Machine learning guided discovery of stable, spin-resolved topological insulators

Alexander C. Tyner
Nordita, KTH Royal Institute of Technology and Stockholm University 106 91 Stockholm, Sweden
and Department of Physics, University of Connecticut, Storrs, Connecticut 06269, USA

(Received 20 February 2024; revised 3 May 2024; accepted 31 May 2024; published 24 June 2024)

Identification of a nontrivial $\mathbb{Z}_2$ index in a spinful two-dimensional insulator indicates the presence of an odd, quantized (pseudo)spin-resolved Chern number, $\mathcal{C}_s = (\mathcal{C}_\uparrow - \mathcal{C}_\downarrow)/2$. However, the statement is not biconditional. An odd spin-Chern number can survive when the familiar $\mathbb{Z}_2$ index vanishes. Identification of solid-state systems hosting an odd, quantized $\mathcal{C}_s$ and trivial $\mathbb{Z}_2$ index is a pressing issue due to the potential for such insulators to admit band gaps optimal for experiments and quantum devices. Nevertheless, they have proven elusive due to the computational expense associated with their discovery. In this work, a neural network capable of identifying the spin-Chern number is developed and used to identify the first solid-state systems hosting a trivial $\mathbb{Z}_2$ index and odd $\mathcal{C}_s$. We demonstrate the potential of one such system, Ti_2CO_2, to support Majorana corner modes via the superconducting proximity effect.

DOI: 10.1103/PhysRevResearch.6.023316

I. INTRODUCTION

The comprehensive cataloguing of materials supporting nontrivial band topology, as diagnosed via elementary band representations (EBRs) and other efficient protocols relying on analysis of symmetry eigenvalues, represents a remarkable milestone [1–5]. However, it is now clear that these works cannot represent a comprehensive analysis of bulk topology. This is due to the existence of symmetry nonindicative phases (SNIPs) [6–15]. Importantly, it has been demonstrated that there exist a number of two-dimensional (2D) higher-order insulators which fall in this category for which the ground-state bulk invariant is a nonzero spin-Chern number, as defined by Prodan [6].

The spin-Chern number, $\mathcal{C}_s = (\mathcal{C}_\uparrow - \mathcal{C}_\downarrow)/2$, has been shown to be robust in both the presence of impurity effects and the absence of spin-rotation symmetry. It is protected by both the energetic bulk gap as well as the spin gap. The spin gap is identified by constructing the projected spin operator (PSO), $PSO = \mathbf{P}\hat{s}\mathbf{P}$, where $\mathbf{P}$ is the projector over occupied states and $\hat{s}$ is the preferred spin direction. In the presence of spin-rotation symmetry, the eigenvalues of the PSO are fixed to be ± 1. When spin-rotation symmetry is broken, the eigenvalues adiabatically deviate, but as long as the spectra of the PSO remains gapped the spin-Chern number is robust. It is further shown by Lin *et al.* [16] that for an odd, ground-state spin-Chern number, only the bulk gap need be maintained to protect the band topology. Under this formulation, the $\mathbb{Z}_2$ index and mirror Chern number both represent special cases of the spin-Chern number, however this statement is not biconditional. A finite spin-Chern number need not imply a finite $\mathbb{Z}_2$ index or mirror Chern number.

Identification of a nontrivial spin-Chern number in systems lacking enhanced symmetry has proven computationally demanding, particularly in the context of *ab initio* simulations of many-band systems [15–17]. While density-functional-theory software exists for efficient symmetry analysis of the wave functions and computation of Wannier center spectra [18–21], these systems often fall under the category of higher-order topological insulators (HOTIs) and can thus be "invisible" to the $\mathbb{Z}_2$ index. In correspondence with a trivialized $\mathbb{Z}_2$ index, the surface spectra does not support gapless modes which can be used to diagnose topology. In certain cases, corner localized states can be used to diagnose topology [22–26], however corner modes are notoriously difficult to identify as their presence depends on the geometry of the sample and in the absence of chiral symmetry they are not pinned to zero energy, often becoming hidden among bulk states.

Alternatively, in a recent work by Tyner and Goswami [17], the proposal of Qi and Zhang [27] and Ran *et al.* [28] to utilize magnetic flux tubes (π-flux vortices) as real-space probes of both Chern and spin-Chern number was expanded to *ab initio* simulations. It was shown to be robust in both symmetry indicative and nonindicative phases. An automated workflow was developed to scan 141 two-dimensional, spinful ($\mathcal{T}^2 = -1$) insulators. The result was identification of 21 quantum spin-Hall insulators, falling under the category of SNIPs, but supporting a nontrivial, even spin-Chern number. While automated and basis agnostic, this work remained computationally demanding. Furthermore, a two-dimensional insulator supporting odd spin-Chern number and trivial $\mathbb{Z}_2$ index remained elusive.

This situation motivates a different search strategy relying on development of a neural network capable of diagnosing the presence of bulk topological order in $\mathcal{T}$-preserving systems

with $\mathcal{T}^2 = -1$. While a machine learning based approach does not eliminate the need for density functional theory computations as they are necessary to validate the results, this approach allows for creation of an optimal list of candidate materials that can be validated within a reasonable timescale. Furthermore, prior intuition or knowledge of the electronic properties of a given material is not necessary to determine whether it is an optimal candidate for nontrivial topology, allowing for a more diverse material landscape to be considered. This neural network should posses the capability to identify topological order of multiple types, including (a) first-order insulators supporting a nontrivial Fu-Kane index, (b) topological crystalline insulators admitting finite mirror Chern numbers, and (c) higher-order or generalized quantum spin-Hall insulators (QSHIs). Such broad capabilities are crucial, as any system supporting a nontrivial Fu-Kane index or mirror Chern number also supports a nontrivial (pseudo)spin-resolved Chern number. Therefore, in the above list, systems belonging to sets (a) and (b) are subsets of (c).

In this work we construct two convolutional neural networks (CNNs), both relying on voxel encoding of the crystal structure [29,30]. The first produces a binary classification, dictating whether the spin-Chern number is zero or finite. The second produces a multiclass classification, dictating whether the spin-Chern number is zero, odd, or nonzero and even. These CNNs are then applied to a set of experimentally synthesized, two-dimensional insulators in the computational two-dimensional materials database (C2DB) [31,32]. Materials of interest predicted to support nontrivial, odd ground-state spin-Chern number are subsequently selected and analyzed in depth, leading to identification of the first two-dimensional insulator supporting $|\mathcal{C}_{s,G}| = 1$ and a trivial $\mathbb{Z}_2$ index. This work serves as a proof that the apparent computational expense present in identification of spin-resolved band topology can be overcome through the use of machine learning techniques.

There is a particular need for such efficient tools to diagnose (pseudo)spin topology given the ongoing experimental interest in two-dimensional heterostructures and twisted materials. In these systems additional degrees of freedom such as valley can give rise to pseudospin-resolved topology [33,34].

II. CONSTRUCTION OF TRAINING DATASET

The training dataset consists of spinful ($\mathcal{T}^2 = -1$), two-dimensional insulators. The corresponding ground-state spin-Chern number is labeled as zero, odd, or finite and even. This dataset is made distinct from existing databases by enhanced criteria for labeling a system as topologically trivial. Following Ref. [17], a compound is labeled topologically trivial only if it is demonstrated that an inserted magnetic flux tube (vortex) admits no midgap bound modes.

It has been shown that for a spin-Hall insulator supporting $|C_{s,G}| = N$, when flux is tuned to $\phi = hc/(2e)$ (π-flux) there exist $2N$ midgap vortex bound modes (VBMs) [15,16,27,28]. The spin-Chern number can then be determined by computing quantized induced charge on the vortex. If the VBMs are half-filled, the vortex acquires induced spin but no induced charge. If we dope by $N_e \in [-N, +N]$ electrons away from half-filling, occupying all VBMs, the vortex acquires induced charge $\delta Q = N_e \times e$. If this condition is satisfied, the spin-Chern number can be directly calculated by fixing $N_e = N$ such that $\delta Q = |\mathcal{C}_{s,G}| \times e$. While the introduction of spin-orbit coupling causes the spin bound to the vortex to become finite but nonquantized, the quantization of bound charge remains robust. For this reason, quantization of bound charge was used in Ref. [17] as the criteria for topological classification. We choose to follow this procedure for construction of the training dataset in this work, screening all materials in the database of Mounet *et al.* [35] with less than ten atoms in the unit cell which have not previously been labeled as topological via existing symmetry based methods.

All first-principles calculations based on density-functional theory (DFT) are carried out using the QUANTUM ESPRESSO software package [36–38]. Exchange-correlation potentials use the Perdew-Burke-Ernzerhof (PBE) parametrization of the generalized gradient approximation (GGA) [39]. We utilize norm-conserving pseudopotentials [40] as obtained on the Pseudo-Dojo site [41]. Spin-orbit coupling is considered in all calculations. The WANNIER90 [19], Z2PACK [20], and BERRYEASY [42] software packages were utilized in calculation of all topological invariants. In order to facilitate automated analysis of the bulk topology, Wannier tight-binding (WTB) models are constructed through use of the selected columns of the density matrix (SCDM) method introduced by Vitale *et al.* [43]. Manipulation of Wannier tight-binding models for vortex insertion is done with a custom PYTHON program which will be made publicly available upon being developed into a stand-alone package. The result is a database of 246 symmetry nonindicative compounds with the full results available at Ref. [44]. Despite this expanded search, in each system analyzed the bulk invariant is found to be either zero or an even spin-Chern number. A two-dimensional insulator supporting odd spin-Chern number and trivial $\mathbb{Z}_2$ index remained elusive.

To expand the dataset, all insulators supporting a nontrivial mirror Chern number or $\mathcal{Z}_2$ index identified in Refs. [45–47] are incorporated into the training set. This is possible due to the fact that a nontrivial $\mathcal{Z}_2$ index or mirror Chern number can be considered as a special case of the (pseudo)spin-Chern number. We remark that we cannot include any materials labeled trivial from Refs. [45–47] in the training data as this may cause mislabeling of SNIPs. The resulting aggregate dataset consists of 443 two-dimensional materials. Further information regarding the composition of the dataset is visible in Fig. 1.

III. DATA AUGMENTATION AND PROCESSING

A number of strategies for presentation of lattice structure to CNNs have been developed and tested in recent years [48–53]. In this work, we utilize the strategy of forming a continuous representation by autoencoding voxel images of the crystal structure to create a 2D crystal graph. We account for the possibility of 79 different elements in the crystal structure, specifically atomic numbers 1–84 removing the noble gases. Details of the training set are shown in Fig. 2. As a result, regardless of the number of elements in a single-crystal structure, 80 voxel images will be produced. An autoencoder then translates each image into a vector. A similar process

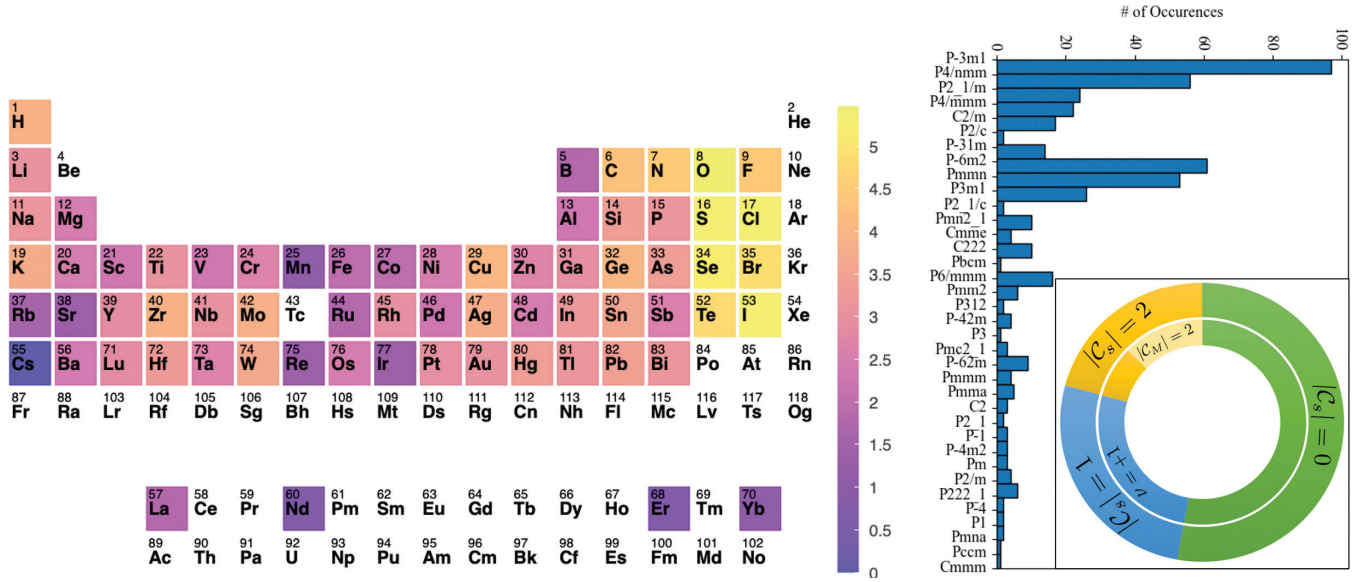

FIG. 1. Statistics of training dataset: Statistics of compounds contained within the training dataset. The periodic table is color coded according to the number of times each element occurs in the dataset on a log scale. Elements in white do not appear in the dataset. The circular plot details the breakdown of how many compounds support $|\mathcal{C}_s| = 2, 1, 0$, and how many of the nontrivial materials are further identified by a $\mathbb{Z}_2$ index or mirror Chern number. The breakdown of space groups present and their occurrences is further provided.

is done to form a voxel image of the lattice, which again is translated into a one-dimensional array through use of an autoencoder. These one-dimensional arrays are then reshaped into 2D crystal images which can be presented to the CNN. In this way the crystal structure obtains a continuous and reversible representation.

Importantly, this process allows for implementation of a data-augmentation strategy, expanding the dataset to include $\sim 10^3$ data points such that the convolutional neural network may achieve a level of accuracy sufficient to benefit our materials search [48]. Namely, we permute the primitive lattice vectors, altering the atomic positions accordingly and

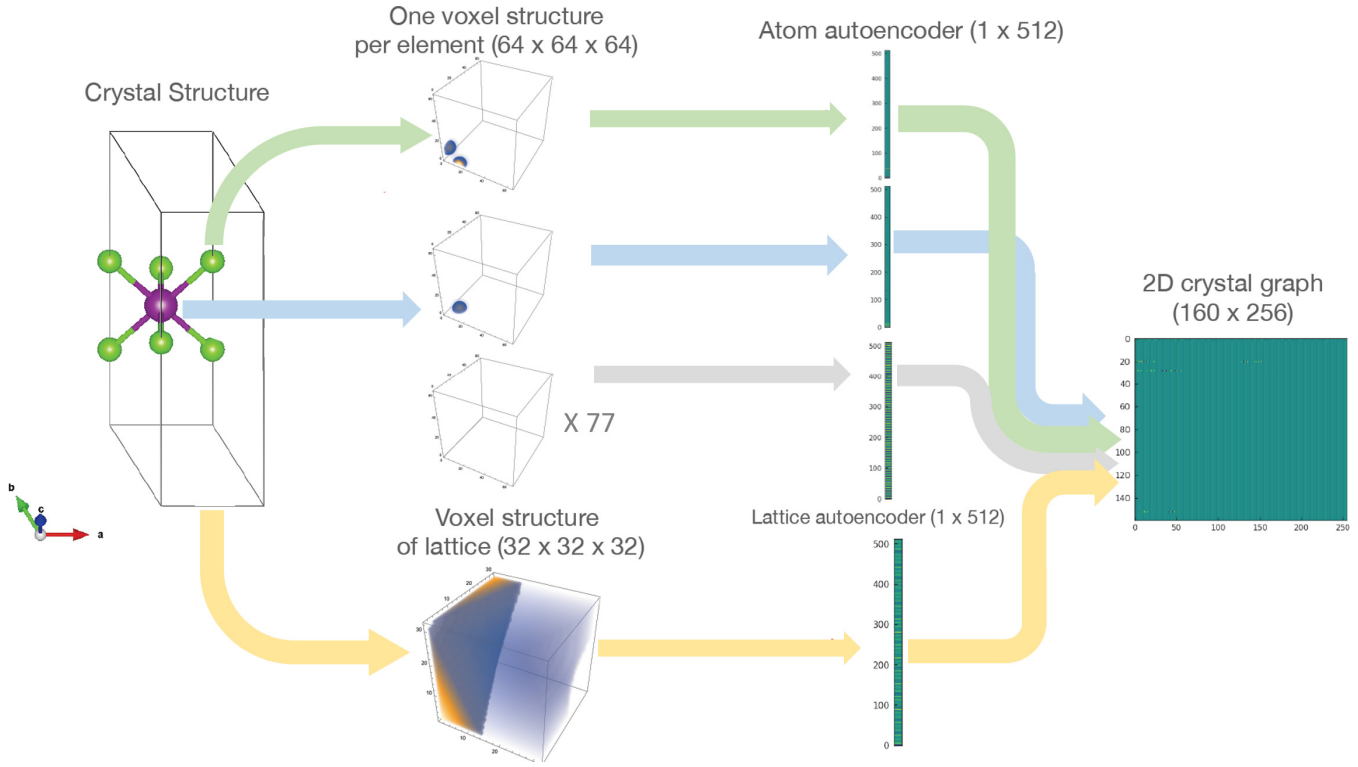

FIG. 2. Construction of crystal graphs: Schematic detailing the construction of two-dimensional crystal graphs for each compound in the training set. This process ensures a continuous representation of the structure that can be presented to the neural network.

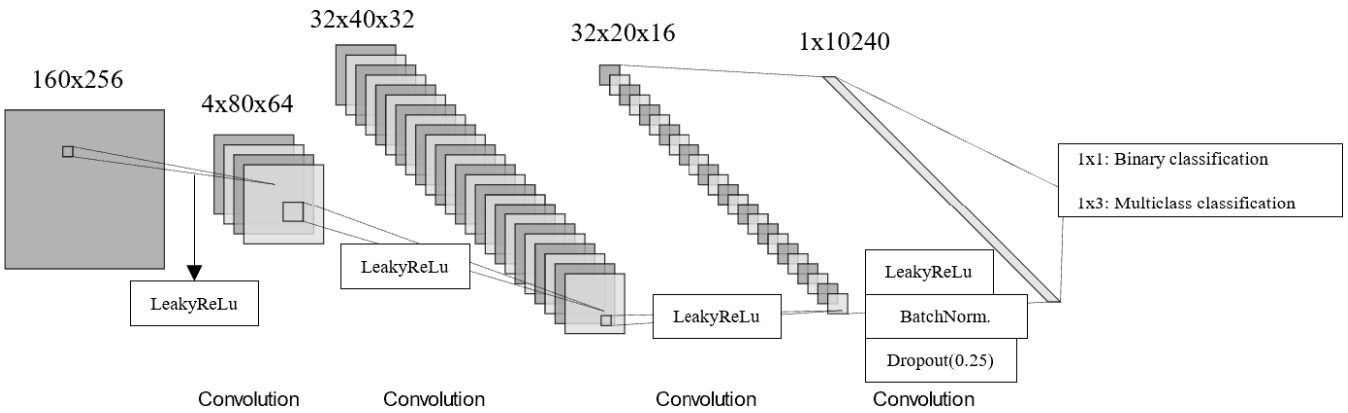

FIG. 3. Convolutional neural network architecture: Architecture of convolutional neural network for binary and multiclass classification. The binary and multiclass classification utilize a sigmoid and softmax activation function, respectively, for the final dense layer. The leakyReLu function is fixed to $\alpha = 0.2$ in each case.

generating new 2D crystal images. This process allows for training data to be augmented by a factor of two if only the a and b primitive vectors are permuted, and up to a factor of six if all are permuted. For further details of voxel image production and autoencoding please consult the Supplemental Material [54]. Details of the CNNs architecture for binary and multiclass classification are visible in Fig. 3.

IV. NEURAL NETWORK PERFORMANCE

We begin with the CNN for binary classification; given the limited training data available, it is expected that this model will achieve higher accuracy. The dataset is randomized and an 80%-10%-10% train-validate-test split is utilized. Early stopping based on validation loss is implemented and the batch size is set to 128. The CNN reaches a train-validation-test accuracy of 96%-88.7%-88.5%. While extraordinary accuracy for neural networks has become commonplace, we note that these values are quite high given the limited training dataset. For context, we compare these values with those produced utilizing a convolutional crystal graph neural network (CGCNN) in the Supplemental Material [48,54,55]. The CGCNN architecture is commonly regarded as state of the art for machine learning based property prediction in computational materials science. It is important to note that CGCNN has not been selected as the primary approach in this work as it disallows the use of the data-augmentation technique of exchanging lattice parameters. This limits the size of the total training dataset. For a larger dataset, the CGCNN method could be advantageous. Other common machine learning architectures for performing topological classification of crystal structures rely primarily on generation of an input vector constructed from features of the constituent atoms as well properties of the nearest neighbors rather than a direct continuous image of the crystal structure [56–58]. In models of this type, details of the crystal structure have also been incorporated by dividing the training dataset such that all constituent compounds correspond to a selected prototypical crystal structure [59]. This approach is again problematic for the purposes of this work as it would result in severely limiting the size of the training dataset. We place emphasis on the use of a continuous and reversible image of the crystal structure as input to lay a foundation for the future incorporation of the model in a generative artificial intelligence architecture [30].

For the multiclass classification model, we adjust the CNN architecture to the form seen in Fig. 2. We again employ the same split for training, test, and validation data and find a train-validation-test categorical accuracy of 95%-89.5%-88%. This represents a significant step toward isolating optimal candidate SNIPs.

At this point it is important to discuss potential sources of bias in the neural network. The most prominent source of bias is likely to be the limited quantity of training data available. This bias is expected as not all elements in the training data set will appear with the same frequency. Those which appear sparsely have the potential to have a significant impact on the model, particularly if all compounds in which the element is found have the same topological classification. Such biases and trends are explored in detail in the Supplemental Material [54].

V. APPLICATION OF CNN

We begin by selecting candidate compounds from the C2DB [31,32] with the criteria that a compound must not be included in the training database or previously identified as topological via symmetry-based methods. Each compound must also admit a bulk band gap greater than 0.1 eV and be dynamically stable. Two-dimensional crystal graphs are then constructed and analyzed using the binary classification CNN. Compounds labeled as nontrivial are isolated and subsequently fed to the multiclass CNN. We filter this final list by number of atoms in the unit cell, selecting two compounds with less than ten atoms in the unit cell predicted to support $|\mathcal{C}_s| = 1$, Os_2Te_4 and Ti_2CO_2.

Plots of the crystal structure and band structure for both compounds are visible in Fig. 4. In order to perform topological analysis of the structures, a Wannier tight-binding model is produced [19], exactly replicating only the Kramers-degenerate bands nearest to the Fermi energy using carefully selected orbitals. As these bands are significantly energetically separated from all other bands, the TB model reproduces

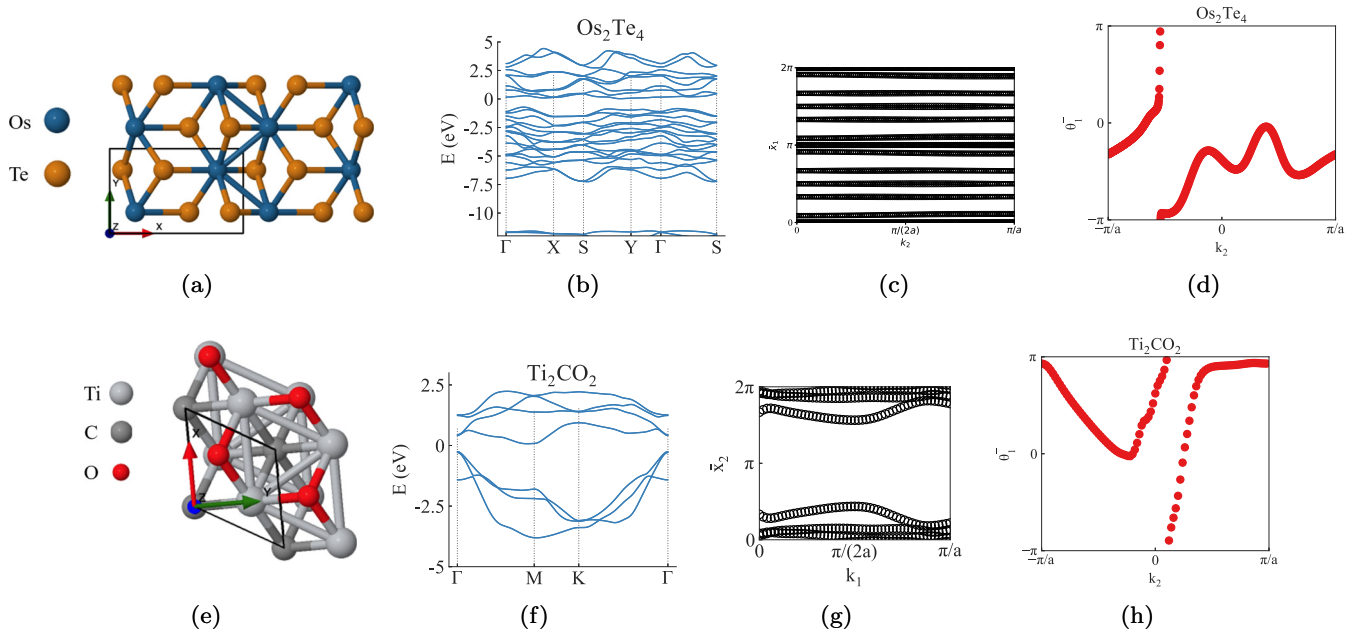

FIG. 4. Analysis of spin-resolved topological insulator candidates: Structure of (a) Os_2Te_4 and (e) Ti_2CO_2 as given via the C2DB database [31,32]. The computed band structure along a high-symmetry path in the Brillouin zone detailing the bands nearest to the Fermi energy can be seen in (b) and (f) for Os_2Te_4 and Ti_2CO_2, respectively. The gapped Wannier center spectra of both compounds, seen in (c) and (g), indicates these systems are trivial under the $\mathbb{Z}_2$ index. The ground-state spin-Chern number is computed via spin-resolved Wilson loop, detailing a single winding for both compounds, confirming $|\mathcal{C}_s| = 1$.

the DFT data precisely. For more computational details please see the Supplemental Material [54]. The spin-Chern number is then computed directly via the method established by Prodan [6,42]. This procedure requires defining the PSO, $P(\mathbf{k})\hat{s}P(\mathbf{k})$, where $P(\mathbf{k})$ is the projector onto occupied bands and $\hat{s}$ is the preferred spin axis. We identify the preferred spin axis supporting a spin gap through a computationally expensive trial-and-error procedure. However, once identified, we are able to produce the results shown in Fig. 4, displaying calculation of the spin-Chern number via a spin-resolved Wilson loop as detailed in Lin *et al.* [16]. Remarkably, both compounds demonstrate proper labeling by the neural network and support of $|C_s| = 1$.

These results are of significance due to the robust nature of the bulk invariant. Unlike other SNIPs which require the preservation of the bulk spin gap, a challenging task experimentally, the bulk topological invariant of these systems requires only preservation of the bulk energetic gap to remain intact. As a result, the bulk invariant is robust to the influence of disorder [6] and other perturbations, much like traditional $\mathbb{Z}_2$ topological insulators. Unlike existing topological insulators, both systems support sizable bulk energetic gaps, 0.325 eV and 0.713 eV for Ti_2CO_2 and Os_2Te_4, respectively. These large energetic gaps make both materials primary candidates for experiments and quantum devices, as we explore in the following section. Furthermore, synthesis of Ti_2CO_2 has been reported in the literature [60] as well as single-crystal Os_2Te_4 [61]. In the case of Ti_2CO_2, synthesis is accelerated by the commercial availability of TiC.

VI. ACCESSING MAJORANA CORNER MODES VIA SUPERCONDUCTING PROXIMITY EFFECT

Prior works have explored the superconducting proximity effect in topological insulators [62–69]. In particular, such proposals have focused on use of the superconducting proximity effect to realize Majorana bound states (MBSs) by layering a two-dimensional topological insulator on the surface of a superconductor. The experimental realization of MBSs is important due to their proposed utilization as a platform for topological quantum computing [70]. A common issue in the proposed platforms for realizing MBSs via the superconducting proximity effect in topological insulators is an extremely small topological gap. Experimental realizations generally rely on known two-dimensional topological insulators [68,69] such as 1T'-WTe_2 [71] and PbTe [72,73] which support normal state topological band gaps on the order of tens of meV.

Similarly, proposals of platforms for the realization of Majorana corner states generally rely on an induced d-wave pairing. Experimental estimates for the induced pairing gap are on the order of several meV [74–76]. Control of this small topological gap poses a significant experimental challenge, particularly as it requires extremely small levels of disorder to be present in a given sample.

In this section we briefly investigate the superconducting proximity effect in one of the proposed spin-resolved topological insulators listed above, Ti_2CO_2. In order to investigate the superconducting proximity effect, we consider a device constructed from a two-dimensional slab of Ti_2CO_2 placed on

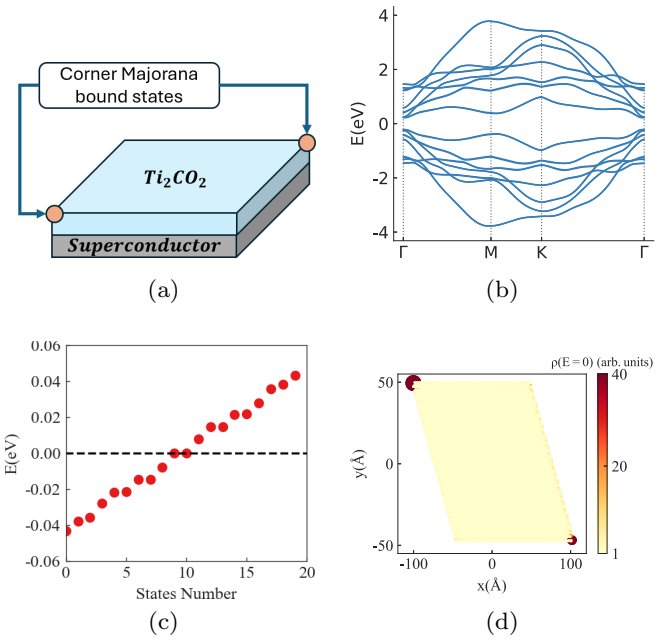

FIG. 5. Majorana corner modes via superconducting proximity effect. (a) Schematic of setup for generating superconducting proximity effect in Ti_2CO_2. (b) Band structure of BdG Hamiltonian H_{eff}. (c) States nearest zero energy for slab of H_{eff} shown in (d) implementing open boundary conditions along both principal axes. (d) Localization of zero energy states shown in (c), demonstrating that they are corner bound. Local density of states on a given site is displayed as a function of color and site size for clarity. A larger site size indicates increased local density of states.

top of a superconductor as shown schematically in Fig. 5(a). To model the system, we construct an effective Bouligobov–de Gennes (BdG) Hamiltonian H_{eff} of the form

$$H_{\text{eff}}(\mathbf{k}) = \begin{bmatrix} H_{\text{WTB}}(\mathbf{k}) & \Delta \\ \Delta^{\dagger} & -H^{*}_{\text{WTB}}(-\mathbf{k}), \end{bmatrix}, \tag{1}$$

where H_{WTB} is the Wannier tight-binding model constructed in the previous section for analysis of Ti_2CO_2. Following Ref. [66], we include the superconducting proximity effect by introducing a simplistic nondissipative pairing term Δ, which we fix as a constant. The resulting band structure of H_{eff} is shown in Fig. 5(b).

We investigate the presence of MBSs by performing exact diagonalization of H_{eff} for a system of 49×37 unit cells, and $\Delta = 180\,\mu\text{eV}$. Fascinatingly, examining the states nearest to zero energy, the results in Fig. 5(c) demonstrate the presence of two zero-energy MBSs. Investigating the localization of the MBSs, we find the results in Fig. 5(d) demonstrating that these states are corner bound and exist on opposite sides of the sample. Importantly, as seen in Fig. 5(b), the bulk energetic gap ΔE_{gap} is on the order of ~ 0.45 eV. This is significantly larger than the typical value for known alternatives and underscores the implications of identifying large band-gap SNIPs. Beyond the intriguing properties which they support on a theoretical level, SNIPs provide a viable route to overcome current practical experimental issues in the identification of MBSs.

Furthermore, we highlight that in this setup, Majorana corner modes survive when the pairing term is suppressed. Such pairing is crucial for the existence of corner modes in alternative proposals based on the proximity effect in two-dimensional first-order topological insulators (FOTIs) [68,69,73]. This result underscores the potential utility of spin-resolved topological insulators in experimental platforms.

VII. SUMMARY

In this work, machine learning guided discovery of symmetry nonindicative topological phases in two dimensions has been shown to be a promising route to the identification of large band-gap quantum spin-Hall insulators. Furthermore, the resulting network can leverage a limited amount of training data through augmentation techniques. In the past, machine learning techniques for identification of nontrivial topology were dismissed due to the computational efficiency of directly calculating the bulk invariant using symmetry indicator techniques. The immense computational expense associated with direct calculation of band topology in SNIPs warrants the use of a different machine learning approach. By limiting the associated computational expense, it is now possible to go beyond analysis of known compounds that have been successfully synthesized and explore large datasets such as the virtual two-dimensional material database (V2DB) [77]. The neural network produced in this work can also serve as a building block in the future development of generative artificial intelligence models for inverse design of topological materials. Finally, we expect that such a network can find extensive use in analyzing the growing number of two-dimensional heterostructures and twisted architectures which have attracted experimental interest but where the normal state band topology is unknown due to the complexity of a many-atom unit cell, which will be explored in a future work.

ACKNOWLEDGMENTS

I am grateful to A. V. Balatsky, P. Goswami, N. Regnault, F. Tafti, and B. Baldassarri for stimulating discussions. Nordita is supported in part by NordForsk. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through Grant Agreement No. 2022-06725.

[1] F. Tang, H. C. Po, A. Vishwanath, and X. Wan, Efficient topological materials discovery using symmetry indicators, Nat. Phys. **15**, 470 (2019).

[2] T. Zhang, Y. Jiang, Z. Song, H. Huang, Y. He, Z. Fang, H. Weng, and C. Fang, Catalogue of topological electronic materials, Nature (London) **566**, 475 (2019).

[3] M. Vergniory, L. Elcoro, C. Felser, N. Regnault, B. A. Bernevig, and Z. Wang, A complete catalogue of high-quality topological materials, Nature (London) **566**, 480 (2019).

[4] F. Tang, H. C. Po, A. Vishwanath, and X. Wan, Comprehensive search for topological materials using symmetry indicators, Nature (London) **566**, 486 (2019).

[5] Y. Xu, L. Elcoro, Z.-D. Song, B. J. Wieder, M. Vergniory, N. Regnault, Y. Chen, C. Felser, and B. A. Bernevig, High-throughput calculations of magnetic topological materials, Nature (London) **586**, 702 (2020).

[6] E. Prodan, Robustness of the spin-Chern number, Phys. Rev. B **80**, 125327 (2009).

[7] J. E. Moore, Y. Ran, and X.-G. Wen, Topological surface states in three-dimensional magnetic insulators, Phys. Rev. Lett. **101**, 186805 (2008).

[8] J. Cano, L. Elcoro, M. I. Aroyo, B. A. Bernevig, and B. Bradlyn, Topology invisible to eigenvalues in obstructed atomic insulators, Phys. Rev. B **105**, 125115 (2022).

[9] A. Nelson, T. Neupert, T. Bzdušek, and A. Alexandradinata, Multicellularity of delicate topological insulators, Phys. Rev. Lett. **126**, 216404 (2021).

[10] B. Lapierre, T. Neupert,and L. Trifunovic, N-band Hopf insulator, Phys. Rev. Res. **3**, 033045 (2021).

[11] G. F. Lange, A. Bouhon, and R.-J. Slager, Spin texture as a bulk indicator of fragile topology, Phys. Rev. Res. **5**, 033013 (2023).

[12] Y. Bai, L. Cai, N. Mao, R. Li, Y. Dai, B. Huang, and C. Niu, Doubled quantum spin Hall effect with high-spin Chern number in α-antimonene and α-bismuthene, Phys. Rev. B **105**, 195142 (2022).

[13] B. Wang, X. Zhou, Y.-C. Lin, H. Lin, and A. Bansil, High spin-Chern-number insulator in α-antimonene with a hidden topological phase, arXiv:2202.04162.

[14] A. C. Tyner, S. Sur, D. Puggioni, J. M. Rondinelli, and P. Goswami, Topology of three-dimensional Dirac semimetals and quantum spin Hall systems without gapless edge modes, Phys. Rev. Res. **5**, L012019 (2023).

[15] A. C. Tyner and P. Goswami, Spin-charge separation and quantum spin Hall effect of β-bismuthene, Sci. Rep. **13**, 11393 (2023).

[16] K.-S. Lin *et al.*, Spin-resolved topology and partial axion angles in three-dimensional insulators, Nat. Commun. **15**, 550 (2024).

[17] A. C. Tyner and P. Goswami, Solitons and real-space screening of bulk topology of quantum materials, arXiv:2304.05424.

[18] M. Taherinejad, K. F. Garrity, and D. Vanderbilt, Wannier center sheets in topological insulators, Phys. Rev. B **89**, 115102 (2014).

[19] G. Pizzi *et al.*, WANNIER90 as a community code: new features and applications, J. Phys.: Condens. Matter **32**, 165902 (2020).

[20] D. Gresch, G. Autès, O. V. Yazyev, M. Troyer, D. Vanderbilt, B. A. Bernevig, and A. A. Soluyanov, Z2PACK: Numerical implementation of hybrid Wannier centers for identifying topological materials, Phys. Rev. B **95**, 075146 (2017).

[21] Q. Wu, S. Zhang, H.-F. Song, M. Troyer, and A. A. Soluyanov, WANNIERTOOLS: An open-source software package for novel topological materials, Comput. Phys. Commun. **224**, 405 (2018).

[22] W. A. Benalcazar, B. A. Bernevig, and T. L. Hughes, Quantized electric multipole insulators, Science **357**, 61 (2017).

[23] W. A. Benalcazar, T. Li, and T. L. Hughes, Quantization of fractional corner charge in C_n-symmetric higher-order topological crystalline insulators, Phys. Rev. B **99**, 245151 (2019).

[24] F. Schindler *et al.*, Higher-order topology in bismuth, Nat. Phys. **14**, 918 (2018).

[25] F. Schindler, A. M. Cook, M. G. Vergniory, Z. Wang, S. S. P. Parkin, B. A. Bernevig, and T. Neupert, Higher-order topological insulators, Sci. Adv. **4**, eaat0346 (2018).

[26] J. Sødequist, U. Petralanda, and T. Olsen, Abundance of second order topology in C_3 symmetric two-dimensional insulators, 2D Mater. **10**, 015009 (2023).

[27] X.-L. Qi and S.-C. Zhang, Spin-charge separation in the quantum spin Hall state, Phys. Rev. Lett. **101**, 086802 (2008).

[28] Y. Ran, A. Vishwanath, and D.-H. Lee, Spin-charge separated solitons in a topological band insulator, Phys. Rev. Lett. **101**, 086801 (2008).

[29] J. Noh, J. Kim, H. S. Stein, B. Sanchez-Lengeling, J. M. Gregoire, A. Aspuru-Guzik, and Y. Jung, Inverse design of solid-state materials via a continuous representation, Matter **1**, 1370 (2019).

[30] T. Long, N. M. Fortunato, I. Opahle, Y. Zhang, I. Samathrakis, C. Shen, O. Gutfleisch, and H. Zhang, Constrained crystals deep convolutional generative adversarial network for the inverse design of crystal structures, npj Comput. Mater. **7**, 66 (2021).

[31] F. A. Rasmussen and K. S. Thygesen, Computational 2D materials database: Electronic structure of transition-metal dichalcogenides and oxides, J. Phys. Chem. C **119**, 13169 (2015).

[32] S. Haastrup *et al.*, The computational 2D materials database: high-throughput modeling and discovery of atomically thin crystals, 2D Mater. **5**, 042002 (2018).

[33] K. F. Mak, K. L. McGill, J. Park, and P. L. McEuen, The valley Hall effect in MoS_2 transistors, Science **344**, 1489 (2014).

[34] Z. Tao *et al.*, Giant spin Hall effect in AB-stacked $MoTe_2/WSe_2$ bilayers, Nat. Nanotechnol. **19**, 28 (2024).

[35] N. Mounet *et al.*, Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds, Nat. Nanotechnol. **13**, 246 (2018).

[36] P. Giannozzi *et al.*, QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, J. Phys.: Condens. Matter **21**, 395502 (19pp) (2009).

[37] P. Giannozzi *et al.*, Advanced capabilities for materials modelling with QUANTUM ESPRESSO, J. Phys.: Condens. Matter **29**, 465901 (2017).

[38] P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. de Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru, and S. Baroni, QUANTUM ESPRESSO toward the exascale, J. Chem. Phys. **152**, 154105 (2020).

[39] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **78**, 1396(E) (1997).

[40] D. R. Hamann, Optimized norm-conserving Vanderbilt pseudopotentials, Phys. Rev. B **88**, 085117 (2013).

[41] M. J. van Setten, M. Giantomassi, E. Bousquet, M. J. Verstraete, D. R. Hamann, X. Gonze, and G.-M. Rignanese, The PSEUDODOJO: Training and grading a 85 element optimized norm-conserving pseudopotential table, Comput. Phys. Commun. **226**, 39 (2018).

[42] A. C. Tyner, BERRYEASY: A GPU enabled PYTHON package for diagnosis of nth-order and spin-resolved topology in the presence of fields and effects, J. Phys.: Condens. Matter **36**, 325902 (2024).

[43] V. Vitale, G. Pizzi, A. Marrazzo, J. R. Yates, N. Marzari, and A. A. Mostofi, Automated high-throughput Wannierisation, npj Comput. Mater. **6**, 66 (2020).

[44] https://github.com/actyner/MLSpinResolvedTopology

[45] D. Wang, F. Tang, J. Ji, W. Zhang, A. Vishwanath, H. C. Po, and X. Wan, Two-dimensional topological materials discovery by symmetry-indicator method, Phys. Rev. B **100**, 195108 (2019).

[46] T. Olsen, E. Andersen, T. Okugawa, D. Torelli, T. Deilmann, and K. S. Thygesen, Discovering two-dimensional topological insulators from high-throughput computations, Phys. Rev. Mater. **3**, 024005 (2019).

[47] A. Marrazzo, M. Gibertini, D. Campi, N. Mounet, and N. Marzari, Relative abundance of Z_2 topological order in exfoliable two-dimensional insulators, Nano Lett. **19**, 8431 (2019).

[48] T. Xie and J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Phys. Rev. Lett. **120**, 145301 (2018).

[49] A. Nouira, N. Sokolovska, and J.-C. Crivello, CrystalGAN: learning to discover crystallographic structures with generative adversarial networks, arXiv:1810.11203.

[50] J. Hoffmann, L. Maestrati, Y. Sawada, J. Tang, J. M. Sellier, and Y. Bengio, Data-driven approach to encoding and decoding 3-D crystal structures, arXiv:1909.00949.

[51] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Comparing molecules and solids across structural and alchemical space, Phys. Chem. Chem. Phys. **18**, 13754 (2016).

[52] K. Kaufmann, C. Zhu, A. S. Rosengarten, D. Maryanovsky, T. J. Harrington, E. Marin, and K. S. Vecchio, Crystal symmetry determination in electron diffraction using machine learning, Science **367**, 564 (2020).

[53] S. Kim, J. Noh, G. H. Gu, A. Aspuru-Guzik, and Y. Jung, Generative adversarial networks for crystal structure prediction, ACS Cent. Sci. **6**, 1412 (2020).

[54] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.6.023316 for details of voxel image production, autoencoding, comparison to CGCNN, analysis of model bias, and details of spin-Chern number computation.

[55] C. W. Park and C. Wolverton, Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery, Phys. Rev. Mater. **4**, 063801 (2020).

[56] N. Claussen, B. A. Bernevig, and N. Regnault, Detection of topological materials with machine learning, Phys. Rev. B **101**, 245117 (2020).

[57] A. Ma, Y. Zhang, T. Christensen, H. C. Po, L. Jing, L. Fu, and M. Soljacic, Topogivity: A machine-learned chemical rule for discovering topological materials, Nano Lett. **23**, 772 (2023).

[58] H. Xu, Y. Jiang, H. Wang, and J. Wang, Discovering two-dimensional magnetic topological insulators by machine learning, Phys. Rev. B **109**, 035122 (2024).

[59] G. R. Schleder, B. Focassio, and A. Fazzio, Machine learning for materials discovery: Two-dimensional topological insulators, Appl. Phys. Rev. **8**, 031409 (2021).

[60] M. Naguib, J. Come, B. Dyatkin, V. Presser, P.-L. Taberna, P. Simon, M. W. Barsoum, and Y. Gogotsi, MXene: a promising transition metal carbide anode for lithium-ion batteries, Electrochem. Commun. **16**, 61 (2012).

[61] B. Müller and H. Lutz, Single crystal Raman studies of pyrite-type RuS_2, RuSe2, OsS_2, $OsSe_2$, PtP_2, and $PtAs_2$, Phys. Chem. Miner. **17**, 716 (1991).

[62] L. Fu and C. L. Kane, Superconducting proximity effect and Majorana fermions at the surface of a topological insulator, Phys. Rev. Lett. **100**, 096407 (2008).

[63] T. D. Stanescu, J. D. Sau, R. M. Lutchyn, and S. Das Sarma, Proximity effect at the superconductor–topological insulator interface, Phys. Rev. B **81**, 241310(R) (2010).

[64] A. Cook and M. Franz, Majorana fermions in a topological-insulator nanowire proximity-coupled to an *s*-wave superconductor, Phys. Rev. B **84**, 201105(R) (2011).

[65] A. M. Cook, M. M. Vazifeh, and M. Franz, Stability of Majorana fermions in proximity-coupled topological insulator nanowires, Phys. Rev. B **86**, 155431 (2012).

[66] M. Hell, M. Leijnse, and K. Flensberg, Two-dimensional platform for networks of Majorana bound states, Phys. Rev. Lett. **118**, 107701 (2017).

[67] C. Reeg, D. Loss, and J. Klinovaja, Proximity effect in a two-dimensional electron gas coupled to a thin superconducting layer, Beilstein J. Nanotechnol. **9**, 1263 (2018).

[68] Z. Yan, F. Song, and Z. Wang, Majorana corner modes in a high-temperature platform, Phys. Rev. Lett. **121**, 096803 (2018).

[69] T. Liu, J. J. He, and F. Nori, Majorana corner states in a two-dimensional magnetic topological insulator on a high-temperature superconductor, Phys. Rev. B **98**, 245413 (2018).

[70] C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. Das Sarma, Non-Abelian anyons and topological quantum computation, Rev. Mod. Phys. **80**, 1083 (2008).

[71] X. Qian, J. Liu, L. Fu, and J. Li, Quantum spin Hall effect in two-dimensional transition metal dichalcogenides, Science **346**, 1344 (2014).

[72] T. H. Hsieh, H. Lin, J. Liu, W. Duan, A. Bansil, and L. Fu, Topological crystalline insulators in the SnTe material class, Nat. Commun. **3**, 982 (2012).

[73] J. Liu, X. Qian, and L. Fu, Crystal field effect induced topological crystalline insulators in monolayer IV-VI semiconductors, Nano Lett. **15**, 2657 (2015).

[74] F. Lüpke, D. Waters, S. C. de la Barrera, M. Widom, D. G. Mandrus, J. Yan, R. M. Feenstra, and B. M. Hunt, Proximity-induced superconducting gap in the quantum spin Hall edge state of monolayer WTe_2, Nat. Phys. **16**, 526 (2020).

[75] N. Shimamura *et al.*, Ultrathin bismuth film on high-temperature cuprate superconductor $Bi_2Sr_2CaCu_2O_8 + \delta$ as a candidate of a topological superconductor, ACS Nano **12**, 10977 (2018).

[76] H. Zhao, B. Rachmilowitz, Z. Ren, R. Han, J. Schneeloch, R. Zhong, G. Gu, Z. Wang, and I. Zeljkovic, Superconducting proximity effect in a topological insulator using Fe(Te, Se), Phys. Rev. B **97**, 224504 (2018).

[77] M. C. Sorkun, S. Astruc, J. V. A. Koelman, and S. Er, An artificial intelligence-aided virtual screening recipe for two-dimensional materials discovery, npj Comput. Mater. **6**, 106 (2020).