

PAPER • OPEN ACCESS

Unsupervised interpretable learning of topological indices invariant under permutations of atomic bands

To cite this article: Oleksandr Balabanov and Mats Granath 2021 *Mach. Learn.: Sci. Technol.* **2** 025008

View the [article online](#) for updates and enhancements.

You may also like

- [Interpretable Neural Network Construction: From Neural Network to Interpretable Neural Tree](#)
Xuming Ouyang and Cunguang Feng
- [Self-energy functional theory with symmetry breaking for disordered lattice bosons](#)
Dario Hugel, Hugo U R Strand and Lode Pollet
- [IDP-PGFE: an interpretable disruption predictor based on physics-guided feature extraction](#)
C. Shen, W. Zheng, Y. Ding et al.



PAPER

OPEN ACCESS

RECEIVED
24 August 2020REVISED
22 October 2020ACCEPTED FOR PUBLICATION
19 November 2020PUBLISHED
29 December 2020

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Unsupervised interpretable learning of topological indices invariant under permutations of atomic bands

Oleksandr Balabanov^{1,2} and Mats Granath¹ ¹ Department of Physics, University of Gothenburg, SE 412 96 Gothenburg, Sweden² Department of Physics, Stockholm University, SE 106 91 Stockholm, SwedenE-mail: oleksandr.balabanov@fysik.su.se and mats.granath@physics.gu.se**Keywords:** machine learning, topological phases, unsupervised learning, interpretable neural networks

Abstract

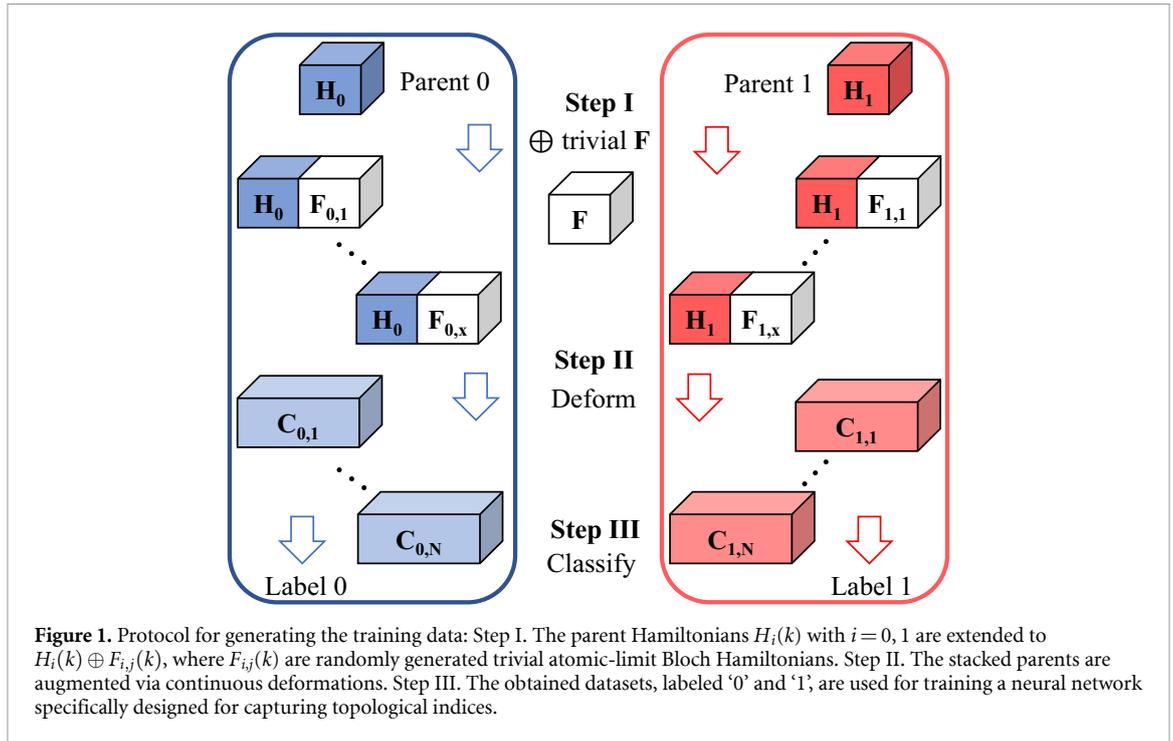
Multi-band insulating Bloch Hamiltonians with internal or spatial symmetries, such as particle-hole or inversion, may have topologically disconnected sectors of trivial atomic-limit (momentum-independent) Hamiltonians. We present a neural-network-based protocol for finding topologically relevant indices that are invariant under transformations between such trivial atomic-limit Hamiltonians, thus corresponding to the standard classification of band insulators. The work extends the method of ‘topological data augmentation’ for unsupervised learning introduced (2020 *Phys. Rev. Res.* **2** 013354) by also generalizing and simplifying the data generation scheme and by introducing a special ‘mod’ layer of the neural network appropriate for Z_n classification. Ensembles of training data are generated by deforming seed objects in a way that preserves a discrete representation of continuity. In order to focus the learning on the topologically relevant indices, prior to the deformation procedure we stack the seed Bloch Hamiltonians with a complete set of symmetry-respecting trivial atomic bands. The obtained datasets are then used for training an interpretable neural network specially designed to capture the topological properties by learning physically relevant momentum space quantities, even in crystalline symmetry classes.

1. Introduction

To find and classify exotic phases of quantum matter is of key importance in modern science. These phases often point to materials with unique properties and high scientific impact. A particular type of quantum matter, the so-called topologically non-trivial quantum phases, are especially at the frontier of current research [2–4]. Topological phase transitions lie outside the standard Landau paradigm of symmetry breaking and highlight topological non-triviality of the underlying state spaces. Despite of all the progress made and understanding gained over the years, the field is still very far from being complete.

Although still in its infancy as applied to topological matter, the use of machine learning (ML), specialized in big data pattern recognition, has the potential to revolutionize the field by circumventing the need for explicit analytic representations of topological markers (indices). Ideally, identifying topology using ML can subsequently guide further theoretical developments. In many areas of quantum physics, applications of ML has become a major research area [5, 6], for example to represent many-body states [7–16], identifying both symmetry breaking and topological phase transitions [17–44], for quantum state tomography [45–47], quantum control [48–50] and quantum error correction [51–56].

The use of artificial neural networks (NNs) or deep learning, in particular, stand out as the premier ML tool, with the power both to find complex patterns in data and to generalize this knowledge to previously unseen data. The capabilities and limitations of NNs applied to the task of classifying phases of topological quantum matter are currently under active research [17–25, 42, 44]. A potential limitation to using NNs is that common protocols are based on supervised learning which requires labeled data (object and corresponding topological marker) from an explicit model. This limits the perspectives for using ML to explore unknown topological structures. To begin to address this issue several *unsupervised* protocols for doing topological classifications have recently been developed [1, 36–40]. The method presented by us in



reference [1] is based on ‘topological data augmentation’, where datasets of topologically equivalent children states are created from single parent states. The created data is then used for training a NN, using the unsupervised ‘learning by confusion’ concept introduced in [36]: only datasets that have some distinguishing feature can be functionally classified by the network. Thus, any two *a priori* unknown datasets can be used for training the network with dummy label targets, with successful training outcome only if the two sets are topologically distinct. The procedure also used interpretable neural networks, designed to learn an intermediate momentum space output which is closely related to integral expressions over the local curvature. A crucial and challenging requirement for the procedure to work is that the data is sufficiently randomized to erase any non-topological information related to the specific parent states. In order to accomplish this randomization without destroying the topological information we construct in reference [1] explicit topology preserving transformations, valid in the discretized Brillouin zone. However, to construct such transformations, one might argue, requires a good understanding of the topological structure. Thus to relax the specificity of the topology preserving deformations would be a step forward in the spirit of the outlined objective.

With this in mind, in this paper we significantly advance the protocol from [1] and reveal new horizons of its applicability. As before, the protocol is split into two main stages, data generation and training of the network. The training data is generated from two parent states via random topology-preserving deformations. Here we always require the states to change slowly with momentum, in this way automatically ensuring continuity of the applied deformations. The procedure to do this while at the same time ensuring sufficient exploration of the space in order to obscure any non-topological features is one of the main methodological points. This also allows us to consider more general quantum systems without requiring any external knowledge on the topological structure of the underlying state space. We also modified the network’s layout, cf. figure 3, allowing the net to generically represent a much broader class of topological indices, including such that are sensitive to high symmetry points, lines, and surfaces in the Brillouin zone. Importantly, as in [1], our net’s layout ensures interpretable output of the learned topological quantity: The weighted output of the last convolutional layer corresponds to the relevant momentum resolved topological measure, as exemplified in figures 4, 5, and 7. By studying this intermediate output we can establish for example a direct connection between the topological indices and high symmetry points; cf. figures 5 and 7.

An additional limitation of the original protocol of topological data augmentation is that network may learn momentum space local invariants that are irrelevant for establishing non-trivial topological features. As an example, consider Hamiltonians with inversion symmetry, where in the atomic limit the individual orbitals may be odd or even under inversion. A half-filled 4-band Hamiltonian in the atomic limit may have two occupied bands either both with even parity, both with odd parity, or one band with even and one with odd. These three sectors of atomic Hamiltonians are topologically trivial, have no edge states [57], but

nevertheless cannot be connected by continuous non gap-closing deformations. Thus, a network trained on dataset based on deformations of single seed objects from one of these three sectors may learn to distinguish different trivial band insulators based on a local index (such as the parity at the Brillouin zone center) rather than trivial from non-trivial, depending on the specifics of the two training datasets.

In order to address this issue we extend the Hilbert space and corresponding matrix dimensions of the Bloch Hamiltonians that we want to classify from n to $2n$ bands, by stacking them with n atomic bands as sketched in figure 1. The children states are produced by first stacking the parents with a number of trivial atomic bands and only then deforming them employing random topology-preserving deformations. The created data is then used for training an interpretable NN allowing us to extract the learned topological quantities. Trivial atomic bands here mean bands that respect the symmetries of the corresponding symmetry class but do not vary in momentum, they represent the intrinsically trivial atomic limit where individual atomic bands do not hybridize. Thus, within the new procedure we purposely assign the same labels to any Bloch Hamiltonians differing by trivial atomic bands, in this way penalizing the NN for outputting topologically irrelevant indices that differentiate between *a priori* trivial objects. This approach is inspired by the commonly employed K -theoretic analytic classification schemes [4, 58, 59], where the topological equivalence between two band insulators is defined up to stacking of trivial bands. All trivial Hamiltonians here constitute a trivial monoid element and any Hamiltonian can then be stacked with atomic bands without changing the topological class.

The paper is organized as follows. The protocol is described in details in section 2, including how to generate parent states that vary sufficiently slowly on the scale of discretized momentum, how these are stacked with atomic bands and subsequently deformed in a manner that conserves the discrete measure of continuity of the Bloch Hamiltonians. The protocol is exemplified on one-dimensional (1d) 4-band insulators from three different symmetry classes in section 3. This includes examples with particle-hole symmetry and inversion symmetry where the network learns to single out the high symmetry points in the Brillouin zone and combine these in the appropriate way to produce the topological index. A summary and outlook is given in section 4.

2. Methods

To set the stage, we describe our protocol in general terms and provide the details applicable for all examples to follow. Any one-dimensional (1d) gaped n -band system can be represented using a n by n Hermitian Bloch Hamiltonian $H(k)$ that is a periodic continuous function of momentum k . Two Bloch Hamiltonians are said to be *topologically equivalent* if the occupied spaces can be continuously deformed into each other. Any Bloch Hamiltonian can therefore be continuously transformed to have energies -1 and 1 for occupied and empty bands respectively, and we consider the space of normalized Bloch Hamiltonians only. The aim is then to find a topological index, call it ν , labeling the Bloch Hamiltonians $H(k)$ by their topological equivalence classes. It is known that the space of all Bloch Hamiltonians in 1d is topologically trivial but the classification becomes nontrivial after imposing various symmetries on the systems, in this way restricting the allowed deformations. In this paper, we always assume the systems to be half filled and explicitly normalize the Bloch Hamiltonians to have equal number of positive and negative bands.

2.1. Prerequisites

For our protocol to be functional we need to choose a distance measure between any Hermitian matrices H_1 and H_2 to later establish a notion of continuity for discretized Bloch Hamiltonians. The distance is here defined through relation:

$$d(H_1, H_2) = 1 - \frac{1}{N_{\text{o.b.}}} \sum_{i,j \in \text{o.b.}} |\langle \psi_{1,i} | \psi_{2,j} \rangle|^2, \quad (1)$$

where $|\psi_{1/2,i}\rangle$ are the normalized eigenstates of $H_{1/2}$, with i running over the $N_{\text{o.b.}}$ occupied bands of $H_{1/2}(k)$. Note that the distance $d(H_1(k), H_2(k)) = 0$ for every momentum k if and only if the (normalized) Bloch Hamiltonians $H_1(k)$ and $H_2(k)$ are identical.

We also need an efficient algorithm to numerically move between the matrices and for this purpose we employ a numeric method introduced by Scheurer *et al* [38]. The idea is to gradually change matrix H_1 towards H_2 via minimizing the distance $d(H_1, H_2)$ employing the gradient descent algorithm. Explicitly, in each numeric step we perform a unitary rotation $H_1 \rightarrow U_j H_1 U_j^\dagger$ using different unitary generators $U_j = \exp(i\phi_j \Lambda_j)$, where Λ_j is a basis of Hermitian matrices. The phase constants ϕ_j are computed using the gradient descent rule [38]:

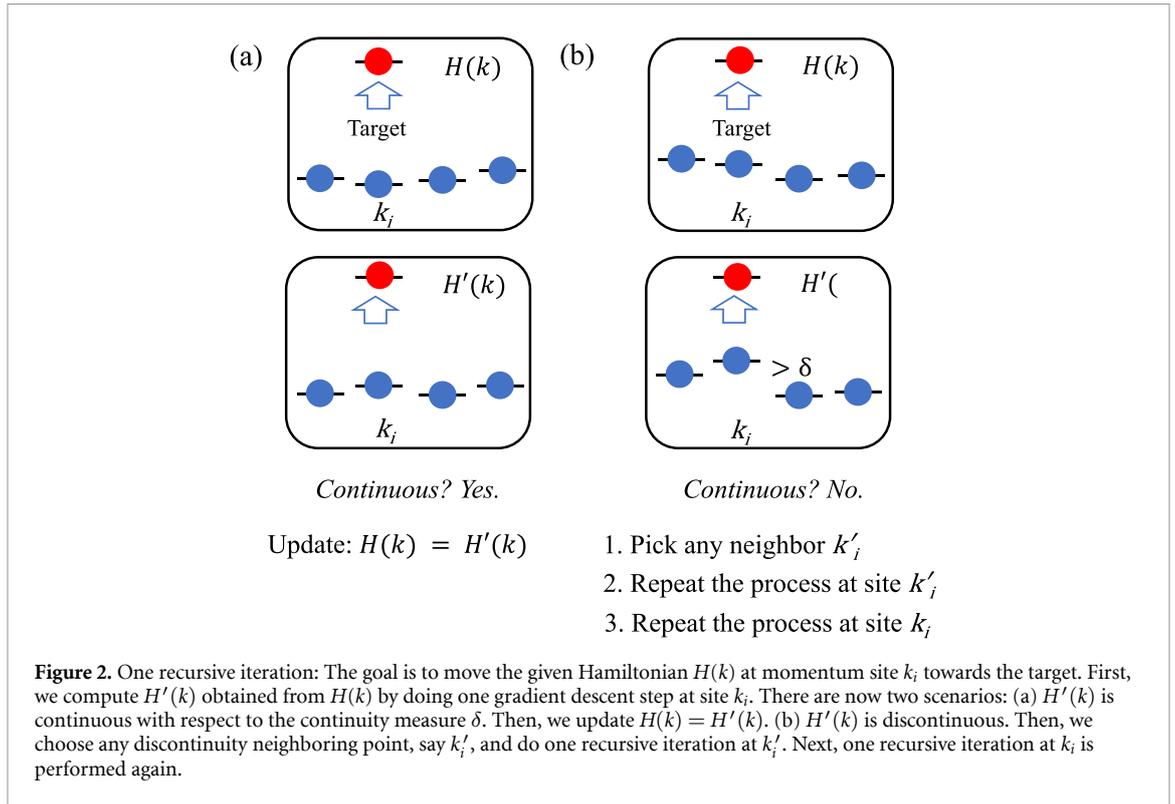


Figure 2. One recursive iteration: The goal is to move the given Hamiltonian $H(k)$ at momentum site k_j towards the target. First, we compute $H'(k)$ obtained from $H(k)$ by doing one gradient descent step at site k_j . There are now two scenarios: (a) $H'(k)$ is continuous with respect to the continuity measure δ . Then, we update $H(k) = H'(k)$. (b) $H'(k)$ is discontinuous. Then, we choose any discontinuity neighboring point, say k'_j , and do one recursive iteration at k'_j . Next, one recursive iteration at k_j is performed again.

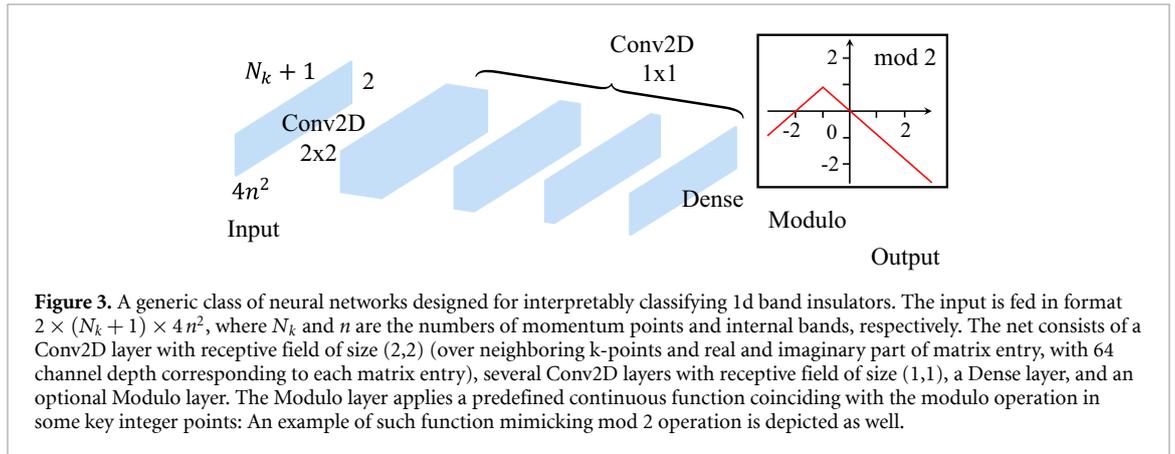


Figure 3. A generic class of neural networks designed for interpretably classifying 1d band insulators. The input is fed in format $2 \times (N_k + 1) \times 4n^2$, where N_k and n are the numbers of momentum points and internal bands, respectively. The net consists of a Conv2D layer with receptive field of size (2,2) (over neighboring k -points and real and imaginary part of matrix entry, with 64 channel depth corresponding to each matrix entry), several Conv2D layers with receptive field of size (1,1), a Dense layer, and an optional Modulo layer. The Modulo layer applies a predefined continuous function coinciding with the modulo operation in some key integer points: An example of such function mimicking mod 2 operation is depicted as well.

$$\phi_j = -\eta \frac{\partial d(U_j H_1 U_j^\dagger, H_2)}{\partial \phi_j} \Big|_{\phi_j=0}, \tag{2}$$

with some small learning rate η . To avoid rapid changes possibly leading to discontinuous deformations we also restrict ϕ_j to not exceed some upper bound value ϕ_{\max} . The symmetries are maintained by doing the unitary rotations in pairs bonded by the symmetries, in this way forbidding the deformed Hamiltonian to go outside the considered symmetry class. In contrast to [38], where the matrix distance gradient descent formalism is used to cluster topologically equivalent data points, we use it to implement topological data augmentation, i.e. to generate new data points; see section 2.4.

2.2. Generating parent states

The training data is generated from two n by n parent states $H_0(k)$ and $H_1(k)$ defined on a grid of N_k momentum points. The baseline trivial state $H_0(k)$ is taken to be any trivial atomic-limit Bloch Hamiltonian satisfying the symmetries. It is produced by repeating a random properly symmetrized n by n Hermitian matrix. The parent $H_1(k)$ is generated by symmetrizing a Bloch Hamiltonian

$$H(k) = \sum_{p=0}^m A_n \sin(pk) + B_n \cos(pk), \tag{3}$$

with A_p and B_p random n by n Hermitian matrices. With this procedure we can efficiently sample the symmetry classes and control state's continuity by choosing $m \ll N_k$. The continuity here is realized by requiring all neighbor to neighbor distances $d(H(k), H(k + \Delta k))$ (with $\Delta k = 2\pi/N_k$) to be below some small threshold δ . Thus, the parent $H_1(k)$ is suggested to be randomly generated via equation (3). Note, however, that in general we are mainly interested in picking $H_1(k)$ that is topologically nontrivial and therefore can facilitate the protocol by applying all available external intuition for purposely choosing a promising candidate for $H_1(k)$. Completely random guessing, however, is also anticipated to eventually lead to the same results according to learning by confusion ideology but one will need to make several attempts until reaching success in training.

2.3. Stacking parents with trivial atomic bands

Within our protocol of topological classification, figure 1, the parents $H_i(k)$ with $i = 0, 1$ are first extended to $H_i(k) \oplus F_{i,j}(k)$ using random trivial atomic-limit Bloch Hamiltonians $F_{i,j}(k)$, and only then the stacked parents are deformed. The extension of the parents is not generally necessary within our protocol, however, the datasets generated in this way facilitate the NN to only look at topologically relevant indices that do not distinguish between different trivial atomic-limit Hamiltonians. We therefore develop a procedure for randomly generating trivial atomic-limit Bloch Hamiltonians $F(k)$ respecting the relevant symmetries. Each trivial Bloch Hamiltonian $F(k)$ is constructed by repeating some base n by n matrix \tilde{F} . The matrix \tilde{F} has to respect the relevant symmetries, and within our protocol it is generated by filling the matrix entries with random complex numbers $(a + ib)$ with $a, b \in [-1, 1]$ (with real numbers $c \in [-1, 1]$ for diagonal entries), symmetrizing the obtained matrix to respect the given symmetries, and then normalizing it to have eigenvalues ± 1 . Depending on symmetry class the space of all symmetry-respecting matrices can be composed of multiple topologically disconnected sectors and these sectors can in general be non-uniformly represented in our randomly created data. Note that trivial Hamiltonians $F(k)$ corresponding to \tilde{F} from disconnected sectors generically cannot be connected using continuous symmetry-preserving deformations. This may mislead our NN to find topologically relevant quantities because some connected parts of trivial atomic-limit Hamiltonians might be underrepresented. For avoiding our algorithm giving preference to a particular disconnected matrix sector we put the generated matrices into datasets, each composed of 10^3 continuously connected random matrices. The connectivity is established by checking if they can be connected via the gradient descent algorithm described in section 2.1 without breaking the symmetries, and the number of obtained disconnected sectors of trivial Hamiltonians depends on the symmetry class. We then pick a base matrix $\tilde{F}_{i,j}$ defining $F_{i,j}(k)$ with equal probability from the created datasets of matrices \tilde{F} . In this way all the disconnected sectors are represented equally within the process. Using this approach we produce $j = 1, \dots, N_p$ stacked parent states $H_i(k) \oplus F_{i,j}(k)$ corresponding to two parents $H_0(k)$ and $H_1(k)$. By using these trivially stacked parents we hint the network to learn only quantities that cannot differentiate between two trivial atomic states even if they cannot be continuously connected.

2.4. Topological data augmentation

The stacked parent states are then augmented by performing random deformations while keeping the states continuous and maintaining the symmetries. The deformations are implemented by doing unitary rotations of discretized $2n$ by $2n$ Bloch Hamiltonians $H(k)$ at distinct momentum points k_i towards randomly generated target matrices \tilde{H} . These rotations are done in small steps by minimizing the distance function $d(H(k_i), \tilde{H})$ following gradient descent algorithm with some learning rate η and rotation cutoff ϕ_{\max} . We also perform the steps in symmetry-bonded pairs to always maintain the symmetries. The unitary matrices are taken to be $U_j = \exp(i\phi_j \Lambda_j)$ with Λ_j a set of basis matrices for $2n$ by $2n$ Hermitian matrices. Importantly, the state is kept continuous by recursively moving also matrices at neighboring momentum sites towards the same target once the distance exceeds the threshold δ , figure 2. In this way any deformation at one momentum site pulls the neighbors along with it, resulting in efficient augmentation procedure that at all times maintains continuity of the state. We terminate the deformation algorithm if: the corresponding matrix reached the target, the state could not be maintained continuous after a number of recursive steps M_{\max} , or if we exceeded a maximum number of gradient descent learning steps N_{\max} . The motivation for this procedure, which is similar to a string being pulled over a large distance at one point, is to be able to traverse large distances in the topological sector in an unbiased fashion. This in contrast to making small random local deformations, just producing a random walk, which in a high dimensional space will provide less efficient exploration.

2.5. Neural network structure and training

The two datasets corresponding to topologically augmented stacked parents are then used to train a neural-network-based classifier specifically designed to represent generic expressions of topological indices.

The NN is trained on the Bloch Hamiltonians $H(k)$ transformed from $N_k \times 2n \times 2n$ complex-valued format to $2 \times (N_k + 1) \times 4n^2$ float format with an extra momentum site added to encode the periodicity. The classifying network consists of several convolution layers and in the earlier work [1] followed by a summation layer. The activation functions associated with each convolution layer are rectified linear units (ReLU) except the last one with linear activation. This network type was designed to perform identical local operations on each site and then sum the obtained outputs over the whole sample, in this way capturing topological numbers that are given by an integral over some local curvature. However, in presence of certain symmetries, high symmetry momentum points, lines or surfaces may be of special importance and the indices are expected to take a different form. We therefore advance the design of the network to cover a much broader class of expressions by changing the last summation layer to a fully connected (dense) layer, see figure 3. To be concrete, we use a dense layer with a single output node, absolute weight values ≤ 1 , zero bias, and linear activation. The reason for avoiding non-linear activation and large weights is to constrain the preceding feature map (the output from last convolution layer) to learn relevant and *interpretable* momentum space quantities. Note that this family of networks, the net in figure 3 and its generalizations to 2d and 3d, can adjust to calculate sums of some local functions over arbitrary symmetry-preserving points, lines, or planes, and then add or subtract them in any order. A coarse grained version of the output before summation in this single node is what is displayed as the momentum resolved images in figures 4, 5, and 7. The network thus processes the local information in k-space convoluted over all bands in the first convolution layer (2 by 2 filters operate on nearest neighbor sites in k-space and real and imaginary part), which after subsequent non-linear operations outputs the relevant momentum resolved local quantities. This is thus the essence of the interpretability of the neural network.

To capture an even larger class of indices we also suggest to use an extra layer applying a predefined operation on the single-valued output to represent a modulo function over some assumed range of integers. The modulo function itself was found to be not applicable because it is periodic and discontinuous, giving convergence problems during the training. A way around this problem is to use a continuous function coinciding with the modulo operation in some key integer points. An example mimicking mod 2 operation is given inside a box in figure 3. This function outputs 0 for input 0 and -2 , and outputs 1 for input -1 . The network illustrated in figure 3 can be generalized to 2d case by analogously modifying the 2d network of reference [1].

3. Results

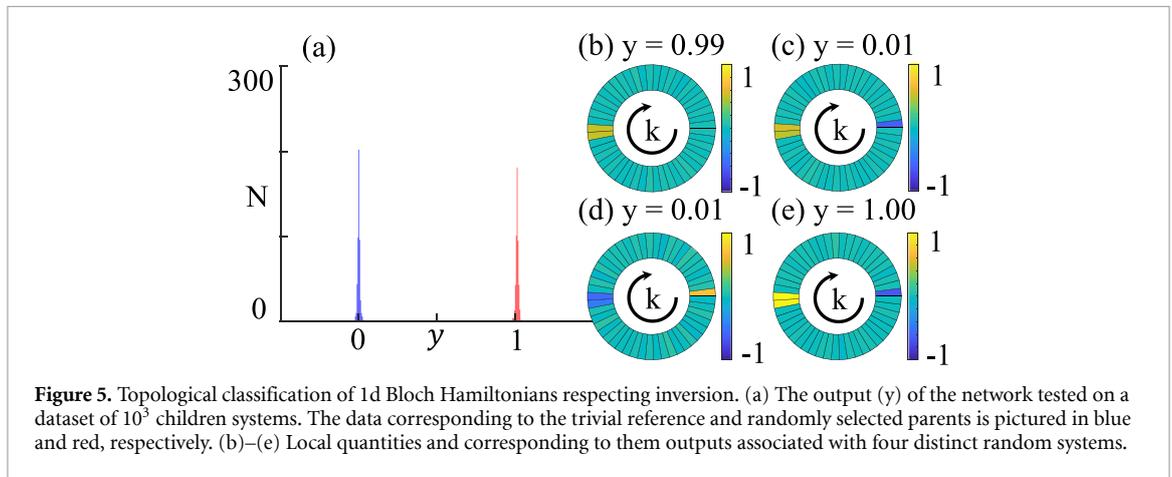
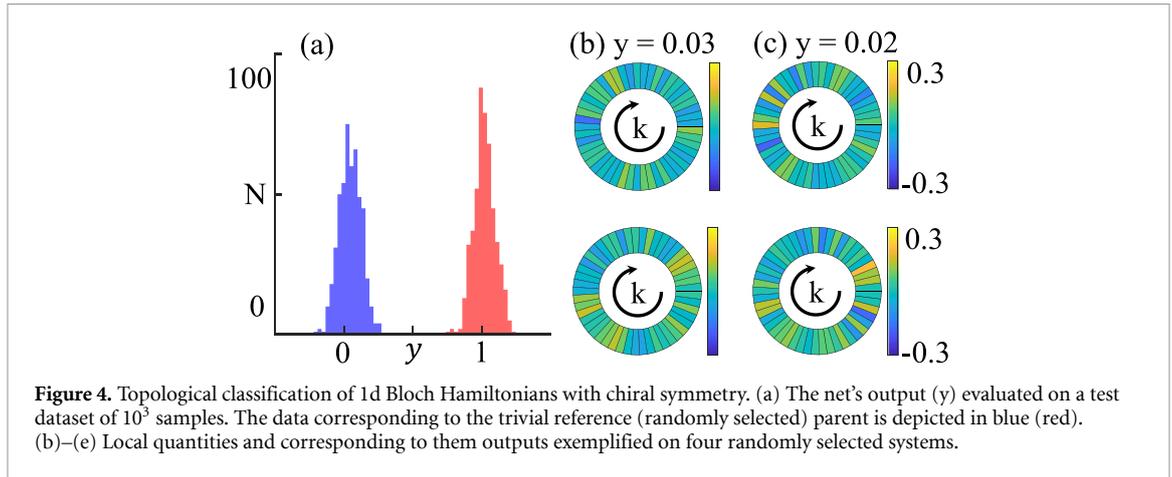
Here we present results of the analysis for three examples of 4-band Hamiltonians in 1d, having chiral symmetry, inversion symmetry, and particle-hole symmetry respectively. We show how the network correctly learns to classify trivial from non-trivial datasets in a way that gives interpretable information corresponding to relevant momentum resolved quantities.

3.1. Hyperparameters

We selected the same hyperparameters throughout all of our examples and they are listed here. The momentum space was discretized by $N_k = 100$ points. The gradient descent learning rate was taken to be $\eta = 0.1$, the rotation cutoff $\phi_{\max} = 0.1$, and the continuity parameter $\delta = 0.1$. The maximum numbers of performed learning and recursive steps before terminating the algorithm were $N_{\max} = 20$ and $M_{\max} = 20$, respectively. The trivial parent $H_0(k)$ was generated by picking a random symmetry-respecting matrix and duplicating it for all N_k points. The second parent $H_1(k)$ was randomly generated by symmetrizing the output of equation (3) with $m = 4$. The two parents were then stacked with $N_p = 10$ randomly chosen symmetry-respecting trivial atomic-limit Hamiltonians to create two sets of $2n$ band Bloch Hamiltonians. Each of the stacked parents was then used to create 10^3 children. For doing so, at first we deformed them by 10 independent single-site unitary rotations, and only then saved a child for each new unitary deformation. In this way we produced two datasets of 10^4 children Bloch Hamiltonians corresponding to two original parents.

3.2. Multiband 1d insulators with chiral symmetry

As our first illustration, we implement the protocol for topological classification of 4-band insulators in 1d with chiral symmetry. It is well known that all gapped systems within this symmetry class are distinguished by a Z topological invariant, the so-called winding number [4]. Here we train a NN to fit into a representation of the winding number without using any externally labeled data: The training data is entirely produced via the topological data augmentation procedure described in section 2.2. This symmetry class was also explored for 2-band systems in reference [1], using a different protocol for generating the training data, but with similar results.



Bloch Hamiltonians satisfying chiral symmetry must obey $H(k) = -U_S^\dagger H(k) U_S$, with a unitary matrix U_S chosen to be a diagonal matrix with first half of entries $+1$ and last half of entries -1 , thus consisting of off-diagonal blocks. In order to symmetrize randomly generated Hermitian matrices A and Bloch Hamiltonians $H(k)$ we then performed the following operations on them:

$$A \rightarrow A - U_S A U_S^\dagger \quad (4)$$

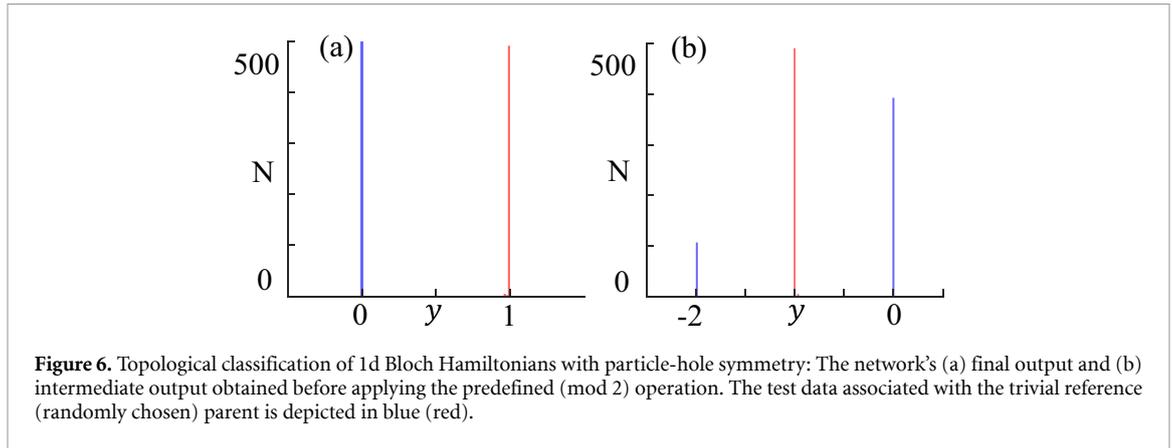
$$H(k) \rightarrow H(k) - U_S H(k) U_S^\dagger. \quad (5)$$

It was found that all generated symmetrized matrices A could be connected via the gradient descent algorithm described in section 2.1, creating a single dataset of interconnected matrices, in agreement with our anticipation. The base matrices defining the chiral atomic-limit Bloch Hamiltonians were therefore uniformly picked from that single dataset, taken to be sampled by 10^3 matrices. (Note that, because there is only a single trivial atomic sector, the stacking procedure is not strictly necessary for this class of systems.) All other specifications needed for topological data augmentation are listed in section 3.1.

At the training stage we employed a NN with a layout thoroughly described in section 2.5 and depicted in figure 3. The convolutional part of the network was explicitly taken to consist of one 2d convolution layer of 512 filters with $(2, 2)$ receptive fields and three 2d convolution layers of 256, 128, 1 filters with $(1, 1)$ receptive fields. The network was successfully trained without any modulo layer in this case, confirming that the learned topological number is a Z invariant. In total our network has 296 038 trainable parameters.

The training was done on $2 \cdot 9500$ and tested on $2 \cdot 500$ samples using absolute mean error cost and Adam optimizer. We trained over 2000 epochs with learning rate $\eta = 10^{-4}$. Importantly, we also effectively augmented the training dataset: Before each epoch every Bloch Hamiltonian was translated by a random number of momentum sites and uniformly rotated using a random symmetry-respecting unitary operator.

The results are presented in figure 4. The network has successfully learned to distinguish the datasets corresponding to two ensembles of topologically equivalent children, figure 4(a): the classification accuracy



evaluated on the test set is 100%. Importantly, the NN layout allows us to interpret the obtained classification and find a local quantity corresponding to the learned topological index. In figures 4(b)–(e) we exemplified the learned local quantity on four representative systems from the test dataset.

3.3. Multiband 1D insulators with inversion symmetry

In our second example we focus on 4-band systems with inversion symmetry and implement our protocol for classifying them. The topological phases within this symmetry class are characterized by a Z invariant $\nu_{IS} = n_0 - n_\pi$ with $n_{0/\pi}$ counting a number of negative parity eigenstates at high-symmetry points $k = 0$ and $k = \pi$ [60, 57].

The inversion relation reads as $H(k) = U_{IS}^\dagger H(-k) U_{IS}$, where U_{IS} is some unitary matrix and it is here chosen to be the same as U_S , i.e. a diagonal matrix with $+1$ and -1 entries. The symmetrization of randomly generated Hermitian matrices A and Bloch Hamiltonians $H(k)$ is done via

$$A \rightarrow A + U_{IS} A U_{IS}^\dagger \quad (6)$$

$$H(k) \rightarrow H(k) + U_{IS} H(-k) U_{IS}^\dagger. \quad (7)$$

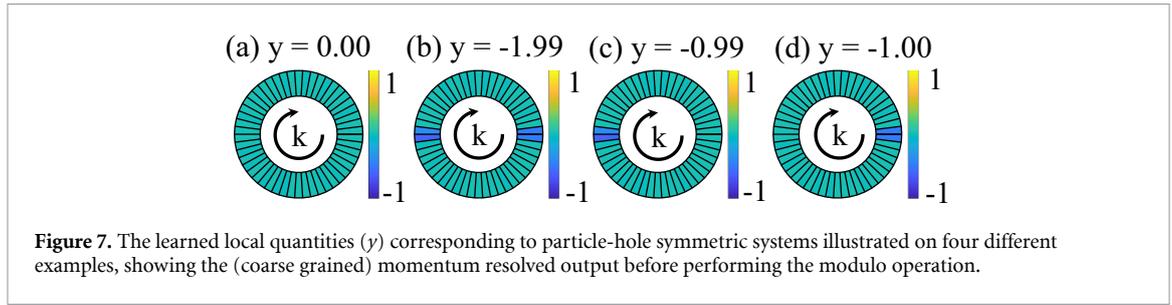
The generated symmetrized matrices A were found to create three distinct blocks of matrices, with all matrices connected via the gradient descent algorithm within each block. Analytically, as discussed in the introduction, the disconnected blocks correspond to three different combinations of inversion-symmetric eigenvalues ± 1 of the occupied eigenstates. Thus, for producing inversion-symmetric trivial atomic-limit Hamiltonians we uniformly picked matrices from these three blocks. Each block was taken to consist of 10^3 generated matrices. Other details of data generation are given in section 3.1.

We then trained a NN with a layout described in section 2.5 and illustrated in figure 3. The convolution layers of the net were chosen to have 128 filters for the first layer with (2, 2) receptive field, and 64, 32, 1 filters for the other ones with (1, 1) receptive field. The same as in section 3.2, the network was successfully trained without any modulo layer. In total there are 43 366 trainable parameters. The network was trained on $2 \cdot 9500$ states and tested on $2 \cdot 500$ states employing absolute mean error cost and Adam optimizer. The training was done over 10^3 epochs with learning rate $\eta = 10^{-3}$. Before every epoch we performed a uniform symmetry-respecting rotation of each training state and by this effectively augmented the dataset.

The outcome of our protocol is depicted in figure 5. The network has learned to differentiate between two ensembles of topologically equivalent children with precision 100%, figure 5(a). By looking at the network's intermediate output we could also retrieve the local quantity corresponding to the learned topological index, as shown in figure 5(b)–(e). Strikingly, the network learned the importance of high-symmetric $k = 0, \pi$ and simply avoids all other momentum points. Important to note that this property was not built-in into the network by hand and it highlights the flexibility of our NN layout, figures 3: The last dense layer in the network allows it to find important points, lines and/or surfaces in momentum space without any external supervision.

3.4. Multiband 1D insulators with particle-hole symmetry

For our last example we implement the protocol for 1d topological classification in symmetry class D composed of systems exhibiting a particle-hole symmetry. It is well known that the topological number distinguishing the phases within this symmetry class is a Z_2 number. The topologically nontrivial phase



within this symmetry class exhibits very rich physics and opens up a road for realization of robust Majorana end modes [61].

In first quantization the particle-hole symmetry is antiunitary anticommuting symmetry explicitly described by relation $H(k) = -U_C^\dagger H^*(-k) U_C$ for some unitary operator U_C . Here we use the conventional representation of particle-hole symmetry that is basic within the BCS theory of superconductivity: We take $U_C = I \otimes \tau_x$ with a Pauli matrix τ_x and identity I . To symmetrize Hermitian matrices A and Bloch Hamiltonians $H(k)$ the following transformations are performed:

$$A \rightarrow A - U_C A^* U_C^\dagger \quad (8)$$

$$H(k) \rightarrow H(k) - U_C H^*(-k) U_C^\dagger. \quad (9)$$

By applying the gradient descent algorithm from section 2.1, it was found that the generated symmetrized matrices A clustered into two blocks of matrices and we uniformly picked matrices from them for producing trivial atomic-limit Bloch Hamiltonians respecting the particle-hole symmetry. These two blocks of trivial Hamiltonians would in the Majorana basis correspond to the Pfaffian at $k=0$ and $k=\pi$ being ± 1 [4, 62]. In total there were 2×10^3 matrices generated. All the details on other aspects of the data generation are provided in section 3.1.

Again, in the training stage we employed a neural net layout shown in figure 3 with convolution layers of 256, 128, 64, 1 filters. The first convolution layer operates with (2, 2) receptive field and all other ones with (1, 1) receptive field. There are 107 110 adjustable parameters in total. Without a modulo layer the network has failed to adjust to separate the classes of the training data, but with the modulo layer depicted in figure 3 it has succeeded. We trained the network on $2 \cdot 9500$ and tested on $2 \cdot 500$ samples using Adam optimizer and absolute mean error cost function. The training was done over 10^3 epochs with learning rate $\eta = 10^{-4}$. The same as in all previous cases we effectively augmented the training dataset by doing a uniform symmetry-respecting rotation of every Bloch Hamiltonian before each epoch.

The network's output evaluated on a test dataset is shown in figure 6: Two ensembles of topologically equivalent children are successfully differentiated by the network with high precision, figure 6(a). In figure 6(b) we show the intermediate output produced by the network before applying the mod 2 operation: Interestingly, we see the data corresponding to topologically trivial ensemble clustered around two different values, -2 and 0 , confirming the Z_2 nature of the topological index. In figure 7 four examples of test states show the local quantity corresponding to the learned topological number. Again, the network has adjusted to look at only high-symmetric $k=0, \pi$ and neglect other momentum points.

These results demonstrate that the network without explicit guidance learns to focus on the pertinent information. In fact, it is known that these Hamiltonians can be classified by product of the sign of the Pfaffian at the high symmetry points $k=0$ and π . Without extending the datasets with different sectors of trivial atomic bands the network would only learn to pick up one of the high symmetry points, and it would fail to classify all Hamiltonians properly. One might train two networks to learn to separately classify the two symmetry points, but the corresponding topological indices will be a composition of topologically irrelevant and relevant invariants. Another alternative to the mod layer for this problem would be to use an additional dense layer with non-linear activation functions to effectuate an xor classification of the output from the last convolution layer. However, the drawback of this is that it would allow for a drift of the last feature map output, with a corresponding loss of the interpretability of the results.

4. Conclusion and outlook

We present a novel protocol for using NNs to learn topological indices of Bloch Hamiltonians, extending previous work on 'topological data augmentation'. The protocol is characterized by the following features. (i)

It is unsupervised, i.e. training data is generated by randomly deforming parents states while ensuring the topological integrity of the data sets, without using any prior knowledge or specific models except a specific representation of the symmetry class to constrain the data. (ii) The NNs are specifically designed to be interpretable, in the sense that the single feature map of the last convolution layer gives momentum resolved information about the learned quantities. In this way the network can for example learn to single out relevant high symmetry momentum points. Note that, as demonstrated in detail in [1], our net's layout implicitly guarantees the network to generalize to systems with values of topological index lying outside of the training regime. This is because the network is trained to represent the topological index via a momentum resolved quantity, and given sufficient exploration all possible local data configurations are represented within the training dataset. (iii) By extending the training data samples with a complete set of symmetry respecting atomic insulators (specifically here, stacking 4 bands on 4 bands) the network learns to pick up only relevant topological invariants that do not discriminate between topologically disconnected but trivial atomic-limit Hamiltonians.

The protocol takes a next step towards the goal of using machine learning and NNs to identify unknown topological invariants that go beyond the already well established non-interacting and translationally invariant systems with spatial or internal symmetries. Areas to explore could be periodically-driven (Floquet) topological phases [63, 64], non-Hermitian topological matter [48, 65], interacting systems [66], and others. The NNs used in this work are very small compared to state of the art deep learning networks which opens up for extending the protocol to systems with less symmetry (e.g. disordered systems) that would require dense matrix input which is not in a block diagonal form. A nice ingredient of the network structure for the examples presented in this work and in [1] is the interpretability that follows from the construction where the network has local operations in momentum space, until the very last layers. It remains to be explored if and how this can be extended to problems where local measures may not be sufficient. Nevertheless, the interpretability is a bonus which is not strictly necessary for the general topological data augmentation procedure and network classification.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://github.com/oleksandr-balabanov/ML-protocol-topo-indices-1d>.

Acknowledgments

This work was supported by the Swedish Research Council through Grant No. 621-2014-5972. The python code using TensorFlow for generating data and training networks can be found in [67].

ORCID iDs

Oleksandr Balabanov  <https://orcid.org/0000-0003-0417-9856>

Mats Granath  <https://orcid.org/0000-0003-3185-2014>

References

- [1] Balabanov O and Granath M 2020 Unsupervised learning using topological data augmentation *Phys. Rev. Res.* **2** 013354
- [2] Hasan M Z and Kane C L 2010 Coll.: topological insulators *Rev. Mod. Phys.* **82** 3045–67
- [3] Xiao-Liang Q and Zhang S-C 2011 Topological insulators and superconductors *Rev. Mod. Phys.* **83** 1057–110
- [4] Ching-Kai C, Teo J C Y, Schnyder A P and Ryu S 2016 Classification of topological quantum matter with symmetries *Rev. Mod. Phys.* **88** 035005
- [5] Dunjko V and Briegel H J 2018 Machine learning & artificial intelligence in the quantum domain: a review of recent progress *Rep. Prog. Phys.* **81** 074001
- [6] Carrasquilla J 2020 Machine learning for quantum matter *Adv. Phys.:* X **5** 1797528
- [7] Carleo G and Troyer M 2017 Solving the quantum many-body problem with artificial neural networks *Science* **355** 602–6
- [8] Gao X and Duan L-M 2017 Efficient representation of quantum many-body states with deep neural networks *Nat. Commun.* **8** 662
- [9] Deng D-L, Xiaopeng Li and Das Sarma S 2017a Quantum entanglement in neural network states *Phys. Rev. X* **7** 021021
- [10] Deng D-L, Xiaopeng Li and Das Sarma S 2017b Machine learning topological states *Phys. Rev. B* **96** 195145
- [11] Nomura Y, Darmawan A S, Yamaji Y and Imada M 2017 Restricted Boltzmann machine learning for solving strongly correlated quantum systems *Phys. Rev. B* **96** 205152
- [12] Kaubruegger R, Pastori L and Budich J C 2018 Chiral topological phases from artificial neural networks *Phys. Rev. B* **97** 195136
- [13] Pastori L, Kaubruegger R and Budich J C 2019 Generalized transfer matrix states from artificial neural networks *Phys. Rev. B* **99** 165123
- [14] Melko R G, Carleo G, Carrasquilla J and Cirac J I 2019 Restricted Boltzmann machines in quantum physics *Nat. Phys.* **15** 887–92
- [15] Glasser I, Pancotti N, August M, Rodriguez I D and Cirac J I 2018 Neural-network quantum states, string-bond states and chiral topological states *Phys. Rev. X* **8** 011006

- [16] Choo K, Carleo G, Regnault N and Neupert T 2018 Symmetries and many-body excitations with neural-network quantum states *Phys. Rev. Lett.* **121** 167204
- [17] Zhang P, Shen H and Zhai H 2018 Machine learning topological invariants with neural networks *Phys. Rev. Lett.* **120** 066401
- [18] Sun N, Jinmin Y, Zhang P, Shen H and Zhai H 2018 Deep learning topological invariants of band insulators *Phys. Rev. B* **98** 085402
- [19] Casert C, Vieijra T, Nys J and Ryckebusch J 2019 Interpretable machine learning for inferring the phase boundaries in a nonequilibrium system *Phys. Rev. E* **99** 023304
- [20] Carvalho D, García-Martínez N A, Lado J L and Fernández-Rossier J 2018 Real-space mapping of topological invariants using artificial neural networks *Phys. Rev. B* **97** 115453
- [21] Ming Y, Lin C-T, Bartlett S D and Zhang W-W 2019 Quantum topology identification with deep neural networks and quantum walks *NPJ Comput. Mater.* **5** 88
- [22] Greplova E, Valenti A, Boschung G, Schäfer F, Lörch N and Huber S D 2020 Unsupervised identification of topological phase transitions using predictive models *New J. Phys.* **22** 045003
- [23] Tsai Y-H, Meng-Zhe Y, Hsu Y-H and Chung M-C 2020 Deep learning of topological phase transitions from entanglement aspects *Phys. Rev. B* **102** 054512
- [24] Caio M D, Caccin M, Baireuther P, Hyart T, and Fruchart M 2019 Machine learning assisted measurement of local topological invariants (<https://ui.adsabs.harvard.edu/abs/2019arXiv190103346C>) (arXiv:1901.03346)
- [25] Zhang Y, Ginsparg P and Kim E-A 2020a Interpreting machine learning of topological quantum phase transitions *Phys. Rev. Res.* **2** 023283
- [26] Carrasquilla J and Melko R G 2017 Machine learning phases of matter *Nat. Phys.* **13** 431–4
- [27] Ch'ng K, Carrasquilla J, Melko R G and Khatami E 2017 Machine learning phases of strongly correlated fermions *Phys. Rev. X* **7** 031038
- [28] Broecker P, Carrasquilla J, Melko R G and Trebst S 2017 Machine learning quantum phases of matter beyond the fermion sign problem *Sci. Rep.* **7** 8823
- [29] Beach M J S, Golubeva A and Melko R G 2018 Machine learning vortices at the kosterlitz-thouless transition *Phys. Rev. B* **97** 045207
- [30] Huembeli P, Dauphin A and Wittek P 2018 Identifying quantum phase transitions with adversarial neural networks *Phys. Rev. B* **97** 134109
- [31] Zhang W, Liu J and Wei T-C 2019 Machine learning of phase transitions in the percolation and xy models *Phys. Rev. E* **99** 032142
- [32] Zhang Y, Melko R G and Kim E-A 2017 Machine learning z_2 quantum spin liquids with quasiparticle statistics *Phys. Rev. B* **96** 245119
- [33] Rem B S, Käming N, Tarnowski M, Asteria L, Fläschner N, Becker C, Sengstock K and Weitenberg C 2019 Identifying quantum phase transitions using artificial neural networks on experimental data *Nat. Phys.* **15** 917–20
- [34] Huembeli P, Dauphin A, Wittek P and Gogolin C 2019 Automated discovery of characteristic features of phase transitions in many-body localization *Phys. Rev. B* **99** 104106
- [35] Goetz J B, Zhang Y and Lawler M J 2020 Detecting nematic order in STM/STS data with artificial intelligence *Sci. Post Phys.* **8** 87
- [36] Van Nieuwenburg E P L, Liu Y-H and Huber S D 2017 Learning phase transitions by confusion *Nat. Phys.* **13** 435–9
- [37] Rodríguez-Nieva J F and Scheurer M S 2019 Identifying topological order through unsupervised machine learning *Nat. Phys.* **15** 790–5
- [38] Scheurer M S and Slager R-J 2020 Unsupervised machine learning and band topology *Phys. Rev. Lett.* **124** 226401
- [39] Long Y, Ren J and Chen H 2020 Unsupervised manifold clustering of topological phononics *Phys. Rev. Lett.* **124** 185501
- [40] Che Y, Gneiting C, Liu T, and Nori F 2020 Topological quantum phase transitions retrieved from manifold learning (arXiv:2002.02363)
- [41] Kharkov Y A, Sotskov V E, Karazeev A A, Kiktenko E O and Fedorov A K 2020 Revealing quantum chaos with machine learning *Phys. Rev. B* **101** 064406
- [42] Zhang L-F, Tang L-Z, Huang Z-H, Zhang G-Q, Huang W, and Zhang D-W 2020b Machine learning topological invariants of non-hermitian systems (arXiv:2009.04058)
- [43] Arnold J, Schäfer F, zonda M, and Lode A U J 2020 Interpretable and unsupervised phase classification (arXiv:2010.04730)
- [44] Holanda N L and Griffith M A R 2020 Machine learning topological phases in real space *Phys. Rev. B* **102** 054107
- [45] Torlai G, Mazzola G, Carrasquilla J, Troyer M, Melko R and Carleo G 2018 Neural-network quantum state tomography *Nat. Phys.* **14** 447–50
- [46] Zhang Y and Kim E-A 2017 Quantum loop topography for machine learning *Phys. Rev. Lett.* **118** 216401
- [47] Torlai G *et al* 2019 Integrating neural networks with a quantum simulator for state reconstruction *Phys. Rev. Lett.* **123** 230504
- [48] Gong Z, Ashida Y, Kawabata K, Takasan K, Higashikawa S and Ueda M 2018 Topological phases of non-hermitian systems *Phys. Rev. X* **8** 031079
- [49] Bukov M, Day A G R, Sels D, Weinberg P, Polkovnikov A and Mehta P 2018 Reinforcement learning in different phases of quantum control *Phys. Rev. X* **8** 031086
- [50] Dalgaard M, Motzoi F, Sorensen J J and Sherson J 2020 Global optimization of quantum dynamics with alphazero deep exploration *Quantum Inform.* **6** 6
- [51] Torlai G and Melko R G 2017 Neural decoder for topological codes *Phys. Rev. Lett.* **119** 030501
- [52] Baireuther P, O'Brien T E, Tarasinski B and Beenakker C W J 2018 Machine-learning-assisted correction of correlated qubit errors in a topological code *Quantum* **2** 48
- [53] Sweke R, Kesselring M S, van Nieuwenburg E P L, and Eisert J 2018 Reinforcement learning decoders for fault-tolerant quantum computation (arXiv:1810.07207)
- [54] Andreasson P, Johansson J, Liljestränd S and Granath M 2019 Quantum error correction for the toric code using deep reinforcement learning *Quantum* **3** 183
- [55] Nautrup H P, Delfosse N, Dunjko V, Briegel H J and Friis N 2019 Optimizing quantum error correction codes with reinforcement learning *Quantum* **3** 215
- [56] Valenti A, van Nieuwenburg E, Huber S and Greplova E 2019 Hamiltonian learning for quantum error correction *Phys. Rev. Res.* **1** 033092
- [57] Hughes T L, Prodan E and Andrei Bernevig B 2011 Inversion-symmetric topological insulators *Phys. Rev. B* **83** 245132
- [58] Freed D S and Moore G W 2013 Twisted equivariant matter *Ann. Henri Poincaré* **14** 1927–2023
- [59] Kitaev A 2009 Periodic table for topological insulators and superconductors *Conf. Proc* vol 1134 pp 22–30
- [60] Yuan-Ming L and Lee D-H 2014 Inversion symmetry protected topological insulators and superconductors (<https://ui.adsabs.harvard.edu/abs/2014arXiv1403.5558L>) (arXiv:1403.5558)

- [61] Lutchyn R M, Bakkers E P A M, Kouwenhoven L P, Krogstrup P, Marcus C M and Oreg Y 2018 Majorana zero modes in superconductor—semiconductor heterostructures *Nat. Rev. Mater.* **3** 52–68
- [62] Budich J C and Ardonne E 2013 Equivalent topological invariants for one-dimensional majorana wires in symmetry class *D* *Phys. Rev. B* **88** 075419
- [63] Kitagawa T, Berg E, Rudner M and Demler E 2010 Topological characterization of periodically driven quantum systems *Phys. Rev. B* **82** 235114
- [64] Lindner N H, Refael G and Galitski V 2011 Floquet topological insulator in semiconductor quantum wells *Nat. Phys.* **7** 490–5
- [65] Kawabata K, Shiozaki K, Ueda M and Sato M 2019 Symmetry and topology in non-Hermitian physics *Phys. Rev. X* **9** 041015
- [66] Fidkowski L and Kitaev A 2011 Topological phases of fermions in one dimension *Phys. Rev. B* **83** 075103
- [67] <https://github.com/oleksandr-balabanov/ML-protocol-topo-indices-1d>