Research Article

Ekaterina Smolina, Lev Smirnov, Daniel Leykam, Franco Nori and Daria Smirnova*

Identifying topology of leaky photonic lattices with machine learning

Abstract: We show how machine learning techniques can be applied for the classification of topological phases in finite leaky photonic lattices using limited measurement data. We propose an approach based solely on a single real-space bulk intensity image, thus exempt from complicated phase retrieval procedures. In particular, we design a fully connected neural network that accurately determines topological properties from the output intensity distribution in dimerized waveguide arrays with leaky channels, after propagation of a spatially localized initial excitation at a finite distance, in a setting that closely emulates realistic experimental conditions.

Keywords: topological photonics; machine learning; non-Hermitian photonics; waveguide arrays

1 Introduction

Machine learning holds great promise for solving a variety of problems in nanophotonics. Rather than attempting to model the system of interest exactly from first principles (e.g., by solving Maxwell's equations), machine learning techniques aim to discover or reproduce key features of a system by optimizing parametrized models using a set of training data [1]. A trained model can often predict the properties of a device faster than conventional simulation techniques [2], [3]. Machine learning can also be used to solve the inverse problems of how to design a nanophotonic structure with desired functionalities, and how to reconstruct the parameters of a device using indirect measurements [4]–[8]. The latter is particularly important for nanophotonic devices, since structural parameters may differ substantially from the nominal design due to fabrication imperfections.

Recently developed topological photonic systems provide a useful testbed for better understanding the capabilities and limitations of machine learning approaches in nanophotonics [9], [10]. Topological photonic structures host robust edge states which are protected against certain classes of fabrication imperfections. This robustness is explained by the bulk-boundary correspondence, which relates the existence of localized boundary modes to non-local topological invariants expressed as integrals of a connection or curvature of the bulk modes [11]. While the direct measurement of a topological invariant entails the reconstruction of both the intensity and phase profiles of the bulk modes of a structure, machine learning models can perform supervised classification of topological phases using a limited set of observables [9].

In general, the performance of machine learning depends on both the quality and quantity of the data used to train the model. Supervised learning approaches, such as deep neural networks, typically require a huge quantity of labelled training data, which may be hard to come by. This has motivated recent interest in the use of unsupervised learning techniques such as manifold learning, which do not require labelled training data to distinguish topological phases [12]–[16]. Broadly speaking, these techniques are sensitive to sharp changes to observables that occur in the vicinity of topological phase transition points, and thus perform best when one has access to measurements from a large set of different model parameters, which is most feasible when the parameter controlling the phase transition is continuously tunable [14].

*Corresponding author: Daria Smirnova, Theoretical Quantum Physics Laboratory, Cluster for Pioneering Research, RIKEN, Wakoshi, Saitama 351-0198, Japan; and Research School of Physics, Australian National University, Canberra, ACT 2601, Australia, E-mail: daria.smirnova@anu.edu.au. https://orcid.org/0000-0001-8033-3427
Ekaterina Smolina and Lev Smirnov, Department of Control Theory, Nizhny Novgorod State University, Gagarin Av. 23, Nizhny Novgorod, 603950, Russia. https://orcid.org/0000-0002-2293-6534 (L. Smirnov)
Daniel Leykam, Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore. https://orcid.org/0000-0002-8588-9886
Franco Nori, Theoretical Quantum Physics Laboratory, Cluster for Pioneering Research, RIKEN, Wakoshi, Saitama 351-0198, Japan; Center for Quantum Computing (RQC), RIKEN, Wako-shi, Saitama 351-0198, Japan; and Physics Department, University of Michigan, Ann Arbor, MI 48109-1040, USA. https://orcid.org/0000-0003-3682-7432

The above methods also rely on prior knowledge of the characteristics of the physical system (such as its sizes, its internal structure and the parameters of the initial excitation), therefore, being not in line with a realistic experimental framework. Data quality and feature selection can have a significant impact on the machine learning-based reconstruction of topological phase diagrams [17]. For example, missing data arising from incomplete measurements or local perturbations to the data can act as adversarial attacks that fool neural network-based classifiers of topological phases into making incorrect predictions [18]. The existence of adversarial examples highlights the importance of taking platform-specific uncertainties and disorder into account in the selection and design of machine learning classifiers of topological phases.

The aim of this study is to investigate how common obstacles encountered in the characterization of nanophotonic devices – disorder, imperfect alignment, and access to a limited set of output observables – affect the performance of machine learning-based classification and clustering methods for topological phases. Specifically, we focus on the case of one-dimensional waveguide arrays which have provided a versatile platform for the investigation of topological effects in nanophotonics [19]–[21], considering the problem of predicting the existence or absence of edge states based on bulk intensity measurements of a finite lattice, i.e., measurements do not include edges of the lattice. First, we show that while curated input data can improve the performance of clustering, ambiguity in the training data (in the form of uncertainty in the alignment of the input waveguide) leads to incorrect cluster assignments, requiring the use of supervised learning techniques. We compare the performance of several supervised classification models, including a convolutional neural network, demonstrating the ability to predict the existence of different edge state configurations with high accuracy using bulk intensity measurements. Finally, we show the feasibility of transfer learning for sufficiently weak disorder strengths, i.e. maintaining accurate predictions of topological edge states using a model trained on disorder-free data. Our numerical results reveal the feasibility using machine learning techniques to distinguish nanophotonic topological phases using incomplete measurements.

The outline of this article is as follows: Section 2 reviews the properties of the leaky Su-Schrieffer-Heeger (SSH) tight binding model and introduces the datasets which will be used in our study. Section 3 presents the results of unsupervised clustering according to the edge state configuration using the t-distributed stochastic neighbor embedding (t-SNE) method. We compare the performance of different supervised learning techniques in Section 4. As an example of the feasibility of transfer learning we consider in Section 5 the classification performance for disordered waveguide arrays. We conclude with Section 6. The Supplementary Materials contain additional details on the tight binding model parameters, training data, and the employed machine learning models.

2 Model and dataset preparation

We consider light propagation in waveguide arrays governed by the paraxial wave equation,

$$i\frac{\partial \mathcal{E}}{\partial z} + \frac{1}{2k_0}\Delta_\perp \mathcal{E} + \frac{k_0 n_L(\mathbf{r}_\perp)}{n_0}\mathcal{E} = 0, \tag{1}$$

where $\mathcal{E}$ is the envelope of the optical wavepacket propagating along the z (waveguide) axis, $\mathbf{r}_\perp = (x, y)$ are the transverse coordinates, $k_0 = 2\pi n_0/\lambda$ is the wave number, $n_L(\mathbf{r}_\perp)$ is a perturbation of the refractive index forming the waveguide lattice, and n_0 is the background refractive index of the medium.

Formally, the final state after a propagation distance L can be obtained by projecting the input ($z = 0$) state $\mathcal{E}(0, \mathbf{r}_\perp)$ onto the propagation-invariant modes of the array $\phi_n(\mathbf{r}_\perp)$ with propagation constant β_n, i.e.

$$\mathcal{E}(L, \mathbf{r}_\perp) = \sum_n A_n e^{-i\beta_n L}\phi_n(\mathbf{r}_\perp), \tag{2}$$

where $A_n = \int d\mathbf{r}_\perp \phi_n^*(\mathbf{r}_\perp)\mathcal{E}(0, \mathbf{r}_\perp)$ are the amplitudes of the modes excited at the input ($z = 0$). The intensity of the final state

$$|\mathcal{E}(L, \mathbf{r}_\perp)|^2 = \sum_{mn} A_n A_m^* \phi_n(\mathbf{r}_\perp)\phi_m^*(\mathbf{r}_\perp)e^{i(\beta_m - \beta_n)L} \tag{3}$$

is sensitive to both the modal excitation amplitudes A_n and the propagation length L, so intensity measurements at a single L are generally insufficient to uniquely reconstruct the modal profiles, propagation constants, and topological invariants of the system.

Conventional schemes for predicting topological properties of the modes $\phi_n(\mathbf{r}_\perp)$ based on measuring field distributions (both amplitude and phase) require either the large L limit [22], [23] or measuring the evolution as a function of z [24], [25]. On the other hand, machine learning approaches can in principle infer topological properties using intensity measurements at a fixed propagation distance [26]–[28], at least given access to a sufficient amount of high-quality training data. However, the latter requires information about the symmetries, including the

Hamiltonian structure and geometry. We will employ no *a priori* knowledge of full lattice geometry in our approach.

As a specific example, in the following we consider the leaky Su-Schrieffer-Heeger waveguide lattice shown in
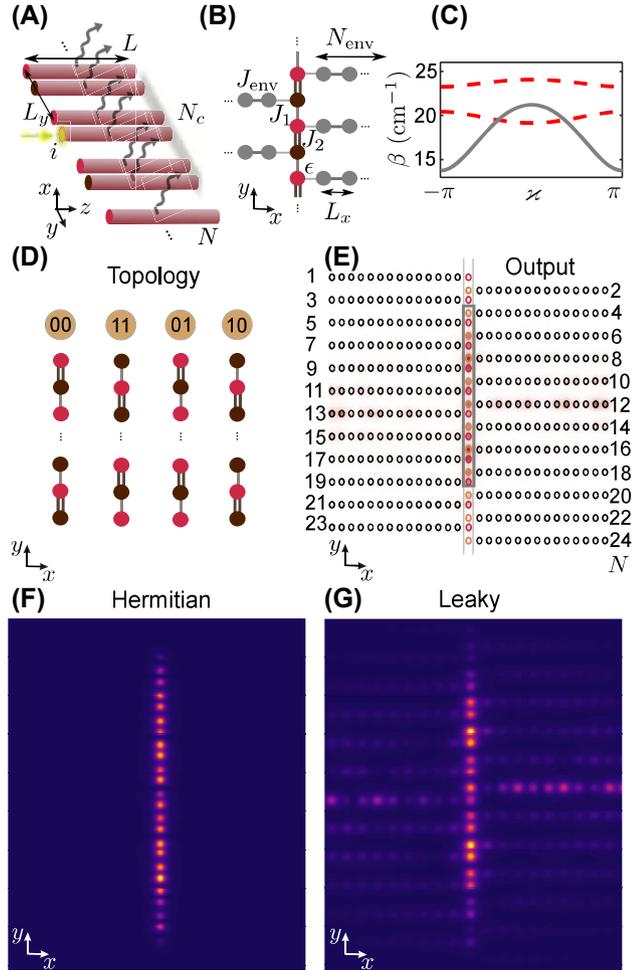
Figure 1: Photonic lattice platform. (A) Schematic of a dimerized lattice of single-mode dielectric waveguides with tunable radiative losses and a possible experiment: the waveguide indexed by *i* is excited at the input as indicated by a yellow circle, the intensity distribution is measured in the central area of N_c elements at the output of the sample (the gray rectangle) to generate a dataset for learning the topological properties. (B) Tight binding model visualization of the photonic lattice in (A). The red and brown circles depict the main array – a one-dimensional dimerised SSH-like array of coupled elements. Gray circles illustrate auxiliary arrays constituting leaky channels attached to the main array. The differing dashing between the elements denote different coupling strengths. (C) Band structures of the main (dashed red lines) and auxiliary (gray solid line) arrays in the designed leaky photonic lattice inscribed in glass. (D) Different configurations of the two edges in a finite SSH lattice. (E) The output intensity distribution (colored) overlaid with the proposed lattice cross-section. (F, G) Intensity distribution, numerically obtained in paraxial modeling at the output facet of the waveguide array for (F) the Hermitian (lossless) lattice and (G) the lattice with leaky channels.

Figure 1(A), a dimerized array composed of N leaky waveguides with elliptical cross-sections of semi-axes $a_{x,y}$ induced by the refractive index perturbations of magnitude n_A [23]. With increasing coupling between the structural elements, some supermodes of the lattice become radiative, acquiring a finite lifetime. The radiation losses can be fine-tuned by optimizing the effective potential of the environment and radiation channels. This will allow us to study how changes to the input dataset affect the performance of machine learning-based classification of the different topological phases of this lattice. One possible implementation of the radiation channels is by coupling the main array to auxiliary arrays, each consisting of N_{env} equidistantly spaced single-mode waveguides with an index contrast n_B, as shown in Figure 1(B and C). Examples of feasible parameters close to those employed in the experimental work Ref. [29] are given in Table 1. Ideally, N_{env} should be sufficiently large to prevent back-reflection from the ends of the environmental array. We set $N_{env} = 14$ that ensures experimental feasibility in terms of the overall sample's size and the propagation distances.

Provided only one band of the main array overlaps with the dispersion curve of side-coupled leaky channels, an initially localised excitation with a broad transverse wavenumber spectrum would undergo gradual radiation and decay during propagation. Therefore, only the top branch will remain populated after a certain propagation distance, making it possible to calculate the topological invariant of the band using the projector of the output

Table 1: Parameters of the designed leaky photonic lattice: semiaxes of elliptical single-mode waveguides $a_{x,y}$; center-to-center distances $d_{1,2}$ between waveguides along the vertical axis; center-to-center distance ρ between waveguides along the horizontal axis. Arrays of auxiliary waveguides are set aside from the main array at a distance d_ϵ. Here, λ is the operating wavelength, n_0 is the background refractive index of silica glass, $n_{A,B}$ are the perturbations of the refractive index inside the waveguides of the main array and arrays of the environment, respectively.

Parameter	Value
a_y	5.4 μm
a_x	4 μm
d_1	17 μm
d_2	23 μm
ρ	17 μm
d_ϵ	19 μm
n_0	1.47
n_A	1.2×10^{-3}
n_B	1.1×10^{-3}
λ	1030 nm

field distribution following the method used in Ref. [23]. However, this recipe generally requires knowing the complex-valued field, whereas phase retrieval could be a challenging task. For example, the commonly used scheme to recover the phase is iterative Gerchberg–Saxton algorithm. It relies on the complex-valued field by taking the intensity measurements in the real-space (picture) plane and the diffraction plane. The primary concern with this algorithm, aside from its high computational demand and resolution requirements on data, is selecting the initial guess appropriately. Moreover, phase retrieval does not converge well in one-dimensional case and behaves even worse for discrete systems. We will demonstrate the possibility to unravel topology of the sample lattice based solely on the output intensity profile in a roughly center-positioned floating window with the use of machine and deep learning methods.

To simplify propagation simulations, we constructed the tight binding model (TBM) corresponding to the schematic in Figure 1(B) and determined parameters of the effective Hamiltonian in compliance with the paraxial modeling,

$$i\frac{\partial \psi_m}{\partial z} = \hat{H}_0 \psi_m + \epsilon c_{m1}, \tag{4a}$$

$$i\frac{\partial c_{m1}}{\partial z} = \Delta c_{m1} + \epsilon \psi_m + J_{\text{env}} c_{m2}, \tag{4b}$$

$$i\frac{\partial c_{ml}}{\partial z} = \Delta c_{ml} + J_{\text{env}}(c_{ml-1} + c_{ml+1}), \; l = 2, \ldots N_{\text{env}} - 1, \tag{4c}$$

$$i\frac{\partial c_{ml}}{\partial z} = \Delta c_{ml} + J_{\text{env}}(c_{ml-1}), \quad l = N_{\text{env}}, \tag{4d}$$

where ψ_m and c_{ml} are the amplitudes of the optical field in the main array and in the leaky channels, respectively, $\hat{H}_0$ is the $N \times N$ Hamiltonian of the main array, made of the alternating nearest-neighbor (NN) coupling coefficients $J_{1,2}$, ϵ is the coupling strength between the main array and the environment, J_{env} is the NN hopping coefficient in leaky channels, and Δ is a detuning of the propagation constants.

The dispersion characteristics of the disconnected (at $\epsilon = 0$) uniform lattices representing the main (SSH) array and environment (env) are given by

$$\beta_{\text{SSH}}^{(\pm)}(\varkappa) = \pm\sqrt{J_1^2 + J_2^2 + 2J_1 J_2 \cos \varkappa}, \tag{5a}$$

$$\beta_{\text{env}}(\varkappa) = \Delta + 2J_{\text{env}} \cos \varkappa \tag{5b}$$

and plotted in Figure 1(C). Here we introduced $\varkappa$ as a variable $\varkappa = \kappa_y L_y$ or $\varkappa = \kappa_x L_x$ along y and x directions, respectively [see detailed derivations in

Section I in Supplementary Materials]. As deliberately ensured by design, the environmental array's dispersion curve fully intersects the lower band of the SSH lattice, meaning that only the lower band becomes lossy. Given dimerization, the main array is known to be topologically nontrivial for $J_1 < J_2$ and topologically trivial for $J_1 > J_2$.

To prepare a dataset, the TBM equations (4) were solved numerically. At the input, we excite a single waveguide designated as i in Figure 1(A). The use of a single-element input is justified by its wide spectrum, which allows populating both bands of the lattice. By iterating over parameters of the photonic lattice in the ranges indicated in Table 2, we accumulated data for the analysis of topology of the main array. We take into account that the lattice ends can be different, so that N can be odd. We select a sample window composed of a finite number N_c of the central waveguides in the main array. Thereby, we aim to solve the classification problem for a finite lattice sample, i.e., to distinguish between different configurations of the two edges based on the intensity distribution measured at the output of N_c central waveguides. The edge of the SSH main array can be either trivial (0) or non-trivial (1), depending on the lattice termination by strong or weak bond. The nontrivial edge supports a midgap topological edge state. This yields four classes in total: 00, 11, 10, 01. The four possible configurations are visualized in Figure 1(D): 01 (left trivial, right non-trivial), 11 (left nontrivial, right non-trivial), 10 (left non-trivial, right trivial), 00 (left trivial, right trivial). Note that such setup of the problem is different from that in Ref. [23], where both edges of the lattice had the same termination. Also, to calculate the field projector, the field distribution over all elements of the main array was used, that is $N_c = N$ with N even.

Table 2: Ranges of parameters used in data set preparation. Average values of the listed TBM parameters correspond to the physical quantities in Table 1, as established in paraxial modeling. $k = 2$, $p = 1$ in the nontrivial lattice ($J_1 < J_2$), and $k = 1$, $p = 2$ in the trivial lattice ($J_1 > J_2$). While preparing the datasets, $J_{1,2}$ were uniformly sampled from within the specified intervals for each vector.

Parameter	Range
J_k	[1.5; 2]
J_p	[0.4; 0.6]
J_{env}	[1.7; 2]
ϵ	[0.8; 1]
Δ	[−3.3; −3.5]
L	[2.6; 10.6]
N	[20; 26]
N_{env}	14
N_c	16

Our previous work [23] presented a proposal for calculating the topological invariant (Zak phase) for this lattice (of classes 00 or 11) using the field projector of the output distribution. This procedure is summarized in Figure 2. By analyzing the complex-valued field distribution [note Figure 2(C and D) only shows the intensity, with remarkable differences apparent in the initial evolution stage and near the edges], we compute the Zak phase, which asymptotically approaches π in the nontrivial configuration [see Figure 2(A)] (the orange curve approaches the black dotted line), provided the leaky channels are introduced. At distances $4\,\text{cm} < z < 9\,\text{cm}$ the upper band is completely depopulated as a result of leakage. This depopulation is also evident in the total wavepacket norm, which converges towards $1/2$. However, when the propagation distance is increased beyond $z > 9$ cm, reflections occur from the ends of the finite environment array and the main lattice, resulting in an increase in the total wavepacket norm [see Figure 2(B)], rendering the method inapplicable. Thus, accurate reconstruction of the topological invariant requires either a large lattice or a well-controlled propagation length to avoid reflections off the ends.

3 Unsupervised learning

To begin, we perform the preliminary analysis of the prepared datasets using the t-SNE (t-distributed stochastic neighbor embedding) method [see Section II in Supplementary Materials]. t-SNE is a nonlinear dimensionality reduction algorithm which learns a low-dimensional embedding of the input data; points within the input data set that are close to each other will remain close to each other in the embedded space [30]. Ideally, a vector will be most similar to others obtained from the same lattice configuration, resulting in visible clustering in the low-dimensional embedding.

In this approach, we work with the intensity distribution within $N_c = N$ elements ($N = 22$ or 23, to be more specific), and assume that the pumped waveguide can be shifted from the center of the lattice. Figure 3 shows t-SNE maps of the system with fixed $L = 7.6$ cm, $N = 22$ (23) and two different positions of the initially excited waveguide. The distributions in Figure 3 represent the clustered embeddings of the high-dimensional data points in a lower-dimensional space, generated using t-SNE. These embeddings visualize the separability and distribution of the data clusters. In the Hermitian case (leakage disabled), the different classes become mixed up in the embedded space; whereas in the case of a lattice with leaky channels, they do not. This qualitatively agrees with the theory in Ref. [23], specifically that the different phases will exhibit distinct intensity distributions in their bulk.

However, as soon as we introduce uncertainty, such as the position of the initial excitation, the topological classes are no longer clearly separable: in the Hermitian case different classes become mixed up [Figure 3(C)], whereas in the leaky lattice too many clusters are obtained [Figure 3(F)]. Consequently, unsupervised methods are no longer applicable.

Figure 4 presents the statistic analysis of the data used for (C, F) panels of Figure 3. This visualization shows that classes 01 and 00, 10 and 11 can be grouped pairwise. However, the classes with dissimilar edge topologies (01 and 10) are differentiated from the classes with the identical edge topologies (00 and 11) by odd N, due to distinct input vector lengths (the 23th waveguide for which case is shown shaded). This postprocessing also reveals significant overlaps of the intensity bars for 00 and 11 classes in each waveguide of the Hermitian SSH lattice, while the bars overlap less in the leaky lattice forming shifted

Figure 2: Propagation dynamics. (A, B) Evolution characteristics of the field in the main array in the lattice with fixed parameters obtained in the TBM of the nontrivial SSH array with (gold curves) and without (green curves) leaky channels. The Zak phase at $4\,\text{cm} < z < 9\,\text{cm}$ converges to the quantised π value, provided $N_\text{env} = 14$ elements in leaky channels. (C, D) Field evolution in N elements of the main array assembled in a trivial (C) and nontrivial (D) configuration with fixed parameters of the lattice. The gray line on the right side marks the area of N_c central waveguides, the intensity of which is fed to the input of the neural network.

Figure 3: t-SNE maps of the system having 4 topological classes depending on its 2 edges: (A–C) Hermitian lattice, (D–F) lattice with leaky channels. The waveguide excited at the input is indexed by i. (A, B, D, E) correspond to the case of single-waveguide excitation: (A, D) $i = 11$ is odd, (B, E) $i = 12$ is even, (C, F) the excited waveguide is randomly chosen within a dimer. For each point in the two-dimensional parameter space there is a corresponding intensity distribution vector of dimension $N = 22$ (or $N = 23$), depending on the topological class. The four classes are color-coded: 00 (blue), 11 (red), 10 (green), 01 (black).

dimerized patterns, a feature to be noticed by the neural network.

4 Supervised learning

For supervised classification of the four topological classes, we apply machine (K-nearest neighbors (KNN), support vector machine (SVM), decision tree) and deep (multilayer perceptron (MLP), convolutional neural network (CNN)) learning methods (see details in the Supplementary Materials, Section III). The numerical experiments were carried out with varying parameters: propagation distance L, total number of waveguides N, number of the central waveguides in a sample window N_c. The input waveguide i can be shifted by 1 from the center of the array, according to the expression ceil($N/2 + l$), where l can be 0 or 1. For each L we obtain a dataset of 32000 intensity vectors. Accordingly by a parameter, subsets from the whole data set can be grouped. Let us examine the accuracy of classification depending on different parameters. The metric we use for this non-binary classification problem is the accuracy, defined as the percentage of correct model predictions,

$$\text{Accuracy} = \frac{\sum_{i=1}^{n} \mathbb{1}[p_i = y_i]}{n}, \qquad (6)$$
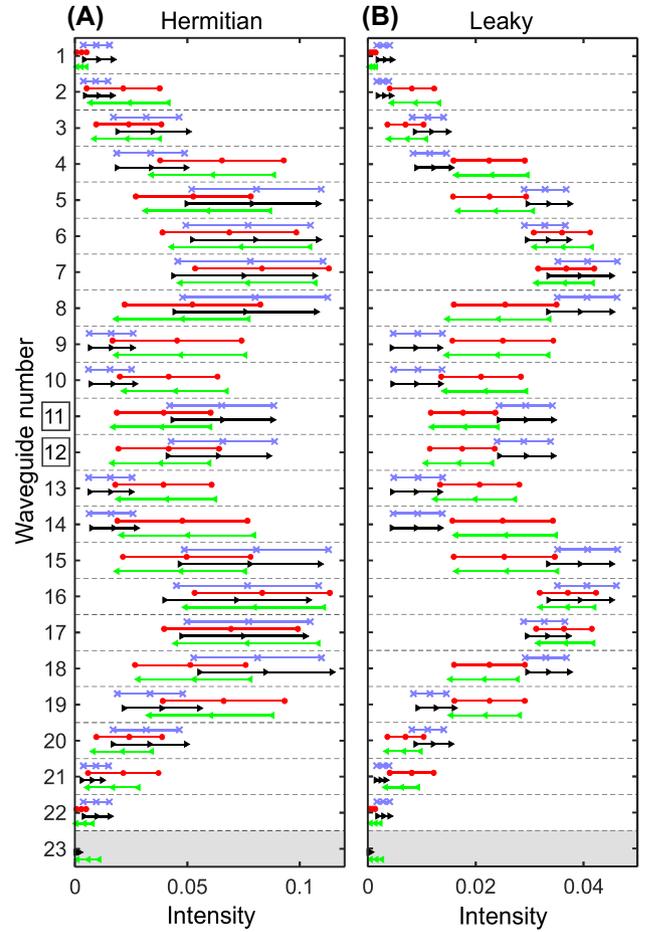
Figure 4: Statistical characteristics of intensity distributions in waveguides. The datasets were prepared for the Hermitian (A) and leaky (B) cases assuming two possible positions $i = 11, 12$ of the initial excitation at $L = 7.6$ cm. The mean value is indicated by markers in the middle of horizontal lines, while the standard deviation is represented by the borders of the lines. The classes are color-coded: 00 (blue crosses), 11 (red circles), 01 (black right-facing triangles), 10 (green left-facing triangles). The total number of waveguides N is 22 (even) for classes 00 and 11, and 23 (odd) for classes 01 an 10.

where p_i and y_i are the predicted and the correct answer, respectively, and $\mathbb{1}$ is an indicator function equal to one if the condition is met and zero otherwise.

Figure 5(A) illustrates how the accuracy of the supervised learning techniques varies with the parameter L. The accuracy increases as the propagation distance increases. When the value of L is small, theoretical predictions cannot distinguish between different topological phases, and all methods show similar accuracy plateaus in their graphs. Further, the accuracy of machine learning methods increases with increasing L, see Figure 5(A). At the same time, the theoretical curve for the Zak phase in the nontrivial case ceases to converge to the quantized invariant value π for $L = 10.6$ cm [see Figure 2(B)], while the power
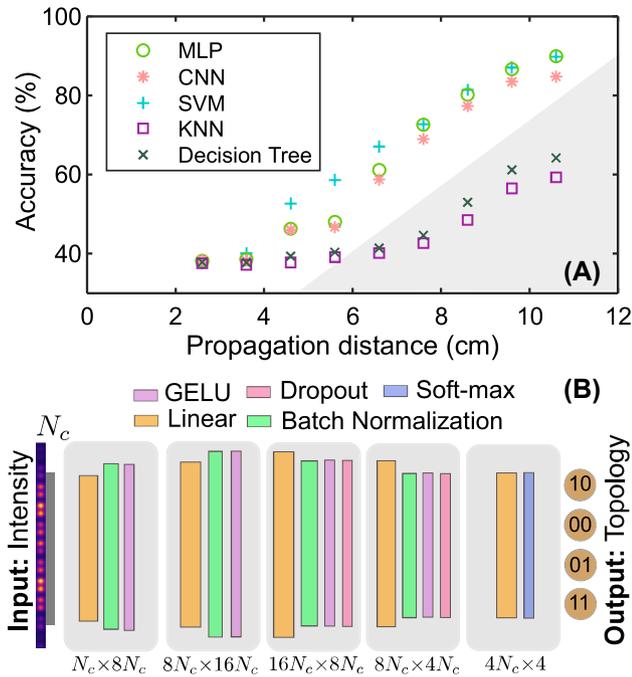
Figure 5: Supervised machine learning for determining topological classes. (A) Accuracy of supervised learning methods as a function of the propagation distance L. (B) Scheme of the convolutional neural network, which takes the intensity distribution at $z = L$ as the input and determines topology of the lattice edges, $N_c = 16$.

Figure 6: Analysis of MLP accuracy. (A) Accuracy of classification by deep learning methods depending on parameters: the total number of waveguides N and the number of the central waveguides N_c involved in the training. (B) Theoretical dependence of the Zak phase on the propagation distance and N_c in the nontrivial lattice of $N = 22$ elements.

in the main array tends to grow and exceeds one half [see Figure 2(A)]. This is explained by reflection from the boundaries of leaky channels, as the field returns back to the main array. The requirement to know both the intensity and phase at the output in the method of Ref. [23] is replaced by statistical information from dynamics, but only intensity distributions at fixed L.

Machine learning methods perform better for larger L. This may be due to the fact that, as soon as the radiation reaches the edges, to distinguish the trivial case from the non-trivial one, we can consider not only bulk properties but also the edges themselves, and machine learning methods allow us to take this effect into account. For instance, the trivial and non-trivial cases are even visually distinguishable in the dynamics shown in Figure 2(C and D): in the non-trivial case the bulk modes poorly couple to waveguides at the edges. Note that if we increase the number of auxiliary waveguides N_{env}, the theoretical power curve will exhibit convergence to 0.5, but the reflection off the main array edges will still manifest at larger propagation distances. Thus, neural network methods are applicable in a wider range of cases than the theoretical scheme based on the projector calculation.

Based on results summarised in Figure 5(A), we conclude that classical machine learning methods show lower

accuracy compared to neural networks and support vector machine (SVM). One of the two most promising models, the MLP method, was chosen for more thorough examination in Figure 6.

As noted above, training was held using $N_c < N$ central waveguides. Figure 6(A) shows the dependence of the classification accuracy on the number of central waveguides while in training batches all L were involved. In the initially proposed theoretical scheme, we calculated the field projector for $N_c = N$ elements, but we can formally calculate it for any $N_c < N$, as shown in Figure 6(B). The Zak phase is seen to converge better to the correct quantised value for larger N_c, and this condition is also necessary to increase the accuracy of machine learning algorithms: in Figure 6(B) the precision increases as the N_c/N ratio increases.

To better understand the performance of the supervised classification approach at distinguishing the different edge types, we compare topological SSH lattice with even number of elements and its non-topological counterpart, where dimerization is stipulated by the alternating difference in propagation constants (Δ_1 and $\Delta_2 = -\Delta_1$), whereas
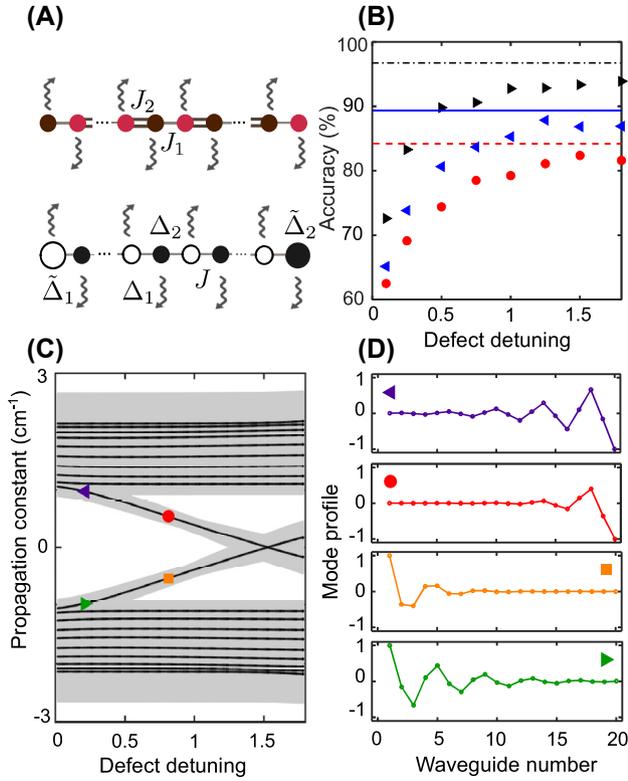
Figure 7: Comparing machine learning classification for topological and non-topological lattices. (A) Schematics of the topological (upper row) dimerized array and the non-topological (lower row) dimer lattice with defect potentials $\tilde{\Delta}_{1,2}$ at the edges. (B) The accuracy of the neural network trained for the non-topological case for different values of the edge defect detuning q_1, introduced as $\tilde{\Delta}_{1,2} = \Delta_{1,2}(1 - q_1)$, and different propagation distances $L = 7.6$ cm (red dots), $L = 8.6$ cm (blue left-facing triangles), $L = 10.6$ cm (black right-facing triangles). For comparison, the colored horizontal lines depict the accuracy in the topological case for the corresponding L. (C) The band structure of the finite non-topological lattice depending on the defect detuning, at the fixed number of elements within the main array $N = 22$. The shading shows bands for all possible coupling coefficients, J, and detunings, $\Delta_1 = -\Delta_2$, that were utilized to generate the datasets. (D) Profiles of the modes bound to the ends of the non-topological lattice. Colors and shapes of the markers in (C) in the representative spectral positions correspond to the profiles in (D).

the coupling between neighboring elements is uniform and equal to J, as schematically shown in Figure 7(A). To prepare the corresponding datasets, parameters of the non-topological lattice (Δ_1 and J) are chosen such that its band structure coincides with the topological one (see Supplementary Materials, Section I). We introduce trivial edge defects as detunings of the propagation constant in the edge elements. Thereby, the defect potential for the left end is $\tilde{\Delta}_1 = \Delta_1(1 - q_1)$, whereas the defect potential for the right end is $\tilde{\Delta}_2 = \Delta_2(1 - q_2)$. We compare the accuracy of the neural network at three propagation distances [see Figure 7(B)]

for the topological SSH array and non-topological array with the edge defects in distinguishing the two classes: both edges either support confined solutions (class 11) or not (class 00). We find that for small amplitudes of the defect the accuracy for the case of the non-topological lattice is small compared to the topological one, since the defect is not connected to its bulk properties (unlike in the topological case), but bulk modes also change when the defect amplitude becomes large, leading to an increase in the model accuracy.

5 Disorder and transfer learning

Transfer learning refers to the use of a model trained on one set of data to make accurate predictions on a new task. Here we consider the performance of models trained on ideal data in classifying data obtained from different model parameters. If the quality metric falls slightly, we can conclude that the model has a generalization ability. This is particularly important in the context of nanophotonic circuits, where inevitable disorder will lead to sample-to-sample variations of device parameters.

First, we note that the generalization ability is not observed for the parameter L, and the accuracy drops significantly when testing on L different from the propagation distance used for the training data. On the other hand, we observe generalization over some N, that corresponds to attaching dimers to both edges of the main array, stipulated by the fact that such an addition of elements does not qualitatively change the topology of the lattice (see the cross-validation control map for parameter N in Supplementary Materials, Section IV).

Next, we examine a transfer learning approach that allows for the reuse of pretrained models at a fixed propagation distance of $L = 10.6$ cm [referring to the last point in Figure 5(A)] on models with disorder. We introduce perturbations into the SSH Hamiltonian coefficients of two types: off-diagonal disorder in the inter-site coupling strengths and on-site disorder in the propagation constants. The former preserves the chiral symmetry, while the latter breaks the symmetry and spoils the topological protection. Incorporating disorder involves adding random variables to the coefficients of the Hamiltonian. For example, the off-diagonal disorder perturbs each coupling coefficient by the random variable $l\langle d\rangle\text{mean}(J_1, J_2)$, where l is uniformly distributed in the range $[-1/2, 1/2]$ and $\langle d\rangle$ is the disorder strength. This is a chiral type disorder in the sense that the Hamiltonian describing the disordered system respects the chiral symmetry, thus its topological edge states will remain at zero energy. We train the neural network using a non-disordered

array and test it on the disordered lattice. We have identified a range of disorder strengths in which the previously trained neural network can operate with high confidence.

To quantify the impact of the disorder on the data, we evaluate the similarity between the output intensities. Specifically, we compute the output fields $\psi_m(\langle d \rangle, i)^{1,2}$, where the superscripts 1 and 2 correspond to diagonal and off-diagonal disorders, respectively, and i represents the number of the specific disorder realization. We then introduce the intensity overlap as $\mathcal{O}^{1,2}(\langle d \rangle, i) = \sum_m |\psi_m(\langle d \rangle, i)^{1,2})|^2 \cdot |\psi_m^0|^2$, where summation is taken over waveguides of the main array and ψ_m^0 is the output distribution in the disorder-free case. This overlap measures the similarity between the two distributions. It is a useful quantity to study the effect of disorder on the output of a system, as it allows us to quantify how much the output changes due to disorder. To plot the overlap measure, we calculate $\mathcal{O}^{1,2}(\langle d \rangle, i)$ over 4000 disorder realizations for each

of the values of $\langle d \rangle$. To standardize the plotted functions, we divide them by the value of $\mathcal{O}^{1,2}(\langle d \rangle, i)$ when $\langle d \rangle$ is zero. This normalization process allows us to compare the variability of the overlap measure across different scenarios. The dotted areas in Figure 8(A) represent the corresponding ranges. Note, we have rescaled the diagonal disorder strength $\langle d_{\text{diag}} \rangle = 4 \langle d_{\text{off-diag}} \rangle$ such that for a given $\langle d \rangle$ the two forms of disorder have a similar effect on the overlap measure.

To demonstrate transfer learning for disordered arrays, we train the neural network using a non-disordered array and test it for the disordered lattice [see Figure 8(B)], the ranges of parameters as in Table 2. The accuracy curves are similar for both types of disorder, showing a decrease in accuracy as the disorder amplitude increases. Expanding the range of the overlap measure results in a significant change in the output intensity, which ultimately leads to a sharp decline in the classification accuracy.

To estimate confidence of the trained neural network, we study the output of the last layer [see Figure 5(B)] in detail. Softmax function returns probabilities of four classes. Here we fix the class 00 (both ends are trivial), but the results are comparable for the other classes as well. If the model assigns a high probability to a particular class, it is more confident in that prediction than if it assigns a lower probability.

We create a set of test vectors for each disorder amplitude and select vectors that have the highest probability of belonging to the class 00. If this vector indeed belongs to the class 00, we label the probability as true; otherwise, it is labeled as false. And then we average false and true answers to plot Figure 8(C and D). Interestingly, as the accuracy of the neural network decreases, its level of certainty in both accurate and inaccurate responses increases. In other words, the neural network will more confidently give the wrong answer as the disorder strength is increased, indicating that the fabrication disorder can act as an adversarial perturbation.

Thus, the neural network does not inherit the robustness of the underlying topological phase. Rather, the similar performance for the symmetry-preserving and symmetry-breaking disorders suggests that the network is picking out features of the intensity as a proxy for the topological invariant, not the topological invariant itself. One potential solution to this problem is to consider a more sophisticated network architecture such as an autoencoder (where the middle layers are narrower, leading to a loss of information), which forces the network to learn the global features characterizing the different classes [31], improving the

Figure 8: The transfer learning approach for disordered lattices. (A) Overlap measure variation induced by the disorder: shaded areas are ranges of variance due to disorder over an ensemble of 4000 disorder realizations (green is for diagonal disorder, gray for off-diagonal disorder), red asterisks and green dots are mean values. All parameters of the lattice are fixed. (B) Transfer learning for the disordered lattice. We train neural network in the absence of disorder $\langle d \rangle = 0$ and test the prediction accuracy for different values of disorder. All parameters of the lattice are varied according to Table 2. (C, D) Probability assigned to false (C) and true (D) answers of the neural network for different values of disorder (green bars are for diagonal disorder, red bars are for off-diagonal disorder).

robustness to noise and experimental imperfections. This will be an interesting direction for future work.

6 Conclusions

We have studied the performance of a variety of machine learning techniques at distinguishing different topological phases of leaky photonic lattices using measurements of the bulk intensity profile after a fixed propagation distance. First, we found that uncertainty in the initial conditions (such as the excited waveguide) reduces the quality of unsupervised clustering, leading to either mixing between different classes or the prediction of too many classes. We then compared the performance of different supervised learning methods, finding that high accuracy can be achieved for sufficiently large propagation distances. The classification accuracy can be further improved by increasing the number of bulk waveguide intensities used. Other approaches to enhancing accuracy may involve acquiring more intensity images of the same system and increasing the number of initially excited waveguides. For example, we may employ a double-input excitation with a varying phase difference and record several intensity measurements. Finally, we studied the transfer learning ability of neural network-based classifiers. While the accuracy drops significantly if the network is trained on data obtained using a different propagation distance, the networks can accurately classify data from systems with sufficiently weak disorder, thus avoiding extensive training on each new system. Our approach for classifying lattices based on incomplete measurements can be further developed to solve a more general problem of reconstruction of the lattice Hamiltonian with some *a priori* knowledge of its symmetries in various fields including photonics, condensed matter physics, and quantum computing.

References

[1] G. Carleo, *et al*., "Machine learning and the physical sciences," *Rev. Mod. Phys.*, vol. 91, no. 4, p. 045002, 2019.

[2] P. R. Wiecha and O. L. Muskens, "Deep learning meets nanophotonics: a generalized accurate predictor for near fields and far fields of arbitrary 3d nanostructures," *Nano Lett.*, vol. 20, no. 1, pp. 329−338, 2020.

[3] M. Chen, *et al*., "High speed simulation and freeform optimization of nanophotonic devices with physics-augmented deep learning," *ACS Photonics*, vol. 9, no. 9, pp. 3110−3123, 2022.

[4] D. Melati, *et al*., "Mapping the global design space of nanophotonic components using machine learning pattern recognition," *Nat. Commun.*, vol. 10, no. 1, p. 4775, 2019.

[5] I. Malkiel, M. Mrejen, L. Wolf, and H. Suchowski, "Machine learning for nanophotonics," *MRS Bull.*, vol. 45, no. 3, pp. 221−229, 2020.

[6] S. So, T. Badloe, J. Noh, J. Bravo-Abad, and J. Rho, "Deep learning enabled inverse design in nanophotonics," *Nanophotonics*, vol. 9, no. 5, pp. 1041−1057, 2020.

[7] P. R. Wiecha, A. Arbouet, C. Girard, and O. L. Muskens, "Deep learning in nano-photonics: inverse design and beyond," *Photonics Res.*, vol. 9, no. 5, pp. B182−B200, 2021.

[8] G.-X. Liu, *et al*., "Inverse design in quantum nanophotonics: combining local-density-of-states and deep learning," *Nanophotonics*, vol. 12, no. 11, pp. 1943−1955, 2023.

[9] J. Yun, S. Kim, S. So, M. Kim, and J. Rho, "Deep learning for topological photonics," *Adv. Phys.: X*, vol. 7, no. 1, p. 2046156, 2022.

[10] L. Pilozzi, F. A. Farrelly, G. Marcucci, and C. Conti, "Machine learning inverse problem for topological photonics," *Commun. Phys.*, vol. 1, no. 1, p. 57, 2018.

[11] T. Ozawa, *et al*., "Topological photonics," *Rev. Mod. Phys.*, vol. 91, no. 1, p. 015006, 2019.

[12] J. F. Rodriguez-Nieva and M. S. Scheurer, "Identifying topological order through unsupervised machine learning," *Nat. Phys.*, vol. 15, no. 8, pp. 790−795, 2019.

[13] M. S. Scheurer and R.-J. Slager, "Unsupervised machine learning and band topology," *Phys. Rev. Lett.*, vol. 124, no. 22, p. 226401, 2020.

[14] E. Lustig, O. Yair, R. Talmon, and M. Segev, "Identifying topological phase transitions in experiments using manifold learning," *Phys. Rev. Lett.*, vol. 125, no. 12, p. 127401, 2020.

[15] Y. Che, C. Gneiting, T. Liu, and F. Nori, "Topological quantum phase transitions retrieved through unsupervised machine learning," *Phys. Rev. B*, vol. 102, no. 13, p. 134213, 2020.

[16] Y. Long and B. Zhang, "Unsupervised data-driven classification of topological gapped systems with symmetries," *Phys. Rev. Lett.*, vol. 130, no. 3, p. 036601, 2023.

[17] Y. Zhang, P. Ginsparg, and E.-A. Kim, "Interpreting machine learning of topological quantum phase transitions," *Phys. Rev. Res.*, vol. 2, no. 2, p. 023283, 2020.

[18] H. Zhang, *et al*., "Experimental demonstration of adversarial examples in learning topological phases," *Nat. Commun.*, vol. 13, no. 1, p. 4993, 2022.

[19] A. Blanco-Redondo, "Topological nanophotonics: toward robust quantum circuits," *Proc. IEEE*, vol. 108, no. 5, pp. 837–849, 2020.

[20] D. T. H. Tan, "Topological silicon photonics," *Adv. Photonics Res.*, vol. 2, no. 9, p. 2100010, 2021.

[21] J. Gao, *et al*., "Observation of Anderson phase in a topological photonic circuit," *Phys. Rev. Res.*, vol. 4, no. 3, p. 033222, 2022.

[22] M. S. Rudner and L. S. Levitov, "Topological transition in a non-Hermitian quantum walk," *Phys. Rev. Lett.*, vol. 102, no. 6, p. 065703, 2009.

[23] D. Leykam and D. A. Smirnova, "Probing bulk topological invariants using leaky photonic lattices," *Nat. Phys.*, vol. 17, no. 5, pp. 632–638, 2021.

[24] Y. Wang, *et al*., "Direct observation of topology from single-photon dynamics," *Phys. Rev. Lett.*, vol. 122, no. 19, p. 193903, 2019.

[25] V. V. Ramasesh, E. Flurin, M. Rudner, I. Siddiqi, and N. Y. Yao, "Direct probe of topological invariants using Bloch oscillating quantum walks," *Phys. Rev. Lett.*, vol. 118, no. 13, p. 130501, 2017.

[26] P. Zhang, H. Shen, and H. Zhai, "Machine learning topological invariants with neural networks," *Phys. Rev. Lett.*, vol. 120, no. 6, p. 066401, 2018.

[27] B. S. Rem, *et al*., "Identifying quantum phase transitions using artificial neural networks on experimental data," *Nat. Phys.*, vol. 15, no. 9, pp. 917–920, 2019.

[28] N. L. Holanda and M. A. R. Griffith, "Machine learning topological phases in real space," *Phys. Rev. B*, vol. 102, no. 5, p. 054107, 2020.

[29] S. Mukherjee and M. C. Rechtsman, "Observation of Floquet solitons in a topological bandgap," *Science*, vol. 368, no. 6493, pp. 856–859, 2020.

[30] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, p. 11, 2008.

[31] N. Käming, *et al*., "Unsupervised machine learning of topological phase transitions from experimental data," *Mach. Learn.: Sci. Technol.*, vol. 2, no. 3, p. 035037, 2021.