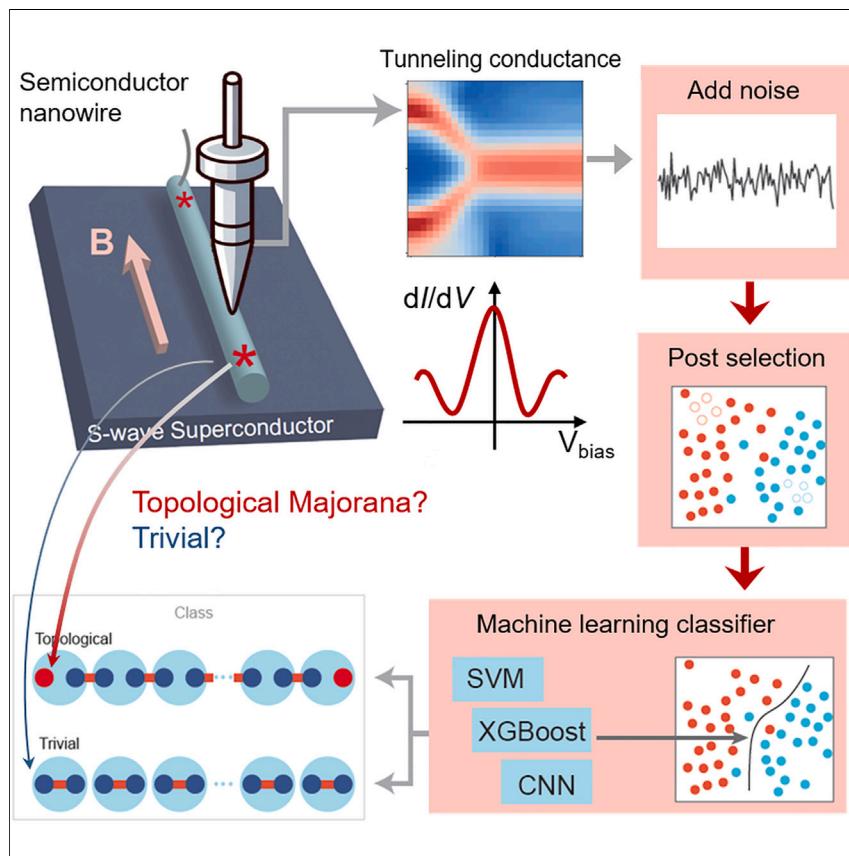


## Article

# Machine learning detection of Majorana zero modes from zero-bias peak measurements



Researchers have developed a machine learning method to detect Majorana zero modes (MZMs) from experimental data, achieving significant accuracy. This approach utilizes quantum transport simulations and topological data analysis, providing a simpler and more effective method to identify these quantum states, crucial for the advancement of fault-tolerant quantum computing. This work demonstrates a significant step toward practical, fault-tolerant quantum systems, making complex quantum behaviors accessible and understandable for non-specialists.

Mouyang Cheng, Ryotaro Okabe, Abhijatmedhi Chotrattanapituk, Mingda Li

vipandyc@mit.edu (M.C.)  
mingda@mit.edu (M.L.)

### Highlights

MZMs show promise for topological quantum computation via zero-bias peaks (ZBPs)

Our machine learning framework distinguishes MZM signals from spurious ZBP signals

XGBoost achieves 85% accuracy for 1D data and 94% for 2D data with Zeeman splitting

2

## Benchmark

First qualification/assessment of material properties and/or performance

Cheng et al., Matter 7, 2507–2520  
July 3, 2024 © 2024 Elsevier Inc. All rights reserved, including those for text and data mining, AI training, and similar technologies.  
<https://doi.org/10.1016/j.matt.2024.05.028>



**Article**

# Machine learning detection of Majorana zero modes from zero-bias peak measurements

Mouyang Cheng,<sup>1,2,\*</sup> Ryotaro Okabe,<sup>1,3</sup> Abhijatmedhi Chotrattanapituk,<sup>1,4</sup> and Mingda Li<sup>1,5,6,\*</sup>

**SUMMARY**

**Majorana zero modes (MZMs) carry non-Abelian statistics and great promise for topological quantum computation. A key signature of MZMs is the zero-bias peaks (ZBPs) in tunneling differential conductance. However, identifying MZMs from ZBPs has been challenging due to topological trivial states generating spurious signals. In this work, we introduce a machine learning framework that can distinguish MZM from other signals using ZBP data. Quantum transport simulation from tight-binding models is used to generate training data, while persistent cohomology analysis confirms the feasibility of machine-based classification. Even with noisy data, the extreme gradient boosting (XGBoost) classifier reaches 85% accuracy for 1D data and 94% for 2D data with Zeeman splitting. Tests on prior experiments show that key observations from some of the prior experiments are more likely to originate from MZMs. Our model offers a quantitative approach to assess MZMs using solely ZBP data. Furthermore, our results highlight the use of machine learning on exotic quantum systems with experimental-computational integration.**

**INTRODUCTION**

The identification of quantum many-body phases from experimental observations is one of the central tasks in condensed matter physics.<sup>1–4</sup> While symmetry-breaking phases can be detected unequivocally using local order parameters, topological phases of matter pose a more complex problem. Unlike the former, the topological phases cannot be characterized by local order parameters but instead carry global topological invariants.<sup>5</sup> As a result, detecting topological phases often requires an indirect measurement where topology can manifest, such as examining bulk excitations or specific boundary states.<sup>6</sup> Successful examples include the quantum anomalous Hall effect with insulating bulk and spin-polarized chiral edge states that can be probed by electrical transport<sup>7–9</sup> and topological Weyl semimetals with bulk Weyl fermions and surface Fermi arcs using photoemission.<sup>10</sup> In other cases, probing topology can become notably more challenging. In quantum spin liquids, for instance, bulk spinon excitations and edge Majorana fermions only leave subtle experimental evidence.<sup>3,11</sup> An enhanced capability to detect topological phases of matter will greatly enrich our understanding of quantum phases and hold paramount importance for next-generation microelectronic and quantum computing applications.

Among the exotic topological phases of matter, Majorana zero modes (MZMs), characterized by the non-Abelian, Ising-type anyonic statistics, have captured significant research and industrial attention over the past decade. Thanks to their unique ability to store information non-locally, and their intrinsic zero energy that guards against

**PROGRESS AND POTENTIAL**

This study presents a machine learning framework capable of distinguishing Majorana zero modes (MZMs)—the building block qubits for topological quantum computation—from spurious signals that mimic the MZM in tunneling conductance measurements. Utilizing quantum transport simulations for training data and persistent cohomology for validation, the model achieves up to 94% accuracy. This approach enhanced our capability to identify MZMs toward fault-tolerant quantum computation.



hybridization, MZMs are deemed a highly promising platform to realize fault-tolerant topological quantum computation.<sup>12–15</sup> Theoretically, MZMs were first proposed in the Kitaev 1D chain model with a p-wave superconductor (SC), where pairs of MZMs can emerge at the ends of the chain.<sup>16</sup> However, the evidence of p-wave SCs has been elusive, with an unclear pathway to lift the double degeneracy of the spin pairing. Several remedies have been proposed in previous literature.<sup>17</sup> Fu and Kane suggest constructing MZMs using the proximity effect at the interface between an s-wave SC and a topological insulator, which resembles a  $p_x + ip_y$  SC with additional time-reversal symmetry.<sup>18</sup> Candidates like 5/2 fractional quantum Hall states<sup>19,20</sup> and other platforms<sup>21–26</sup> are also potential candidates for hosting MZMs. Another milestone was reached to construct MZMs on a 1D nanowire with semiconductor (SM) coupled with a proximity s-wave SC.<sup>27–29</sup> Under strong Rashba spin-orbit coupling and an external Zeeman field, MZMs can emerge from an effective p-wave SC with the double degeneracy lifted. This SM/SC nanowire system has been considered extremely feasible to realize MZMs, with numerous experimental reports demonstrated in the past decade.<sup>30–40</sup> In these cases, the zero-bias peaks (ZBPs) of the differential tunneling conductance under scanning tunneling spectroscopy (STS) provide a strong experimental signature for MZMs.<sup>41</sup> However, there has long been a concern that there are other topologically trivial states that can also produce ZBPs, such as Andreev bound states (ABSs), Yu-Shiba-Rusinov states, or simply large disorders.<sup>42–46</sup> To make the detection of MZMs even more elusive, it is difficult to actually define what a Majorana mode in topological SCs because there are several low-energy localized states, e.g., the so-called quasi-Majorana states, with intermediate properties between topological Majorana modes and trivial low-energy states.<sup>47–51</sup> Various practical approaches to distinguish MZMs from trivial modes have been proposed, such as the topological gap protocol<sup>52,53</sup> and interferometry with floating Majorana islands.<sup>54</sup> However, they both rely on more involved operations such as non-local conductance measurements or embedding a Majorana island into an A-B interferometer. Even under such intricate designs, it is still hard to eliminate false positives mimicking MZMs. Therefore, a more direct approach to identify topological MZMs based solely on traditional experimental ZBP measurements would be highly desirable, and the power of interpretation of machine learning (ML) could facilitate such a task, which is beyond the imagination of conventional protocols based on principles of physics.

In this work, we develop a ML pipeline that aims to differentiate topological MZMs from other topologically trivial states using experimental ZBP signals. The primary obstacles are the scarcity of experimental data and the absence of a universally acknowledged MZM ground truth. However, thanks to the STS technique, which can provide direct access to the single-particle density of states and further enables quantitative comparisons between experiments and computations, we were able to generate the ZBP training data computationally. Using effective Hamiltonian and quantum transport simulations, we cover a broad spectrum of physical parameters and mechanisms and further add data noises to mimic experiments. Although distinguishing MZMs has created challenges due to the spectral similarity of ZBPs between topological MZMs and topologically trivial states, from a ML perspective, this complexity is transformed into a classification task. Persistent cohomology analysis shows that the hidden global features of different topological classes remain robust, indicating that such a classification task is fundamentally machine classifiable. By further implementing various ML methods, such as linear classifiers, convolutional neural networks (CNNs), and extreme gradient boosting (XGBoost), excellent accuracy is finally reached even with a reasonable level of data noise. We carry out additional tests on the experimental ZBP data from existing literature and find that some

<sup>1</sup>Quantum Measurement Group, MIT, Cambridge, MA 02139, USA

<sup>2</sup>School of Physics, Peking University, Beijing 100084, China

<sup>3</sup>Department of Chemistry, MIT, Cambridge, MA 02139, USA

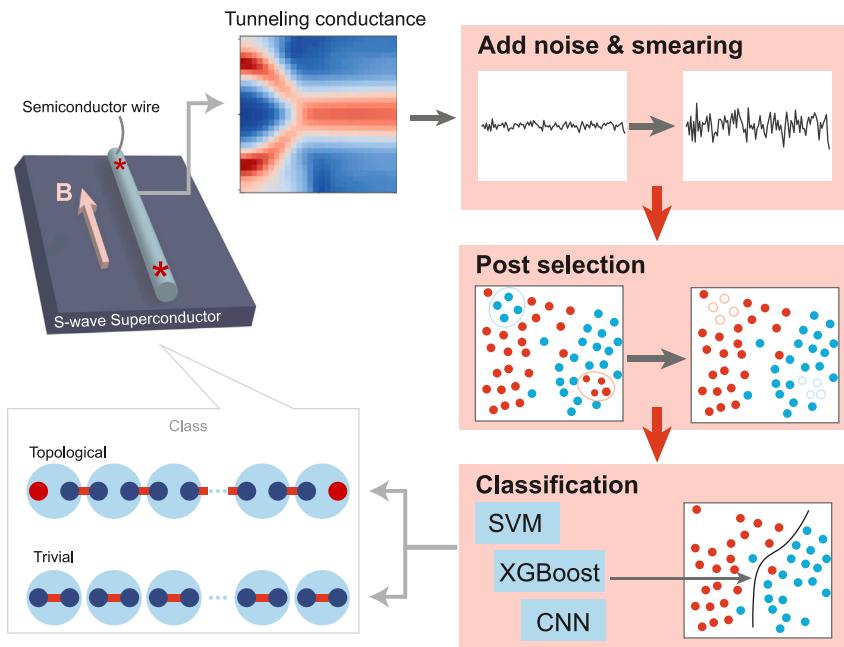
<sup>4</sup>Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139, USA

<sup>5</sup>Department of Nuclear Science and Engineering, MIT, Cambridge, MA 02139, USA

<sup>6</sup>Lead contact

\*Correspondence: [vipandyc@mit.edu](mailto:vipandyc@mit.edu) (M.C.), [mingda@mit.edu](mailto:mingda@mit.edu) (M.L.)

<https://doi.org/10.1016/j.matt.2024.05.028>



**Figure 1. The overview of the machine learning workflow to detect MZMs from ZBPs**

For a Figure360 author presentation of this figure, see <https://doi.org/10.1016/j.matt.2024.05.028>.  
 The system consists of a 1D semiconducting nanowire coupled in proximity with an s-wave superconductor, which resembles a 1D p-wave superconducting Hamiltonian under a parallel magnetic field  $B$ . Training data are generated by an effective Hamiltonian approach. By modifying the on-site potential landscape, states that host topological MZMs and topologically trivial states are generated and labeled by the topological class. The tunneling conductance  $dI/dV$  signals from the scanning tunneling spectroscopy are further computed using the tight-binding and quantum transport approach and are used as input data. Optionally, noise and smearing are added to the dataset to better mimic experimental data, and post-selection is performed to ensure the topological/trivial labels are assigned appropriately. Various machine learning models are established to achieve the MZM classification, with additional tests performed using existing experimental data.

ZBP data are more likely to arise from MZMs, while others are not. This does not rule out the potential presence of MZMs in any of the reported experimental systems, given the limitation of the effective Hamiltonian approach and other experimental complexities not considered in this work. Our model offers an attempt to solve the MZM detection problem with ML. This work can also shed light on the application of ML in other exotic many-body quantum systems with very limited training data and a lack of ground truth.

## RESULTS

### Model setup

The general ML workflow is shown in Figure 1. We consider the popular 1D SC/SM nanowire discussed earlier as the modeled system. The pristine nanowire system can be described by the 1D Bogoliubov-de Gennes Hamiltonian following the s-wave pairing Oreg-Lutchyn model  $H = \frac{1}{2} \int \Psi^\dagger(x) H_{\text{tot}} \Psi(x) dx$ ,<sup>26-28</sup> where

$$H_{\text{tot}} = T + H_{\text{soc}} + U + H_Z + H_{\text{couple}}$$

$$= \left( -\frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial x^2} - i\alpha \frac{\partial}{\partial x} \sigma_y - \mu \right) \tau_z + E_Z \sigma_x + \Delta \tau_x. \quad (\text{Equation 1})$$

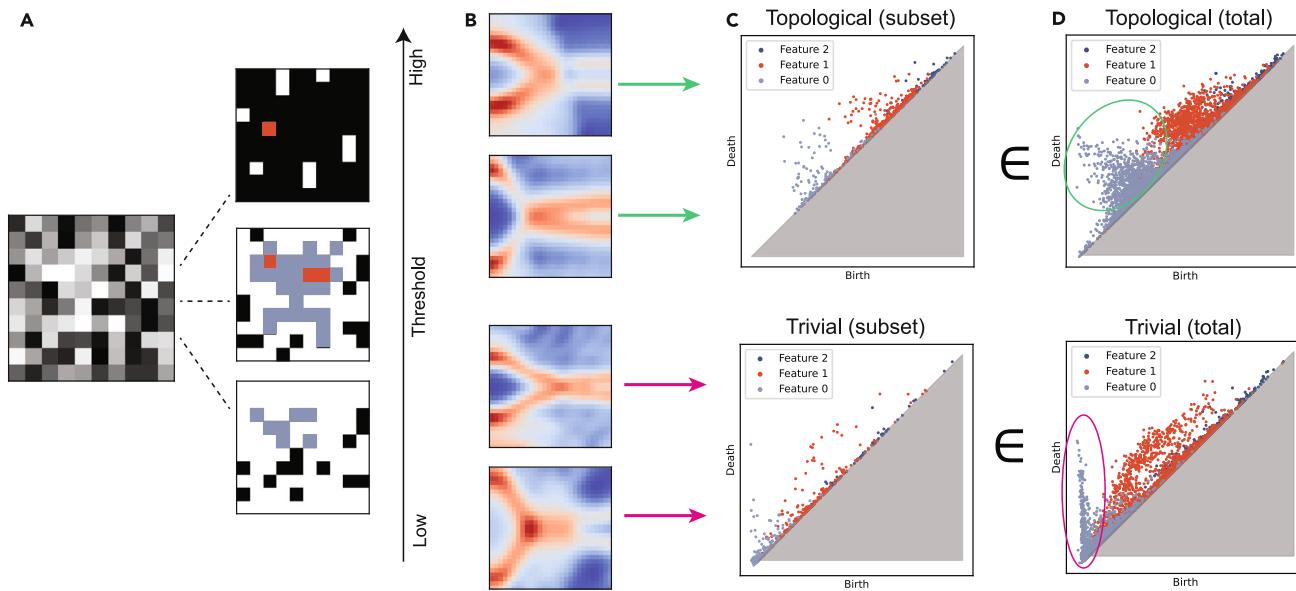
Here,  $\widehat{\Psi}(x) = (\widehat{\psi}_\uparrow(x), \widehat{\psi}_\downarrow(x), \widehat{\psi}_\downarrow^\dagger(x), \widehat{\psi}_\uparrow^\dagger(x))^T$  spans a Nambu space with four spinors, and  $\vec{\sigma}$  and  $\vec{\tau}$  stand for Pauli matrices in the spin and particle-hole space, respectively. The five terms  $T, H_{\text{soc}}, U, H_Z$ , and  $H_{\text{couple}}$  denote the kinetic energy, spin-orbit coupling, on-site potential, Zeeman coupling, and SC-SM coupling, respectively. Detailed information about the choice of parameters is shown in [supplemental experimental procedures S1.1](#).

The Hamiltonian [Equation 1](#) is the pristine Hamiltonian that leads to MZMs. We further apply weak diagonal disorder  $V_{\text{imp}}(x) \sim \epsilon N(0, 1)$  sampled from a normal distribution to mimic the noise but without destroying the topology. In real experiments, trivial ZBPs may arise from a non-ideal potential landscape on the nanowire. At least two scenarios can lead to topologically trivial states, including (1) quantum dots located at ends of the nanowire and (2) large fluctuating disorder spread on the whole nanowire.<sup>44</sup> Therefore, for the topological trivial classes without MZMs, we construct the Hamiltonian in two ways: for scenario 1, we add a Gaussian potential as an incommensurate on-site perturbation to the diagonal Hamiltonian. It has been shown that such smearing potential could be the culprit of the ABS. When the two Andreev bias peaks come closer under tuned parameters, these peaks will merge and form a trivial ZBP.<sup>55,56</sup> For scenario 2, we amplify the disorder strength so that the fluctuation energy is comparable to the original chemical potential  $\mu$ . This can also give rise to topologically trivial states with ZBPs, creating a challenge for MZM identification.<sup>44</sup> Overall, to justify our approach, we also include an extra check on whether smearing potential or large disorder will generally create trivial modes with observable ZBPs in our dataset (see [supplemental experimental procedures S1.1](#)).

To generate the training data for ML, we cover a wide range of input Hamiltonian parameters (see [supplemental experimental procedures S1.1](#)). The continuous Hamiltonian is discretized in real space to a finite tight-binding matrix. Then, we perform tight-binding simulations on this discretized system to calculate the tunneling conductance  $G = dI/dV$  via the  $S$  matrix formalism (see [experimental procedures](#) for more details). A total of 12,000 labeled Hamiltonians are generated, with 4,000 for topological MZMs, 4,000 for trivial ABSs, and 4,000 for trivial large disordered states. The tunneling conductance signal can thus be calculated under sweeping a 2D parameter space composed of bias voltage ( $V_{\text{bias}}$ ) and Zeeman splitting ( $E_Z$ ), each with 28 different values. This leads to the use of a  $28 \times 28$  image to represent the tunneling conductance data, labeled by either topological (hosting MZMs) or trivial states (either ABSs or large disorders) for ML classification. In addition, since some experimental works focus on 1D  $dI/dV$  data without sweeping the Zeeman splitting, we single out the 1D data with zero Zeeman splitting for additional training. This can be done by searching the ZBPs while sweeping through  $E_Z$  horizontally. Lastly, to improve the training robustness and bridge the theoretical-experimental gap, we perform preprocessing on the raw data, including Gaussian smearing, additional noise, and anomaly detection on the dataset. We can also refine our workflow by carrying out a post-selection on the dataset to ensure that the topological/trivial labels are assigned properly, and the ZBPs are generally present in our dataset. More details on the Hamiltonian model, data generation, processing, and selection can be found in the [experimental procedures](#) and [supplemental experimental procedures S1.1](#).

### Global pattern with TDA

We first display typical tunneling conductance  $dI/dV$  data generated from the workflow above in [Figure 2B](#) for topological MZM and trivial classes (see [supplemental](#)



**Figure 2. Persistent cohomology analysis with the training data**

(A) Schematic of the principles on persistent cohomology using simplified 2D data as an example. Given a fixed threshold value, each original pixel can be masked to binary black or white. Isolated black clusters and white loops are highlighted out with blue and red colors as key topological features. As the masking threshold is tuned continuously from minimum to maximum magnitudes, these distinct topological features emerge and die out. Only features near a centered area are colored for better visualization.

(B) Typical computationally generated tunneling conductance data used for machine learning training, for topological MZM and trivial classes. The 2D heatmap plots are tunneling conductance  $dl/dV$  as a function of bias voltage  $V_{bias}$  and Zeeman energy  $E_z$ .

(C and D) Topological data analysis on topological MZM and trivial classes, respectively, using persistent cohomology analysis. (C) shows results for a small random subset containing the conductance data shown in (B), where a distinct feature difference can be seen. (D) shows the full results for the whole topological/trivial dataset. Green and orange ovals highlight the key feature difference between these two classes. Although the individual raw data in (B) are barely distinguishable with bare eyes, an obvious difference is shown between the topological MZM class and the other topological trivial classes through topological data analysis.

experimental procedures S1.2 for more examples). It can be seen that the 2D  $dl/dV$  data from topological and trivial states have similar patterns. One earlier approach to achieve MZM pattern recognition<sup>44</sup> is finding the phase boundary between the topological and trivial classes. By pointing out the difference in the position of the topological phase transition compared to the pristine data, it was concluded that quantum dots and large disorder destroy the topology of the system, thus creating trivial ZBPs. However, this approach is performed with fixed Hamiltonian parameters; when the parameters are unknown, discerning the topological MZM phase is still challenging for human eyes.

To investigate the potential intrinsic separability between the topological MZM and trivial classes, we employ the persistent cohomology analysis on a portion of the training dataset for all classes. Persistent cohomology is a type of topological data analysis (TDA) that studies the global feature difference at various scales. Figure 2A shows an example of persistent cohomology analysis on simplified 2D data. Starting from a grayscale image, a threshold value is tuned from the lowest pixel value to the highest. For a given threshold value, each pixel can be masked to binary black/white (lower/higher than threshold). Then, two topological features emerge: feature 0 identifies isolated black clusters in data (partially marked with light blue), and feature 1 focuses on closed loops encircled by a black cluster (partially filled with red). By sweeping the threshold values, different patterns assigned with different features emerge and annihilate, which creates a birth-death scattered plot.<sup>57</sup> Therefore, persistent

cohomology provides insights into the robustness and significance of these topological characteristics in the data.

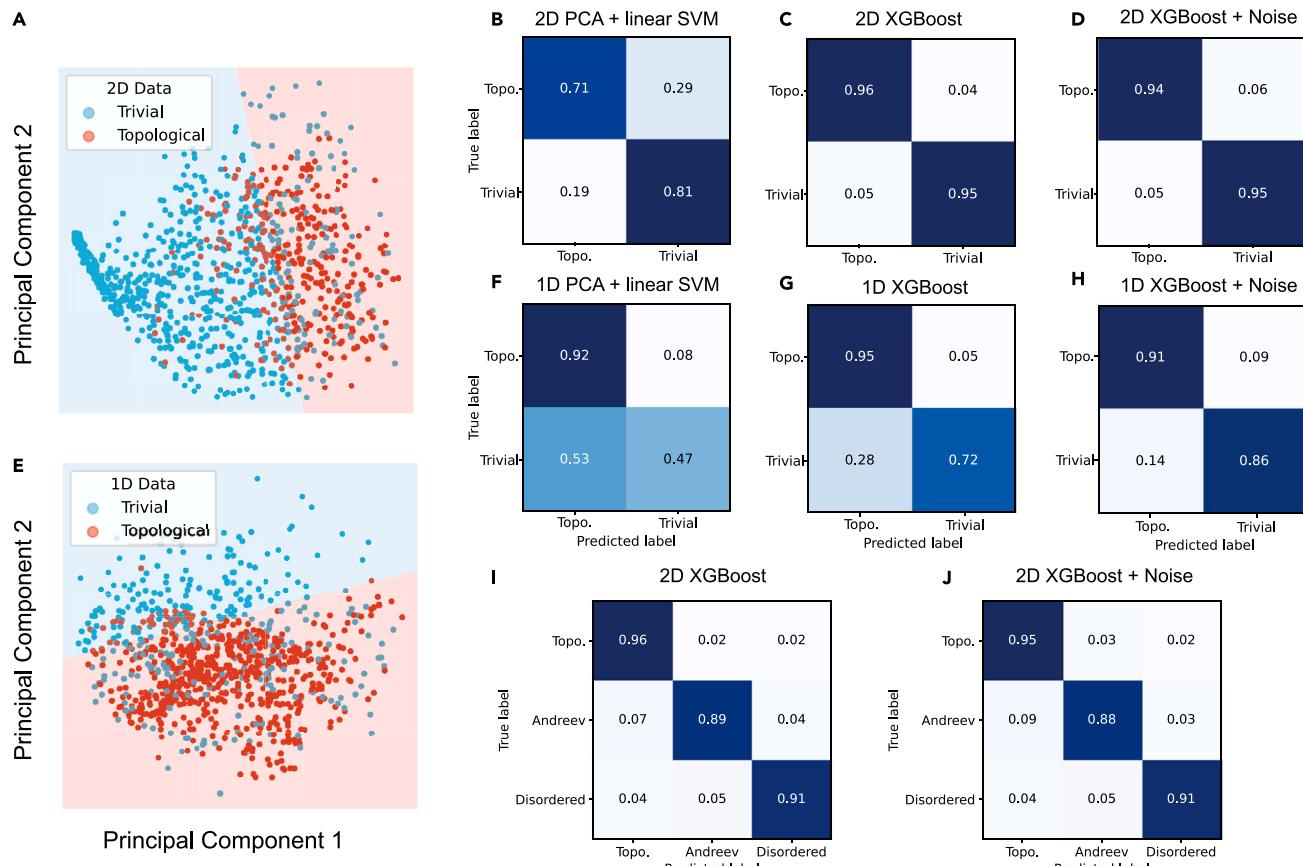
Our analysis involves the 3D data composed of bias voltage, Zeeman splitting, and other Hamiltonian parameters as one dimension. As a result, there is an additional feature 2, which captures voids or cavities entirely enclosed by surfaces. The persistent cohomology analysis is performed on our datasets using the GUDHI package with a cubical complex.<sup>58</sup> The results are shown in Figures 2C and 2D, where the difference between the topological MZM class and the trivial class can be seen clearly. Take feature 0 as an example; on the one hand, for the topological MZM dataset, there are very few clusters (light blue) that emerge near zero birth and annihilate early. On the other hand, for the trivial dataset, there is a continuous distribution of clusters that creates at zero birth and annihilates, evenly distributing on the near zero birth axis. The orange and green ovals marked in Figure 2D clearly highlight such distinct topological feature differences between topological and trivial data. The results indicate that though the individual conductance data shown in Figure 2B could be hard to classify by the human eye, TDA can show the relationships for more than one Hamiltonian parameter for each data class, giving us crucial information on the connectivity of varying similar Hamiltonians. Such collective information allows us to spot feature differences and feel confident that such a binary classification task is machine separable. More detailed insights brought about by persistent cohomology analysis are shown in [supplemental experimental procedures S1.1](#).

Therefore, the persistent cohomology analysis implies that although human eyes cannot readily distinguish topological MZM states from trivial states, there exists a global topological feature difference between them. Such a difference builds confidence that the MZM classification problem with ZBPs is machine separable prior to any design of ML models.

### ML results

We employ a few ML models to perform the topological MZM classification task. For the model inputs, 2D data of tunneling conductance images with 28×28 pixels are flattened into 1D arrays except for the CNN, which directly receives the 2D data. As a baseline check, we first perform linear principal-component analysis (PCA) to compress the data dimension. We reduce the 2D and 1D datasets' complexity to 2 dimensions for better visualization, and the reduced results with a label of 0 or 1 are shown in Figures 3A and 3E. On the scattered plot for the first two leading principal components, there is no clear boundary between two separated clusters with different labels. Figures 3A and 3E show a linear support vector machine (SVM) boundary line that separates two regions (shaded blue and red). However, there is a notable portion of data points crossing the boundary, indicating the limited power of linear classification at least on the PCA dimensionality-reduced dataset (performance shown in Figures 3B and 3F). Particularly, for 1D PCA, the prediction of data labeled topological with 0.47 accuracy is close to a random guess. Further attempts to use linear methods consistently provide lower accuracy than 90% (see [supplemental experimental procedures S1.3](#)), indicating the intrinsic data non-linearity and calling on the necessity of non-linear ML methods.

We carry out non-linear classification methods and ensemble methods including kernel-SVM, random forest, CNN, and XGBoost. The results as well as the hyperparameter tuning process are described in [supplemental experimental procedures S1.3](#). Among them, XGBoost, which combines ensemble models and improved gradient boosting, gives an overall better performance than other methods for both 1D and



**Figure 3. Machine learning classifications to identify the topological MZMs using 1D and 2D tunneling conductance data**

(A and E) PCA on the generated 2D (A) and 1D (E) data projected on the first two principal components. The SVM linear boundary roughly separates the topological MZM and trivial classes.

(B and F) Confusion matrices for PCA + linear SVM learning results for 2D (B) and 1D (F) training data.

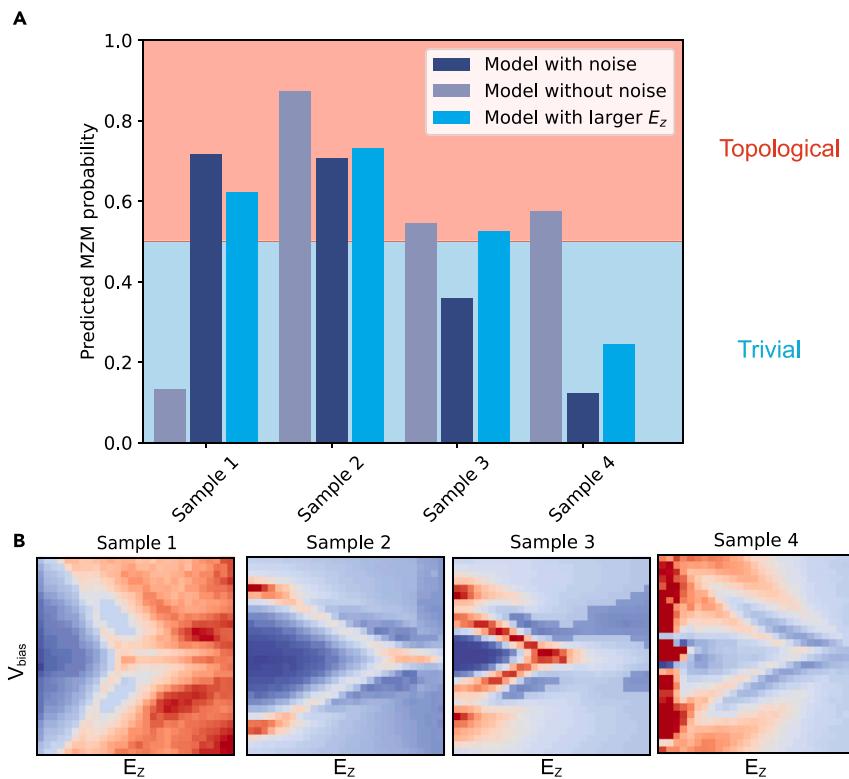
(C and G) Confusion matrices from XGBoost for 2D (C) and 1D (G) training model without noise.

(D and H) Confusion matrices from XGBoost for 2D (C) and 1D (G) training model with added data noise. Note that in all cases, the model with 2D data outperforms the model with 1D data, indicating the advantage of collecting data with Zeeman energy sweeping.

(I and J) Ternary classification confusion matrices from XGBoost for 2D training model with/without added data noise.

2D tasks. The confusion matrix results for XGBoost training are shown in Figures 3C, 3D, 3G, and 3H for 2D and 1D data with and without data noise, respectively. It is worthwhile mentioning that the binary classification with 2D tunneling conductance data for the topological MZM class reaches  $\sim 94\%$  accuracy, even in the presence of data noise. Additionally, although the 1D classifier gives a  $\sim 28\%$  false positive rate for the topological MZM class, it still gives high, 95% confidence in true positives, and the overall accuracy still reaches 85%. While adding noise reduces the accuracy of identifying trivial classes, it significantly improves the performance of detecting trivial classes from  $\sim 72\%$  to  $\sim 86\%$ , which may be attributed to the large data variance and better data generalization.

The success in ML classification agrees well with the persistent cohomology observation. Also, the introduction of the Zeeman energy sweeping in 2D data outperforms the 1D data, indicating the benefit and possible necessity of taking data with sweeping Zeeman energy. Additionally, we also test the capability of our model for multi-classification. Figures 3I and 3J show results for ternary classification results for the most accurate XGBoost model. They show that even for ternary classification,



**Figure 4.** Tests on experimental 2D ZBP data based on our machine learning models

(A) The probability of positive prediction for the XGBoost model is plotted as a histogram. Prediction results from the 2D XGBoost model with/without noise and the 2D XGBoost model trained on a dataset including a larger  $E_z = 4.48$  meV range are shown for more information. Only 4 results, labeled from sample 1 to sample 4,<sup>31,34–36</sup> imply that the system possibly hosts real MZMs, while all the others suggest that they are unlikely to host MZMs.

(B) Patterns of tunneling signal for sample 1 to sample 4 that pass the test.

the model can still perform at > 90% accuracy with an at most 11% false positive rate for the Andreev class.

We also evaluate the robustness of our model by testing it against untrained categories of trivial disordered data, such as nanowires with a large disorder in the  $g$  factor and superconducting gap  $\Delta$ . We claim that the model has a roughly ~ 20% false positive rate for these trivial testing data, indicating that it is still moderately robust (see [supplemental experimental procedures S1.3](#)).

### Experimental tests

For the final part, we use our trained classifiers on real experimental ZBP data from recent literature. Since our classifier with 2D data input gives overall higher accuracy than the 1D classifier, we focus the tests on 2D ZBP data testing. Additional 1D datasets are shown in [supplemental experimental procedures S1.3](#). We extract 16 ZBP data images from 10 references over the past decade.<sup>31–40</sup> The images are cropped online and processed to fit properly within our model input format (see [supplemental experimental procedures S1.4](#) for more details). Since XGBoost returns the continuous probability  $p \in [0, 1]$  before the final binary classification, here we show the returned continuous probability since it carries more information than binary value, with a cutoff value  $p_{\text{crit}} = 0.5$ . The positive result probability, i.e., the probability that the model suggests that the system hosts topologically MZMs, for the test

set is shown in [Figure 4A](#). Here, we only emphasize the examples that manage to pass the trial test either with or without noise in the figure.

Four experimental samples from four prior works pass the test from the 2D model either with or without data noise. The patterns of these samples are shown in [Figure 4B](#) from sample 1 to sample 4 in order.<sup>31,34–36</sup> We also included a prediction result for another XGBoost model with noise but trained on a larger dataset including a larger Zeeman energy window, i.e., larger max range  $E_z = 4.48$  meV. The results indicate that our model prediction is somewhat robust against the choice of energy window (see details in [supplemental experimental procedures S1.1](#)). Among them, the most robust sample, upon which both models with and without noise imply a positive MZM presence, has been retracted.<sup>34</sup> For the other 12 samples, the predicted probability for the existence of MZMs always lies consistently below 0.5, indicating that those systems are unlikely to host MZMs. The complete test results are shown in [supplemental experimental procedures S1.4](#). It is also worth mentioning that due to the moderate false positive rate for the untrained disordered categories of our model, even if an example passes the filter of our model, it might still arise from disordered landscapes beyond our model's consideration. Thus, we comment that the actual number of topological MZMs is likely even less than the 4 out of 16 candidates in [Figure 4](#). Overall, our model predicts that a dominant portion of experimental measurements are unlikely to host MZMs on SM-SC coupling nanowires.

## DISCUSSION

In this work, we propose a ML pipeline to detect MZMs in experimentally measured differential tunneling conductance signals. Our work constructively aligns with the recent efforts to identify topological MZMs from trivial states, replacing human eyes with ML-based visual aids. It offers a few potential advantages, including less bias and the possibility to quantify the performance.

It is important to note that our model is only valid under a number of assumptions. It assumes that (1) the experimental nanowire system can be well described within the 1D s-wave Oreg-Lutchyn model framework, (2) the physical mechanisms for impurity and disorder can be mimicked by modifying the diagonal potential landscape and they are the only false positive sources for misleading ZBPs, and (3) the finite temperature effect can be described by Gaussian smearing (see [supplemental experimental procedures S1.1](#)).

To summarize, though our work is still limited by various assumptions and false positives, compared to other more complicated protocols where more sophisticated measurements are required, our work offers the first framework as a ML attempt to identify MZMs directly from experimentally measured ZBP signals. Our classifier model could easily be generalized to suit other quantum property predictions, as long as the system can be well captured by effective model Hamiltonians. In the context of methodology, our ML model uses a mean-field approximation to capture the topological MZM feature under the condition of suppressed quantum fluctuations. This approach, in a broader setting, could inspire more ML works integrated with experiments to tackle strongly correlated systems as a starting point. The model could also be further generalized to conduct parameter extraction on experimental STS data, which resembles the philosophy of other ML parameter extraction methods from experimental data such as time-resolved diffraction<sup>59</sup> or neutron scattering.<sup>60</sup> We also note a recent work extracting the disorder landscape of an SM/SC nanowire with ML.<sup>61</sup>

## EXPERIMENTAL PROCEDURES

### Resource availability

#### Lead contact

Questions or requests for further information about this work should be directed to the lead contact, Prof. Mingda Li ([mingda@mit.edu](mailto:mingda@mit.edu)).

#### Materials availability

This study did not generate new materials.

#### Data and code availability

The data used in this study are numerically generated using our code implementing KWANT, and the code used in this study is available at [https://github.com/vipandyc/ML\\_majorana](https://github.com/vipandyc/ML_majorana).

### Tunneling conductance simulation

The training data of our work are generated by tight-binding simulation on transport properties using the KWANT package.<sup>62</sup> To calculate the scattering matrix, we attach a normal SM nanowire with a lead to the end of the nanowire. The normal SM nanowire has the same form of Hamiltonian as the SC/SM system except for the SC coupling, i.e.,

$$H_{\text{normal}} = T + H_{\text{soc}} + U + H_Z$$

$$= \left( -\frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial x^2} - i\alpha \frac{\partial}{\partial x} \sigma_y - \mu_{\text{normal}} \right) \tau_z + E_Z \sigma_x.$$

Note that there is a finite difference between the normal wire and the SC/SM nanowire in a chemical potential  $\mu_{\text{normal}} - \mu = eV_{\text{gate}}$ , which represents the gate voltage added to the scattering region. As for the lead, the on-site Hamiltonian is the same as the normal nanowire except for an additional potential barrier,  $V_{\text{barrier}}$ :

$$H_{\text{normal}} = \left( -\frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial x^2} - i\alpha \frac{\partial}{\partial x} \sigma_y - \mu_{\text{normal}} + V_{\text{barrier}} \right) \tau_z + E_Z \sigma_x.$$

All relevant physical parameters in the Hamiltonian can be found in [supplemental experimental procedures S1.1](#). After constructing such a system, KWANT allows convenient calculation on the scattering matrix  $S$  on the defined scattering region, i.e., the connecting junction on the lead:

$$G_0(E) = 2 + \sum_{\sigma, \sigma'=\uparrow, \downarrow} \left( |r_{eh}^{\sigma\sigma'}|^2 - |r_{ee}^{\sigma\sigma'}|^2 \right),$$

where  $r_{eh}$  and  $r_{ee}$  are the Andreev and normal reflection amplitudes from the  $S$  matrix, respectively. The calculated tunneling conductance is energy dependent, and by sweeping the Zeeman energy  $E_Z$ , we can obtain a diagram with  $dI/dV$  versus  $E_Z$  and bias energy (voltage)  $V_{\text{bias}}$ , which finally gives an image of our 2D data. Such a numerical method is extensively performed in the relevant area of literature, and we refer readers to references like Pan et al.,<sup>43</sup> Pan and Das Sarma,<sup>44</sup> Prada et al.,<sup>63</sup> and Liu et al.<sup>64</sup> for more details.

### Data processing before training

After generating these raw data, we add Gaussian smearing by adding Gaussian function convolution to our 2D image:

$$F(G) \sim \exp(-G^2 / 2\sigma^2),$$

where  $\sigma = 1\text{pixel}^{-1}$ . The reasons for such processing are 2-fold: first, such Gaussian smearing mimics the finite temperature effect of experimental measurements based on our zero-temperature simulation (see [supplemental experimental procedures S1.1](#)). Secondly, our smearing also smooths out the experimental STS measurement signal, mimicking the resolution function resembling the Gaussian convolution.

In addition to such smearing, to ensure the robustness of our model and emulate the measurement noise, we further add a small noise to the tunneling conductance signal subject to the normal distribution  $\delta G \sim 0.2N(0,1)e^2/h$ .

### ML

Each ML model, depending on the design sophistication, is more suitable for some types of problems than the others. However, especially for problems like MZM detection that are not well explored through the ML perspective, it is better to start approaching the problem with multiple model architectures. This method would not only provide us with the best model for the job but one could also later utilize multiple models for a better performance through a boosting technique. As a result, in our study, we perform in total five ML architectures: PCA, SVM, random forest, XGBoost, and CNNs. All of these essentially have the same framework in which they are trying to search for classifying criteria that divide the input data high-dimensional space ( $28 \times 28$ ) into regions of trivial and topological labeling.

In PCA, the input undergoes a basis transform such that the first basis (principal component) represents the axis in which there is the largest variation of data among the input data, i.e., a projection of data that separates the data by greatest distance. The second component is the same as the first but for the remaining dimensions of the data space. The subsequent components follow the same idea recursively. We then keep the first two components, which capture two largest main features present in the inputs. Of course, as can be seen from our result, the principal component that defines the classification of the data need not be one of the first two. In fact, there might not be one if the relation is not linear between the input and the MZM class. After that, one can utilize any model to try to subdivide the regions into their respective labels. In this work, we use the linear version of SVM, which will be described next.

For SVM, the model focuses on finding the hyperplane in input space that can separate trivial and topological data points. The method is actually used in the last step of PCA. However, one can also directly apply it to the high-dimensional input space directly. Furthermore, it is also common to transform the input with a predetermined non-linear map, called the kernel method, which improves the performance by allowing the non-linear hyperplane.

Both random forest and XGBoost methods use an ensemble of multiple decision tree models. Each decision tree is a collection of hyperplanes that are mostly perpendicular to the input space axes. Hence, each hyperplane is not as powerful as the one from SVM. However, by having multiple of these, the collection can separate the spaces into many regions, which, if fine enough, can separate the data points into trivial and topological accurately. The difference between these two are that random forest performs majority voting between the trees, i.e., each tree performs the same task, while XGBoost manages each tree to perform the smaller task that is the weakness of the other trees.

Lastly, the CNN model takes advantage of the input being images in which the defining feature of the MZM should be captured with some local patterns inside the images. This

means that there should be local correlation between nearby pixels as well as those invariant to translation. Hence, it suggests that the classifying criteria should be in the form of convolution between the input image and a collection of patterns (kernels). From that, the model needs to find the kernel that only appears in the real MZM data.

The methods and models mentioned are well implemented in Python open-source packages. We use the scikit-learn package<sup>65</sup> for PCA, SVM, random forest, and XGBoost classification, and we implement the Pytorch<sup>66</sup> package for building up the simple CNN network for classification. We also include the model and hyper-parameter settings in supplemental experimental procedures S1.3.

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.matt.2024.05.028>.

## ACKNOWLEDGMENTS

The authors thank J. Sau, B. November, and J. Hoffman for the helpful discussions. R.O. and A.C. acknowledge support from the US Department of Energy (DOE), Office of Science (SC), Basic Energy Sciences (BES), award no. DE-SC0021940, and the National Science Foundation (NSF) Designing Materials to Revolutionize and Engineer Our Future (DMREF) Program with award no. DMR-2118448. A.C. is partially supported by the MIT-QEC Seed Fund. M.L. acknowledges support from NSF ITE-2345084, the Class of 1947 Career Development Chair and support from R. Wachnik.

## AUTHOR CONTRIBUTIONS

M.L. conceived and supervised the project. M.C. built the workflow, performed transport calculations to generate training data, and performed ML and topological data analysis. R.O. and A.C. interpreted the key results in the main text figures and helped with ML analysis. All authors took part in preparing the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 13, 2023

Revised: March 15, 2024

Accepted: May 21, 2024

Published: June 17, 2024

## REFERENCES

1. Wang, Y., Wu, H., McCandless, G.T., Chan, J.Y., and Ali, M.N. (2023). Quantum states and intertwining phases in kagome materials. *Nat. Rev. Phys.* 5, 635–658.
2. Zhou, X., Lee, W.-S., Imada, M., Trivedi, N., Phillips, P., Kee, H.-Y., Törmä, P., and Eremets, M. (2021). High-temperature superconductivity. *Nat. Rev. Phys.* 3, 462–465.
3. Wen, J., Yu, S.-L., Li, S., Yu, W., and Li, J.-X. (2019). Experimental identification of quantum spin liquids. *npj Quantum Mater.* 4, 12.
4. von Klitzing, K., Chakraborty, T., Kim, P., Madhavan, V., Dai, X., McIver, J., Tokura, Y., Savary, L., Smirnova, D., Rey, A.M., et al. (2020). 40 years of the quantum Hall effect. *Nat. Rev. Phys.* 2, 397–401.
5. Wen, X.-G. (2017). Colloquium: Zoo of quantum-topological phases of matter. *Rev. Mod. Phys.* 89, 041004.
6. Qi, X.-L., and Zhang, S.-C. (2011). Topological insulators and superconductors. *Rev. Mod. Phys.* 83, 1057–1110.
7. Yu, R., Zhang, W., Zhang, H.-J., Zhang, S.-C., Dai, X., and Fang, Z. (2010). Quantized anomalous Hall effect in magnetic topological insulators. *Science* 329, 61–64.
8. Chang, C.-Z., Zhang, J., Feng, X., Shen, J., Zhang, Z., Guo, M., Li, K., Ou, Y., Wei, P., Wang, L.-L., et al. (2013). Experimental observation of the quantum anomalous Hall effect in a magnetic topological insulator. *Science* 340, 167–170.
9. Deng, Y., Yu, Y., Shi, M.Z., Guo, Z., Xu, Z., Wang, J., Chen, X.H., and Zhang, Y. (2020). Quantum anomalous Hall effect in intrinsic magnetic topological insulator MnBi<sub>2</sub>Te<sub>4</sub>. *Science* 367, 895–900.
10. Armitage, N., Mele, E., and Vishwanath, A. (2018). Weyl and Dirac semimetals in three-dimensional solids. *Rev. Mod. Phys.* 90, 015001.
11. Zhou, Y., Kanoda, K., and Ng, T.-K. (2017). Quantum spin liquid states. *Rev. Mod. Phys.* 89, 025003.
12. Nayak, C., Simon, S.H., Stern, A., Freedman, M., and Das Sarma, S. (2008). Non-Abelian anyons and topological quantum computation. *Rev. Mod. Phys.* 80, 1083–1159.

13. Kitaev, A. (2006). Anyons in an exactly solved model and beyond. *Annals of Physics* 321, 2–111.
14. Alicea, J., Oreg, Y., Refael, G., Von Oppen, F., and Fisher, M.P.A. (2011). Non-Abelian statistics and topological quantum information processing in 1D wire networks. *Nat. Phys.* 7, 412–417.
15. Marra, P. (2022). Majorana nanowires for topological quantum computation. *Journal of Applied Physics* 132.
16. Kitaev, A.Y. (2001). Unpaired Majorana fermions in quantum wires. *Phys. -Usp.* 44, 131–136.
17. Flensberg, K., von Oppen, F., and Stern, A. (2021). Engineered platforms for topological superconductivity and Majorana zero modes. *Nat. Rev. Mater.* 6, 944–958.
18. Fu, L., and Kane, C.L. (2008). Superconducting proximity effect and Majorana fermions at the surface of a topological insulator. *Phys. Rev. Lett.* 100, 096407.
19. Read, N., and Green, D. (2000). Paired states of fermions in two dimensions with breaking of parity and time-reversal symmetries and the fractional quantum Hall effect. *Phys. Rev. B* 61, 10267–10297.
20. Moore, G., and Read, N. (1991). Nonabelions in the fractional quantum Hall effect. *Nuclear Physics B* 360, 362–396.
21. Linder, J., Tanaka, Y., Yokoyama, T., Sudbø, A., and Nagaosa, N. (2010). Unconventional superconductivity on a topological insulator. *Phys. Rev. Lett.* 104, 067001.
22. Ghosh, P., Sau, J.D., Tewari, S., and Das Sarma, S. (2010). Non-Abelian topological order in noncentrosymmetric superconductors with broken time-reversal symmetry. *Phys. Rev. B* 82, 184525.
23. Alicea, J. (2010). Majorana fermions in a tunable semiconductor device. *Phys. Rev. B* 81, 125318.
24. Qi, X.-L., Hughes, T.L., and Zhang, S.-C. (2010). Chiral topological superconductor from the quantum Hall state. *Phys. Rev. B* 82, 184516.
25. Sato, M., and Fujimoto, S. (2009). Topological phases of noncentrosymmetric superconductors: Edge states, Majorana fermions, and non-Abelian statistics. *Phys. Rev. B* 79, 094504.
26. Sau, J.D., Lutchyn, R.M., Tewari, S., and Das Sarma, S. (2010). Generic new platform for topological quantum computation using semiconductor heterostructures. *Phys. Rev. Lett.* 104, 040502.
27. Lutchyn, R.M., Sau, J.D., and Das Sarma, S. (2010). Majorana fermions and a topological phase transition in semiconductor-superconductor heterostructures. *Phys. Rev. Lett.* 105, 077001.
28. Oreg, Y., Refael, G., and Von Oppen, F. (2010). Helical liquids and Majorana bound states in quantum wires. *Phys. Rev. Lett.* 105, 177002.
29. Lutchyn, R.M., Bakkers, E.P.A.M., Kouwenhoven, L.P., Krogstrup, P., Marcus, C.M., and Oreg, Y. (2018). Majorana zero modes in superconductor-semiconductor heterostructures. *Nat. Rev. Mater.* 3, 52–58.
30. Das, A., Ronen, Y., Most, Y., Oreg, Y., Heiblum, M., and Shtrikman, H. (2012). Zero-bias peaks and splitting in an Al-InAs nanowire topological superconductor as a signature of Majorana fermions. *Nat. Phys.* 8, 887–895.
31. Güçlü, Ö., Zhang, H., Bommer, J.D., Moor, M.W., de Car, D., Plissard, S.R., Bakkers, E.P., Geresdi, A., Watanabe, K., Taniguchi, T., et al. (2018). Ballistic Majorana nanowire devices. *Nature nanotechnology* 13, 192–197.
32. Yu, P., Chen, J., Gomanko, M., Badawy, G., Bakkers, E.P.A.M., Zuo, K., Mourik, V., and Frolov, S.M. (2021a). Non-Majorana states yield nearly quantized conductance in proximatised nanowires. *Nat. Phys.* 17, 482–488.
33. Deng, M.T., Yu, C.L., Huang, G.Y., Larsson, M., Caroff, P., and Xu, H.Q. (2014). Parity independence of the zero-bias conductance peak in a nanowire based topological superconductor quantum dot hybrid device. *Sci. Rep.* 4, 7261.
34. Zhang, H., Liu, C.-X., Gazibegovic, S., Xu, D., Logan, J.A., Wang, G., Van Loo, N., Bommer, J.D.S., de Moor, M.W.A., Car, D., et al. (2018). Retracted article: Quantized majorana conductance. *Nature* 556, 74–79.
35. Deng, M.-T., Vaitiekėnas, S., Prada, E., San-Jose, P., Nygård, J., Krogstrup, P., Aguado, R., and Marcus, C.M. (2018). Nonlocality of Majorana modes in hybrid nanowires. *Phys. Rev. B* 98, 085125.
36. Nichele, F., Drachmann, A.C.C., Whiticar, A.M., O'Farrell, E.C., Suominen, H.J., Fornieri, A., Wang, T., Gardner, G.C., Thomas, C., Hatke, A.T., et al. (2017). Scaling of Majorana zero-bias conductance peaks. *Phys. Rev. Lett.* 119, 136803.
37. Chen, J., Woods, B.D., Yu, P., Hocevar, M., Car, D., Plissard, S.R., Bakkers, E.P.A.M., Stanescu, T.D., and Frolov, S.M. (2019). Ubiquitous non-Majorana zero-bias conductance peaks in nanowire devices. *Phys. Rev. Lett.* 123, 107703.
38. Chen, J., Yu, P., Stenger, J., Hocevar, M., Car, D., Plissard, S.R., Bakkers, E.P.A.M., Stanescu, T.D., and Frolov, S.M. (2017). Experimental phase diagram of zero-bias conductance peaks in superconductor/semiconductor nanowire devices. *Sci. Adv.* 3, e1701476.
39. Mourik, V., Zuo, K., Frolov, S.M., Plissard, S.R., Bakkers, E.P.A.M., and Kouwenhoven, L.P. (2012). Signatures of Majorana fermions in hybrid superconductor-semiconductor nanowire devices. *Science* 336, 1003–1007.
40. Deng, M.T., Vaitiekėnas, S., Hansen, E.B., Danon, J., Leijnse, M., Flensberg, K., Nygård, J., Krogstrup, P., and Marcus, C.M. (2016). Majorana bound state in a coupled quantum-dot hybrid-nanowire system. *Science* 354, 1557–1562.
41. Jäck, B., Xie, Y., and Yazdani, A. (2021). Detecting and distinguishing Majorana zero modes with the scanning tunnelling microscope. *Nature Reviews Physics* 3, 541–554.
42. Yin, J.-X., Wu, Z., Wang, J.H., Ye, Z.Y., Gong, J., Hou, X.Y., Shan, L., Li, A., Liang, X.J., Wu, X.X., et al. (2015). Observation of a robust zero-energy bound state in iron-based superconductor Fe (Te, Se). *Nat. Phys.* 11, 543–546.
43. Pan, H., Cole, W.S., Sau, J.D., and Das Sarma, S. (2020). Generic quantized zero-bias conductance peaks in superconductor-semiconductor hybrid structures. *Phys. Rev. B* 101, 024506.
44. Pan, H., and Das Sarma, S. (2020). Physical mechanisms for zero-bias conductance peaks in Majorana nanowires. *Phys. Rev. Res.* 2, 013377.
45. Frolov, S.M., Manfra, M.J., and Sau, J.D. (2020). Topological superconductivity in hybrid devices. *Nat. Phys.* 16, 718–724.
46. Valentini, M., Peñaranda, F., Hofmann, A., Brauns, M., Hauschild, R., Krogstrup, P., San-Jose, P., Prada, E., Aguado, R., and Katsaros, G. (2021). Flux-tunable Andreev bound states in hybrid full-shell nanowires. *APS March Meeting Abstracts* 2021, X58-X5002.
47. Prada, E., San-Jose, P., de Moor, M.W.A., Geresdi, A., Lee, E.J.H., Klinovaja, J., Loss, D., Nygård, J., Aguado, R., and Kouwenhoven, L.P. (2020). From Andreev to Majorana bound states in hybrid superconductor-semiconductor nanowires. *Nat. Rev. Phys.* 2, 575–594.
48. Vuik, A., Nijholt, B., Akhmerov, A., and Wimmer, M. (2019). Reproducing topological properties with quasi-Majorana states. *SciPost Phys.* 7, 061.
49. Avila, J., Peñaranda, F., Prada, E., San-Jose, P., and Aguado, R. (2019). Non-hermitian topology as a unifying framework for the Andreev versus Majorana states controversy. *Commun. Phys.* 2, 133.
50. Yu, P., Chen, J., Gomanko, M., Badawy, G., Bakkers, E.P.A.M., Zuo, K., Mourik, V., and Frolov, S.M. (2021b). Non-Majorana states yield nearly quantized conductance in proximatised nanowires. *Nat. Phys.* 17, 482–488.
51. Marra, P., and Nigro, A. (2022). Majorana/Andreev crossover and the fate of the topological phase transition in inhomogeneous nanowires. *J. Phys. Condens. Matter* 34, 124001.
52. Pikulin, D.I., Heck, B. van, Karzig, T., Martinez, E.A., Nijholt, B., Laeven, T., Winkler, G.W., Watson, J.D., Heedt, S., Temurhan, M., et al. (2021). Protocol to identify a topological superconducting phase in a three-terminal device. Preprint at arXiv. <https://doi.org/10.48550/arXiv:2103.12217>.
53. Aghaei, M., Akkala, A., Alam, Z., Ali, R., Alcaraz Ramírez, A., Andrzejczuk, M., Antipov, A.E., Aseev, P., Astafiev, M., Bauer, B., et al. (2023). InAs-Al hybrid devices passing the topological gap protocol. *Phys. Rev. B* 107, 245423.
54. Whiticar, A.M., Fornieri, A., O'Farrell, E., Drachmann, A.C.C., Wang, T., Thomas, C., Gronin, S., Kallahe, R., Gardner, G.C., Manfra, M.J., et al. (2020). Coherent transport through a Majorana island in an Aharonov-Bohm interferometer. *Nat. Commun.* 11, 3212.
55. Stanescu, T.D., and Tewari, S. (2019). Robust low-energy Andreev bound states in semiconductorsuperconductor structures: Importance of partial separation of component

Majorana bound states. *Phys. Rev. B* 100, 155429.

56. Liu, C.-X., Sau, J.D., Stanescu, T.D., and Das Sarma, S. (2017). Andreev bound states versus Majorana bound states in quantum dot-nanowire-superconductor hybrid structures: Trivial versus topological zero-bias conductance peaks. *Phys. Rev. B* 96, 075161.
57. Carlsson, G. (2020). Topological methods for data modelling. *Nat. Rev. Phys.* 2, 697–708.
58. The GUDHI Project (2015). GUDHI User and Reference Manual (GUDHI Editorial Board).
59. Chen, Z., Shen, X., Andrejevic, N., Liu, T., Luo, D., Nguyen, T., Drucker, N.C., Kozina, M.E., Song, Q., Hua, C., et al. (2023). Panoramic Mapping of Phonon Transport from Ultrafast

Electron Diffraction and Scientific Machine Learning. *Adv. Mater.* 35, 2206997.

60. Samarakoon, A.M., Barros, K., Li, Y.W., Eisenbach, M., Zhang, Q., Ye, F., Sharma, V., Dun, Z.L., Zhou, H., Grigera, S.A., et al. (2020). Machine-learning-assisted insight into spin ice Dy<sub>2</sub>Ti<sub>2</sub>O<sub>7</sub>. *Nat. Commun.* 11, 892.
61. Taylor, J.R., Sau, J.D., and Sarma, S.D. (2023). Machine learning Majorana nanowire disorder landscape. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2307.11068>.
62. Groth, C.W., Wimmer, M., Akhmerov, A.R., and Waintal, X. (2014). Kwant: a software package for quantum transport. *New J. Phys.* 16, 063065.
63. Prada, E., San-Jose, P., and Aguado, R. (2012). Transport spectroscopy of N S nanowire junctions with Majorana fermions. *Phys. Rev. B* 86, 180503.
64. Liu, C.-X., Sau, J.D., and Das Sarma, S. (2017). Role of dissipation in realistic Majorana nanowires. *Phys. Rev. B* 95, 054502.
65. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
66. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32.