Print ISSN: 3104-932X
Online ISSN: 3104-9338

Machine Learning Hamiltonians for Topological Surface State Prediction and Visualization

Zhaoyu Zhu*, Han Sun, Zean Wang

School of Physics, Hangzhou Normal University, Hangzhou, 311121, Zhejiang, China

Received: Nov 11, 2026

Revised: Nov 19, 2026

Accepted: Nov 23, 2026

Published online: November 27, 2026

To appear in: *International Journal of Advanced AI Applications,* Vol. 3, No. 2 (February 2027)

* Corresponding Author: Zhaoyu Zhu (1975533422@qq.com)

Abstract. Machine learning models of electronic Hamiltonians provide a route to accelerating the analysis of topological materials while retaining access to wavefunctions and boundary spectra. Their usefulness for surface states, however, depends on more than agreement with bulk band energies. This review examines the connection between Hamiltonian learning, topological verification, and surface spectral visualization, with emphasis on topological semimetals. We distinguish topology classification, interatomic potential learning, and electronic Hamiltonian prediction, and discuss the roles of DeepH, DeepH-E3, HamGNN, and recent extensions including universal spin–orbit-coupled models and many-body message passing. We then outline how predicted operators can be coupled to established tight-binding and Green function methods. Particular attention is given to basis consistency, spin–orbit coupling, magnetic symmetry, surface termination, and error propagation near band crossings. TaAs, ZrSiS, and Co3Sn2S2 are discussed as physically motivated validation systems rather than as evidence of unverified machine learning performance. We propose a validation workflow that separates matrix accuracy, low-energy electronic structure, topology, and boundary observables. Machine learning Hamiltonians are promising components of such workflows, but reliable surface-state prediction requires explicit tests of the relevant surface environment and topological quantities.

Keywords: *electronic hamiltonian; graph neural network; topological semimetal; surface spectral function; fermi arc; scientific visualization*

1. Introduction

Topological materials require an electronic-structure description that resolves both energy

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

dispersion and wavefunction geometry. The relevant distinction is not simply between materials with and without band crossings. A topological insulator has an insulating bulk gap, whereas a topological semimetal can host symmetry-protected or topologically charged band crossings. Boundary states depend on the bulk phase and on the symmetries and geometry retained at the boundary [1, 2]. This review focuses on semimetals while using the broader language of topological band theory where needed.

Conventional calculations already provide a substantial infrastructure for this problem. Wannier90 constructs localized representations and supports interpolation of electronic quantities [3]. WannierTools evaluates properties including surface spectra, nodal structures, and topological diagnostics within tight-binding models [4]. The motivation for machine learning is therefore to reduce repeated electronic-structure work and improve access to complex configurations within this infrastructure. It does not follow that extracting an isosurface or rendering a band plot requires an artificial intelligence method.

DeepH established a route from atomic structure to localized density functional theory (DFT) Hamiltonian matrices [5]. DeepH-E3 and HamGNN developed equivariant representations of these operators [6, 7]. Subsequent work has explored broader chemical coverage and more expressive architectures [8–10]. These advances support a review centered on learned operators and their downstream analysis. They should not be conflated with a general demonstration of accurate Fermi arcs, spin textures, or transport responses in every topological material.

We organize the discussion around an explicit chain: atomic structure, learned electronic operator, topology verification, boundary calculation, and visualization. The literature selection is a focused methodological review rather than an exhaustive systematic survey. We separate reported model capabilities from the validation procedures and prospective applications proposed here. No new material-specific training results are reported.

2. Physical quantities and the boundary problem

2.1. Bulk Fermi surfaces and surface Fermi arcs

For a periodic independent-electron description, the bulk Fermi surface is the set of wavevectors satisfying $E_n(\boldsymbol{k}) = \mu$ for at least one band n, where μ is the chemical potential. Its geometry depends on filling and dispersion. Topological information additionally involves eigenvectors and symmetry. Consequently, two visually similar constant-energy surfaces need not have the same topology in the band-theoretical sense [1, 2].

A surface Fermi arc is a feature of the boundary spectral function in the two-dimensional

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

surface Brillouin zone. In a Weyl semimetal, its relation to the bulk follows from the topological charge of Weyl nodes. At suitable energy and surface orientation, arc branches connect projected bulk features associated with opposite net chirality. Nodes at different energies, overlapping projections, and surface–bulk hybridization complicate this simple picture [2, 11]. An arc should therefore not be described as a segment created merely by intersecting a three-dimensional Fermi surface with a node.

Nodal lines require a separate symmetry analysis. Their existence and protection depend on the symmetry class and on whether spin–orbit coupling (SOC) is included [12]. Surface bands associated with a nodal-line phase are often described using the drumhead picture, but their detailed dispersion and visibility are boundary dependent. Figure 1 distinguishes the spaces in which bulk pockets and surface contours are represented.

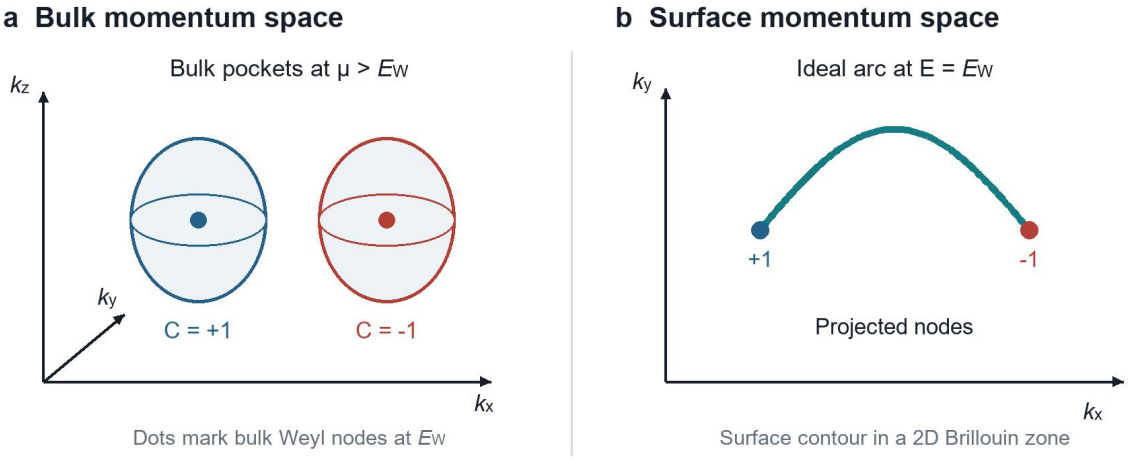


Figure 1. Bulk and surface momentum spaces in an idealized Weyl system. (a) Illustrative bulk pockets at a chemical potential above equal-energy Weyl nodes. (b) An ideal surface arc at the node energy, connecting distinct projections of opposite chiral charge. The common node energy is . The panels use different energy cuts to distinguish the quantities; no material-specific data or calculated spectrum is shown. Original schematic based on the bulk–boundary concepts in [2, 11].

2.2. From a localized Hamiltonian to a surface spectrum

In a localized basis, the bulk Hamiltonian can be assembled from real-space matrix blocks. With a cell-periodic Fourier convention, its elements take the form

$$H_{\alpha\beta}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} H_{\alpha\beta}(\boldsymbol{R}) \exp(i\boldsymbol{k}\cdot\boldsymbol{R}) \quad (1)$$

In Eq. (1), α and β label basis orbitals and $\boldsymbol{R}$ is a lattice translation. Basis-position phases must be treated consistently if a different Bloch convention is used. For nonorthogonal atomic orbitals, an overlap matrix $\mathbf{S}(\boldsymbol{k})$ is also required and the eigenproblem is generalized to $\mathbf{H}(\boldsymbol{k})\boldsymbol{c} = E\,\mathbf{S}(\boldsymbol{k})\boldsymbol{c}$. A learned Hamiltonian is meaningful only together with its basis definition

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

and overlap treatment. Bold lowercase symbols denote vectors, and bold uppercase symbols denote matrices. Scalar variables are italic; the subscripts s and b label the surface region and bulk embedding, respectively. The upright superscript R denotes a retarded quantity.

Boundary spectra can be obtained from a finite slab or a semi-infinite Green function. In an orthonormal representation, the retarded surface Green function and surface spectral function can be written as

$$\mathbf{G}_{\mathrm{s}}^{\mathrm{R}}(\boldsymbol{k}_{\parallel}, E) = \left[(E + \mathrm{i}\eta)\mathbf{I} - \mathbf{H}_{\mathrm{s}}(\boldsymbol{k}_{\parallel}) - \boldsymbol{\Sigma}_{\mathrm{b}}^{\mathrm{R}}(\boldsymbol{k}_{\parallel}, E)\right]^{-1} \tag{2}$$

$$A_{\mathrm{s}}(\boldsymbol{k}_{\parallel}, E) = -\frac{1}{\pi}\,\mathrm{Im}\,\mathrm{Tr}\left[\mathbf{P}_{\mathrm{s}}\,\mathbf{G}_{\mathrm{s}}^{\mathrm{R}}(\boldsymbol{k}_{\parallel}, E)\right] \tag{3}$$

In Eqs. (2) and (3), $\mathbf{H}_{\mathrm{s}}$ describes the chosen surface region, $\boldsymbol{\Sigma}_{\mathrm{b}}$ couples that region to the remaining bulk, η is a positive broadening, and $\mathbf{P}_{\mathrm{s}}$ selects the surface layers. Established iterative methods evaluate the Green function without diagonalizing a macroscopically thick slab [4, 13]. Nonorthogonal implementations require corresponding overlap and projection conventions. A bright spectral feature must still be assessed for its surface localization and relation to the projected bulk continuum.

2.3. Spin information and magnetic symmetry

Spin-resolved analysis requires the appropriate spin operators and spinor eigenstates or Green functions in the same basis. Neither a scalar energy model nor a visually plausible arc determines spin polarization. SOC also links spin transformations to spatial rotations, making the treatment of orbital and spin degrees of freedom part of the operator representation [6].

For a magnetic crystal, reversing or rotating a magnetic moment can change the physical state. Spatial equivariance describes a consistent transformation of the full input configuration and output operator. It does not make a rotation of magnetization relative to a fixed lattice physically equivalent to a rigid rotation of the entire system. Time-reversal invariance of a fixed magnetic Hamiltonian must be distinguished from time-reversal equivariance of the mapping between configurations. The latter can relate a magnetic state to its reversed partner even when each state breaks time-reversal symmetry. The xDeepH framework explicitly models dependence on both atomic and magnetic structures and encodes the corresponding transformation rules [14].

3. What different machine learning models predict

3.1. Classification and interatomic potentials

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

Topology classification can reduce the number of candidates requiring detailed calculations. For example, Claussen and colleagues used gradient-boosted decision trees to classify materials using information associated with composition and symmetry [15]. A classification label does not specify a surface termination, a boundary Hamiltonian, or a momentum-resolved spectral function. Classification is therefore useful for screening but is insufficient for constructing surface-state images. Table 1 summarizes the distinct prediction tasks and their roles in surface analysis.

Table 1. Distinct prediction tasks and their roles. Performance measures are specific to each task.

Task	Typical output	Role in a surface workflow	Necessary additional evidence
Topology classification [15]	Phase label or probability	Candidate screening	Boundary operator and explicit topology checks
Interatomic potential [16]	Energy and atomic forces	Geometry generation or relaxation	Electronic operator and spinor states
Hamiltonian learning [5–10, 14]	Localized matrix elements	Input to electronic calculations	Basis compatibility and downstream validation

Interatomic potentials address another target. NequIP learns energies and forces for atomistic simulations using equivariant features [16]. Such models can help generate or relax candidate surface geometries, subject to their training coverage. Their success does not itself establish a prediction of electronic Hamiltonians or Fermi-arc spin textures. Table 1 separates these tasks so that errors in energy per atom, force, electronic eigenvalues, and class labels are not ranked as interchangeable measures.

3.2. Learning electronic Hamiltonians

DeepH learns localized Hamiltonian blocks from atomic configurations and demonstrates downstream electronic-structure calculations [5]. The important methodological feature is the operator output: the same predicted operator can be used to calculate multiple electronic observables. Its locality assumptions and representation choices define the circumstances in which transfer to larger configurations can be expected.

DeepH-E3 incorporates Euclidean equivariance in a framework that accommodates SOC [6]. HamGNN likewise targets Hamiltonian matrices through an equivariant graph architecture [7]. These approaches encode how matrix elements transform when local environments rotate. Hermiticity, basis ordering, and the symmetries appropriate to a specific magnetic state still need explicit treatment and verification. Equivariance alone is not a certificate of correct band

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

topology.

The original HamGNN study also provides a boundary example: it compares Bi_2Se_3 slab band structures and thickness-dependent gaps with DFT and computes spin textures from the predicted Hamiltonian (Fig. 6 of Ref. [7]). This is direct evidence for those slab observables; it does not establish performance for Weyl-semimetal Fermi arcs or arbitrary surface reconstructions.

Nor does a small matrix-element error provide a universal bound on every observable. An error concentrated in a small coupling near a band inversion can be more consequential than a larger error far from the energy window of interest. For surface applications, evaluation should therefore extend from operator blocks to the low-energy bands and boundary spectrum. This is a validation requirement, rather than a claim that all published Hamiltonian models have already been tested on those targets.

3.3. Broader chemical coverage and recent architectures

Table 2. Representative Hamiltonian models. Entries summarize methodological scope, not a common accuracy ranking.

Method	Operator representation or scope	Evidence relevant here	Surface application requirement
DeepH [5]	Localized DFT Hamiltonian	Learned operators and electronic quantities	Test boundary environments
DeepH-E3 [6]	Equivariant Hamiltonian with SOC	Spatial covariance with spinor treatment	Verify topology and magnetic conditions
HamGNN [7]	Equivariant Hamiltonian blocks	Bi2Se3 slab bands and model-derived spin textures	Check basis and boundary transfer
Universal DeepH [8]	Broader materials coverage	Universal model and fine-tuning strategy	Assess coverage of surface coordination
Uni-HamGNN [9]	Universal SOC Hamiltonian	SOC operator learning across chemistries	Validate target surface and magnetic state
MACE-H [10]	Many-body equivariant messages	Operator and electronic-property benchmarks	Evaluate new geometries independently

The universal DeepH study extended Hamiltonian learning toward models spanning diverse compositions and structures and investigated fine-tuning for specific materials [8]. More recently, Uni-HamGNN targeted universal SOC Hamiltonians across the periodic table [9]. Such developments address the cost of training a separate model for every composition. Their

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

relevance to a surface calculation depends on whether the local bonding, coordination, and magnetic environments of that surface are represented adequately. Table 2 summarizes representative models by their methodological roles.

MACE-H introduces many-body message passing into Hamiltonian prediction and assesses electronic properties as well as matrix elements [10]. Its reported tests also illustrate why a model ranking depends on the split: improved performance on held-out shifted bilayers does not automatically imply improved prediction for unseen twisted configurations. This distinction matters for surface reconstructions, which introduce systematic changes in local environments. Heterogeneous benchmark errors should therefore not be presented as a single ranking of these architectures.

4. Data construction and physical consistency

4.1. Reference operators and compatible labels

Training labels may be obtained directly in localized atomic-orbital bases or through a Wannier representation. These are distinct routes. A learned atomic-orbital Hamiltonian does not necessarily require Wannierization, while a Wannier-based route must establish a consistent gauge, orbital character, and energy window across configurations. Wannier90 provides tools for constructing and interpolating such representations [3].

Automated Wannierization has been demonstrated using the selected columns of the density matrix method within reproducible workflows [17]. This reduces manual intervention but does not establish a hardware-independent throughput for arbitrary compounds. For a surface-oriented dataset, we recommend retaining convergence settings, SOC and magnetic configurations, basis definitions, reconstruction geometry, and any interpolation diagnostics alongside the operator labels.

Entangled bands and topological obstruction should be distinguished. Entanglement concerns the selection and interpolation of a target subspace. A topological obstruction concerns whether a specified band subspace admits localized Wannier functions with the required properties. Nonzero Chern numbers obstruct exponentially localized Wannier functions for the corresponding isolated subspace [18]. This does not imply that every topological material lacks a useful localized Hamiltonian including a broader orbital space. Practical interpolation failure should not be attributed to topology without checking the subspace and constraints.

4.2. Surface coverage and transfer

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

Cleavage changes coordination, charge redistribution, and sometimes atomic positions. Cutting a bulk tight-binding model provides a well-defined ideal termination, but cannot generally reproduce all self-consistent surface effects. Direct inference on a slab also needs an environment that the model can represent. A model trained only on small bulk distortions should not be presumed transferable to a reconstructed or adsorbate-covered surface.

We recommend grouping dataset splits by termination, reconstruction, strain regime, or chemical family. Closely related frames from one configuration should not dominate both training and test sets. A bulk-trained baseline, a surface-adapted model, and a reference slab calculation would then clarify what is gained through additional surface data. This is a proposed assessment design, not a reported result for the materials discussed below.

4.3. Learning objectives and uncertainty

Operator regression may be supplemented by band or observable losses, provided their definitions and weights are reported. A topology-related objective is attractive near band inversion, but its value must be established through an ablation using the same data split. Improvements are task and dataset dependent and should be assessed together with computational cost.

Uncertainty-guided acquisition has already been implemented for electronic Hamiltonian learning. Zhu, Rinke, and Egger coupled MACE, which supplies atomic energies and forces, with DeepH-E3, which predicts the electronic Hamiltonian, in an active-learning framework for defective GaAs [19]. An ensemble of three independently trained models was used for each network: force disagreement among the MACE models selected configurations for retraining, and disagreement among the DeepH-E3 Hamiltonians ranked configurations for additional DFT labeling. Each round added the 30 structures of highest Hamiltonian uncertainty, balanced across defect type and temperature, and the cycle was stopped when the eigenvalue accuracy near the band gap saturated. After twelve rounds and 460 structures, the Hamiltonian matrix elements reached a mean absolute error of about 1 meV and the eigenvalues about 10 meV. This provides a concrete precedent for ensemble-based acquisition in electronic-structure learning. Its extension to topological surface states still requires validation for the chosen boundary and observable.

For topological surface analysis, we propose prioritizing configurations that are uncertain and relevant to surface observables. Ensemble disagreement must be checked against held-out operator and observable errors before it is interpreted as a quantitative uncertainty estimate.

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

Agreement among models alone does not establish reliable extrapolation. A useful acquisition strategy would also account for proximity to a small gap and changes in surface localization. The stopping criterion should concern the intended observable, such as stable node identification or converged surface spectra. These surface-specific criteria remain proposals for validation.

5. Verification and visualization

5.1. A sequence of independent checks

Figure 2 outlines the workflow proposed in this review. First, check operator consistency, including Hermiticity, overlap conditioning, and symmetry residuals. Second, compare low-energy dispersions and relevant orbital character against reference calculations. Energy alignment and band matching must be specified. Third, evaluate topology using wavefunctions and a converged momentum mesh. Z2Pack provides a framework based on hybrid Wannier centers for relevant topological classifications [20].

For a Weyl node, a small band separation identifies a candidate crossing but does not establish its chiral charge. Berry-flux or equivalent calculations are needed around an isolated candidate. Nodal-line verification similarly requires the relevant symmetry and loop diagnostic [4, 12]. Visual inspection of a crossing is insufficient. Finally, calculate the chosen boundary and test slab thickness or Green function convergence. Each step addresses a different failure mode, so passing an earlier step should not replace a later one.

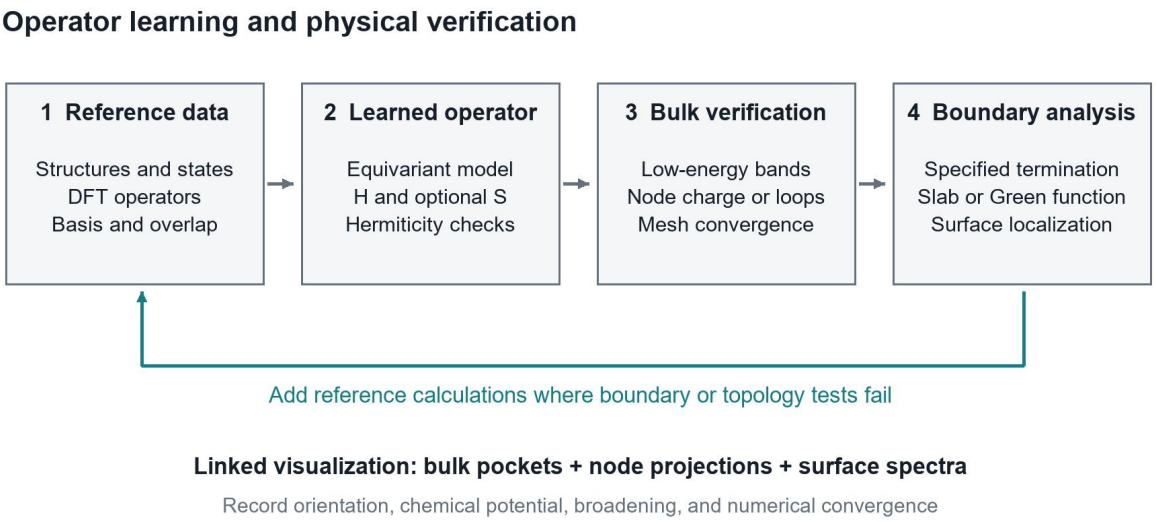


Figure 2. Proposed workflow connecting Hamiltonian learning to surface analysis. Operator, bulk, topology, and boundary checks are separate stages. Failed tests motivate additional reference calculations before interpreting linked visualizations. This is an original workflow synthesis, not a claim that every listed model implements all stages [3–7, 13, 20].

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

5.2. Showing the quantities being compared

A visualization should state whether it depicts a three-dimensional bulk isosurface, a projected bulk continuum, or a surface spectral function. Node positions should be projected using the same reciprocal coordinates as the surface spectrum. Surface orientation, termination, chemical potential, energy window, and broadening should be stated. This information allows a reader to interpret changes in apparent contour connectivity.

Linked views can display bulk pockets and nodes alongside a surface spectral map and selected dispersion cuts. Such coordination is a useful interface design proposal. It can be implemented with established rendering methods; machine learning is needed only where a separately identified inference task justifies it. Automatic contour labeling may assist inspection, but the physical classification must remain tied to the underlying electronic calculation.

5.3. Reporting errors at the observable level

We recommend reporting matrix and eigenvalue errors separately from topological agreement. For boundary spectra, useful comparisons include surface localization, spectral-peak energy, and contour displacement in a stated coordinate system. A momentum error and an energy error measure different quantities. Broadening should be matched when comparing images because an excessively large η can hide discrepancies.

Transport requires additional validation. Berry-curvature-dependent responses are sensitive to wavefunctions and integration near small gaps [21]; visual agreement of Fermi contours alone cannot establish the accuracy of a Hall conductivity. Any acceleration claim should state whether it includes reference-data generation, training, operator inference, diagonalization, and surface post-processing. Machine learning does not remove the cost of these downstream operations.

6. Material systems for targeted validation

6.1. TaAs and surface Fermi arcs

TaAs offers an established setting for testing the relation between bulk Weyl physics and surface arcs. Experimental work has also shown the importance of distinguishing arc states from other surface features [11]. For a learned Hamiltonian, a meaningful validation would combine bulk node identification with a surface-resolved spectrum for a specified termination. Strain-dependent calculations would additionally require reference points on both sides of any proposed transition.

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

These tasks should be evaluated independently: correct bulk nodes do not determine every detail of a termination-dependent contour, and a plausible contour does not establish the correct charge. A spin-resolved benchmark would additionally compare the predicted spin expectation values along matched surface branches and account for weak polarization where an angular error becomes ill-defined.

6.2. ZrSiS and symmetry-sensitive crossings

ZrSiS is useful for testing the distinction between bulk crossings and surface-derived bands. The original combined angle-resolved photoemission spectroscopy (ARPES) and electronic-structure study identified nonsymmorphic features and surface-derived states hybridized with bulk bands; it also discussed small SOC-induced gaps [22]. This is a more specific physical picture than treating every observed surface feature as an ideal flat drumhead band.

A model benchmark should identify which crossing is protected under the chosen symmetry and SOC treatment, then inspect the surface projection and localization. The material also illustrates why visual linearity over part of a dispersion is insufficient to establish exact gaplessness. Comparative predictions across ZrSiS-related compounds would be a useful prospective test of chemical transfer, provided consistent reference calculations and surface geometries are available.

6.3. Co3Sn2S2 and magnetic surfaces

$Co_3Sn_2S_2$ supplies a boundary-sensitive magnetic test. Morali and colleagues observed different Fermi-arc connectivity on different surface terminations [23]. This result motivates evaluating termination explicitly rather than attaching one universal arc map to a bulk crystal structure.

A proposed learned-operator study should separate errors from the magnetic bulk state, the surface environment, and the spectral calculation. Changing magnetization relative to the lattice would require an appropriate SOC and magnetic representation and reference validation. No specific Hall-conductivity error or magnetization-rotation accuracy follows from a model trained only to classify magnetic order.

7. Outlook

The next step is a benchmark that couples operator learning to reproducible boundary calculations. We propose a set of materials with specified bulk structures and terminations, common energy conventions, reference operator files, and converged topology diagnostics.

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

Held-out reconstructions and magnetic configurations would test transfer more directly than random splits alone. Reporting failed Wannierization, unreliable interpolation, and unresolved crossings would make the usable domain of each approach clearer.

Universal SOC models and many-body architectures expand the available methods [9, 10], but boundary transfer should be demonstrated separately. A particularly informative experiment would compare bulk-only inference with a model adapted using a limited set of surface calculations. The assessment should ask whether adaptation improves both spectra and topology, and whether the added reference cost is recovered over repeated applications.

Experimental integration is another prospective direction. ARPES intensities contain matrix-element, resolution, and surface-condition effects, so they should not be treated as direct labels for bare band energies without a measurement model. Likewise, adding many-body corrections requires appropriate reference information. A surrogate trained on one DFT approximation can reproduce that approximation efficiently; exceeding its physical accuracy needs independently justified corrections and experimental tests.

8. Conclusion

Machine learning Hamiltonians offer an operator-based route to the electronic analysis of topological materials. DeepH, equivariant extensions, and recent universal and many-body models establish relevant computational ingredients, while conventional topology and Green function tools provide the downstream analysis. Connecting these ingredients to reliable surface-state prediction requires compatible bases, explicit boundary conditions, and validation of wavefunctions and topology in addition to energies. The material examples considered here define concrete tests rather than unsupported demonstrations of model performance. A workflow that retains these distinctions can make surface visualization more informative and computationally accessible while keeping its physical interpretation verifiable.

References

[1] Qi, X.-L., & Zhang, S.-C. (2011). Topological insulators and superconductors. *Reviews of Modern Physics*, *83*(4), 1057–1110. https://doi.org/10.1103/RevModPhys.83.1057

[2] Hasan, M. Z., Xu, S.-Y., Belopolski, I., & Huang, S.-M. (2017). Discovery of Weyl fermion semimetals and topological Fermi arc states. *Annual Review of Condensed Matter Physics*, *8*(1), 289–309. https://doi.org/10.1146/annurev-conmatphys-031016-025225

[3] Pizzi, G., Vitale, V., Arita, R., Blügel, S., Freimuth, F., Géranton, G., Gibertini, M., Gresch, D., Johnson, C., Koretsune, T., Ibañez-Azpiroz, J., Lee, H., Lihm, J.-M., Marchand, D., Marrazzo, A., Mokrousov, Y., Mustafa, J. I., Nohara, Y., Nomura, Y., ... Yates, J. R. (2020). Wannier90 as a community code: New features and applications. *Journal of Physics: Condensed Matter*, *32*(16), Article 165902.

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

https://doi.org/10.1088/1361-648X/ab51ff
[4] Wu, Q., Zhang, S., Song, H.-F., Troyer, M., & Soluyanov, A. A. (2018). WannierTools: An open-source software package for novel topological materials. *Computer Physics Communications*, *224*, 405–416. https://doi.org/10.1016/j.cpc.2017.09.033
[5] Li, H., Wang, Z., Zou, N., Ye, M., Xu, R., Gong, X., Duan, W., & Xu, Y. (2022). Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation. *Nature Computational Science*, *2*(6), 367–377. https://doi.org/10.1038/s43588-022-00265-6
[6] Gong, X., Li, H., Zou, N., Xu, R., Duan, W., & Xu, Y. (2023). General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian. *Nature Communications*, *14*, Article 2848. https://doi.org/10.1038/s41467-023-38468-8
[7] Zhong, Y., Yu, H., Su, M., Gong, X., & Xiang, H. (2023). Transferable equivariant graph neural networks for the Hamiltonians of molecules and solids. *npj Computational Materials*, *9*, Article 182. https://doi.org/10.1038/s41524-023-01130-4
[8] Wang, Y., Li, Y., Tang, Z., Li, H., Yuan, Z., Tao, H., Zou, N., Bao, T., Liang, X., Chen, Z., Xu, S., Bian, C., Xu, Z., Wang, C., Si, C., Duan, W., & Xu, Y. (2024). Universal materials model of deep-learning density functional theory Hamiltonian. *Science Bulletin*, *69*(16), 2514–2521. https://doi.org/10.1016/j.scib.2024.06.011
[9] Zhong, Y., Wang, R., Gong, X., & Xiang, H. (2026). A universal spin–orbit-coupled Hamiltonian model for accelerated quantum material discovery. *Nature Machine Intelligence*, *8*(3), 403–414. https://doi.org/10.1038/s42256-026-01196-x
[10] Qian, C., Vitartas, V., Kermode, J. R., & Maurer, R. J. (2026). Equivariant electronic Hamiltonian prediction with many-body message passing. *npj Computational Materials*, *12*, Article 169. https://doi.org/10.1038/s41524-026-02020-1
[11] Batabyal, R., Morali, N., Avraham, N., Sun, Y., Schmidt, M., Felser, C., Stern, A., Yan, B., & Beidenkopf, H. (2016). Visualizing weakly bound surface Fermi arcs and their correspondence to bulk Weyl fermions. *Science Advances*, *2*(8), Article e1600709. https://doi.org/10.1126/sciadv.1600709
[12] Fang, C., Chen, Y., Kee, H.-Y., & Fu, L. (2015). Topological nodal line semimetals with and without spin-orbital coupling. *Physical Review B*, *92*(8), Article 081201. https://doi.org/10.1103/PhysRevB.92.081201
[13] López Sancho, M. P., López Sancho, J. M., & Rubio, J. (1985). Highly convergent schemes for the calculation of bulk and surface Green functions. *Journal of Physics F: Metal Physics*, *15*(4), 851–858. https://doi.org/10.1088/0305-4608/15/4/009
[14] Li, H., Tang, Z., Gong, X., Zou, N., Duan, W., & Xu, Y. (2023). Deep-learning electronic-structure calculation of magnetic superstructures. *Nature Computational Science*, *3*(4), 321–327. https://doi.org/10.1038/s43588-023-00424-3
[15] Claussen, N., Bernevig, B. A., & Regnault, N. (2020). Detection of topological materials with machine learning. *Physical Review B*, *101*(24), Article 245117. https://doi.org/10.1103/PhysRevB.101.245117
[16] Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., & Kozinsky, B. (2022). E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, *13*, Article 2453. https://doi.org/10.1038/s41467-022-29939-5
[17] Vitale, V., Pizzi, G., Marrazzo, A., Yates, J. R., Marzari, N., & Mostofi, A. A. (2020). Automated high-throughput Wannierisation. *npj Computational Materials*, *6*, Article 66. https://doi.org/10.1038/s41524-020-0312-y
[18] Brouder, C., Panati, G., Calandra, M., Mourougane, C., & Marzari, N. (2007). Exponential localization of Wannier functions in insulators. *Physical Review Letters*, *98*(4), Article 046402. https://doi.org/10.1103/PhysRevLett.98.046402

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

[19] Zhu, X., Rinke, P., & Egger, D. A. (2026). Predicting temperature-dependent optoelectronic properties of semiconductor defects with equivariant neural networks. *npj Computational Materials*, *12*, Article 176. https://doi.org/10.1038/s41524-026-02111-z

[20] Gresch, D., Autès, G., Yazyev, O. V., Troyer, M., Vanderbilt, D., Bernevig, B. A., & Soluyanov, A. A. (2017). Z2Pack: Numerical implementation of hybrid Wannier centers for identifying topological materials. *Physical Review B*, *95*(7), Article 075146. https://doi.org/10.1103/PhysRevB.95.075146

[21] Xiao, D., Chang, M.-C., & Niu, Q. (2010). Berry phase effects on electronic properties. *Reviews of Modern Physics*, *82*(3), 1959–2007. https://doi.org/10.1103/RevModPhys.82.1959

[22] Schoop, L. M., Ali, M. N., Straßer, C., Topp, A., Varykhalov, A., Marchenko, D., Duppel, V., Parkin, S. S. P., Lotsch, B. V., & Ast, C. R. (2016). Dirac cone protected by non-symmorphic symmetry and three-dimensional Dirac line node in ZrSiS. *Nature Communications*, *7*, Article 11696. https://doi.org/10.1038/ncomms11696

[23] Morali, N., Batabyal, R., Nag, P. K., Liu, E., Xu, Q., Sun, Y., Yan, B., Felser, C., Avraham, N., & Beidenkopf, H. (2019). Fermi-arc diversity on surface terminations of the magnetic Weyl semimetal Co3Sn2S2. *Science*, *365*(6459), 1286–1291. https://doi.org/10.1126/science.aav2334

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.