials, a very fine grid is necessary to get high-accuracy statistical results (much finer than the grid used for training). Moreover, even for a fine grid the method might fail in a handful of statistically irrelevant cases. For this reason, it would be difficult to obtain the results shown in the main text without relying on the speed of our NN.
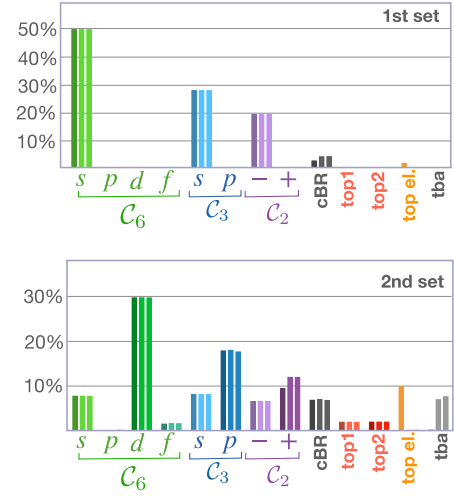

FIG. 9. (Quasi-) BR distribution calculated using three different methods to determine the bands connectivity. The methods are (from left to right for each entry) (i) standard method of topological quantum chemistry which neglects connections away from high-symmetry points, (ii) taking into account the connections away from high-symmetry points which are predicted by the parity compatibility relation, and (iii) looking for π defects in the Berry flux. This method is able to detect all connections (if a fine enough grid is used). The column top1 and top2 refer to the two fragile topological phases discovered by our NN; cf. Fig. 3 of the main text. The standard method would predict other types of topological bands (top. el.). These sets of bands turn out not to be isolated once the connections away from the high-symmetry points are taken into account with the other two methods. All results are calculated using the first six bands. The set of bands that is connected to the seventh band can not be assigned to any (quasi-) BR based on the available information and is grouped in the last entry [tba (to be assigned)].

A much faster method to calculate the connectivity is to infer it from the compatibility relations, including also the additional parity compatibility relation. We compare the results obtained with this method to the one obtained with the Berry flux method using a fine grid containing approximately 10 000 plaquettes; cf. Fig. 9. We find a disagreement in only less than 0.1% (1%) of the cases for
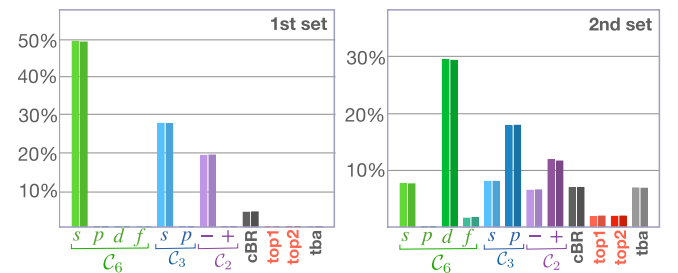

FIG. 10. Comparison of the (quasi-) BR distribution calculated using the NN and the Schrödinger equation. The probability bars for a given (quasi-) BR but different methods are shown side by side. The results obtained using the Schrödinger equation are plotted in slightly darker colors.
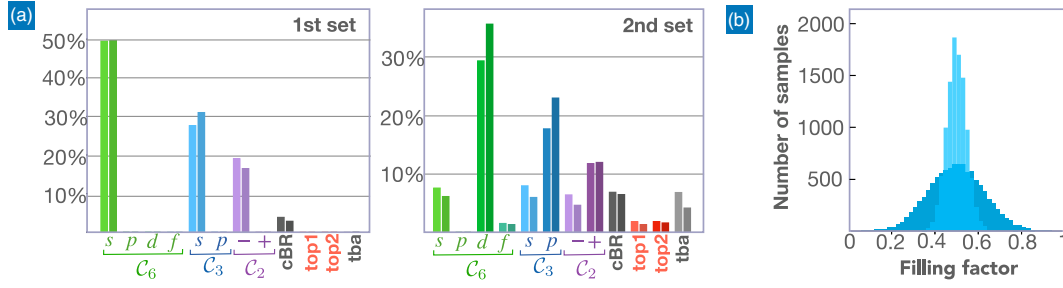
FIG. 11.　(a) Comparison of the (quasi-) BR distributions for two different distributions of potentials. The probability bars for a given (quasi-) BR but different distributions are shown side by side. (b) Number of samples as a function of the filling factor [portion of the unit cell where $V(\mathbf{x}) = V_{\max}$] for the two distributions of interest. In all panels, the results referring to the training distribution are plotted in slightly darker colors.

the first (second) group of bands. This indicates that the compatibility relations (when also the additional parity compatibility relation is taken into account) are sufficient to identify connected bands for the overwhelming majority of the samples.

The speed and reliability of the compatibility method offer the possibility to validate the results obtained using the NN with results obtained directly solving the Schrödinger equation; cf. Fig. 10.

In order to give an idea how the distribution of (quasi-) BRs depends on the underlying potential distribution, we calculate the statistics also for the training distribution. A comparison between the two statistics is shown in Fig. 11.

APPENDIX J: SUPPLEMENTAL INFORMATION FOR THE 3D NONSYMMORPHIC CASE STUDY

1. Symmetry-enhanced TB model

For our two symmorphic examples, we construct symmetry-enhanced TB models that are defined on the Bravais lattice of the underlying space group. As a consequence, the orbitals are representations of its point group. This is not possible for the space group $p4_222$ (or any other nonsymmorphic group), as it is well known that TB models with nonsymmorphic space groups should be defined on a lattice with a basis. Instead, we generalize our approach by considering as a lattice a single so-called crystallographic orbit (the set of sites obtained by applying all space-group transformations to a single site). In this case, the number of sublattices is the so-called multiplicity of the underlying Wyckoff position, while the orbitals are representations of the corresponding site-symmetry group (the group of transformations that leaves a site invariant), which is (isomorphic to) a subgroup of the point group to which the space group belongs. For nonsymmorphic groups, the multiplicity is larger than 1 for all Wyckoff positions leading to a lattice with a basis. (This in contrast to symmorphic groups where there is at least one position with multiplicity 1.)

A natural generalization of the approach we adopt so far is to pick (one of) the Wyckoff position(s) with smallest

multiplicity. For the space group $p4_222$, we choose the Wyckoff position $2a$ (here, the number 2 indicates the multiplicity). We construct our TB model starting from a set of orbitals $\{|n, m\rangle\}$ localized about the origin (n is the principal quantum number and m labels the irrep of the site-symmetry group). Here, the site-symmetry group is the point group $\mathcal{D}_2$ (twofold rotations about the x, y, and z axes). This point group has four inequivalent irreps, $m = A, B_1, B_2, B_3$ with the atomic s, p_z, p_y, and p_x orbitals, respectively, being representative states transforming under these irreps. From each orbital localized about the origin, we construct another orbital localized about the position $\mathbf{a}_3/2$ by applying a screw rotation. Finally, we obtain a basis of Wannier states by applying all lattice translations,

$$|n, \mathbf{j}, s, m\rangle = (\mathbb{1}, \mathbf{a_j})|n, s, m\rangle.$$

Here, $\mathbf{j}$ indicates the unit cell and s the sublattice. Thus, $\mathbf{a_j} = j_1\mathbf{a}_1 + j_2\mathbf{a}_2 + j_3\mathbf{a}_3$, $|n, 0, m\rangle = |n, m\rangle$, and $|n, 1, m\rangle = (R_{\pi/2,z}, \mathbf{a}_3/2)|n, m\rangle$. We note that according to our definition, the states $|n, \mathbf{j}, 1, B_{2/3}\rangle$ are labeled according to the irreps of the state $|0, m\rangle$, not their own. For example, the state $|\mathbf{j}, 0, B_2\rangle$ transforms under symmetry as a p_y orbital, while the state $|\mathbf{j}, 1, B_2\rangle$ being rotated by 90° will rather transform as a p_x orbital.

With the above definitions in hand, we find

$$H_{n,m;n',m'} \equiv \langle n', m'|\hat{H}|n, m\rangle = \langle n', \mathbf{j}, s, m'|\hat{H}|n', \mathbf{j}, s, m'\rangle$$
$$= \omega_{n,m}\delta_{n,n'}\delta_{mm'}.$$

Thus, we have $\tilde{n}_A^{(TB)} + \tilde{n}_{B_1}^{(TB)} + \tilde{n}_{B_2}^{(TB)} + \tilde{n}_{B_3}^{(TB)}$ independent on-site energies where $\tilde{n}_{j=A,B_1,B_2,B_3}^{(TB)}$ are the number of orbitals of each type. Here we assume without loss of generality that the principal basis diagonalizes the Hamiltonian projected onto the Wannier states localized about the origin. Next, we define the matrix containing the hopping amplitudes for nearest-neighbor vertical transitions within the same unit cell

$$J^{(z)}_{n'm';nm} \equiv \langle n', 1, m'|\hat{H}|n, 0, m\rangle. \qquad (J1)$$

We note that $J^{(z)}_{nm;n'm'}$ can be chosen real because the orbitals $|n, s, m\rangle$ are invariant under time-reversal symmetry. Moreover, the symmetry under twofold rotations about the z axis gives rise to a selection rule forbidding vertical hopping transitions between states of opposite parity. Thus, for example, an excitation can hop vertically from an A orbital to another A orbital or a B_1 orbital (as both A and B_1 orbitals are even under twofold rotations about the z axis) but not to a B_2 or a B_3 orbital. Thus, the matrix $J^{(z)}_{n'm';nm}$ is block diagonal with one block spanned by the A and B_1 orbitals and the other Block spanned by the B_2 and B_3 orbitals. By applying a screw rotation followed by a 180 rotation about the x axis, we find

$$J^{(z)}_{n'm';nm} = m'_z m'_x m_y J^{(z)}_{nm;n'm'}.$$

Here, $m_{x/y/z}$ is the parity of the orbitals $|n, m\rangle$ under $x/y/z$ twofold rotations. Thus, $m_{z/y/x} = 1$ for $m = A, B_{1/2/3}$ and -1 otherwise. We can conclude that the hopping matrix $J^{(z)}_{n'm';nm}$ has $(\tilde{n}^{(TB)}_A + \tilde{n}^{(TB)}_{B_1} + 1)(\tilde{n}^{(TB)}_A + \tilde{n}^{(TB)}_{B_1})/2 + (\tilde{n}^{(TB)}_{B_2} + \tilde{n}^{(TB)}_{B_3} + 1)(\tilde{n}^{(TB)}_{B_2} + \tilde{n}^{(TB)}_{B_3})/2$ independent matrix elements. Applying the symmetry under twofold transitions about the y axis, one finds the hopping amplitudes for vertical NN transitions between sites in different unit cells

$$\langle n, \boldsymbol{a}_z, 0, m|\hat{H}|n', 1, m'\rangle = J^{(z)}_{n'm';nm} m'_x m_y.$$

Thus, all nearest-neighbor vertical hopping transitions are encoded in the matrix $J^{(z)}_{n'm';nm}$.

Next, we define the horizontal hopping rates on sublattice $s = 0$,

$$J^{(x/y,0)}_{n'm';nm} \equiv \langle n', \mathbf{e}_{x/y}, 0, m'|\hat{H}|n, 0, m\rangle. \qquad (J2)$$

Analogous to what we discuss above for the vertical hopping rates, the matrix $J^{(x,0)}_{n'm';nm}$ $(J^{(y,0)}_{n'm';nm})$ is block diagonal, this time with one block spanned by the A and the B_3 (B_2) orbitals and the other by the B_1 and B_2 (B_3) orbitals. As their vertical counterparts, the blocks of $J^{(x/y)}_{n'm';nm}$ are constrained by the symmetry

$$J^{(x/y)}_{n'm;nm} = J^{(x/y)}_{nm;n'm},$$
$$J^{(x/y)}_{n'm';nm} = -J^{(x/y)}_{nm;n'm'} \quad \text{for } m \neq m'.$$

Thus, the hopping matrices $J^{(x/y)}_{n'm';nm}$ are defined by $(\tilde{n}^{(TB)}_A + \tilde{n}^{(TB)}_{B_{3/2}} + 1)(\tilde{n}^{(TB)}_A + \tilde{n}^{(TB)}_{B_{3/2}})/2 + (\tilde{n}^{(TB)}_{B_1} + \tilde{n}^{(TB)}_{B_{2/3}} + 1)(\tilde{n}^{(TB)}_{B_1} + \tilde{n}^{(TB)}_{B_{2/3}})/2$ independent parameters. By using a screw

TABLE I. Character table of the irreps of the little group $G_{\mathbf{k}}$ for $\mathbf{k} = \Gamma, M$. For the first column, each row corresponds to an EBR as identified by the orbital type (i.e., the corresponding irreps of $\mathcal{D}_2$) on sublattice $s = 0$. For the remaining columns, each row corresponds to a different irrep of $G_{\mathbf{k}}$. For the cases in which the same EBR gives rise to two one-dimensional irreps of $G_{\mathbf{k}}$, the wave function is either a symmetric (indicated with a + sign in the table) or an antisymmetric (− in the table) superposition of sublattice-polarized plane waves.

Orb.		$\mathbb{1}$	$(R_{\pi,z}, \boldsymbol{a}_3)$	$R_{\pi,y}$	$(R_{\pi,x+y}, -\boldsymbol{a}_3/2)$	$(R_{\pi/2,z}, \boldsymbol{a}_3/2)$
A	+	1	1	1	1	1
	−	1	1	1	−1	−1
B_1	+	1	1	−1	−1	1
	−	1	1	−1	1	−1
B_2 B_3		2	−2	0	0	0

rotation, we can also obtain the horizontal hopping matrices $J^{(x/y,1)}_{n'm';nm}$ on sublattice $s = 1$,

$$J^{(y,1)}_{n'm';nm} \equiv \langle n', \mathbf{e}_y, 1, m'|\hat{H}|n, 1, m\rangle = J^{(x,0)}_{n'm';nm}, \quad (J3)$$

$$J^{(x,1)}_{n'm';nm} \equiv \langle n', \mathbf{e}_x, 1, m'|\hat{H}|n, 1, m\rangle = J^{(y,0)}_{nm;n'm'}. \quad (J4)$$

After setting the appropriate constraints to the hopping matrices, our symmetry-enhanced TB Hamiltonian will have automatically a block diagonal form at the maximal $\mathbf{k}$ points, with each block corresponding to an irrep of the so-called little group $G_{\mathbf{k}}$ (the group of symmetries g that leave $\mathbf{k}$ invariant modolus a vector of the reciprocal lattice). The same property, obviously, will hold for the Hamiltonian of the 3D Schrödinger equation. A crucial step in implementing our method is to require (via the appropriate cost function contribution) that matching blocks (corresponding to the same irrep) of both Hamiltonians have the same spectrum. Thus, we need to identify the underlying irrep for each such block. For this purpose, we construct the characters Tables I–IV. The first column in every table is the irrep (orbital type) of an orbital at the origin (A, B_1, B_2, or B_3). This orbital is then used to construct the corresponding EBR. At each maximal $\mathbf{k}$ point ($\mathbf{k} = \Gamma, X, M, Z, R$, or A),

TABLE II. Characters table of the representations of the little group $G_{\mathbf{k}}$ for $\mathbf{k} = Z, A$. See caption of Table I.

Orb.		$\mathbb{1}$	$(R_{\pi,z}, \boldsymbol{a}_3)$	$R_{\pi,y}$	$(R_{\pi,x+y}, -\boldsymbol{a}_3/2)$	$(R_{\pi/2,z}, \boldsymbol{a}_3/2)$
A B_1		2	−2	0	0	0
B_2	+	1	1	1	1	1
	−	1	1	1	−1	−1
B_3	+	1	1	−1	−1	1
	−	1	1	−1	1	−1

TABLE III. Characters table of the representations of the little group $G_\mathbf{k}$ for $\mathbf{k} = X$. Each row corresponds to an EBR as identified by the orbital type on sublattice $s = 0$. For the cases in which the same EBR gives rise to two one-dimensional irreps of $G_\mathbf{k}$, the Bloch waves are sublattice polarized. The sublattice (0 or 1) is indicated in the first column. For the maximal $\mathbf{k}$ point X and R, the little group can be decomposed as a sum of lattice translations and the point group $\mathcal{D}_2$. This property allows us to identify each irrep of $G_\mathbf{k}$ with an irrep of $\mathcal{D}_2$, indicated in the last column.

Orb.		$\mathbb{1}$	$R_{\pi,z}$	$R_{\pi,y}$	$R_{\pi,x}$	Irrep
A		1	1	1	1	A
B_1		1	1	−1	−1	B_1
B_2	0	1	−1	1	−1	B_2
	1	1	−1	−1	1	B_3
B_3	0	1	−1	−1	1	B_3
	1	1	−1	1	−1	B_2

TABLE IV. Characters table of the representations of the little group $G_\mathbf{k}$ for $\mathbf{k} = R$. See caption of Table III.

Orb.		$\mathbb{1}$	$R_{\pi,z}$	$R_{\pi,y}$	$R_{\pi,x}$	Irrep
A	0	1	1	1	1	A
	1	1	1	−1	−1	B_1
B_1	0	1	1	−1	−1	B_1
	1	1	1	1	1	A
B_2		1	−1	1	−1	B_2
B_3		1	−1	−1	1	B_3

such an EBR will give rise to a representation of the little group $G_\mathbf{k}$ spanned by the sublattice plane waves

$$|n, \mathbf{k}, s, m\rangle = \sum_\mathbf{j} e^{i\mathbf{k}\cdot\mathbf{a}_\mathbf{j}} |n, \mathbf{j}, s, m\rangle, \quad s = 0, 1.$$

This representation is either a two-dimensional irrep of $G_\mathbf{k}$ or can be decomposed in two one-dimensional irreps. In the latter case, we also give the underlying superposition of sublattice plane waves for each irrep (also in the first column). As usual, it is possible to uniquely identify any irrep of the little group by listing the characters (traces) $\chi[\rho(g)]$ of the matrix representative $\rho(g)$ for an appropriate finite set of transformations g. This set is finite because it is enough to consider a single transformation for each infinite set of transformations that differ by a lattice translation (as the traces for different elements will differ only by a phase that does not depend on the irrep). In addition, it is enough to consider a single transformation for each conjugacy class (because the trace is the same for all elements of the same conjugacy class).

We use the Tables I–IV for three purposes: (i) We identify the subset of Wannier orbitals (or superposition thereof) that span the same block of the symmetry-enhanced TB model at a specific high-symmetry point.

Layer	Type	Dim.	Filt. size	Ker. size	Strides	Activ.
0		20^3				
1	Conv3D	20^3	128	2	N/A	ReLu
2	Conv3D	20^3	128	2	N/A	ReLu
3	Pool3D	10^3	128	2	2	N/A
4	Conv3D	10^3	256	2	N/A	ReLu
5	Conv3D	10^3	256	2	N/A	ReLu
6	Conv3D	10^3	256	2	N/A	ReLu
7	Pool3D	5^3	256	2	2	N/A
8	Conv3D	5^3	256	2	N/A	ReLu
9	Conv3D	5^3	256	2	N/A	ReLu
10	Conv3D	5^3	256	2	N/A	ReLu
11	Pool3D	3^3	256	2	2	N/A
12	Dense	3456	N/A	N/A	N/A	ReLu
13	Dense	3456	N/A	N/A	N/A	ReLu
14	Dense	234	N/A	N/A	N/A	Lin

FIG. 12. Sketch of the NN layout for the 3D nonsymmorphic case study.

For example, according to Table I, all B_2 and B_3 orbitals give rise to the same irrep at the Γ and M point and, thus, will span a single block there. (ii) We assign each exact solution of the 3D Schrödinger equation (at a maximal $\mathbf{k}$ point) to the correct irreps by checking its behavior under the transformations listed in the relevant table. (iii) We calculate the symmetry fingerprints for the EBRs generated by each orbital type. These are then used to calculate lower bounds for $\tilde{n}_l^{(\mathrm{TB})}$; see discussion in Appendix C. As usual, the lower bound depends on the number of target bands. For four bands, the lower bounds are $\tilde{n}_A^{(\mathrm{TB})} = 3$, $\tilde{n}_{B_1}^{(\mathrm{TB})} = 2$, $\tilde{n}_{B_2}^{(\mathrm{TB})} = 1$, $\tilde{n}_{B_3}^{(\mathrm{TB})} = 2$. Empirically, we find that it is advantageous to use a larger Hilbert space with $\tilde{n}_A^{(\mathrm{TB})} = 5$, $\tilde{n}_{B_1}^{(\mathrm{TB})} = 4$, $\tilde{n}_{B_2}^{(\mathrm{TB})} = 3$, $\tilde{n}_{B_3}^{(\mathrm{TB})} = 4$.

2. NN layout

The NN layout for the 3D case study is sketched in Fig. 12. In this case, we use conv3D layers with ReLu activation, kernel size (2,2,2) using the option padding = same. With the aim of progressively reducing the image size, the convolutional layers are alternated with max_pooling3D layers with pool size (2,2,2) and stride = 2. As usual, the convolutional and pooling layers are followed by a series of dense layers. The dropout (0.15) is applied between each pair of subsequent dense layers (not shown in the sketch).

APPENDIX K: SUPPLEMENTAL INFORMATION FOR THE TWO-COMPONENT TOPOLOGICAL METAMATERIAL CASE STUDY

1. Details of the microscopic model

As we discuss in the main text, we model one spin sector of our two-component metamaterial using "one-half" of the Bernevig-Hughes-Zhang (BHZ) model with a site-dependent mass $M_\mathbf{j}$,

$$H = \sum_{\mathbf{j}} (M_{\mathbf{j}} + 2J)\hat{\sigma}_z |\mathbf{j}\rangle\langle\mathbf{j}|$$
$$- \frac{J}{2} \sum_{\langle \mathbf{l}, \mathbf{j}\rangle} (\hat{\sigma}_z - i\hat{\sigma}_{d(\mathbf{l},\mathbf{j})}) |\mathbf{l}\rangle\langle\mathbf{j}| + \text{H.c.} \qquad \text{(K1)}$$

Here, $M_{\mathbf{j}} = \pm m$ where the multi-index $\mathbf{j} = (j_x, j_y)$ parametrizes the sites of an $N \times N$ square grid, $\langle \mathbf{l}, \mathbf{j}\rangle$ indicates the sum over nearest neighbors, $d(\mathbf{l}, \mathbf{j}) = x$, y is the hopping direction, and J is the hopping rate. In our simulation, we choose $N = 70$ and $m/J = 10/N \approx 0.14$ with $J > 0$. For this sign of J, any positive mass $M > 0$ corresponds to the trivial phase, $\mathcal{C} = 0$. On the other hand, for the topological region, the mass $M = -m$ falls in the interval $-2 < M/J < 0$ where $\mathcal{C} = 1$ [45]. We note that the Hamiltonian (K1) applies not only to a scenario in which the spin is conserved but also to a more general case in which the spin is not necessarily conserved but the out-of-plane mirror transformation $\hat{M}_z$ is still a symmetry. Thus, also our results apply to this more general scenario. In this framework, the "half" HBZ Hamiltonian as well as the corresponding Chern numbers refer to a mirror symmetry sector (one of the two possible eigenvalues of the mirror symmetry, $M_z = i$ or $M_z = -i$).

Taking the large-wavelength limit of the half BHZ Hamiltonian Eq. (K1), one arrives at the Dirac Hamiltonian Eq. (3) with $v = Ja/N$. It is of fundamental interest to investigate the large-wavelength limit because it is of higher generality going beyond our specific microscopic model. In particular, the physics becomes independent of N (once v/m is held fixed). Our particle-hole-conserving Dirac Hamiltonian would also emerge as a large-wavelength description of a more general form of the BHZ Hamiltonian that does not have this symmetry (as in the original formulation of this model). More generally, it will describe any situation where the two-component materials differ in Chern number by one unit and their valence and conduction bands have minimal splitting at the Γ point where they support gapped Dirac cones, irrespective of the underlying microscopic lattice. To make sure that our results are not model specific and really apply to the most general setting described above, we check that they converge to the limit $N \to \infty$ comparing simulations with equal v/M but different N.

2. Symmetry-enhanced TB model

We construct our symmetry-enhanced TB model using orbitals localized about the Wyckoff position $1a$ (one of the two fourfold rotocenters). The orbitals will then be irreps of the point group $\mathcal{C}_4$. In this case, the relevant irreps are not time-reversal symmetric and, thus, are simply parametrized by the quasiangular momentum m, $m = 0, \pm 1, 2 \mod 4$, or s, $p_{\pm}$, and d. The particle-hole symmetry sets the additional constraint that pairs of orbitals with quasiangular momentum $m \mod 4$ and $1 - m \mod 4$, respectively, have

opposite energy. Thus, there are only two different types of orbital pairs (or equivalently, EBRs): For one EBR, the quasiangular momenta of the two particle-hole-connected orbitals are $m = 0$ and $m = -1$, for the other $m = 2$ and $m = 1$. The constraints described above lead to a symmetry-enhanced TB Hamiltonian in the form

$$\hat{H}_{\mathbf{k};n,m;n'm'}(\mathbf{k})/\hbar = \delta_{m,m}\delta_{n,n'}\omega_{n,m} + J_{n,m;n',m'} f_{m-m'}(\mathbf{k}).$$

Because of the $\mathcal{C}_4$ symmetry, we find

$$f_0(\mathbf{k}) = \cos(k_x a) + \cos(k_y a),$$
$$f_1(\mathbf{k}) = -f_{-1}^*(\mathbf{k}) = -i[\sin(k_x a) + e^{-i\pi/2}\sin(k_y a)],$$
$$f_2(\mathbf{k}) = \cos(k_x a) - \cos(k_y a),$$

where $J_{n,m;n',m'}$ are the hoppings in the rightward direction. Note that from the inversion symmetry, it follows that

$$J_{n,m;n',m'} = \pm J_{n',m';n,m}^*,$$

where the positive (negative) sign applies if $\Delta m = m - m'$ is even (odd). [This constraint ensures that the Hamiltonian is Hermitian.] Because of the particle-hole symmetry, we have the additional constraints

$$\omega_{n,m} = -\omega_{n,-1-m}, \qquad J_{n,m;n,m'} = -J_{n,-1-m;n',-1-m'}^*.$$

Thus, the independent coefficients can be chosen to be $\omega_{n,m}$ (real) for $m = 0, 2$, $J_{n,m;n',m'}$ (complex) for

$$\begin{array}{lll} m = 0, & m' = 0, & \text{for } n > n', \\ m = 2, & m' = 2, & \text{for } n > n', \\ m = 0, & m' = -1, & \text{for } n \geq n', \\ m = 1, & m' = 2, & \text{for } n \geq n', \\ m = 0, & m' = 2, & \\ m = 0, & m' = 1, & \end{array}$$

and $J_{n,m;n,m}$ (real) for $m = 0, 2$. The number of independent parameters is thus, $2(\tilde{n}_s^{(\text{TB})} + \tilde{n}_d^{(\text{TB})})(1 + \tilde{n}_s^{(\text{TB})} + \tilde{n}_d^{(\text{TB})})$ where $(\tilde{n}_d^{(\text{TB})})$ $\tilde{n}_s^{(\text{TB})}$ is the number of (d) s orbitals. For training on the eight central bands, we use $\tilde{n}_s^{(\text{TB})} = 9$ and $\tilde{n}_d^{(\text{TB})} = 8$.

In this case study, it is straightforward to match different blocks of the TB model Hamiltonian with the corresponding blocks of the Dirac equation by checking for the quasiangular momentum (parity) at the Γ and M points (X point). However, there remains an outstanding challenge: Since the two models have a different numbers of bands and we are trying to predict the band structure in the middle of the spectrum, it is still not obvious which energy level should correspond to which. For the global cost

function Eq. (2), this ambiguity is readily eliminated by the particle-hole symmetry: By symmetry, there is always an equal number of positive- and negative-energy states, and one can simply match the first positive bands of the symmetry-enhanced TB model to the corresponding levels of the Dirac equation. At a maximal **k** point, however, this symmetry does not apply separately to each block. Since the particle-hole symmetry maps onto each state of different symmetry, e.g., s and p_-, there is no guarantee that the number of positive- and negative-energy states of a given symmetry, e.g., p_-, are the same. In this case, the matching requires an additional assumption. We assume that a band touching between the higher-energy negative band and the lowest-energy positive band (leading to a band inversion) can occur only at the Γ point between an s and a p_- Bloch wave. In other words, we assume that for each irrep the number of positive and negative bands is the same with a single exception: Whenever the p_- Bloch wave is the lowest positive-energy state at the Γ point, we allow for one additional positive (negative) energy p_- state (s state). We have not proven this assumption, but we believe that the nearly perfect Chern number predictions by postselected NNs is a good indication of its validity.

In the main text, we compare the prediction of our NN to microscopic simulations for the band Chern number. This topological invariant can readily be calculated as the sum of the Berry fluxes [cf. Eq. (I1)] over all quasimomentum plaquettes of a fine grid (here a 62×62 grid) covering the whole BZ (Fukui *et al.* [46]). Alternatively, one might have considered the band-gap Chern number, this is the sum of the band Chern numbers of all bands below a certain band gap. We focus on the band Chern number mainly for two reasons discussed below. The first reason is of fundamental nature: The band Chern numbers of the central bands are well defined within the Dirac Hamiltonian large-wavelength description. On the other hand, the band-gap Chern number goes beyond the long-wavelength limit in that it involves also bands with large negative energy outside of the bandwidth where the large-wavelength limit is expected to apply. The second motivation to focus on the band Chern numbers is of practical nature: The band Chern numbers are easily accessible by calculating the band structure and eigenvectors in the central region of the spectrum using the Lanczos algorithm. On the other hand, the calculation of the band-gap Chern numbers would be numerically very expensive because of the large number of bands of our microscopic model (we have $2 \times N^2$ bands with $N = 70$ in our simulations).

While, for the reasons discussed above, we focused on the band Chern number so far, we wish to address at least briefly the question whether the band-gap Chern numbers predicted by our symmetry-enhanced TB model do coincide with those of the microscopic model. This is an important question because the band-gap Chern numbers are relevant for the bulk-boundary correspondence and,

thus, determine the behavior of the edge states in a system with boundary. We note that our assumption that a band touching between the lowest-energy positive band and the highest-energy negative band occurs only between an s and a p_- band at the Γ point (used while matching the bands of our symmetry-enhanced TB model to the bands of the microscopic model) represents a stringent constraint to the Chern number $\mathcal{C}_{-/+}$ for the band gap separating the positive bands from the negative bands (or, equivalently, to the sum of the Chern numbers over all negative-energy bands). If this assumption holds true, there remain only two scenarios: (i) $\mathcal{C}_{-/+} = 0$ if an s state is the lowest positive-energy state at the Γ point (no band inversion); (ii) $\mathcal{C}_{-/+} = 1$ if the lowest positive-energy eigenstate is a p_- Bloch wave. These are the same scenarios that occur in the special case of a homogeneous bulk. In the framework of this hypothesis, we can easily derive any band-gap Chern number (for the central bands) from the band Chern numbers of the first few positive bands and the symmetry label of the lowest-energy positive state at the Γ point, which are both very accurately predicted by our NNs.

We verify our hypothesis regarding the Chern number $\mathcal{C}_{-/+}$ for a statistically relevant number of samples using the bulk-boundary correspondence. This is achieved by deducing $\mathcal{C}_{-/+}$ from the slopes and position (upper or lower edge) of the edge states in the central band gap calculated from strip simulations. With the aim of reducing the computational effort, we consider a low-resolution grid ($N = 10$) and a narrow strip (ten-unit-cell width). While we do not expect our strip simulations for such a low resolution to have already converged to the continuous limit, we still expect them to reproduce at least qualitatively the band

Layer	Type	Dim.	Filt. size	Ker. size	Strides	Activ.
0		70^2				
1	Conv2D	70^2	128	3	N/A	ReLu
2	Conv2D	70^2	128	3	N/A	ReLu
3	Pool2D	35^2	128	2	2	N/A
4	Conv2D	35^2	256	3	N/A	ReLu
5	Conv2D	35^2	256	3	N/A	ReLu
6	Conv2D	35^2	256	3	N/A	ReLu
7	Pool2D	18^2	256	2	2	N/A
8	Conv2D	18^2	256	3	N/A	ReLu
9	Conv2D	18^2	256	3	N/A	ReLu
10	Conv2D	18^2	256	3	N/A	ReLu
11	Pool2D	18^2	256	2	2	N/A
12	Conv2D	18^2	256	3	N/A	ReLu
13	Conv2D	18^2	256	3	N/A	ReLu
14	Conv2D	18^2	256	3	N/A	ReLu
15	Pool2D	9^2	256	2	2	N/A
16	Dense	4096	N/A	N/A	N/A	ReLu
17	Dense	4096	N/A	N/A	N/A	ReLu
18	Dense	612	N/A	N/A	N/A	Lin

FIG. 13. Sketch of the NN layout for the two-component topological insulator case study.

structure and to correctly reproduce the robust topological features of interest here.

3. NN layout

The NN layout is sketched in Fig. 13. conv2D layers with ReLu activation, kernel size (2,2) using the option padding = same are alternated to max_pooling2D layers with pool size (2,2) and stride = 2. These layers are followed by three dense layers. The dropout (0.15) is applied between each pair of subsequent dense layers (not shown in the sketch).

APPENDIX L: DETAILS OF THE OPTIMIZATION METHOD

Implementing an optimization task starts by specifying some goal to achieve for the band structure: e.g., maximizing some band gap, or matching the predicted band structure as well as possible to a fixed given band structure. This goal has to be expressed in terms of a reward function (a function of the predicted band structure). Furthermore, the geometry has to be parametrized; in our case, we choose to describe a completely general geometry via its Fourier coefficients. Afterward, one can do gradient ascent on the reward, with respect to the geometry, exploiting the fact that backpropagation through the full network or tight-binding pipeline is possible.

The illustrative example we treat in the main text is a kind of "inverse problem," where we want to target a given band structure (calculated from some selected simple tight-binding model, in our case).

1. Creation of sharp potentials from smooth functions

A difficulty in our case is the constraints on the potential, which should assume only two discrete values, besides being $C6$ symmetric. To guarantee these properties, we optimize not directly the potential defined on a lattice, but instead the Fourier coefficients from which the potentials can be generated. This is possible since our approach for creating potentials, applying a sigmoid to a smooth scalar field, is differentiable for finite (nonzero) "temperatures" of the sigmoid. The step from Fourier coefficients to potential can be implemented directly in TENSORFLOW.

In this way, the potentials already obey the required symmetry. However, even though the sigmoid constrains the potential values between 0 and V_{max}, it cannot reliably enforce the potential to take only these values. In order to enforce the discreteness of the potential, we define a new term for the cost function which is proportional to $\sum_{\mathbf{x}} v(\mathbf{x})[v(\mathbf{x}) - 1]$, where $\mathbf{x}$ runs over the unit cell and $v(\mathbf{x}) = V(\mathbf{x})/V_{max}$ is the rescaled potential. The cost function for optimizing the input is hence the old cost function plus this new potential cost function. Again, it is important to choose the right weighting between both contributions. Without the potential cost function, the optimized potential might become less discrete, while a weight that is too large leads to stagnation of training. We observe that a good weight for the potential cost function is around 0.0001. In this way, the neural network focuses on making the potentials discrete only after it has reached already a low-loss value. As an optimizer, we again use ADAM with a learning rate of around 0.5 and otherwise default settings.

2. Our optimization procedure

Since it can happen that certain starting conditions may lead to stagnation, we recommend running several optimization trials with different random starting conditions. In our case, we choose a uniform distribution for all coefficients, with 0 as the mean. The suitable limits for the uniform distribution depends on the choice of inverse temperature β [cf. Eq. (D4)]. In the case of $\beta = 1$, we observe good results for random amplitudes in the interval $[-0.5, 0.5]$. Since the optimization can be performed on a GPU, we are able to optimize many samples in parallel. For example, we can optimize for one target band structure with 200 different starting conditions at the same time. On our hardware (mentioned in the training section) ten update steps on 200 samples take roughly 3 s. As usual, the number of update steps we recommend is 400–500. Out of the 200 trials for one band structure, we can then choose the best trials and check the results of the Schrödinger equation on the predicted potentials.

3. Details on the calculation of the loss map

For the optimization loss map (measuring the quadratic deviation between the optimized band structure and the target for different target band structures), we use parallelization to optimize for different band structures at the same time. By this, we can quickly produce one potential for each point in our grid. By repeating this procedure and updating the loss map such that the loss and the corresponding potential are replaced if the new version is better, we can reduce the noise of the loss map over time. To assess the quality, we always compare against the loss of the neural-network predictions on validation data (this is the square of the rms band-structure deviation measured in units of V_{max}). For the final optimization loss map, we distinguish between "relatively good" results, with an optimization loss about twice the network loss, and very good results, with an optimization loss below the network loss (here roughly 6×10^{-6}).

APPENDIX M: TIGHT-BINDING MODEL SUPPORTING FRAGILE TOPOLOGICAL PHASES

Here, we give more details on the TB model implemented using our optimization method. For a detailed

investigation of the model with only nearest-neighbor hopping, we refer to Refs. [55,64]. There are two main reasons why this model has attracted huge attention in the field of topological physics: (i) It is the simplest toy model that describes the band-folding p-d band-inversion transition that underlies the designs of a large number of topological photonics and topological phononics experiments demonstrating helical edge states; see, e.g., Refs. [56,57,59]. In this case, the edge states are localized about domain walls separating a "trivial" region (without band inversion, $J' < J$) and a "topological" region (with band inversion, $J' > J$). (ii) It is possible to describe its topology in terms of mirror winding numbers. Such mirror winding numbers are connected via a bulk-boundary correspondence to the edge states at the physical boundary of systems with a selected shape (decoration) [64]. We emphasize, however, that from the point of view of topological quantum chemistry (which focus on the orbitals in real space rather than the bulk-boundary correspondence), the band-folding phase transition is actually a hybridization transition and not a topological one; see Ref. [35]. For $J > J'$, the Wannier orbitals are formed by the hybridization of six atoms within one unit cell and are, thus, localized about the $\mathcal{C}_6$ rotocenters. For $J > J'$, on the other hand, the Wannier orbitals are formed by the dimerization of pairs of nearest-neighbor orbitals belonging to different unit cells and are, thus, localized about the $\mathcal{C}_2$ rotocenters. In both cases, there is a well-defined "atomic limit."

Here, we show that by adding a next-nearest-neighbor hopping modulated in amplitude, it is possible to induce a topological phase transition where a set of isolated bands does not admit an atomic limit. More precisely, we implement topological quasi-BRs with the same symmetry fingerprints as those we discover in our topological exploration; cf. Fig. 3(d) of the main text. Thus, this finding is yet another example of discovery that is stimulated by the rapid exploration allowed by our NN.

The Hamiltonian for our TB model reads

$$H_{\text{TB}}/\hbar = \sum_j \omega_0 \hat{a}_j^\dagger \hat{a}_j - \sum_{\langle j,j' \rangle} J_{j,j'}(\hat{a}_j^\dagger \hat{a}_{j'} + \hat{a}_{j'}^\dagger \hat{a}_j)$$
$$- \sum_{\langle\langle j,j' \rangle\rangle} L_{j,j'}(\hat{a}_j^\dagger \hat{a}_{j'} + \hat{a}_{j'}^\dagger \hat{a}_j),$$

where $\hat{a}_j$ is the annihilation operator on site j, $(\langle\langle j,j' \rangle\rangle)\langle j,j' \rangle$ indicates the sum over (next-) nearest neighbors, and $(L_{j,j'})J_{j,j'}$ are the (next-) nearest-neighbors hopping amplitudes. We choose $J_{j,j'} = J$ $(J_{j,j'} = J')$ for j and j' within the same unit cell (in different unit cells); cf. sketch in Fig. 6(a). Likewise, we choose $L_{j,j'} = L$ $(L_{j,j'} = L')$ for j and j' within the same unit cell (in different unit cells); cf. sketch in Fig. 6(c).

In order to analyze the (quasi-) BR as a function of the parameters for our TB model, we need to calculate the spectrum and symmetry at the high-symmetry points. It is possible to find simple close formulas because at each high-symmetry point there are at most two orbitals for each irrep. At the Γ point, the six sites combine to generate one orbital for each quasiangular momentum; the corresponding energies are

$$E_s/\hbar = \omega_0 + 3\bar{J} - \frac{\delta J}{2} - 6\bar{L} - \delta L,$$
$$E_p/\hbar = \omega_0 + \delta J + 3\bar{L} + \frac{\delta L}{2},$$
$$E_d/\hbar = \omega_0 - \delta J + 3\bar{L} + \frac{\delta L}{2},$$
$$E_f/\hbar = \omega_0 + 3\bar{J} - \frac{\delta J}{2} - 6\bar{L} - \delta L.$$

At the K points, we have two orbitals for each value of the quasiangular momentum $m = 0, \pm 1$,

$$E_{0/1,0}/\hbar = \omega_0 + 2\delta L \mp \sqrt{3\bar{J}^2 - 3\bar{J}\delta J + \frac{7}{4}\delta J^2},$$
$$E_{0/1,1}/\hbar = \omega_0 - \delta L \mp \sqrt{3\bar{J}^2 + \frac{1}{4}\delta J^2}. \tag{M1}$$

At the M points, where the proper group is $\mathcal{C}_{2v}$, it is useful to label the orbitals according to the parity under the two mirror symmetries. [We thus implicitly take into account that the space group of the TB model is actually the Wallpaper group $p6m$ and not just $p6$.] We then find

$$E_{+-}/\hbar = \omega_0 - 2\bar{J} - \bar{L} - \frac{3}{2}\delta L,$$
$$E_{--}/\hbar = \omega_0 + 2\bar{J} - \bar{L} - \frac{3}{2}\delta L,$$
$$E_{0/1,-+}/\hbar = \omega_0 + \frac{\bar{J}}{2} + \frac{\bar{L}}{2} - \frac{\delta J}{4} + \frac{3}{4}\delta L$$
$$\pm \left(\frac{9}{4}(\bar{J} - \bar{L})^2 - \frac{5}{4}(\bar{J} - \bar{L})(\delta J - \delta L) \right.$$
$$\left. + \frac{17}{16}(\delta J - \delta L)^2 \right)^{1/2},$$
$$E_{0/1,++}/\hbar = \omega_0 - \frac{\bar{J}}{2} + \frac{\bar{L}}{2} + \frac{\delta J}{4} + \frac{3}{4}\delta L$$
$$\pm \left(\frac{9}{4}(\bar{J} + \bar{L})^2 - \frac{5}{4}(\bar{J} + \bar{L})(\delta J + \delta L) \right.$$
$$\left. + \frac{17}{16}(\delta J + \delta L)^2 \right)^{1/2}. \tag{M2}$$

Here, we define

$$\bar{J} = (J + J')/2, \qquad \delta J = J' - J, \qquad \bar{L} = (L + L')/2,$$
$$\delta L = L' - L.$$

While it would be interesting to use the above analytical expressions to derive all possible topological and hybridization phases supported by the our TB model, below we focus on the parameter regime where the TB model describes a small perturbation about the graphene TB model, and, thus, $\bar{J} > 0$ is the largest hopping amplitude, $\bar{J} \gg \bar{L}, \delta J, \delta L$. In this framework, it is not necessary anymore to distinguish between states with different mirror symmetry at the M point. Instead, we adopt the same convention used in the main text of ordering the three odd (even) states under $\mathcal{C}_2$ rotations by increasing energy,

$$E_{0,-}/\hbar \equiv E_{+-}/\hbar = \omega_0 - 2\bar{J} - \bar{L} - \frac{3}{2}\delta L,$$

$$E_{0,+}/\hbar \equiv E_{0,++}/\hbar \approx \omega_0 - 2\bar{J} - \bar{L} + \frac{2}{3}\delta J - \frac{7}{6}\delta L,$$

$$E_{1,-}/\hbar \equiv E_{0,-+}/\hbar \approx \omega_0 - \bar{J} + 2\bar{L} + \frac{1}{6}\delta J + \frac{1}{3}\delta L,$$

$$E_{1,+}/\hbar \equiv E_{1,++}/\hbar \approx \omega_0 + \bar{J} + 2\bar{L} - \frac{1}{6}\delta J + \frac{1}{3}\delta L,$$

$$E_{2,-}/\hbar \equiv E_{1,-+}/\hbar \approx \omega_0 + 2\bar{J} - \bar{L} - \frac{2}{3}\delta J + \frac{7}{6}\delta L,$$

$$E_{2,+}/\hbar \equiv E_{--}/\hbar = \omega_0 + 2\bar{J} - \bar{L} - \frac{3}{2}\delta L. \tag{M3}$$

Here, we expand Eq. (M2) up to leading order. Likewise, expanding Eq. (M1), we find

$$E_{0/1,0}/\hbar = \omega_0 + 2\delta L \mp \sqrt{3}\left(\bar{J} - \frac{\delta J}{2}\right),$$

$$E_{0/1,1}/\hbar = \omega_0 - \delta L \mp \sqrt{3}\bar{J}. \tag{M4}$$

We first analyze the special case $\delta J = \delta L = 0$. In this case, the smallest possible unit cell contains only two sites, and we recover the band structure of graphene (with next-nearest-neighbor hopping) but folded into a smaller Brillouin zone (because in real space, we are using a larger unit cell containing six atoms). The two connected bands of graphene give rise to six connected bands after folding. Because of this underlying symmetry, we also expect two triply degenerate levels at the K point. The reason is that three quasimomenta of the larger BZ that are mapped onto each other by $\mathcal{C}_3$ rotations are projected onto the same quasimomentum of the smaller BZ. Indeed, from Eq. (M1) we see that $E_{0,0} = E_{0,1}$ ($E_{1,0} = E_{1,1}$) corresponding to a triple degeneracy because $E_{0,1}$ ($E_{1,1}$) is a doubly degenerate level. With similar arguments, one can prove that at the M point, two doubly degenerate levels are to be expected. Indeed, from Eq. (M2) we see that $E_{0,-} = E_{0,+}$ and $E_{2,-} = E_{2,+}$.

The first step toward constructing the topological quasi-BRs is to create an imbalance in the nearest-neighbor hopping by choosing $\delta J > 0$ (the external hopping is larger). In this scenario (discussed also above for

$\bar{L} = 0$), the folded graphene band structure is split into two sets of three connected bands each with dimerized Wannier orbitals localized about the $\mathcal{C}_2$ rotocenters; cf. Fig. 14 (central panel). At the Γ point, the p orbital is lifted above the d orbital. At the same time, at the M point, the lowest- (highest-) energy even band is lifted above the lowest- (highest-) energy odd band, $E_{0,+} > E_{0,-}$ ($E_{2,+} > E_{2,-}$). Likewise, at the K point, the lowest (highest) s-orbital wave is lifted above (lowered below) the lowest (highest) p level, $E_{0,0} > E_{0,1}$ ($E_{1,0} < E_{1,1}$).

Next, we tweak the band structure described above to obtain topological fragile bands. This is achieved by creating an imbalance between the next-nearest-neighbor hopping, $\delta J \neq 0$. From Eq. (M3), we see that at the M point, a positive δJ decreases the energy $E_{2,+}$ of the highest even orbital while increasing the energy of the odd orbital $E_{2,-}$. Meanwhile, at the K point, the energy $E_{1,1}$ of the highest p-Bloch wave is also decreased, while the energy $E_{1,0}$ of the corresponding s orbital is increased; cf. Eq. (M4). For sufficiently large δL, $\delta L > \delta J/\sqrt{12}$, the order of the highest two bands have been inverted compared to the situation where $\delta L = 0$ at both high-symmetry points K and M (at the M point, the band inversion occurs already for $\delta L > \delta J/4$). As a consequence, the highest three bands are split into a pair of topological bands, and an f orbital localized about the $\mathcal{C}_6$ rotocenter; cf. Figs. 14 (right panel) and 4(c) of the main text.

A similar analysis shows that for δL negative, $\delta L < -\delta J/\sqrt{12}$, the lowest three bands are split into an s orbital localized about the $\mathcal{C}_6$ rotocenters and a pair of topological bands; cf. Fig. 14 (left panel). A similar bands configuration is observed for the lowest three bands of the randomly generated potential 2 in Fig. 3 of the main text.
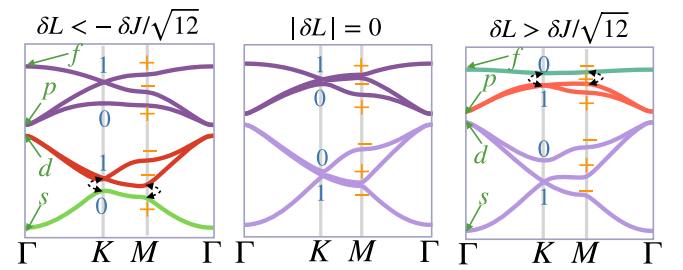

FIG. 14. Three different phases for the graphene TB model with next-nearest-neighbor hopping. For $\delta J > 0$, $|\delta L| = 0$ the folded graphene band structure is split into two kagomelike band structures (central panel). For $\delta L < -\delta J/\sqrt{12}$, after band-inversion transitions at both the M and K points (the orbitals that are exchanged are marked by dashed lines), the lower kagome bands split into a triangularlike s band and a pair of topological bands (left panel). For $\delta L > \delta J/\sqrt{12}$, similar band inversions lead to the splitting of the higher kagome bands into a topological pair of bands and triangularlike f band (right panel).

[1] J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade, *Photonic Crystals: Molding the Flow of Light*, 2nd ed. (Princeton University Press, Princeton, NJ, 2008).

[2] M. Maldovan, *Sound and Heat Revolutions in Phononics*, Nature (London) **503**, 209 (2013).

[3] N. R. Cooper, J. Dalibard, and I. B. Spielman, *Topological Bands for Ultracold Atoms*, Rev. Mod. Phys. **91**, 015005 (2019).

[4] M. Z. Hasan and C. L. Kane, *Colloquium: Topological Insulators*, Rev. Mod. Phys. **82**, 3045 (2010).

[5] T. Ozawa, H. M. Price, A. Amo, N. Goldman, M. Hafezi, L. Lu, M. C. Rechtsman, D. Schuster, J. Simon, O. Zilberberg, and I. Carusotto, *Topological Photonics*, Rev. Mod. Phys. **91**, 015006 (2019).

[6] B. Bradlyn, L. Elcoro, J. Cano, M. G. Vergniory, Z. Wang, C. Felser, M. I. Aroyo, and B. A. Bernevig, *Topological Quantum Chemistry*, Nature (London) **547**, 298 (2017).

[7] H. C. Po, A. Vishwanath, and H. Watanabe, *Symmetry-Based Indicators of Band Topology in the 230 Space Groups*, Nat. Commun. **8**, 1 (2017).

[8] M. G. Vergniory, L. Elcoro, C. Felser, N. Regnault, B. A. Bernevig, and Z. Wang, *A Complete Catalogue of High-Quality Topological Materials*, Nature (London) **566**, 480 (2019).

[9] G. N. Malheiros-Silveira and H. E. Hernandez-Figueroa, *Prediction of Dispersion Relation and PBGs in 2-D PCs by Using Artificial Neural Networks*, IEEE Photonics Technol. Lett. **24**, 1799 (2012).

[10] L. Pilozzi, F. A. Farrelly, G. Marcucci, and C. Conti, *Machine Learning Inverse Problem for Topological Photonics*, Commun. Phys. **1**, 57 (2018).

[11] A. da Silva Ferreira, G. N. Malheiros-Silveira, and H. E. Hernández-Figueroa, *Computing Optical Properties of Photonic Crystals by Using Multilayer Perceptron and Extreme Learning Machine*, J. Lightwave Technol. **36**, 4066 (2018).

[12] Z. Shi, E. Tsymbalov, M. Dao, S. Suresh, A. Shapeev, and J. Li, *Deep Elastic Strain Engineering of Bandgap through Machine Learning*, Proc. Natl. Acad. Sci. U.S.A. **116**, 4117 (2019).

[13] Z. Zhaochun, P. Ruiwu, and C. Nianyi, *Artificial Neural Network Prediction of the Band Gap and Melting Point of Binary and Ternary Compound Semiconductors*, Mater. Sci. Eng. **54**, 149 (1998).

[14] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, *Recent Advances and Applications of Machine Learning in Solid-State Materials Science*, npj Comput. Mater. **5**, 83 (2019).

[15] P. Zhang, H. Shen, and H. Zhai, *Machine Learning Topological Invariants with Neural Networks*, Phys. Rev. Lett. **120**, 066401 (2018).

[16] N. Sun, J. Yi, P. Zhang, H. Shen, and H. Zhai, *Deep Learning Topological Invariants of Band Insulators*, Phys. Rev. B **98**, 085402 (2018).

[17] N. Claussen, B. A. Bernevig, and N. Regnault, *Detection of Topological Materials with Machine Learning*, Phys. Rev. B **101**, 245117 (2020).

[18] A. K. Maier, C. Syben, B. Stimpel, T. Würfl, M. Hoffmann, F. Schebesch, W. Fu, L. Mill, L. Kling, and S. Christiansen, *Learning with Known Operators Reduces Maximum Error Bounds*, Nat. Mach. Intell. **1**, 373 (2019).

[19] A. Altland and M. R. Zirnbauer, *Nonstandard Symmetry Classes in Mesoscopic Normal-Superconducting Hybrid Structures*, Phys. Rev. B **55**, 1142 (1997).

[20] Y. E. Kraus, Y. Lahini, Z. Ringel, M. Verbin, and O. Zilberberg, *Topological States and Adiabatic Pumping in Quasicrystals*, Phys. Rev. Lett. **109**, 106402 (2012).

[21] M. C. Rechtsman, J. M. Zeuner, Y. Plotnik, Y. Lumer, D. Podolsky, F. Dreisow, S. Nolte, M. Segev, and A. Szameit, *Photonic Floquet Topological Insulators*, Nature (London) **496**, 196 (2013).

[22] T. Weiss, G. K. L. Wong, F. Biancalana, S. M. Barnett, X. M. Xi, and P. S. J. Russell, *Topological Zeeman Effect and Circular Birefringence in Twisted Photonic Crystal Fibers*, J. Opt. Soc. Am B **30**, 2921 (2013).

[23] M. A. Bandres, M. C. Rechtsman, and M. Segev, *Topological Photonic Quasicrystals: Fractal Topological Spectrum and Protected Transport*, Phys. Rev. X **6**, 011016 (2016).

[24] P. S. J. Russell, R. Beravat, and G. K. L. Wong, *Helically Twisted Photonic Crystal Fibres*, Phil. Trans. R. Soc. A **375**, 20150440 (2017).

[25] H. C. Po, H. Watanabe, and A. Vishwanath, *Fragile Topology and Wannier Obstructions*, Phys. Rev. Lett. **121**, 126402 (2018).

[26] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, *Quantum Spin Hall Insulator State in HgTe Quantum Wells*, Science **318**, 766 (2007).

[27] J. Carrasquilla and R. G. Melko, *Machine Learning Phases of Matter*, Nat. Phys. **13**, 431 (2017).

[28] D.-L. Deng, X. Li, and S. D. Sarma, *Machine Learning Topological States*, Phys. Rev. B **96**, 195145 (2017).

[29] Y. Zhang and E.-A. Kim, *Quantum Loop Topography for Machine Learning*, Phys. Rev. Lett. **118**, 216401 (2017).

[30] W. Lian, S.-T. Wang, S. Lu, Y. Huang, F. Wang, X. Yuan, W. Zhang, X. Ouyang, X. Wang, X. Huang, L. He, X. Chang, D.-L. Deng, and L. Duan, *Machine Learning Topological Phases with a Solid-State Quantum Simulator*, Phys. Rev. Lett. **122**, 210503 (2019).

[31] F. Schäfer and N. Lörch, *Vector Field Divergence of Predictive Model Output as Indication of Phase Transitions*, Phys. Rev. E **99**, 062107 (2019).

[32] L.-H. Wu and X. Hu, *Scheme for Achieving a Topological Photonic Crystal by Using Dielectric Material*, Phys. Rev. Lett. **114**, 223901 (2015).

[33] J. Cano, B. Bradlyn, Z. Wang, L. Elcoro, M. G. Vergniory, C. Felser, M. I. Aroyo, and B. A. Bernevig, *Building Blocks of Topological Quantum Chemistry: Elementary Band Representations*, Phys. Rev. B **97**, 035139 (2018).

[34] J. Zak, *Symmetry Specification of Bands in Solids*, Phys. Rev. Lett. **45**, 1025 (1980).

[35] M. B. de Paz, M. G. Vergniory, D. Bercioux, A. García-Etxarri, and B. Bradlyn, *Engineering Fragile Topology in Photonic Crystals: Topological Quantum Chemistry of Light*, Phys. Rev. Research **1**, 032005(R) (2019).

[36] L. J. P. van der Maaten and G. E. Hinton, *Visualizing Data Using t-SNE*, J. Mach. Learn. Res. **9**, 2579 (2008).

[37] G. van Miert and C. Morais Smith, *Dirac Cones beyond the Honeycomb Lattice: A Symmetry-Based Approach*, Phys. Rev. B **93**, 035401 (2016).

[38] C. Fang, M. J. Gilbert, and B. A. Bernevig, *Bulk Topological Invariants in Noninteracting Point Group Symmetric Insulators*, Phys. Rev. B **86**, 115112 (2012).

[39] J. Kruthoff, J. de Boer, J. van Wezel, C. L. Kane, and R.-J. Slager, *Topological Classification of Crystalline Insulators through Band Structure Combinatorics*, Phys. Rev. X **7**, 041069 (2017).

[40] H. C. Po, L. Zou, T. Senthil, and A. Vishwanath, *Faithful Tight-Binding Models and Fragile Topology of Magic-Angle Bilayer Graphene*, Phys. Rev. B **99**, 195455 (2019).

[41] V. Peri, Z.-D. Song, M. Serra-Garcia, P. Engeler, R. Queiroz, X. Huang, W. Deng, Z. Liu, B. A. Bernevig, and S. D. Huber, *Experimental Characterization of Fragile Topology in an Acoustic Metamaterial*, Science **367**, 797 (2020).

[42] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. (Butterworth-Heinemann, Amsterdam, 1980), Vol. 5.

[43] S. M. Young and C. L. Kane, *Dirac Semimetals in Two Dimensions*, Phys. Rev. Lett. **115**, 126803 (2015).

[44] C. L. Kane and E. J. Mele, *Quantum Spin Hall Effect in Graphene*, Phys. Rev. Lett. **95**, 226801 (2005).

[45] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, *Quantum Spin Hall Effect and Topological Phase Transition in HgTe Quantum Wells*, Science **314**, 1757 (2006).

[46] T. Fukui, Y. Hatsugai, and H. Suzuki, *Chern Numbers in Discretized Brillouin Zone: Efficient Method of Computing (Spin) Hall Conductances*, J. Phys. Soc. Jpn. **74**, 1674 (2005).

[47] S. J. Cox and D. C. Dobson, *Band Structure Optimization of Two-Dimensional Photonic Crystals in H-Polarization*, J. Comput. Phys. **158**, 214 (2000).

[48] H. Men, K. Y. K. Lee, R. M. Freund, J. Peraire, and S. G. Johnson, *Robust Topology Optimization of Three-Dimensional Photonic-Crystal Band-Gap Structures*, Opt. Express **22**, 22632 (2014).

[49] S. S. Nanthakumar, X. Zhuang, H. S. Park, C. Nguyen, Y. Chen, and T. Rabczuk, *Inverse Design of Quantum Spin Hall-Based Phononic Topological Insulators*, J. Mech. Phys. Solids **125**, 550 (2019).

[50] R. E. Christiansen, F. Wang, and O. Sigmund, *Topological Insulators by Topology Optimization*, Phys. Rev. Lett. **122**, 234502 (2019).

[51] T. Asano and S. Noda, *Optimization of Photonic Crystal Nanocavities Based on Deep Learning*, Opt. Express **26**, 32704 (2018).

[52] J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark, and M. Soljacic, *Nanophotonic Particle Simulation and Inverse Design Using Artificial Neural Networks*, Sci. Adv. **4**, eaar4206 (2018).

[53] T. Asano and S. Noda, *Iterative Optimization of Photonic Crystal Nanocavity Designs by Using Deep Neural Networks, Nanophotonics*, https://doi.org/10.1515/nanoph-2019-0308 (2019).

[54] K. H. Matlack, M. Serra-Garcia, A. Palermo, S. D. Huber, and C. Daraio, *Designing Perturbative Metamaterials from Discrete Models*, Nat. Mater. **17**, 323 (2018).

[55] L.-H. Wu and X. Hu, *Topological Properties of Electrons in Honeycomb Lattice with Detuned Hopping Energy*, Sci. Rep. **6**, 1 (2016).

[56] S. Barik, A. Karasahin, C. Flower, T. Cai, H. Miyake, W. DeGottardi, M. Hafezi, and E. Waks, *A Topological Quantum Optics Interface*, Science **359**, 666 (2018).

[57] N. Parappurath, F. Alpeggiani, L. Kuipers, and E. Verhagen, *Direct Observation of Topological Edge States in Silicon Photonic Crystals: Spin, Dispersion, and Chiral Routing*, arXiv:1811.10739.

[58] C. Brendel, V. Peano, O. Painter, and F. Marquardt, *Snowflake Phononic Topological Insulator at the Nanoscale*, Phys. Rev. B **97**, 020102(R) (2018).

[59] J. Cha, K. W. Kim, and C. Daraio, *Experimental Realization of On-Chip Topological Nanoelectromechanical Metamaterials*, Nature (London) **564**, 229 (2018).

[60] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.11.021052 for gradient-descent based optimization. A randomly initialized geometry is evolved until the NN-predicted band structure (dark lines) approaches a target band structure (bright lines). The band colors encode the underlying Band Representations.

[61] S. Jiang, S. Lu, and D.-L. Deng, *Vulnerability of Machine Learning Phases of Matter*, arXiv:1910.13453.

[62] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, SCHNET—*A Deep Learning Architecture for Molecules and Materials*, J. Chem. Phys. **148**, 241722 (2018).

[63] B. Bradlyn, L. Elcoro, M. G. Vergniory, J. Cano, Z. Wang, C. Felser, M. I. Aroyo, and B. A. Bernevig, *Band Connectivity for Topological Quantum Chemistry: Band Structures as a Graph Theory Problem*, Phys. Rev. B **97**, 035138 (2018).

[64] T. Kariyado and X. Hu, *Topological States Characterized by Mirror Winding Numbers in Graphene with Bond Modulation*, Sci. Rep. **7**, 16515 (2017).