Machine Prediction of Topological Transitions in Photonic Crystals

Bei Wu,[1,†] Kun Ding,[2,3,†] C.T. Chan,[2] and Yuntian Chen[1,4,*]

[1] *School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan 430074, China*

[2] *Department of Physics and William Mong Institute of Nano Science and Technology, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China*

[3] *The Blackett Laboratory, Department of Physics, Imperial College London, London SW7 2AZ, United Kingdom*

[4] *Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, China*

(Received 19 July 2020; revised 2 September 2020; accepted 8 September 2020; published 19 October 2020)

Neural networks based on machine learning can interpolate well within the training dataset, but their ability to extrapolate is severely limited by fundamental issues such as the bias-variance trade-off. Here we introduce the concept of an operator parameter space consisting of physical entities encoded with Maxwell's equations to improve the networks' capability to generalize beyond their training set. We illustrate the idea with photonic crystals, and show that the network trained with operator parameters yields remarkably accurate predictions of the topological transitions both within and beyond the training physical space. Such concepts can be generalized to higher-dimensional wave structures by choosing the appropriate operator parameters.

DOI: 10.1103/PhysRevApplied.14.044032

I. INTRODUCTION

Neural network (NN) based machine learning (ML), benefitting from the development of statistical physics as well as recent advances in big data and computer science, has far-reaching implications and applications in various branches of physics [1–6]. Notably, NN is not only useful in exploring the known data without knowledge of underlying laws but also has relevant roles in presenting the physical laws or inverse-design macroscopic problems across broad branches in physics, including optics and condensed-matter physics [7–15]. As for supervised learning, the generalization ability (i.e., interpolation and extrapolation) is rather fundamental yet desirable in ML. The ability to extrapolate is highly desirable for many applications, in particular the discovery of alternative materials and inverse design.

The fundamental challenge of achieving extrapolation in supervised learning is the bias-variance trade-off between model complexity and training dataset size [16], where the extrapolation region contains zero training-data points. The bias represents the intrinsic error for a given model and can be reduced by the increased model complexity, while the variance is the error caused by sampling noises and can be alleviated with simpler models and larger amounts of the training data. Therefore, for a given problem with a finite number of training data, the prediction capability of a NN is determined by the model complexity that balances the bias and variance, and thus begs for an optimal model to get the best prediction ability. With a limited number of input data carrying simple input information, even a model that optimizes the bias-variance trade-off does not carry sufficient information to capture the complexity of a physical problem. To mitigate this problem, we can increase the information content of the input dataset so that the complexity of the input dataset matches the complexity of the physical problem at hand. In this regard, the best-possible model may be the one that carries parameters that have physical meanings. A few attempts of predicting physical quantities beyond the training dataset have been made in the context of condensed-matter physics by using explicit Hamiltonians, with emphasis on the classifications of the topological order or phase transitions [17–26]. These studies focused mainly on discrete systems with a finite number of bands by using simplified tight-binding models, and thus may not be applicable for realistic physical systems where the tight-binding approximation is inadequate or the explicit Hamiltonians may not be obtained easily. More importantly, the underlying principle that leads to the successful extrapolation based on a Hamiltonian is not clarified from the perspective of the model complexity and the finite data volume.

In this paper, we introduce the concept of an operator parameter space and a physics-adapted NN to enrich the complexity of the input training data, which significantly

*yuntian@hust.edu.cn

†Bei Wu and Kun Ding contributed equally to this work.

improve the prediction capability of a NN, especially the extrapolation capability. We demonstrate in one-dimensional (1D) photonic crystals (PCs) that within the scheme of supervised learning the physics-adapted NN trained in the operator parameter space is able to predict the topological invariant of photonic bands with an accuracy rate 99.72% for the training dataset and 97.32% for the test dataset. Our method can predict topological phase transitions of 1D PCs with relative prediction errors below 0.5% for geometric configurations that lie outside the physical parameter space of training dataset. Moreover, more-sophisticated physics-adapted NNs can detect the boundaries of the topological transition in the physical parameter space, where a portion of the trained data near that boundary is excluded purposely. All these show that with the operator parameter space and physics-adapted NN, the connection between training and test datasets, which have different physical parameters but obey the same physical equations, is restored by the underlying physical laws embedded in the operator parameters.

II. OPERATOR PARAMETER SPACE AND PHYSICS-ADAPTED NEURAL NETWORK

We illustrate the idea of operator parameters through the example of learning the topological invariants of photonic bands. The PC considered is shown at the top left of Fig. 1(a) with its unit cell marked by a dashed yellow box, where $\varepsilon_a(\varepsilon_b)$ and $d_a(d_b)$ are the dielectric constant and thickness of layer a (layer b), $\Lambda = d_a + d_b$ is the lattice constant. The physical parameters $(\varepsilon_a, \varepsilon_b, d_a, d_b)$ have three degrees of freedom and constitute the physical parameter space of PCs, as shown at the bottom left of Fig. 1(a). To avoid the band crossing and a small band gap, two constraints on the range of the parameter space are imposed: $\varepsilon_a \geq \varepsilon_b + 0.5$ and $0.1 \leq d_a/\Lambda \leq 0.9$. We divide the space into four nonoverlapping sectors labelled as "Train-1 (Train-2)," "Test-1," "Test-2," and "Test-3" datasets for later use (see details in Supplemental Material, Sec. I [27]). The dataset Train-i is for training the NN and the dataset Test-i is for testing the well-trained NN. The right panel in Fig. 1(a) shows the dispersion of the lowest four bands calculated by the transfer-matrix method (TMM) [28,29]. The geometric Zak phase [30] of each band is either 0 or π, which corresponds to a winding number of 0 or 1 due to inversion symmetry. For training purposes, the four-binary-number labeling of the winding numbers of the lowest four bands is translated into a decimal integer; for example, 1011 in Fig. 1(a) to 11 for the one-hot labeling of the output.

As shown in the middle of Fig. 1(a), we first use the physical parameters together with the calculated labels to train the fully connected NN (FNN), which works reasonably well inside the domain of the training dataset and fails completely beyond the training-data range (see results in Sec. II in Supplemental Material [27]). The

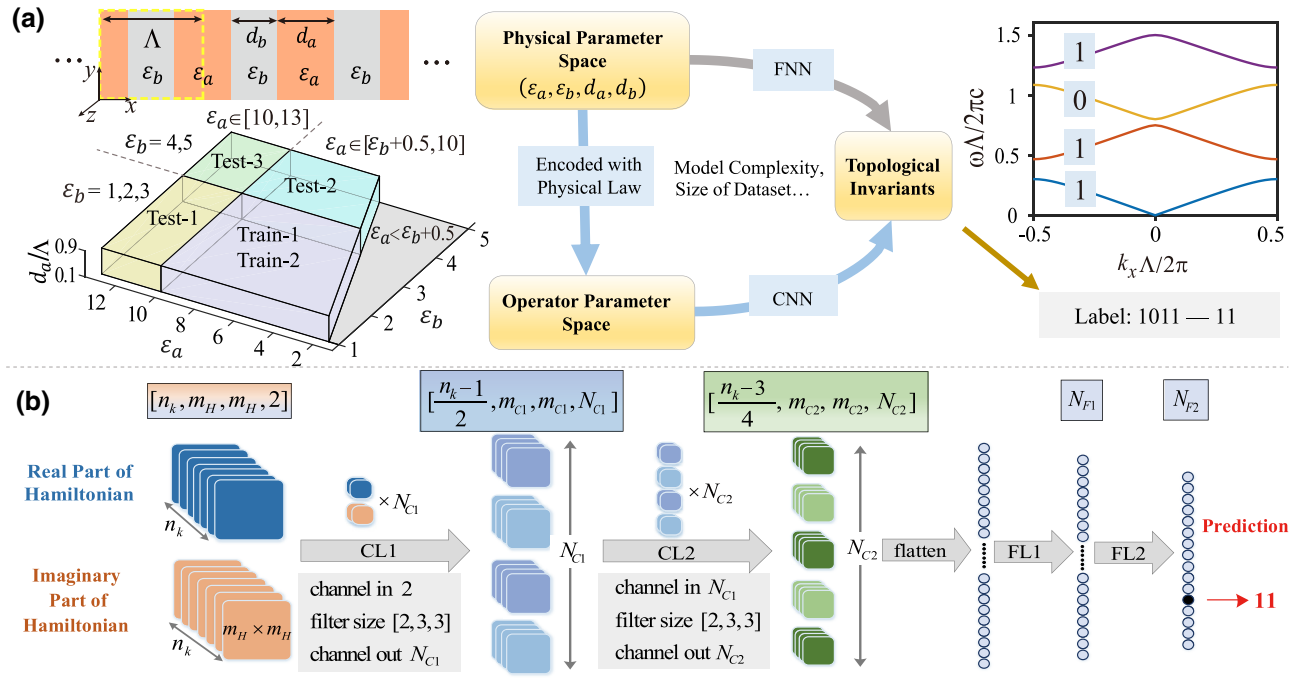

FIG. 1. (a) Workflow of the NNs to predict the topological invariants (right panel) of PCs (left panel). The FNN is trained with the physical parameters $(\varepsilon_a,\ \varepsilon_b, d_a, d_b)$, while the CNN is trained by the operator parameters that are encoded with a physical law based on the physical parameter space. (b) The CNN with two CLs and two FLs.

poor extrapolation indicates that the FNN trained by the physical parameters with a limited volume of data (10 800 samples) does not match the complexity of the given problem; namely, the band topology here. To solve the aforementioned problem, we encode the governing physical laws into the physical parameters, coined the "operator parameters." We use a physics-adapted NN to achieve the augmented prediction capability of the NN, as shown in the middle of Fig. 1(a). The problem considered concerns the band topology, and thus a straightforward choice of the operator parameters is a set of k-parameterized Bloch Hamiltonians based on the physical parameters of PCs encoded with the structure of Maxwell's equations in momentum space. Here we use the plane-wave method (PWM) [31,32] to generate the operator parameter space as the input training data, which includes all the information on the PCs, that is, physical parameters $(\varepsilon_a, \varepsilon_b, d_a, d_b)$, as well as the inner structure dictated by Maxwell's equations. In the following, we give the concrete form of the Hamiltonian of the PCs generated by the PWM.

In 1D PCs with the dielectric function as $\varepsilon(x)$ as shown in the left of Fig. 1(a), the band properties can be obtained by solving the wave equation. Because of the periodicity of the material, that is, $\varepsilon(x+\Lambda)=\varepsilon(x)$, $\varepsilon^{-1}(x)$ is also periodic and can be expanded in a Fourier series: $\varepsilon^{-1}(x)=\sum_{m=-m_H}^{m_H} a_m \exp[i(2\pi m/\Lambda)x]$, where $\{a_m\}$ are the Fourier coefficients and $m=0, \pm1, \ldots, \pm m_H$. With Bloch's theorem, the magnetic field of the eigenmodes in the 1D crystal can be expressed by $H(x)=H_{k_x}(x)=v_{k_x}(x)\exp(ik_x x)$, where v_{k_x} is the periodic Bloch wave function and satisfies the relationship $v_{k_x}(x+\Lambda)=v_{k_x}(x)$ with one fixed-wave Bloch k vector k_x, and $n_k=11$ is the number of the discrete sampled Bloch k vectors within the Brillouin zone. The magnetic field can also be described as $H(x)=\sum_{n=-p}^{p} v_n \exp[i(k_x+(2\pi n/\Lambda))x]$ on the basis of the plane-wave expansion, where $n=0, \pm1, \ldots, \pm p$, and $m_H=2p+1$ is the number of truncated plane waves. Substituting the Fourier expansions of $\varepsilon^{-1}(x)$ and $H(x)$ into the wave equation, we obtain the following eigenvalue equations for $\{v_n\}$:

$$H_{m_H\times m_H}\begin{bmatrix} v_{-p} \\ v_{-p+1} \\ \vdots \\ v_{p-1} \\ v_p \end{bmatrix} = \frac{\omega^2}{c^2}\begin{bmatrix} v_{-p} \\ v_{-p+1} \\ \vdots \\ v_{p-1} \\ v_p \end{bmatrix}, \qquad (1)$$

where h_{ij} is the matrix element of the Hamiltonian $H_{m_H\times m_H}$ and $h_{ij}=a_{i-j}[k_x+2\pi(-p+i)/\Lambda][k_x+2\pi(-p+j)/\Lambda]$. The Zak phases of the isolated bands can be calculated by the fields on the basis of the PWM. In our work, we truncate the number of plane waves $m_H=61$ to balance the numerical errors between the PWM and the TMM (see details in Sec. III in Supplemental Material [27]) as well as the size of the NN input. Thus, by encoding with the physical law, the Hamiltonian represented by a rank-4 tensor $[n_k, m_H, m_H, 2]$ is the operator parameters of one PC, and the last dimension, 2, represents the real and imaginary parts of the Hamiltonian. In such an operator parameter space, classifying topological invariants of photonic bands is equivalent to the image classification of Hamiltonian patterns, in which the convolutional NN (CNN) can be used to capture both local and global features of the k-space "image." As a simple example shown in Fig. 1(b), we consider a CNN workflow containing two convolutional layers (CLs) and two fully connected layer (FLs). The two CLs contain N_{C1} and N_{C2} kernels of size $[2,3,3]$, as well as a max-pooling operation with size $[1,2,2]$, and the two FLs have N_{F1} and N_{F2} neurons, with $N_{F2}=16$ meaning the output layer has 16 neurons. The input is the Hamiltonian generated by the PWM represented as a rank-4 tensor $[n_k, m_H, m_H, 2]$. The one-hot encoding output is a rank-1 tensor with shape [16], which has one-to-one correspondence to the binary labelling. For instance, 11 means that only the 12th neuron of the output is 1, while the other neurons are 0. The loss function to evaluate whether the network has been trained well is the cross entropy between the output of the CNN and the labeling. In principle, more layers could be used to train the network to obtain better performance.

Once the CNN is set up, the Hamiltonians of PCs and the four-band-Zak-phase labeling obtained by the TMM within two different training datasets, Train-1 and Train-2, are fed into this supervised network. As an example, Fig. 2(a) shows the distribution of the labelling in one training dataset, Train-1, plotted as functions of d_a and ε_a for the $\varepsilon_b=2$ plane within the Train-1 dataset. The six colors of the solid circles represent the six labels defined in Fig. 2, corresponding to the following six cases 1-0001, 5-0101, 7-0111, 9-1001, 11-1011, 15-1111. Each case has 600 randomly distributed samples within one fixed ε_b plane and there are three planes $(\varepsilon_b=1,2,3)$ within the Train-1 dataset, leading to 10 800 $(600\times6\times3)$ samples in total. The five solid lines depict the boundaries separating different labels determined by the TMM, labeled

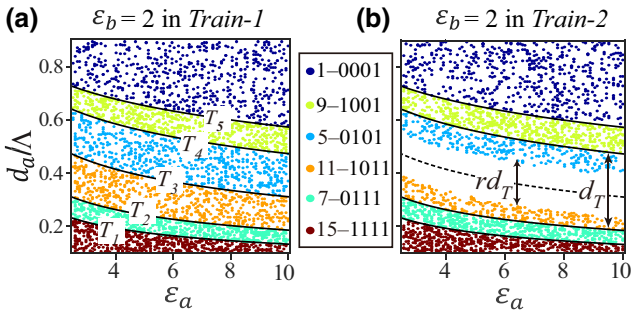

FIG. 2. Distribution of different topological phase sequences of samples in the $\varepsilon_b=2$ plane within (a) the Train-1 dataset and (b) the Train-2 dataset with $r=0.5$.

as T_1 to T_5. These lines perfectly match the borders of the six different colors calculated with the PWM, showing that the TMM and the PWM give consistent results for winding numbers for the lowest four bands. To examine the capability and possibility of our proposed CNN, we conceive a different training dataset; that is, the Train-2 dataset, which is identical to the Train-1 dataset except that a portion of samples in the neighborhood of the T_3 line (dashed black line) is omitted purposely as shown in Fig. 2(b). The fraction of the omitted data in Fig. 2(b) is $r = 0.5$, which is defined as the width of rd_T of the omitted data divided by the difference of d_a at T_2 and T_4 (d_T) for any ε_a. For self-testing purposes, each training dataset is separated into two subsets, 70% for training and 30% for testing. We follow the standard routine of supervised learning to train the network [33–35]; that is, running the learning and testing procedure simultaneously and iteratively until the desired accuracy is fulfilled, which can be saved and used to predict topological transitions with parameters outside the training dataset (i.e., the three test datasets).

III. PREDICTING TOPOLOGICAL TRANSITIONS BEYOND TRAINING DATASETS

Firstly, we use the Train-1 dataset to train the network, which consists of three CLs with $[N_{C1}, N_{C2}, N_{C3}] = [10, 20, 20]$, and three FLs with $[N_{F1}, N_{F2}, N_{F3}] = [300, 100, 16]$. The trained network is used to predict the boundaries of the parameter space [for examples, see the black lines $T_1 - T_5$ in Fig. 2(a)], where the topological transition occurs. The well-trained CNN is able to predict the topological invariants of the PCs in all datasets with an accuracy of 99.72% for the training dataset and 97.32% for the test datasets. Figure 3(a) by the purple background shows the physical parameter range for the $\varepsilon_b = 2$

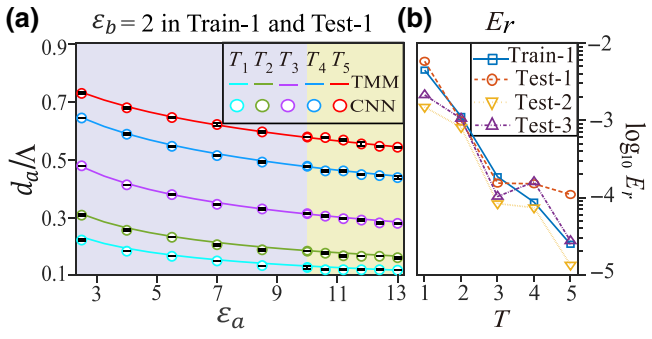

FIG. 3. (a) The transition lines in the $\varepsilon_b = 2$ plane calculated by the well-trained CNN (circles) and the TMM (solid lines). The purple and yellow backgrounds show the physical parameter ranges of Train-1 and Test-1, respectively. (b) The relative errors E_r between the CNN predictions and the TMM results for the training dataset and three test datasets.

plane within the Train-1 dataset, which has six topological labels, 1-0001, 5-0101, 7-0111, 9-1001, 11-1011, and 15-1111, and five topological transition boundaries, $T_1 - T_5$, determined by the TMM. To examine the ability of the trained network, we use the CNN to predict the boundaries $T_1 - T_5$ as follows. We scan d_a for fixed $(\varepsilon_a, \varepsilon_b)$ and take the mean value of two adjacent d_a with different Zak-phase prediction, labeling the predicted transition point as d_{CNN}. To avoid numerical instability, we repeat such calculations 10 times and take the mean value $\overline{d_{\mathrm{CNN}}}$ as the predicted results and the standard deviation as the prediction error. The topological transition points within the Train-1 dataset predicted by the well-trained CNN in the $\varepsilon_b = 2$ plane are shown by open circles $(\varepsilon_a, \overline{d_{\mathrm{CNN}}})$ with error bars in Fig. 3(a). The size of the open circles can cover the error bar, indicating that the predicted results obtained with the well-trained CNN are stable. For comparison, the topological transition lines calculated by the TMM in the $\varepsilon_b = 2$ plane are plotted by solid lines in Fig. 3(a). The excellent agreement shows the well-trained CNN can reproduce the five topological phase transitions within the training dataset.

We then examine the extrapolation ability of the trained network by predicting $T_1 - T_5$ beyond the training range of the physical parameter space. If the trained CNN can predict four-band-Zak-phase labeling and associated phase transitions in all test datasets, Test-1, Test-2, and Test-3, then we are extrapolating to parameter values outside the physical parameter space $(\varepsilon_a, \varepsilon_b, d_a)$ of the training dataset. Figure 3(a) by the yellow background shows a comparison of the predicted transition lines (open circles) and the TMM results (solid lines) within the Test-1 dataset. The detailed comparison results of all test datasets are depicted in Fig. S5 in Supplementary Material [27], which shows prediction accuracy comparable to that in Fig. 3(a). To quantify the prediction accuracy of our network, we define the relative prediction error of the topological transitions of our network with reference to the TMM results as $E_r = (1/N)\sum_{i=1}^{N}[|(\overline{d_{\mathrm{CNN}}} - d_{\mathrm{TMM}})|/d_{\mathrm{TMM}}]^2$, where N is the total number of data points in each phase-transition surface T_i. $\overline{d_{\mathrm{CNN}}}$ (d_{TMM}) corresponds to the topological transition critical point calculated by the CNN (TMM). In Fig. 3(b), the relative errors for the Train-1, Test-1, Test-2, and Test-3 datasets as a function of five transition surfaces are plotted by squares, circles, inverted triangles, and upright triangles, respectively. The overall prediction errors are less than 0.5%, indicating high predication accuracy and stability of the trained CNN in the training and test datasets. Evidently, the predication in the test datasets Test-1, Test-2, and Test-3 is not a simple interpolation based on the topological invariants within the Train-1 dataset because the physical parameter spaces of these test datasets are beyond the training data. The remarkable agreement between the TMM and the CNN results reveals that our network indeed decodes the characteristics

of Maxwell's equations encoded in the Hamiltonians and captures the essence of topological properties of photonic bands of different PCs.

IV. PREDICTION OF UNCERTAIN TRANSITIONS

We now examine whether the CNN can predict the uncertain topological transitions in the physical parameter space, and thus we conceive the Train-2 dataset defined in Fig. 2(b), which is identical to the Train-1 dataset except that a portion of samples in the neighborhood of the T_3 line (dashed black line) is omitted purposely, as shown by the white area in Fig. 2(b). The fraction of the omitted data in Fig. 4(a) is controlled by the parameter $r = 0.2$ as illustrated by the white region.

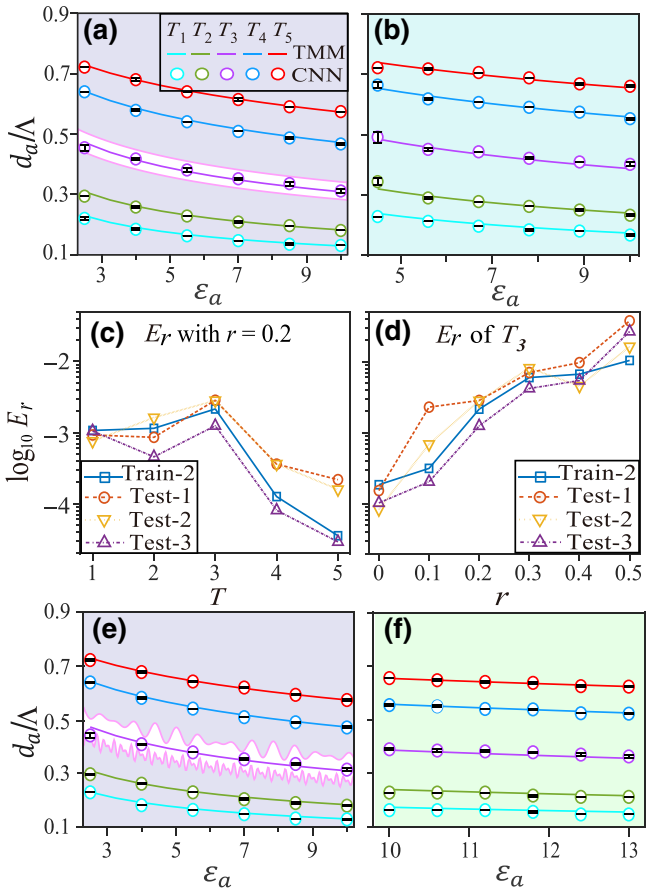

FIG. 4. (a),(b)The topological-phase-transition lines for (a) $\varepsilon_b = 2$ in the Train-2 dataset with $r = 0.2$ and (b) $\varepsilon_b = 4$ in the Test-2 dataset predicted by the CNN (trained by Train-2) are shown by open circles with error bars. (c),(d) The relative error E_r for (c) all transitions and (d) T_3 as a function of the fraction r. (e),(f) The topological-phase-transition lines for (e) $\varepsilon_b = 2$ in the training dataset and (f) $\varepsilon_b = 4$ in Test-3 dataset predicted by the CNN (trained by Train-2 with $r = 0.3$ and random disturbance on the boundary) are shown by open circles with error bars. All the solid lines in (a),(b),(e),(f) are calculated by the TMM.

The architecture of the CNN used here is similar to but slightly more complex than the one trained by Train-1, and it contains three CLs with $[N_{C1}, N_{C2}, N_{C3}] = [24, 32, 48]$ and three FLs with $[N_{F1}, N_{F2}, N_{F3}] = [500, 300, 16]$. We train this CNN by the Train-2 dataset with $r = 0.2$ and apply it to the four datasets. Figure 4(a) displays the phase-transition lines $T_1 - T_5$ predicted by the CNN (open circles) and calculated by the TMM (solid lines) in the $\varepsilon_b = 2$ plane within the Train-2 dataset, and Fig. 4(b) depicts the results in the $\varepsilon_b = 4$ plane within the Test-2 dataset. Remarkably, the transition lines predicted by our network agree well with the TMM results, especially the accurate predication of the T_3 transition line, highlighting the extrapolation capability of our network. It can also be seen that the predicted results have a better performance on T_1, T_2, T_4, and T_5 than on T_3 due to the missing data near the T_3 line in the Train-2 dataset (see Fig. S8 in Supplemental Materials [30]). Like Fig. 3(b), Fig. 4(c) shows the relative errors E_r of the phase-transition surfaces calculated by the CNN trained by the Train-2 dataset with $r = 0.2$. E_r at T_3 is indeed the largest due to the omitted data near T_3. Nevertheless, the CNN trained by the Train-2 dataset also works with the overall E_r smaller than 0.2%. For comparison, we also use a FNN for the same task, and the FNN works well in the parameter space of the training dataset, but the predicted results beyond the training dataset deviate greatly from the TMM results, indicating the FNN trained by physical parameters has poor extrapolation ability (see Supplemental Material, Sec. II, for details [27]). The failure of the FNN in extrapolation further confirms the fact that a NN trained by the limited data is hard to really match the problem even with good performance in the training range.

To further verify the extrapolation ability of our method, we first calculate the relative error of the T_3 transition surface shown in Fig. 4(d) as the fraction r of the omitted training data varies. As expected, the training error is relatively small for a small value of r. However, the relative error of the overall predicted transition surface for the same CNN architecture increases with r, and the extrapolation capability is compromised since more data are omitted (see the detailed error analysis in Sec. IV b in Supplemental Material [27]). Secondly, we add random perturbations to the boundary of the omitted data area of the Train-2 dataset with $r = 0.3$ (called "Train-3"), as shown by the edge of the white area in Fig. 4(e). We then train the CNN with the Train-3 dataset, and Figs. 4(e) and 4(f) show the topological-phase-transition lines predicted by this trained CNN (open circles) and calculated by the TMM (solid lines) in the $\varepsilon_b = 2$ plane within the training dataset and in the $\varepsilon_b = 4$ plane within the Test-3 dataset. We also see that the predicted transition lines $T_1 - T_5$ obtained with our network trained by Train-3 agree well with the TMM results, indicating our network is robust and stable, even though there are complex disturbances.

V. DISCUSSION AND CONCLUSION

The extrapolation capability demonstrated in Figs. 3 and 4 highlights the advantage of using the operator parameter space, which encodes physical parameters with the Fourier-transformed Maxwell's equations. Such an ability is not achievable with a FNN trained directly from physical parameters. While our demonstration is focused on the lowest four bands with the help of the PWM, our approach is neither limited to the lowest bands nor tied to the basis used. The extension to complex systems is straightforward, and higher predication accuracy is achievable by optimizing the modal basis and the complexity of the NN. For example, cylindrical waves or vector spherical waves are more-efficient operator parameters for two-dimensional cylinders or three-dimensional spheres than plane waves. However, because of the supervised-learning framework, our method can predict the topological invariants known in the training dataset for a wider range in the physical parameter space, but cannot recognize new patterns that are not given in the training dataset. In other words, we realize the extrapolation of known topological invariants in the physical parameter space rather than the recognition of k-space patterns representing new topological invariants (see more details in Sec. V in Supplemental Material [30]).

As a concluding remark, the physics-adapted NN together with the operator parameters provides an alternative solution to deal with realistic photonic structures and paves the way toward the intelligent design of photonic devices. In broader scenarios, the underlying principle of achieving a greater prediction capability or extrapolation for limited training data is to increase the data complexity to match the optimal network model, which should be matched to the target problem itself too. Such augmented prediction capability is particularly useful in condensed-matter physics, optics, and many other inverse-design problems, where the governing laws at the microscopic level are known yet the macroscopic behaviors are untrackable.

ACKNOWLEDGMENTS

This work was supported by Natural National Science Foundation of China (Grant No. 11874026), and the Research Grants Council of Hong Kong (Grants No. AoE/P-02/12 and No. 16304717). K.D. acknowledges funding from the Gordon and Betty Moore Foundation.

[1] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Boston, MA, USA, 2015), pp. 1–9.

[2] K. He, X. Zhang, S. Ren, and J. Sun, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Las Vegas, NV, USA, 2016), pp. 770–778.

[3] D. Silver, et al., Mastering the game of go with deep neural networks and tree search, Nature **529**, 484 (2016).

[4] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, Mastering the game of go without human knowledge, Nature **550**, 354 (2017).

[5] A. Graves, A. Mohamed, and G. Hinton, in *2013 IEEE International Conference on Acoustics* (Speech and Signal Processing, 2013), pp. 6645–6649.

[6] C. Chen, A. Seff, A. Kornhauser, and J. Xiao, in *2015 IEEE International Conference on Computer Vision (ICCV)* (IEEE, Santiago, Chile, 2015), pp. 2722–2730.

[7] Z. Liu, D. Zhu, S. P. Rodrigues, K.-T. Lee, and W. Cai, Generative model for the inverse design of metasurfaces, Nano Lett. **18**, 6570 (2018).

[8] J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark, and M. Soljačić, Nanophotonic particle simulation and inverse design using artificial neural networks, Sci. Adv. **4**, eaar4206 (2018).

[9] D. Liu, Y. Tan, E. Khoram, and Z. Yu, Training deep neural networks for the inverse design of nanophotonic structures, ACS Photonics **5**, 1365 (2018).

[10] I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, and H. Suchowski, Plasmonic nanostructure design and characterization via deep learning, Light Sci. Appl. **7**, 1 (2018).

[11] L. Pilozzi, F. A. Farrelly, G. Marcucci, and C. Conti, Machine learning inverse problem for topological photonics, Commun. Phys. **1**, 57 (2018).

[12] Y. Long, J. Ren, Y. Li, and H. Chen, Inverse design of photonic topological state via machine learning, Appl. Phys. Lett. **114**, 181105 (2019).

[13] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, Deep learning with coherent nanophotonic circuits, Nature Photon **11**, 441 (2017).

[14] Y. Qu, L. Jing, Y. Shen, M. Qiu, and M. Soljačić, Migrating knowledge between physical scenarios based on artificial neural networks, ACS Photonics **6**, 1168 (2019).

[15] T. Yan, J. Wu, T. Zhou, H. Xie, F. Xu, J. Fan, L. Fang, X. Lin, and Q. Dai, Fourier-Space Diffractive Deep Neural Network, Phys. Rev. Lett. **123**, 023901 (2019).

[16] P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, Low-variance introduction to machine learning for physicists, Phys. Rep. **810**, 1 (2019).

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

[18] N. Sun, J. Yi, P. Zhang, H. Shen, and H. Zhai, Deep learning topological invariants of band insulators, Phys. Rev. B **98**, 085402 (2018).

[19] T. Ohtsuki and T. Ohtsuki, Deep learning the quantum phase transitions in random two-dimensional electron systems, J. Phys. Soc. Jpn. **85**, 123706 (2016).

[20] T. Ohtsuki and T. Ohtsuki, Deep learning the quantum phase transitions in random electron systems: Applications to three dimensions, J. Phys. Soc. Jpn. **86**, 044708 (2017).
[21] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Learning phase transitions by confusion, Nat. Phys. **13**, 435 (2017).
[22] Y. Zhang and E.-A. Kim, Quantum Loop Topography for Machine Learning, Phys. Rev. Lett. **118**, 216401 (2017).
[23] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).
[24] C. Wang, H. Zhai, and Y.-Z. You, Emergent schrödinger equation in an introspective machine learning architecture, Sci. Bull. **64**, 1228 (2019).
[25] D.-L. Deng, X. Li, and S. Das Sarma, Machine learning topological states, Phys. Rev. B **96**, 195145 (2017).
[26] D.-L. Deng, X. Li, and S. Das Sarma, Quantum Entanglement in Neural Network States, Phys. Rev. X **7**, 021021 (2017).
[27] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevApplied.14.044032 for more computational details.
[28] A. Yariv and P. Yeh, *Optical Waves in Crystals: Propagation and Control of Laser Radiation* (Wiley, New York, 1984).
[29] M. Xiao, Z. Q. Zhang, and C. T. Chan, Surface Impedance and Bulk Band Geometric Phases in One-Dimensional Systems, Phys. Rev. X **4**, 021017 (2014).
[30] J. Zak, Berry's Phase for Energy Bands in Solids, Phys. Rev. Lett. **62**, 2747 (1989).
[31] K. M. Ho, C. T. Chan, and C. M. Soukoulis, Existence of a Photonic Gap in Periodic Dielectric Structures, Phys. Rev. Lett. **65**, 3152 (1990).
[32] K. Sakoda, *Optical Properties of Photonic Crystals*, 2nd ed. (Springer, Berlin; New York, 2005).
[33] N. Buduma and N. Locascio, *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*, First edition (O'Reilly Media, Sebastopol, CA, 2017).
[34] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature **521**, 436 (2015).
[35] A. Géron, *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, First edition (O'Reilly, Beijing Boston Farnham Sebastopol Tokyo, 2017).