Machine Learning the Disorder Landscape of Majorana Nanowires

Jacob R. Taylor, Jay D. Sau, and Sankar Das Sarma
Condensed Matter Theory Center and Joint Quantum Institute, Department of Physics, University of Maryland, College Park, Maryland 20742-4111, USA

(Received 24 July 2023; accepted 6 April 2024; published 16 May 2024)

We develop a practical machine learning approach to determine the disorder landscape of Majorana nanowires by using training of the conductance matrix and inverting the conductance data in order to obtain the disorder details in the system. The inversion carried out through machine learning using different disorder parametrizations turns out to be unique in the sense that any input tunnel conductance as a function of chemical potential and Zeeman energy can indeed be inverted to provide the correct disorder landscape. Our work opens up a qualitatively new direction of directly determining the topological invariant and the Majorana wave-function structure corresponding to a transport profile of a device using simulations that quantitatively match the specific conductance profile. In addition, this also opens up the possibility for optimizing Majorana systems by figuring out the (generally unknown) underlying disorder only through the conductance data. An accurate estimate of the applicable spin-orbit coupling in the system can also be obtained within the same scheme.

DOI: 10.1103/PhysRevLett.132.206602

Introduction.—Hybrid superconductor-semiconductor (SC-SM) nanowire structures are the most extensively studied systems for creating laboratory Majorana zero modes (MZMs), which are localized non-Abelian excitations that can be used for creating a topological quantum computer [1–3]. A recent experiment [1] has provided extensive tunnel conductance measurements to make the case for the existence of small and fragile topological regimes where MZMs should exist at the wire ends. Earlier experiments in such nanowires were strongly adversely affected by unintentional disorder in the system with the MZM-like conductance signatures arising from disorder-induced trivial Andreev bound states (and not from topological MZMs) [2,4–10]. Although cleaner samples are used in [1], the disorder situation even for this latest experiment is not yet completely clarified, but the small fragile gaps reported in [1] indicate that random disorder is likely still playing a role [11,12]. Disorder has thus emerged as the single most important physical mechanism suppressing topology, certainly in Majorana nanowires, but likely in most solid state topological platforms. Although the possible importance of disorder in suppressing MZM physics was pointed out early [13–24], we still do not have any direct information about the disorder in actual samples, hindering progress in the field. Understanding and controlling unintentional (and thus, unknown) random disorder has become by far the most important problem in the search for MZMs in solid state platforms [2]. Progress toward the realization of topological MZMs necessitates an understanding of the underlying disorder leading to cleaner and better samples [2,11,12]. This leads to the other key problem in the field of identifying MZMs in a device based on its transport characterization. This is because, as already mentioned, the unknown disorder leads to transport signatures that are often misinterpreted as MZMs, which often exist in a very limited part of parameter space in moderately disordered devices [12,17]. Ultimately, this motivates the other key challenge in the field, which is to identify when an SC-SM nanowire device has been tuned to support MZMs.

In the current work, we introduce an intuitively appealing machine learning (ML) approach for figuring out and understanding the disorder landscape in Majorana nanowires using the tunnel conductance data, which are the standard measurements for Majorana nanowires carried out in every experiment (and simulated in the corresponding theories). The disorder potential together with other parameters can then be used to quantitatively verify the validity of the model, which in turn can be used to determine if the transport signature indicates a topological superconducting device. The idea is deceptively simple: the tunnel conductance depends crucially on disorder, and therefore, it should be possible to invert the measured conductance to extract the underlying disorder. Of course, the uniqueness of such an inverse scattering problem is a key question since, in principle, it is possible for different disorder landscapes to give similar conductance data. We solve the uniqueness problem *a posteriori* by showing that input disorder producing the conductance agrees with the output disorder obtained from our ML procedure. Our ML approach is powerful, and can be used to obtain other quantities entering the MZM physics, and we estimate the applicable spin-orbit coupling (a key parameter directly determining the nanowire topological gap) using our theory. As a matter

of principle, the approach we use can be extended to improve estimates of other parameters of the model. The transport profile of such a complete model can be used to verify the model and provide the most direct correspondence so far of whether the specific transport profile corresponds to a topological device. We mention that other types of ML approaches have been used for optimizing gate operations in Majorana nanowires [25], but our work is totally different since we use ML to solve the inverse scattering problem of extracting the disorder by using tunnel conductance itself as our training data. More significantly, our ML approach, by providing the unknown parameters for the system, enables a direct realistic estimate for the topological invariant as a function of system parameters to decisively ascertain where the system is topological with MZMs and where it is trivial with no topological properties. This eliminates all subjective judgments about the topological nature of the system as ML itself provides an answer provided sufficient data are used for the training.

Theory.—We use the extensively used minimal 1D model for Majorana nanowires where the system Hamiltonian is given by a 1D Bogoliubov–de Gennes (BdG) equation and the minimal input parameters are the effective mass (m), the spin-orbit coupling (α), the Lande g factor (g), the parent SC gap, the SC-SM coupling, the self-consistent chemical potential μ, the magnetic field B, and the disorder potential $V_{\rm dis}(x)$. We follow Ref. [12] where the details (which are standard in the literature to study SC-SM Majorana nanowires [3,9]) can be found. Note that all quantities other than $V_{\rm dis}$ and α are taken to be fixed and known although this constraint can be relaxed in future works (at the cost of needing much more data for training since the more unknown parameters there are, the larger the training dataset must be). We use, following Ref. [1], a 3 μm long wire for all our results and discretize the BdG equation using a lattice size of 10 nm as in Ref. [12]. We provide the details for the theory and the model in Sec. 1 in the Supplemental Material [26].

Among these SC-SM nanowire parameters, the specific spatial dependence of the disorder potential $V_{\rm dis}(x)$ is completely uncontrolled and varies among devices as well as slowly over time in the same device. In addition, some system parameters such as α depend on the inversion symmetry breaking of the final device structure, which in general is unknown (and most likely varies from device to device). Given a $V_{\rm dis}(x)$, we can solve the BdG equation and generate the four-component tunnel conductance matrix using the KWANT scattering matrix approach [29]. In addition, we use the spin-orbit (SO) coupling α also as an unknown parameter. We vary the disorder potential and α to generate, using KWANT, training transport data (i.e., the four-component conductance matrix G as a function of the magnetic field B and the chemical potential μ) set for our ML algorithm. We provide details of how the conductance

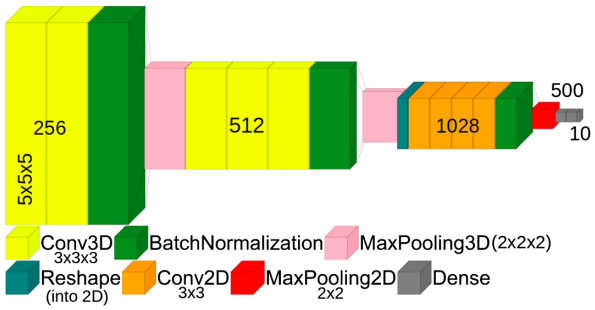

FIG. 1. Neural network diagram. The model combines a CNN with dense layers. The CNN has three convolutional layer sets: two with 256, three with 512, and four with 1028 filters, respectively. Each set is followed by batch normalization and max pooling ($2 \times 2 \times 2$ for 3D, 2×2 for 2D layers). The first two sets are 3D, starting with a $5 \times 5 \times 5$ kernel, then $3 \times 3 \times 3$; the third set contains four 2D layers. A reshaping layer maps to 2D before ending with two 500-size dense layers and an output-sized dense layer. Rectified linear unit (ReLU) functions add nonlinearity. The output Y includes α only for the $5 + 1$ component model.

matrix G is computed as a function of B, μ and other parameters in Sec. 1 of SM. The ML algorithm then predicts $V_{\rm dis}$ and α that can be used using KWANT to reproduce a test transport dataset. Ideally, in the test transport data would come from experiments. In our proof of principle demonstration, our test transport data is generated using KWANT with a random choice of $V_{\rm dis}(x)$ and α. The real test of our ML success is whether the conductance generated using the output $V_{\rm dis}(x)$ and α is close to the test transport dataset.

Method.—Our ML model consists of a convolutional neural network (CNN) created by the package KERAS [30], which builds upon tensorflow [31]. In our specific CNN (Fig. 1), each input step consists of a measurement operation, seven parameters consisting of a row of X (four parameters), which are the components of the conductance matrix $G_{\alpha\beta=L,R}$ and of K (three parameters), which are the parameters bias voltage $V_{\rm bias}$, chemical potential μ, and magnetic field B. The conductance measurements are reshaped into a 3D array based on the values of K columns. The five Fourier components of $V_{\rm dis}(x)$ and α are organized into the output vector Y of the CNN in Fig. 1. As described in Sec. 1 of SM, the conductance matrix X is generated for each instance of K and Y using a KWANT simulation. Later we will also discuss data where we use a 10 Fourier component model for the disorder. In this case Y would be an 10 component vector. The convolutional neural network that was chosen is based on AlexNet [32] to process the conductance plots visually and to detect hidden structures within them. The device consists of sets of 3D convolutional layers followed by a set of 2D convolutional layers, aggregating conductance measurements for different K parameters (μ, B, $V_{\rm bias}$) (in 3D layers) and (μ, B (in the 2D layers). The CNN is sufficient for proof of principle but can likely be enhanced with more elaborate methods, such

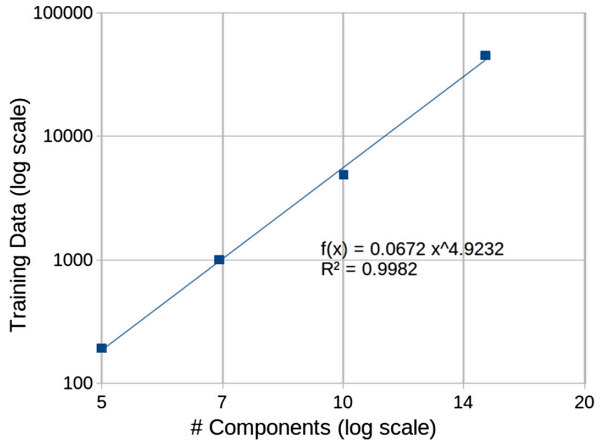

FIG. 2. Training (Y) realizations required for $R_2 = 0.77$ by disorder component count. It's a log-log plot, indicating polynomial scaling with the fit function $f(x) = 0.0672x^{4.9232}$ ($R^2 = 0.9982$), confirming such scaling. A fixed neural network architecture or size is used, shown in Fig. 1, with a constant number of measurements (20 B, 20 μ, and 5 V_{bias} values) except for the five components where, to prevent overfitting, $15 \times 15 \times 5$ is used. The deviation of the ten component point is explained by the NN being originally fine-tuned for it, removing this point suggests near-perfect fitting ($R^2 > 0.999\,95$). Polynomial scaling indicates potential for $n = 30$, $n = 50$, and $n = 100$. For more on disorder components required for certain fidelities, see Sec. II-3 in SM.

as those based on transformers [33], or an encoder-decoder setup [34], which may scale better (See Fig. 2 for CNN scaling). (More ML technical details are given in Sec. II in SM.)

Results.—We use both five-component and 10-component representations for our Gaussian random disorder, obtaining very similar results. We measure the ML effectiveness using the standard scaled R_2 fidelity parameter between the Y_{test} and $Y_{\text{predicted}}$ predicted by the trained CNN. The standard scaling converts the data used in the training processes into a zero mean, unit variance form. $R_2 = 1$ implies perfect prediction, while $R_2 = 0$ is equivalent to just taking the average output of the testing data, i.e., no prediction. This scaling and R_2 can be found in details on the package scikit-learn [35].

To train the model for five disorder components we generate a total of 8000 Y realizations for a fixed disorder strength. We split our data such that 90% is used in the training process and the remaining 10% used for testing. Using the 800 test realizations we achieve a standard scaled $R_2 = 0.991$, which is an excellent fidelity. The standard deviation error in the prediction is $\Delta V_{\text{dis}}(x_i) = \pm 0.0350$ meV and $\Delta\alpha = \pm 0.11$ meV nm. For this model we find it sufficient to use only five bias values of conductance data (i.e., we only use five V_{bias} values in K) to obtain the 0.991 fidelity already. We emphasize that the ML algorithm predicts V_{dis}, α accurately enough that the predicted values can be used in KWANT to match a test conductance quite accurately, as seen in Fig. 3. The good R_2 value of the predictions of the CNN are apparent from comparing the test and predicted Fourier components of disorder and SO coupling seen in Figs. 4 and 5. (More results are given in the Sec. III in the SM.) The accuracy of the predicted parameters implies that the Y values predicted by the CNN

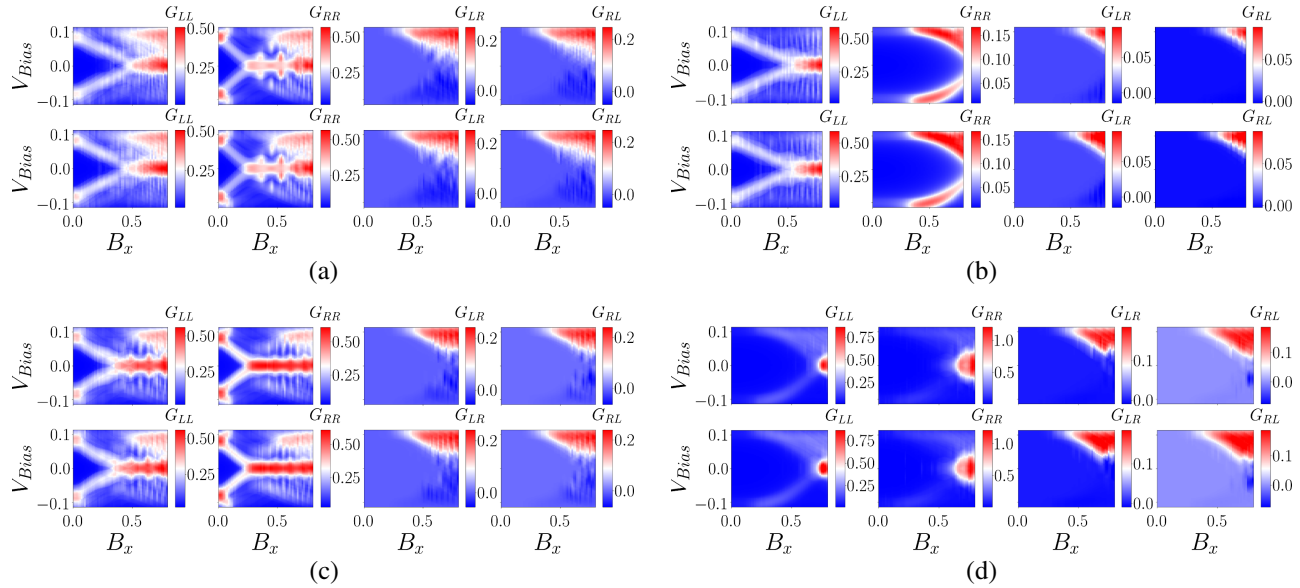

FIG. 3. Representative input and predictions for five-component disorder potentials and α in neural network tests. (a)–(d) Two rows: first for expected conductance (G_{LL}, G_{RR}, G_{LR}, G_{RL}) from input disorder potentials, and second for conductance measurements from predicted potentials. Each subfigure represents different disorder and α realizations (Fig. 4). The associated topological invariant is in Fig. S7 in SM. Inputs and predictions are based on five-component truncated disorder potentials, with conductance simulated across various V_{bias} (meV) and B_x (T) values, at $\mu = 0$.

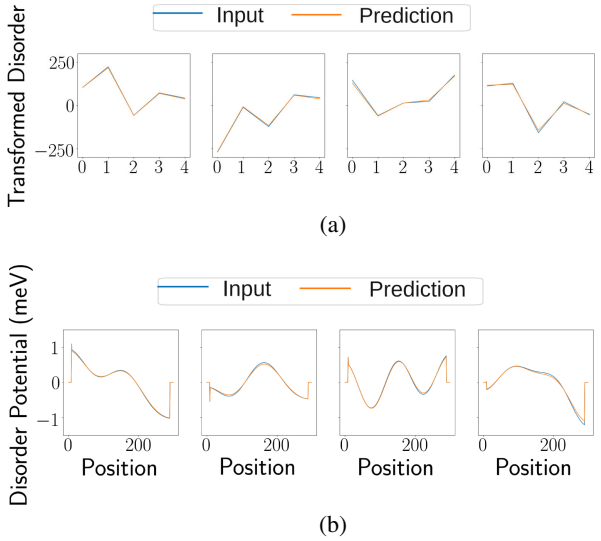

FIG. 4. Representative ML input and predictions for five-component disorder potentials. (a) Shows discrete cosine-transformed components for input and predictions. (b) Real-space disorder potentials with boundary effects. The corresponding topological invariant is found in Fig. S7 in SM. Each sample had a disorder amplitude of 1.5 meV and correlation length of three sites (30 nm), with (b) having longer apparent length due to Fourier component truncation.

can also be used to generate the topological invariant in parameter (i.e., K) space. To test this we compared the topological visibility in $B-\mu$ space between a test case and predicted case and found good agreement as seen in Figs. S7, S8, and S9 in the SM. Further, we find that our model is extremely resistant to experimental errors in conductance measurements to the point where a Gaussian error of magnitude 0.01 e^2/h yields a $R_2 = 0.99$ and

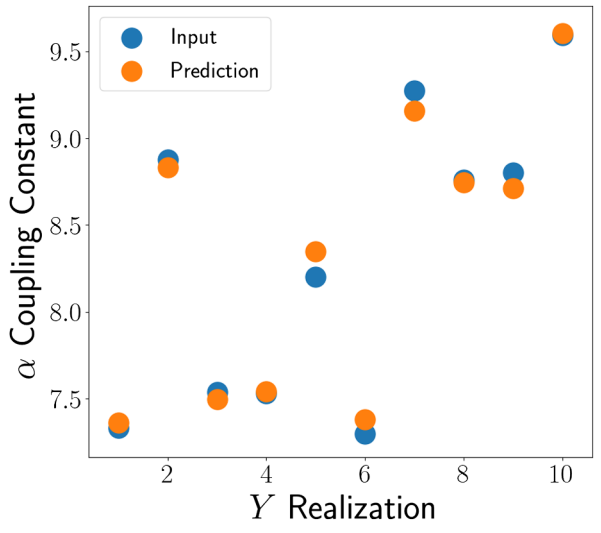

FIG. 5. Representative samples of α neural network test input and predictions. The blue points representing the input α values (in meV nm) and the orange represent the predicted α values.

0.05 e^2/h yields $R_2 = 0.98$, even going to 0.1 e^2/h only decreases to $R_2 = 0.97$. In context the noise floor in experiment [1] is about 0.001 e^2/h, making conductance errors effectively irrelevant.

In Fig. S4 in SM we show the results of our ten-component disorder ML results. To train for ten disorder components we generate a total of 50 000 disorder realizations, 90% of which is used in the training process and the remaining 10% used to test. We use a fixed $\alpha = 8.0$ meV nm. Using the 5000 test realizations we achieve a standard scaled $R_2 = 0.95$, which is great, but can be improved with more datasets. The standard deviation error in the prediction is $\Delta V_{\mathrm{dis}}(x_i) = \pm 0.117$ meV. Because of computational limitations we only use five V_{bias} values (using more values of V_{bias} should increase R_2). This lower fidelity is likely due to the need for more training and potentially a slightly larger K matrix since differentiating among ten component potentials is a more computationally challenging task. Representative results are shown in Fig. S4 in SM, and again, visual inspection shows excellent agreement between the input and output data, verifying the success of our ML approach. When considering further scalability, a couple of factors warrant attention. First, we show that the data requirements do not increase exponentially, making it feasible to compute additional components. Furthermore, since we anticipate that the low-frequency disorder components will be dominant, the number of extra components needed to accurately represent the device's physics may be limited. We do an extended assessment of the number of components required within Sec. II.3 in SM.

Additionally, our method's disorder prediction can be verified through the use of additional conductance measurements. While we do not expect additional effects to be necessary, in principle, if our model proves inadequate for describing the physics of an experimental device, it is possible that our learning scheme could produce an invalid disorder prediction. This possibility can be significantly mitigated by conducting measurements with a new $\mathbf{K}'_{\mathrm{full}}$ (see Sec. II in SM), which is not integrated into the ML scheme, and comparing the results to the theoretical conductance values generated by KWANT for the respective disorder prediction. This allows for the experimental validation of the machine-learned system parameters. Subsequently, one could iteratively introduce additional potential physical effects into the model until it successfully passes all $\mathbf{K}'_{\mathrm{full}}$ tests, leading to a better understanding of the device's underlying physics.

Conclusion.—We introduce a ML approach to extract unknown system parameters, particularly the disorder landscape, from the simulated (or measured) tunnel conductance data in hybrid SC-SM Majorana nanowire structures. We validate the approach by using simulated conductance data as the training set, establishing that such training leads to strongly predictive results for both disorder and spin-orbit coupling using as few as three parameters

(i.e., K) for the conductance data used in our simulations. The accuracy of the predicted results can be verified by comparing the predicted conductance as well as topological visibility profiles to the test profiles. The predicted profiles, which are functions of multiple parameters such as V_{bias} and μ are generated by KWANT from the parameters that are predicted by ML to match a given test conductance data. The topological visibility profile as a function of B and μ is also generated by KWANT for test conductance data generated by KWANT. Using experimental conductance data (and a bigger computer), one should be able to generalize our approach to obtain all the relevant parameters for Majorana nanowires, not only the disorder and the SO coupling as we do, but also the g factor, the superconductor, the number of occupied sub-bands, etc., since our ML approach is general, and only requires as inputs sufficient amount of conductance training data that are easy to obtain both experimentally and theoretically.

Our method, given enough training data and computing resources, should in principle be able to decisively indicate, just through our ML protocol, whether a set of conductance data in a particular sample indicates an underlying topological (or trivial) system with MZMs (or not). In particular, by determining all the relevant unknowns, one can simulate a device and calculate its topological invariant directly using the output parameters, decisively indicating whether the sample is or is not topological.

We thank Maissam Barkeshli for teaching a machine learning course that helped us in developing our approach presented in this Letter. This work is supported by the Laboratory for Physical Sciences.

[1] M. Aghaee, A. Akkala, Z. Alam, R. Ali, A. A. Ramirez, M. Andrzejczuk, A. E. Antipov, P. Aseev, M. Astafev, B. Bauer *et al.*, Phys. Rev. B **107**, 245423 (2023).
[2] S. Das Sarma, Nat. Phys. **19**, 165 (2023).
[3] S. D. Sarma, M. Freedman, and C. Nayak, npj Quantum Inf. **1**, 1 (2015).
[4] H. Zhang, M. W. de Moor, J. D. Bommer, D. Xu, G. Wang, N. van Loo, C.-X. Liu, S. Gazibegovic, J. A. Logan, D. Car *et al.*, arXiv:2101.11456.
[5] S. Ahn, H. Pan, B. Woods, T. D. Stanescu, and S. Das Sarma, Phys. Rev. Mater. **5**, 124602 (2021).
[6] B. D. Woods, S. Das Sarma, and T. D. Stanescu, Phys. Rev. Appl. **16**, 054053 (2021).
[7] H. Pan, C.-X. Liu, M. Wimmer, and S. Das Sarma, Phys. Rev. B **103**, 214502 (2021).
[8] S. Das Sarma and H. Pan, Phys. Rev. B **103**, 195158 (2021).
[9] H. Pan, J. D. Sau, and S. Das Sarma, Phys. Rev. B **103**, 014513 (2021).
[10] H. Pan and S. Das Sarma, Phys. Rev. Res. **2**, 013377 (2020).
[11] S. D. Sarma and H. Pan, Phys. Rev. B **108**, 085415 (2023).
[12] S. D. Sarma, J. D. Sau, and T. D. Stanescu, Phys. Rev. B **108**, 085416 (2023).
[13] O. Motrunich, K. Damle, and D. A. Huse, Phys. Rev. B **63**, 224204 (2001).
[14] A. M. Lobos, R. M. Lutchyn, and S. Das Sarma, Phys. Rev. Lett. **109**, 146403 (2012).
[15] A. R. Akhmerov, J. P. Dahlhaus, F. Hassler, M. Wimmer, and C. W. J. Beenakker, Phys. Rev. Lett. **106**, 057001 (2011).
[16] J. Liu, A. C. Potter, K. T. Law, and P. A. Lee, Phys. Rev. Lett. **109**, 267002 (2012).
[17] İ. Adagideli, M. Wimmer, and A. Teker, Phys. Rev. B **89**, 144506 (2014).
[18] D. Bagrets and A. Altland, Phys. Rev. Lett. **109**, 227005 (2012).
[19] S. Takei, B. M. Fregoso, H.-Y. Hui, A. M. Lobos, and S. Das Sarma, Phys. Rev. Lett. **110**, 186803 (2013).
[20] P. W. Brouwer, M. Duckheim, A. Romito, and F. Von Oppen, Phys. Rev. Lett. **107**, 196804 (2011).
[21] M.-T. Rieder, P. W. Brouwer, and I. Adagideli, Phys. Rev. B **88**, 060509(R) (2013).
[22] D. I. Pikulin, J. Dahlhaus, M. Wimmer, H. Schomerus, and C. Beenakker, New J. Phys. **14**, 125011 (2012).
[23] J. D. Sau and S. Das Sarma, Phys. Rev. B **88**, 064506 (2013).
[24] J. D. Sau, S. Tewari, and S. Das Sarma, Phys. Rev. B **85**, 064512 (2012).
[25] M. Thamm and B. Rosenow, Phys. Rev. Lett. **130**, 116202 (2023).
[26] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.132.206602 for additional information regarding the specifics of the theoretical model, training generation, and numerical results, which includes Refs. [27,28].
[27] R. M. Lutchyn, J. D. Sau, and S. Das Sarma, Phys. Rev. Lett. **105**, 077001 (2010).
[28] J. D. Sau, R. M. Lutchyn, S. Tewari, and S. D. Sarma, Phys. Rev. B **82**, 094522 (2010).
[29] C. W. Groth, M. Wimmer, A. R. Akhmerov, and X. Waintal, New J. Phys. **16**, 063065 (2014).
[30] F. Chollet *et al.*, KERAS (2015), https://keras.io.
[31] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from 10.5281/zenodo.4724125.
[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Adv. Neural Inf. Process. Syst. **25** (2012).
[33] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, in *Proceedings of the IEEE/CVF international conference on computer vision* (2021), pp. 10012–10022, 10.1109/ICCV48922.2021.00986.
[34] G. J. Percebois, A. Lacerda-Santos, B. Brun, B. Hackens, X. Waintal, and D. Weinmann, SciPost Phys. **15**, 242 (2023).
[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, J. Mach. Learn. Res. **12**, 2825 (2011).