

# Phase Diagram of Disordered Higher Order Topological Insulator: a Machine Learning Study

Hiromu Araki,<sup>1,\*</sup> Tomonari Mizoguchi,<sup>2</sup> and Yasuhiro Hatsugai<sup>1,2</sup>

<sup>1</sup>Graduate School of Pure and Applied Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8571, Japan

<sup>2</sup>Department of Physics, University of Tsukuba, Tsukuba, Ibaraki 305-8571, Japan

A higher order topological insulator is a new concept of topological states of matter, which is characterized by the emergent boundary states whose dimensionality is lower by more than two compared with that of the bulk, and draws a considerable interest. Yet, its robustness against disorders is still unclear. Here we investigate a phase diagram of higher order topological insulator phases in a breathing Kagome model in the presence of disorders, by using a state-of-the-art machine learning technique. We find that the corner states survive against the finite strength of disorder potential as long as the energy gap is not closed, indicating the stability of the higher order topological phases against the disorders.

**Introduction.**— Understanding and classifying topological states of matter are central issues in today's condensed matter physics [1–3]. Among a number of such systems, topological states in gapped free-fermion systems are the most well-understood example. There, the keen interplay between symmetry and topology gives rise to various non-trivial states, and the elegant mathematical tools such as group theory, Clifford algebra, and K-theory allow us to classify topological insulators (TIs) and superconductors (TSCs) under the internal [4–6], space-group [7–9], and magnetic space-group [10] symmetries.

However, this is not the end of the story. Very recently, a new class of topological insulators has been found, which is called a higher order topological insulator (HOTI) [11–19]. In this newly-introduced concept, the “order” means the dimensionality of the boundary state. Namely, in conventional TIs in  $d$ -dimensions, the gapless boundary states emerge at the  $(d-1)$ -dimensional boundary [20, 21], and in that sense, conventional TIs are the first order TIs. Similarly, in  $n$ th order TIs ( $2 \leq n \leq d$ ), gapless boundary states emerge at the  $(d-n)$ -dimensional boundary. In many cases, the emergence of such gapless states is protected by the spacial symmetries such as mirror symmetries and rotational symmetries.

To reveal the physical properties of HOTIs, studying the effects of disorders is important. Indeed, for the first order TIs protected by the time-reversal symmetry, the most essential property is the prohibition of back-scatterings of the helical edge/surface states. As a results, the phase diagrams of the disordered TIs can be obtained by investigating the conductance [22–25], and their validity is evidenced by the direct calculation of topological indices [26–30]. For HOTIs, on the other hand, the definition of the disordered HOTIs, relying on either physical quantities or topological indices, is still unclear. Nevertheless, we expect that the boundary states of the HOTI are localized even in the presence of weak disorders, and that the existence of the boundary states may serve as an indicator of the HOTI phase

in the disordered systems. To examine this scenario, it is desirable to employ a method that can distinguish the disordered HOTI from other phases in a systematic manner.

To this end, in this Letter, we investigate the robustness of the HOTI against disorders by using a machine learning method. The machine learning method, especially the neural network, is widely used in the various fields in physics, including quantum many-body systems [31–33], Monte Carlo simulations [34], high-energy

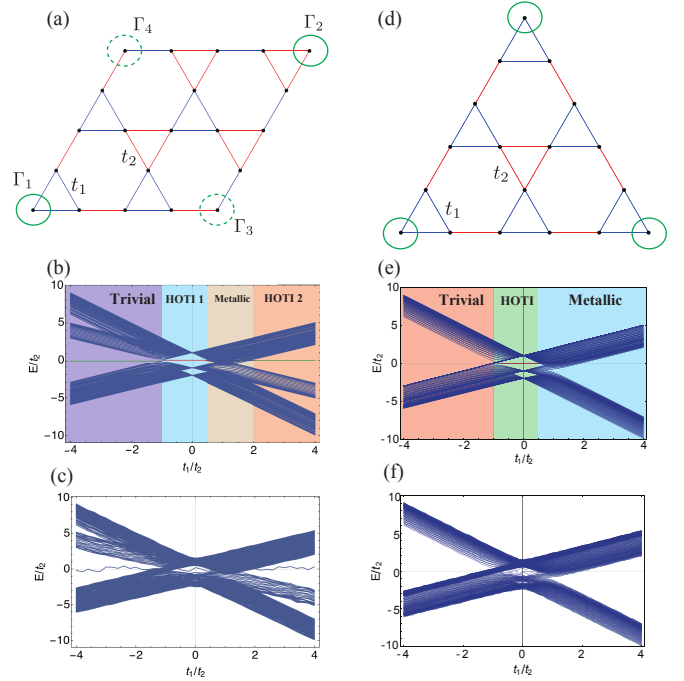


FIG. 1. A Kagome flake with (a) a rhombus geometry and (d) a triangle geometry. In each geometry, corners are denoted by green circles. The energy spectrum of the rhombus as a function of  $t_1/t_2$  for (b) a rhombus geometry and (e) a triangle geometry. The energy spectrum with the disorder  $W = 1.1$  for (c) a rhombus geometry and (f) a triangle geometry.

physics [35, 36] and astrophysics [37]. As for the topological states, the first order TIs and TSCs have been successfully classified in the presence of disorders, and the resulting phase diagrams reproduce those obtained by the other methods. [38–40]. This motivates us to apply this method to disordered HOTIs.

In the present work, we employ a tight-binding Hamiltonian on a breathing Kagome lattice [17–19, 41]. The Hamiltonian, as we will explain later, is known to possess the second order TI phases protected by symmetries [17–19]. For this system, we first perform supervised learning in the clean limit, where the phase diagram has already been known [17–19]. Once the supervised “model” for classification is obtained, we then introduce on-site disorders and classify the phases by using that model. We obtain phase diagrams for the disordered breathing Kagome model in two kinds of global geometries. For both of them, we find that the HOTI phase survives as far as the energy gap does not collapse, due to the robustness of the corner states against the disorders.

*Model and phase diagram in the clean limit.*— We consider a tight binding Hamiltonian for spinless fermions on a breathing Kagome lattice:

$$\mathcal{H}_0 = - \sum_{i,j} t_{ij} c_i^\dagger c_j \quad (1)$$

where  $c_i^\dagger$  and  $c_i$  are, respectively, the creation and annihilation operators of an electron on a site  $i$ , and  $t_{ij} = t_1(t_2)$  if the bond between  $i$  and  $j$  belongs to the nearest-neighbor bond and it lives on the upward (downward) triangles [Figs. 1(a) and 1(d)]. In the following, we take  $t_2$  to be positive.

The band structure of the bulk is obtained by diagonalizing  $\mathcal{H}_0$  in the Fourier space. There exists a flat band with energy  $t_1 + t_2$ . The other two bands are gapped when  $|t_1| \neq |t_2|$ ; the gapless linear dispersion appears at  $K$  and  $K'$  for  $t_2 = t_1$ , and at  $\Gamma$  point for  $t_2 = -t_1$ .

To study HOTI phases, we have to consider finite system with an open boundary condition. So far, it is known that there are two choices of global geometries, namely, rhombus and triangular geometries [see Fig. 1(a) and 1(d), respectively]. In both of these geometries, the model has HOTI phases with zero dimensional localized states, i.e. corner states [17–19]. This can be understood by considering the limit of  $|t_1| = 0$  or  $|t_2| = 0$ . For instance, if  $|t_1| = 0$ , the “trimers” are formed on all downward triangles, and the “dimers” at all edges which do not belong to the downward triangles. For a rhombus geometry, only the site at  $\Gamma_1$  corner is isolated and serves as a zero energy mode, and it survives even for finite  $t_1$ , as long as  $|t_1| < |t_2|$  is satisfied. Similarly, for  $|t_1| > |t_2|$ , the corner state appears at  $\Gamma_2$  corner. Indeed, in Ref. 17, it was shown that the corner zero modes can be explicitly constructed when  $|t_1| \neq |t_2|$ . Note that the other two corners,  $\Gamma_3$  and  $\Gamma_4$ , do not possess corner states at

any parameters. For a triangular geometry, on the other hand, all three corners belong to upward triangles, so the corner states exist only when  $-1 \leq t_1/t_2 \leq 1/2$ , and those corner states have three-fold degeneracy [18].

The difference of corner states with different geometry is also crucial when defining the phases. In a rhombus geometry, the corner state exists for every parameter, but it matters whether the corner state are in-gap state or not. So, we classify the phases in such a way that we fix the electron number as  $N = \nu_R = M^2$ , where we consider the system contains  $M \times M$  rhombuses ( $3M^2 - 2M$  sites), and see what the highest occupied state is. With  $t_1/t_2$  as a tuning parameter, there are four phases [19] [Fig. 1(b)]. For  $t_1/t_2 > 2$ , the system is in a HOTI phase, in which a zero energy state localized at the  $\Gamma_2$  corner exists in a band gap. We referred to this state as “HOTI 2”, named after the position of the corner state. For  $1/2 < t_1/t_2 < 2$ , the system is in a metallic phase, which has zero energy bulk or edge states, and the corner state is buried into those states. For  $-1 < t_1/t_2 < 1/2$ , there is another HOTI phase, in which a zero energy state localized at the  $\Gamma_1$  corner and it is in the band gap. We label this state “HOTI 1”. For  $t_1/t_2 < -1$ , the system is in a “trivial” phase. In this phase, there is a zero energy state localized at the  $\Gamma_2$  corner in the gap, but it is not the highest occupied state. Rather, one of the degenerated flat modes is the highest occupied state and it masks the corner state.

In contrast to a rhombus geometry, there are only three phases for a triangle geometry since there is only one HOTI phase for  $-1 \leq t_1/t_2 \leq 1/2$  [18]. In this phase, the corner states appear at  $(\nu_T - 1)$ -th,  $\nu_T$ -th, and  $(\nu_T + 1)$ -th states, with  $\nu_T = \frac{M(M+3)}{2}$ , for the system consists of  $\frac{M(M-1)}{2}$  upward triangles [i.e.  $\frac{3M(M-1)}{2}$  sites]. We therefore fix the electron number as  $N = \nu_T$  so that the highest occupied state is the corner state in this region. The other two phases are trivial ( $t_1/t_2 \leq -1$ ) and metallic ( $t_1/t_2 \geq 1/2$ ) phases, which are essentially the same as those for the rhombus geometry.

To study the effects of disorders, we introduce a on-site random potential,  $\mathcal{H}_R$ , where

$$\mathcal{H}_R = \sum_i w_i c_i^\dagger c_i. \quad (2)$$

Here  $w_i$  are uniform randomnesses in  $[-W/2, W/2]$  with fixed  $W$ . The total Hamiltonian we consider is  $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_R$ .

*Methods.*— We classify the phases of the disordered system by using a machine learning technique. Here a neural-network-based algorithm is implemented by using the open-source library PyTorch [42]. For the details of the architecture of the neural network, see Supplemental Material [43].

Our input data is the electron density of the highest occupied single-particle state. This choice is similar to the previous works on the first order TIs [38]. For a

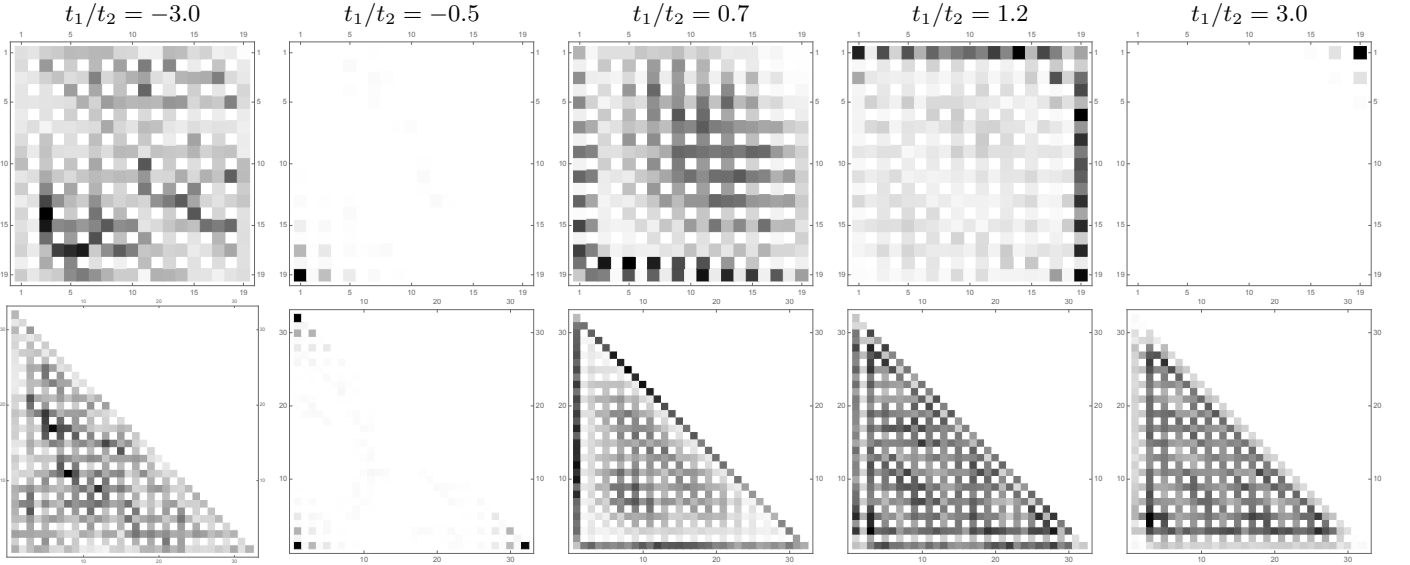


FIG. 2. The averaged wave functions of the highest occupied state with  $W = 1$  on a rhombus geometry (the upper row) and a triangle geometry (the lower row). The average is taken over 30 samples. For a triangle geometry, the average over three corner states is also taken (see the main text).

rhombus geometry, the input data is  $|\phi_i^{\nu_R}|^2$  defined as  $c_{\nu_R} = \sum_i \phi_i^{\nu_R} c_i$ , with  $c_{\nu_R}$  being the annihilation operator of the state  $\nu_R$ . For a triangle geometry, the input data is  $(|\phi_i^{\nu_T-1}|^2 + |\phi_i^{\nu_T}|^2 + |\phi_i^{\nu_T+1}|^2)/3$ , which is the average electron density among three corner states. This particular choice is convenient since it preserves the three-fold rotational symmetry inherent in this geometry. To make the input data as two-dimensional square-shaped images, we transform the geometry from a Kagome lattice to a square lattice with vacancy sites [43].

To classify the phases, we first perform a supervised learning to create a model. To do this, we first prepare 5000 images with randomly-chosen values of  $t_1/t_2$  in the clean limit. Each wave function is tied with the label of the phases which is represented by a four- (three-) dimensional vector for a rhombus (triangular) geometry. For a vector corresponding to the  $n$ -th phase, only  $n$ -th component is unity and all the others are zero. Then the model is trained such that the deviation from the answer and the output (or a cost function) is minimized. This can be archived by updating the parameters in the model on the basis of a gradient method. Note that the output data are also in the form of four- (three-) dimensional vectors, and that  $n$ th component of output can be interpreted as a probability that the state belongs to  $n$ th phase. In the present system, the obtained model can reproduce the data over 99% of accuracy.

Then, we prepare data for the disordered system. Here we take an average over 30 samples at each parameter set. As is pointed out in Ref. 40, this procedure is necessary to restore the translational symmetry that is broken by disorders. Indeed, the wave function in the presence of

disorders for a single snapshot is localized randomly for large  $W$ , which may make it difficult for the model to distinguish the corner states from disordered localized states.

The averaged wave functions at  $W = 1$  for several parameters are shown in Fig. 2. Clearly, the corner states exist even in the presence of disorders at the HOTI phase, which serve as an indicator of the HOTI phase for the neural networks, as we will see later. Interestingly, for the metallic phases, we see the mixture of bulk and edge states, which is attributed to the fact that bulk and edge states are degenerate in the clean limit. Note that, in the supervised learning, both bulk and edge states are taught to be as a sign of a metallic phase, thus even the mixture of them can be recognized as a metallic phase.

*Results for rhombus geometry.*— We first consider the rhombus geometry with the system size  $M = 10$ . The phase diagram is shown in Fig. 3(a). The colors represent the phases: Blue, right-blue, yellow, and orange regions, respectively, correspond to trivial, HOTI 1, metallic, and HOTI 2 phases.

The HOTI 1 phase turns into the trivial phase for  $W > W_c^{(1)}$ , while HOTI 2 phase into the gapless phase for  $W > W_c^{(2)}$ , indicating that the level crossing between corner states and the states nearby zero-energy occurs in these regimes. If it is the case, one can estimate  $W_c^{(1,2)}$  with respect to the gap between corner states and nearby bulk/edge states. To be specific, the formulae for  $W_c^{(1)}$  are  $2t_1/t_2 + 2$  for  $-1 \leq t_1/t_2 \leq 0$ , and  $-2t_1/t_2 + 2$  for  $0 \leq t_1/t_2 \leq 1/2$ ; similarly, those for  $W_c^{(2)}$  are  $2t_1/t_2 - 2$  for  $t_1/t_2 > 2$ . These estimated values are denoted by

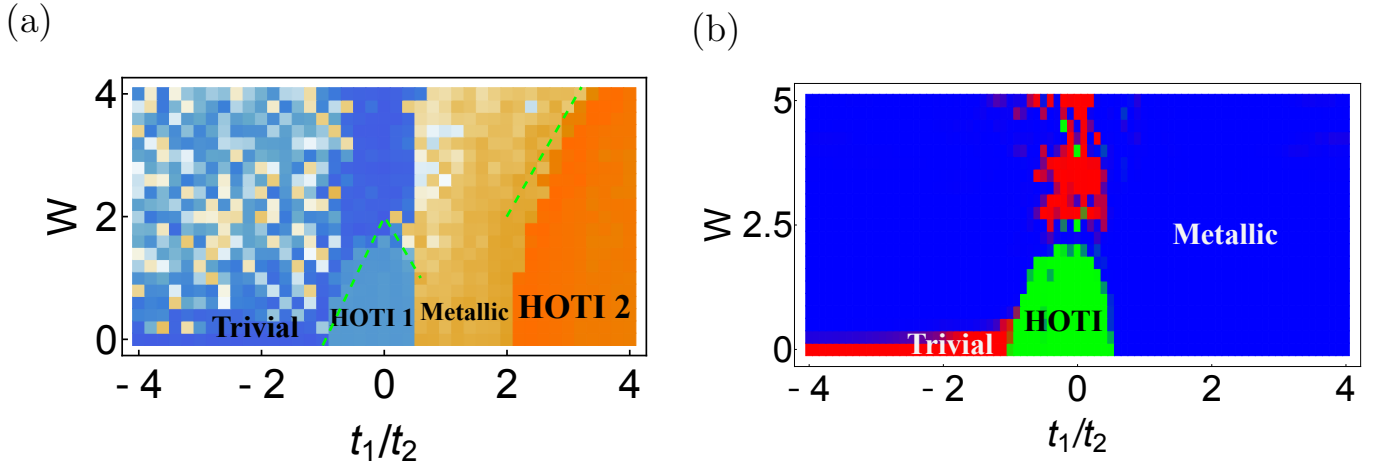


FIG. 3. (a) The phase diagram of the breathing Kagome lattice with disorders for the rhombus geometry. The blue, right-blue, yellow and orange areas denote trivial, HOTI 1, HOTI 2 and metallic phases, respectively. Green dashed lines denote the energy gap between the corner states and the bulk/edge states. (b) The phase diagram for the triangle geometry. The red, blue and green areas denote trivial, HOTI, and metallic phases, respectively.

green lines, which turn out to be good approximations.

Looking at the other phases, the trivial phase is the most fragile against disorders. Indeed, we see in Fig. 3(a) that there is a mosaic-like region, which means that the phase is “unknown” for the trained neural network. (To be more specific, in this region, the confidence is less than 98% for all four phases [43].) We expect that this phase is likely to be an Anderson-localized phase, because of the following reason: In this regime, the highest occupied state belongs to a flat band, which has a massive degeneracy. Then, the disorders immediately lift the degeneracy and pick one of the localized state. This tendency to the localization can be captured by looking at the wave function in a single snapshot (see Supplemental Material [43]).

*Results for triangle geometry.*— Next, we show the results for the triangle geometry. The phase diagram for  $M = 16$  is shown in Fig. 3(b). Here, the output in the form of three-dimensional vectors is mapped onto the RGB component of the color map; red, blue, and green regions respectively, correspond to the trivial, HOTI, and metallic phases.

We see that the HOTI phase turns into the trivial phase above a critical strength of the disorder potential, which is again roughly estimated by the energy gap. This behavior is similar to the rhombus geometry, so one may speculate that the HOTI phase is stable against the disorders as far as the corner states survive.

We also see that the “unknown” phase above the trivial phase is extremely small compared with the rhombus geometry, although the single-shot wave function suggests that the Anderson localization occurs, as is in the case of the rhombus geometry [43]. In the machine learning method, the state is classified as a metallic phase rather

than the “unknown” phase. The reason for this is unclear, but considering the result for the rhombus geometry, it may be due to the insertion of “redundant” sites in making input data [44].

*Summary.*— We have investigated the phase diagram of the disordered breathing Kagome model by using a machine learning method. We use the wave function of the highest occupied state as input data. By doing so, the neural network can distinguish the HOTI phase from other phases by the existence of the corner states. Our result reveals that the HOTI phase is robust against the disorders as far as the disorder strength does not exceed the energy gap.

So far, the disordered HOTI has not been studied in alternative approaches, such as a field-theoretical method and/or numerical calculations. One possible way to determine the phase diagram of the disordered system is to use the transfer matrix method. For the conventional TIs/TSCs, the finite-size scaling of eigenvalues of transfer matrices distinguish the topological phase from the others [25, 38–40, 45]. However, due to the dimensionality of the corner state, the finite-size scaling can not be applied to the present model, so we need to develop alternative methods. One of the candidates of such methods is to use the quantized Berry phases. For the Kagome HOTI model, the  $\mathbb{Z}_3$  Berry phase is defined for the additional local twist parameters and is quantized in the clean limit [41]. Thus, it may serve as a topological order parameter even in the disordered system. Developing such theories and examining the present results will be an interesting and important future problem.

*Acknowledgment.*— The authors would like to thank Nobuyuki Yoshioka for drawing our attention to application of machine learning techniques to topological phases.

They also thank Yutaka Akagi and Hosho Katsura for the useful comments. This work is partly supported by JSPS KAKENHI Grant Number JP17H06138 and JP16K13845.

---

\* araki@rhodia.ph.tsukuba.ac.jp

- [1] M. Z. Hasan and C. L. Kane, *Rev. Mod. Phys.* **82**, 3045 (2010).
- [2] X.-L. Qi and S.-C. Zhang, *Rev. Mod. Phys.* **83**, 1057 (2011).
- [3] X.-G. Wen, *Rev. Mod. Phys.* **89**, 041004 (2017).
- [4] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, *Phys. Rev. B* **78**, 195125 (2008).
- [5] A. Kitaev, *AIP Conference Proceedings* **1134**, 22 (2009).
- [6] S. Ryu, A. P. Schnyder, A. Furusaki, and A. W. W. Ludwig, *New Journal of Physics* **12**, 065010 (2010).
- [7] R.-J. Slager, A. Mesaros, V. Juričić, and J. Zaanen, *Nature Physics* **9**, 98 (2012).
- [8] T. Morimoto and A. Furusaki, *Phys. Rev. B* **88**, 125129 (2013).
- [9] H. C. Po, A. Vishwanath, and H. Watanabe, *Nature Communications* **8**, 50 (2017).
- [10] H. C. Po, H. Watanabe, and A. Vishwanath, *arXiv:1707.01903 [cond-mat.str-el]*.
- [11] S. Hayashi, *arXiv:1611.09680 [math-ph]*.
- [12] K. Hashimoto, X. Wu, and T. Kimura, *Phys. Rev. B* **95**, 165443 (2017).
- [13] W. A. Benalcazar, B. A. Bernevig, and T. L. Hughes, *Phys. Rev. B* **96**, 245115 (2017).
- [14] W. A. Benalcazar, B. A. Bernevig, and T. L. Hughes, *Science* **357**, 61 (2017).
- [15] Z. Song, Z. Fang, and C. Fang, *Phys. Rev. Lett.* **119**, 246402 (2017).
- [16] F. Schindler, A. M. Cook, M. G. Vergniory, Z. Wang, S. S. P. Parkin, B. A. Bernevig, and T. Neupert, *Science Advances* **4** (2018), 10.1126/sciadv.aat0346.
- [17] F. K. Kunst, G. van Miert, and E. J. Bergholtz, *Phys. Rev. B* **97**, 241405 (2018).
- [18] M. Ezawa, *Phys. Rev. Lett.* **120**, 026801 (2018).
- [19] Y. Xu, R. Xue, and S. Wan, *arXiv:1711.09202 [cond-mat.str-el]*.
- [20] Y. Hatsugai, *Phys. Rev. Lett.* **71**, 3697 (1993).
- [21] Y. Hatsugai, *Phys. Rev. B* **48**, 11851 (1993).
- [22] R. Shindou and S. Murakami, *Phys. Rev. B* **79**, 045321 (2009).
- [23] R. Shindou, R. Nakai, and S. Murakami, *New Journal of Physics* **12**, 065008 (2010).
- [24] S. Ryu and K. Nomura, *Phys. Rev. B* **85**, 155138 (2012).
- [25] K. Kobayashi, T. Ohtsuki, and K.-I. Imura, *Phys. Rev. Lett.* **110**, 236803 (2013).
- [26] Q. Niu, D. J. Thouless, and Y.-S. Wu, *Phys. Rev. B* **31**, 3372 (1985).
- [27] E. Prodan, *Phys. Rev. B* **83**, 195119 (2011).
- [28] H. Katsura and T. Koma, *J. Math. Phys.* **57**, 021903 (2016).
- [29] H. Katsura and T. Koma, *J. Math. Phys.* **59**, 031903 (2018).
- [30] Y. Akagi, H. Katsura, and T. Koma, *J. Phys. Soc. Jpn.* **86**, 123710 (2017).
- [31] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [32] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, *Phys. Rev. B* **96**, 205152 (2017).
- [33] H. Fujita, Y. O. Nakagawa, S. Sugiura, and M. Oshikawa, *Phys. Rev. B* **97**, 075114 (2018).
- [34] H. Shen, J. Liu, and L. Fu, *Phys. Rev. B* **97**, 205140 (2018).
- [35] S. Whiteson and D. Whiteson, *Engineering Applications of Artificial Intelligence* **22**, 1203 (2009).
- [36] P. Baldi, P. Sadowski, and D. Whiteson, *Nature Communications* **5**, 430 (2014).
- [37] M. Morii, S. Ikeda, N. Tominaga, M. Tanaka, T. Morokuma, K. Ishiguro, J. Yamato, N. Ueda, N. Suzuki, N. Yasuda, and N. Yoshida, *Publications of the Astronomical Society of Japan* **68**, 104 (2016).
- [38] T. Ohtsuki and T. Ohtsuki, *J. Phys. Soc. Jpn.* **85**, 123706 (2016).
- [39] T. Ohtsuki and T. Ohtsuki, *J. Phys. Soc. Jpn.* **86**, 044708 (2017).
- [40] N. Yoshioka, Y. Akagi, and H. Katsura, *Phys. Rev. B* **97**, 205110 (2018).
- [41] Y. Hatsugai and I. Maruyama, *Europhys. Lett.* **95**, 20003 (2011).
- [42] <https://pytorch.org>.
- [43] See Supplemental Material.
- [44] For a triangle geometry, more than half of the sites are “redundant” (empty) sites which are inserted to make the wave functions in the form of a square image (see Supplemental Material A). We speculate that this makes it difficult to distinguish one phase from another, especially the Anderson localization phase from the metallic phase.
- [45] A. MacKinnon and B. Kramer, *Zeitschrift für Physik B Condensed Matter* **53**, 1 (1983).

# Supplemental material for “Phase Diagram of Disordered Higher Order Topological Insulator: a Machine Learning Study”

## A Input Data for the Neural Network

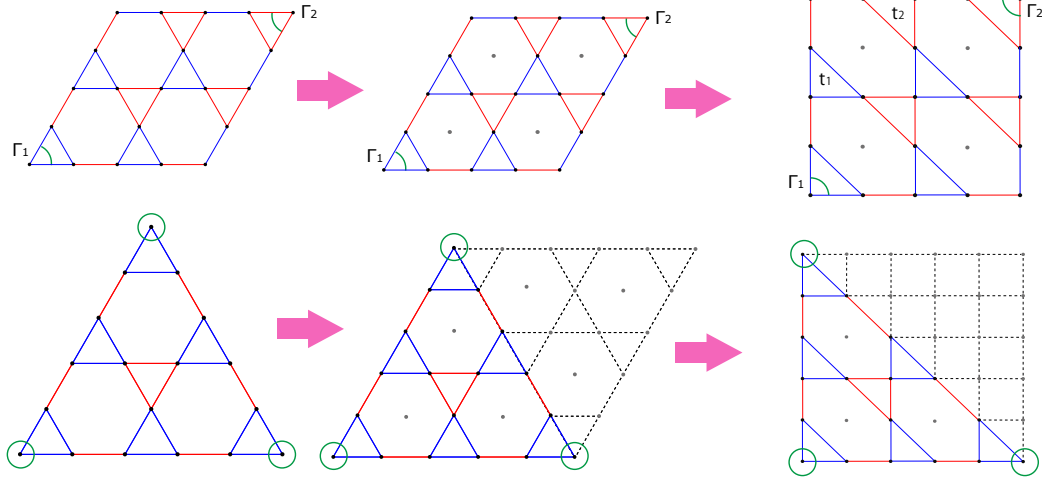


FIG. S1. The modification of the Kagome lattice to the square lattice for a rhombus geometry (upper) and a triangle geometry (lower). We add empty sites denoted by gray circles.

To make input data, we deform the lattice structure of the Kagome lattice to the square lattice (Fig. S1).

For a rhombus geometry, this is archived by adding redundant sites (denoted by gray circles) at the center of hexagonal plaquettes, and deform equilateral triangles into right triangles. Then we make input data from a single-particle wave function, in which probabilities on the added sites are always zero. In this way, the size of image data is increased from about  $3N$  to  $4N$ , where  $N$  is the number of unit cells. Although the computational cost is increased, we do not lose any information during this process.

Similarly, for a triangle geometry, we add the redundant sites at the center of the plaquettes. In addition, we also need to add a large downward triangle composed of the redundant sites to make the image in a square form. Then, in the resulting image, more than half of the sites are always empty. Nevertheless, we successfully detect the higher order topological phase (see the main text).

## B The Architecture of the Neural Network

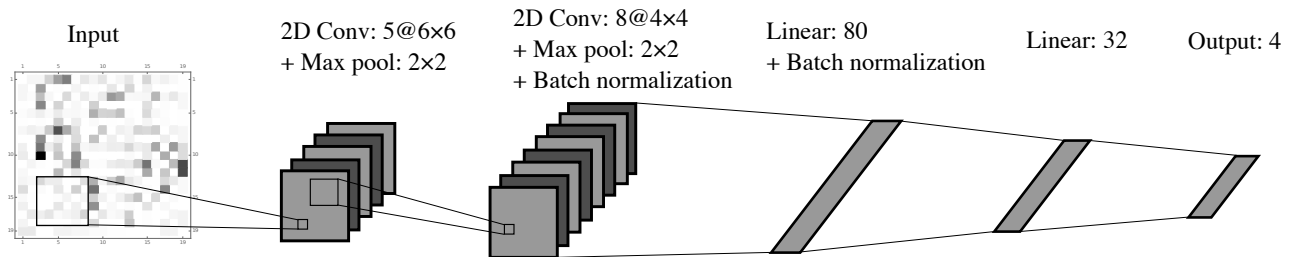


FIG. S2. The architecture of the neural network. The first two layers are two-dimensional convolutional layers and the rests are linear layers. We use some max pooling layers and batch normalization. The size of the layers are shown in the figure.

Here we show the architecture of the neural network in this study (Fig. S2). The size of input data is  $(2M - 1) \times (2M - 1)$  with  $M = 10$  for the rhombus geometry;  $2M \times 2M$  with  $M = 16$  for the triangle geometry. The correct labels are the one-hot representation of the phases. We use the convolutional neural network with five layers, in which the first two layers are two dimensional convolutional layers and the rests are linear layers. All the activation functions are the ReLU functions  $f(x) = \max(0, x)$ . After the first convolutional layer, we use the max pooling layer. After the second convolutional layer, we use the max pooling layer and the batch normalization. The max pooling layer takes the max value in  $2 \times 2$  sectors. Next layer is a linear layer. We use the batch normalization for it. Then the neural network outputs the probabilities of phases. The number of output is four for rhombus geometry and three for triangle geometry, corresponding the number of phases (see the main text).

### C Disordered phases

In this section, we elucidate the “unknown phase” appears above the trivial phase in a rhombus geometry. To this end, we first plot the phases with the highest probability over 98% in Fig.S3(a); the white region corresponds to the “unknown phase”. Clearly, the “mosaic-like” region in the original phase diagram now turns into white.

Next, we investigate a wave function of a single snapshot in a clean and dirty regimes, shown in Figs. S3(b) and S3(c), respectively. For the clean limit, the wave function corresponds to the flat band, and the real-space distribution of the wave function looks somewhat random. On the other hand, in the disordered system, one clearly sees the localized state, indicating that the stated is an Anderson-localized state. As explained in the main text, this is attributed to the fact that the degeneracy of the flat band is lifted by the small disorders. Even after taking the average over 30 samples, this pattern of localization is still distinguishable from other phases, resulting in the “unknown phase” in this region.

We also investigate a triangle geometry (Fig. S4). We see a small white region in Fig. S4(a); instead, we see a huge blue region (i.e., a metallic state). Looking at the wave functions, they look similar to the case of a rhombus geometry [Figs. S4(b) and S4(c)]. Thus we believe that the Anderson localization indeed occurs in this region, although the machine learning method does not successfully classify it.

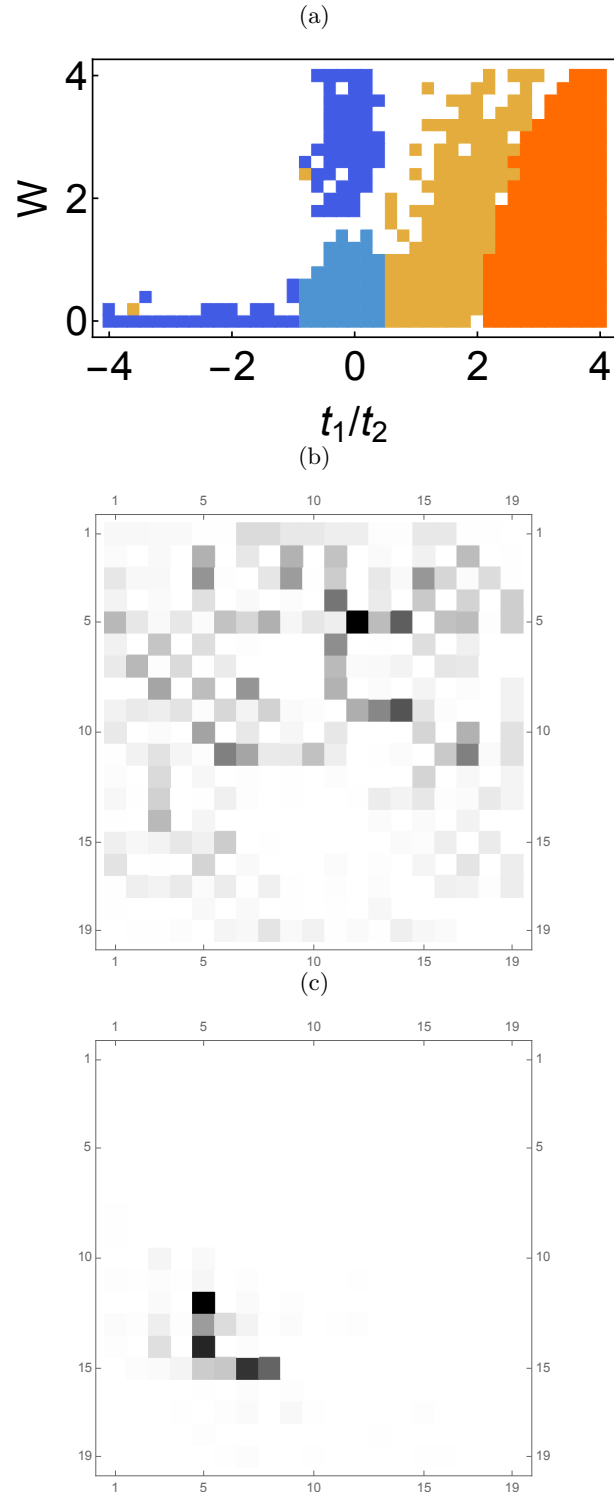


FIG. S3. (a) The phase diagram colored only if the probability is over 98%. The colors are corresponding with a phase that have the highest probability. In the white region, the probability less than 98% for any of four phases. (b) A highest occupied state for trivial phase ( $t_1/t_2 = -3, W = 0$ ). (c) A single snapshot for the disordered phase ( $t_1/t_2 = -3, W = 0.1$ ).



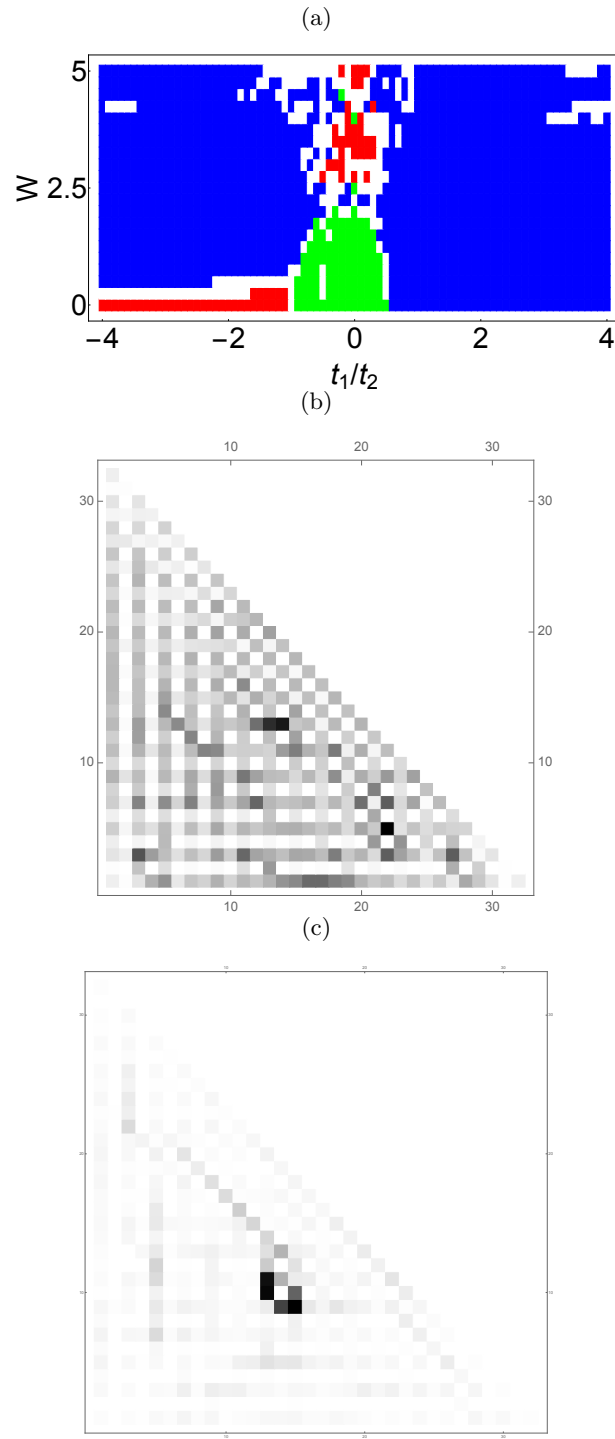


FIG. S4. The same plots as in Fig. 3 for a triangle geometry.