

Active learning algorithm for computational physics

Juan Yao¹,^{*} Yadong Wu,² Jahyun Koo,³ Binghai Yan,³ and Hui Zhai^{2,1,*}¹Center for Quantum Computing, Peng Cheng Laboratory, Shenzhen 518005, China²Institute for Advanced Study, Tsinghua University, Beijing 100084, China³Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel

(Received 16 May 2019; accepted 21 February 2020; published 10 March 2020)

In large-scale computations of physical problems, one often encounters the situation of having to determine a multidimensional function, which can be numerically costly when computing each point in this multidimensional space is already time-demanding. In the work, we propose that the active learning algorithm can speed up such calculations. The basic idea is to fit a multidimensional function by neural networks, and the key point is to make the query of labeled data more economical by using a strategy called “query by committee.” We present the general protocol of this fitting scheme, as well as the procedure of how to further compute physical observables with the fitted functions. We show that this method can work well with two examples, which are the quantum three-body problem in atomic physics and the anomalous Hall conductivity in condensed matter physics, respectively. In these examples, we show that one reaches an accuracy of a few percent error in computing physical observables, all the while using fewer than 10% of total data points compared with uniform sampling. With these two examples, we also visualize that by using the active learning algorithm, the required amount of data points are added mostly in the regime where the function varies most rapidly, which explains the mechanism for the efficiency of the algorithm. We expect broad applications of our method to various kinds of computational physical problems.

DOI: [10.1103/PhysRevResearch.2.013287](https://doi.org/10.1103/PhysRevResearch.2.013287)

I. BACKGROUND

Neural network (NN) based supervised learning methods have nowadays found broad applications in studying quantum physics in condensed matter materials and atomic, molecular and optical systems [1,2]. On the theoretical side, applications include finding orders and topological invariants in quantum phases [3–9], generating variational wave functions for quantum many-body states [10–14] and speeding up quantum Monte Carlo sampling [15,16]. On the experimental side, these methods can help optimizing experimental protocols [17,18] and analyzing experimental data [19–21]. Usually the supervised learning scheme requires a huge set of labeled data. However, in many physical applications, labeling data can be quite expensive in terms of computational resources. For instance, performing computation or experiments repeatedly can be time- and resources-demanding. Therefore, in many cases, labeled data are not abundant, which is a challenge that has prevented many applications.

The active learning is a scheme to solve this problem [22]. It starts from training a NN with a small initial data set, and then actively queries the labeled data based on the prediction of the NN and iteratively improves the performance of the NN

until the goal of the task is reached. With this approach, sampling the large parameter space can be made more efficiently, and the demand of labeled data is usually much less than normal supervised learning methods. Recently a few works have applied the active learning algorithm to determination of the interatomic potentials in quantum materials [23–25] and to optimal control in quantum experiments [26,27]. In these situations, labeled data have to be obtained either by *ab initio* calculation or by repeating experiments, which are both time consuming.

In this work, we focus on a class of general and common task in computational physics that is to numerically determine a multidimensional function, say, $\mathcal{F}(\alpha_1, \alpha_2, \dots, \alpha_n)$, where α_i are parameters. Supposing that we uniformly discretize each parameter into L points, there are then L^n data points in total that need to be calculated. In many cases, calculation of each point already takes quite some time, and thus the total computational cost can be massive. Nevertheless, for most functions, there are regions where the function varies smoothly and regions where the function varies rapidly. Ideally, the goal is to sample more points in the steep regions and fewer points in the smooth regions in order to efficiently obtain a good fitting in the entire parameter space. However, it seems to be a paradox because one does not know which regions the function varies more rapidly prior to computing the function.

Here we show that this goal can be achieved by using the active learning algorithm and the “query by committee” strategy [28] where data points are added iteratively. Below we will first introduce the general protocol and then demonstrate the algorithm in two concrete problems: the quantum three-body problem and the anomalous Hall conductivity problem.

^{*}hzhai@tsinghua.edu.cn

These are two very representative examples in atomic and condensed matter physics, respectively. Using these two examples, we will illustrate how the active learning and “query by committee” strategy can steer the addition of data points to steep regions of the fitting function. We will also discuss how to compute physical observables from the fitted functions.

II. GENERAL PROTOCOL

The main procedure is summarized in Fig. 1 and explained as follows. (1) We start with an initial data set with the number of data points $S_0 \ll L^n$. The values of \mathcal{F} at these points have been computed exactly. We use these data set to train a number of different NNs. (2) We ask all NNs to make predictions of \mathcal{F} on all L^n data points, and for each point, we compute the variance among the predictions made by different NNs. (3) We select S_t data points with the largest variances, again with $S_t \ll L^n$. We then query the accurate values of \mathcal{F} at these points by numerical calculation. (4) We add S_t new data points into the training set to train all NNs again, and then repeat from step 2. The steps 2–4 are repeated for m -epochs until the results from all NNs converge. (5) We use NN to calculate the value of \mathcal{F} at all L^n data points.

In this protocol, one needs only to query the values of \mathcal{F} at $S_0 + mS_t$ data points, which is always kept to be much smaller than L^n . The trade-off is that we need to train a number of different NNs and keep using all NNs to make predictions on all data points. However, it can save computational cost if training NN and making predication with NN is much less time-consuming than fully evaluating \mathcal{F} , which is often the case in many applications.

In this protocol, the key idea is the “query by committee” strategy [28]. Here the committee is made of different NNs, which contain different number of layers, different number of nodes at each layer and different activation functions. Thanks to the great expressive power of NN, we do not need to assume a specific form of the fitting functions. In regions where the function varies smoothly, different NNs can quickly reach a consensus and the variance will be small. In regions where

the function varies rapidly, in comparison, it is difficult for different NNs to converge on the same values and the variance will be large. Hence, we can use this variance to provide guidance on the sampling data point. In fact, as we will see in the examples below, the data points added in later epoches are all added in those regions where the function varies rapidly.

III. THREE-BODY PROBLEM

In the first example, we consider a quantum three-boson problem. These bosons interact with a two-body pairwise potential described by an s -wave scattering length a_s and a high-energy cutoff Λ . When a_s is positive, there exists a two-body bound dimer state, and it is important to compute the atom-dimer scattering length a_{ad} by solving the three-body problem. Here a key quantity is the scattering kernel $\mathcal{U}(\mathbf{k}, \mathbf{k}')$. Focusing on the s -wave scattering only, \mathcal{U} depends solely on the amplitude of momentum $k = |\mathbf{k}|$ and $k' = |\mathbf{k}'|$, and can be simplified as $\mathcal{U}(k, k')$. Once we know $\mathcal{U}(k, k')$, we can calculate a_{ad} through the Skorniakov-Ter-Martirosian (STM) equation [29]. The details of how to compute $\mathcal{U}(k, k')$ and how to obtain a_{ad} from $\mathcal{U}(k, k')$ are summarized in Appendix A.

In conventional methods, we uniformly discretize both k and k' into $L = 100$ points between zero and Λ , and we need to compute \mathcal{U} at all 10^4 data points. We then solve the STM equation with these $\mathcal{U}(k, k')$ to obtain a_{ad}^{exact} . We have checked that such a discretization can ensure the reaching of the convergence, which is referred to as the exact result and is shown by the dash line as reference in Fig. 3. Here in comparison we will follow the general procedure described above to fit $\mathcal{U}(k, k')$ using NNs. We will show that (i) only 300 data points are needed at most in order to obtain a reliable fitting, (ii) most of the queried data occupy the parameter region where the function $\mathcal{U}(k, k')$ is steep, and (iii) the trained NN is used to generate $\mathcal{U}(k, k')$ on all 10^4 data points for solving the STM equation, with typical errors of a few percent compared with the exact result.

To be more concrete, we start with an initial data set with uniformly sampled $S_0 = 100$ points. Here we design five different fully connected NNs, whose structures are schematically shown in Fig. 2. The input of all NNs are two numbers k and k' , and the output is \mathcal{U} . Each layer of a NN is characterized by the number of nodes N_α and an activation function f_α , and we specify each NN by the combination $(N_1, f_1; N_2, f_2; \dots)$.

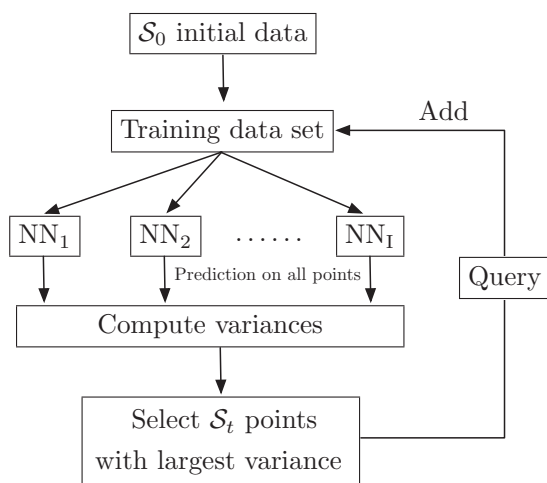


FIG. 1. Active learning protocol for fitting a multidimensional function. Here “NN” represents “neural network.”

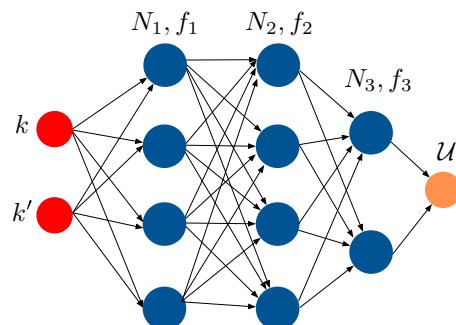


FIG. 2. Schematic of the fully connected neural network used in these two examples.

The five different NNs used in this work are

- 1 : (20, tanh; 20, tanh; 6, LS),
- 2 : (20, tanh; 20, LS; 6, tanh),
- 3 : (30, tanh; 20, tanh; 6, LS),
- 4 : (30, tanh; 20, LS; 4, tanh),
- 5 : (30, tanh; 20, tanh; 10, LS; 4, tanh),

where LS denotes the logistic sigmoid activation function with $f(a) = 1/(1 + e^{-a})$. It is important to keep these NNs different but their detailed structures are not important for final results. For each NN, the training results also depend on the initialization, which is particularly so when the number of data points is not enough. Therefore, for each NN, we also consider 20 different initializations.

At each iteration, we have 5×20 trained NNs in total to predict \mathcal{U} on all $\{k, k'\}$ points in the set \mathcal{M} . For each point $i \equiv \{k, k'\}$, we can compute the variance σ_i for all 100 predictions. We will select $\mathcal{S}_i = 10$ points with the largest variance to compute \mathcal{U} at these points, and add them into the training set. At each iteration, we can compute the mean variance $\sigma_{\text{mean}} = 1/L^2 \sum_{i \in \mathcal{M}} \sigma_i$. When the mean variance σ_{mean} is sufficiently small, we start to evaluate the physical quantity a_{ad} .

Now we come to another important point of our method. At the beginning, we have utilized the discrepancy between NNs to guide the query process more efficiently. When we start to compute observables, however, we need to properly average out these variances between different NNs to obtain a converged result. Hence, we compute a_{ad} as follows.

(1) Since there are 20 copies depending on different initializations, we first average the predictions over these 20 copies to obtain a mean value $\bar{\mathcal{U}}_t(k, k')$ for each point and for each NN. We then choose one of $\mathcal{U}_t^j(k, k')$ that is the closest to $\bar{\mathcal{U}}_t(k, k')$, and we use this $\mathcal{U}_t(k, k')$ (ignoring the upper index) as the representative of the t th NN.

(2) We solve the STM equation with $\mathcal{U}_t(k, k')$ and obtain a_{ad}^t . Among all five a_{ad}^t , we discard the largest and the smallest one, and perform an average over the rest of the three, which is taken as a_{ad} predicted by the active learning approach. The results are shown in Fig. 3 with blue circles. One can see that the results fluctuate around the exact value, and the fluctuation decreases as \mathcal{S} increases.

(3) To further suppress the fluctuation, we can further take a self-average of a_{ad} . That is to average over a_{ad} for five successive iterations. This result is denoted by \bar{a}_{ad} and shown in Fig. 3 with yellow solid dots. Indeed, one can see that the fluctuation is already suppressed strongly at $\mathcal{S} \sim 200$.

When the result does not change with training epoch, we take a self-average over the last five iterations to obtain \bar{a}_{ad} , and we compare this result with $a_{\text{ad}}^{\text{exact}}$. For instance, when $a_s \Lambda = 10$, the relative error $(\bar{a}_{\text{ad}} - a_{\text{ad}}^{\text{exact}})/a_{\text{ad}}^{\text{exact}}$ is about 1%.

We apply this method to scan different values of a_s . We stop the iteration at $\mathcal{S} = 300$, a number much smaller than the number of data points 10^4 required in a uniform sampling. Afterwards we take a self-average over the last five iterations to obtain \bar{a}_{ad} . The results are shown in Fig. 4. Our calculation can obtain the right location for the atom-dimer resonances, and our active learning method can well reproduce the Efimov

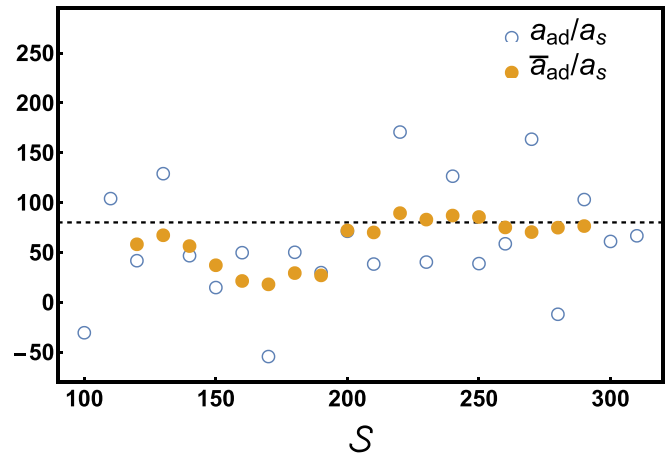


FIG. 3. The atom-dimer scattering length a_{ad} calculated with the active learning method. a_{ad} converges with the increasing number of queried data set \mathcal{S} . The blue empty circles are the results averaged over different NNs at each step, and the yellow solid dots are results averaged over the adjacent five steps. The dashed line denotes the exact results obtained with all 10^4 data points. Here we take the number of initial data set $\mathcal{S}_0 = 100$ and at each step $\mathcal{S}_i = 10$ data points are added. $a_s \Lambda = 10$.

scaling law. By comparing with the exact result, we define a relative error as $(\bar{a}_{\text{ad}} - a_{\text{ad}}^{\text{exact}})/a_{\text{ad}}^{\text{exact}} \equiv \epsilon\%$. The ϵ around the first resonance is shown in the inset of Fig. 4, which shows that the relative error does increase close to the Efimov resonance, but overall it is kept within a few thousandths.

To further reveal the mechanism of our method, we plot in Fig. 5(a) the function $\mathcal{U}(k, k')$ generated by the uniform sampling of all 10^4 points. All the data points added during iterations (with $100 < \mathcal{S} \leq 300$) are shown in Fig. 3(b). It is very clear that nearly all points are added in the regions around $k \sim 0$ and $k' \sim 0$, where the variation of $\mathcal{U}(k, k')$ is the steepest [see Figs. 5(a) and 5(b)]. In Figs. 5(c) and

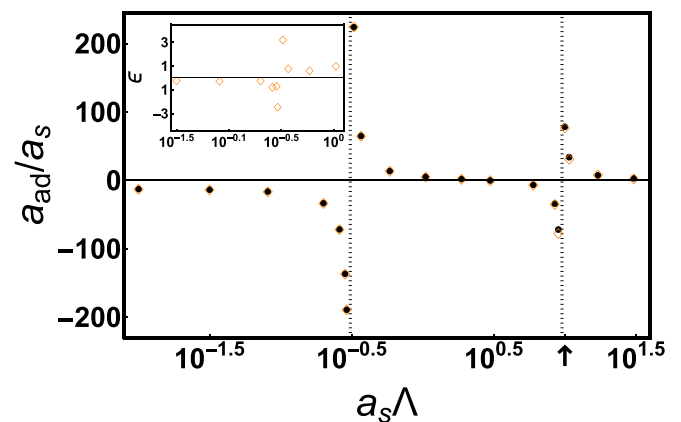


FIG. 4. a_{ad} as a function of $a_s \Lambda$. The divergence of a_{ad} is the atom-dimer resonance due to the Efimov effect. The black dots are calculated by evaluating \mathcal{U} with uniform sampling all data points, and the yellow diamonds are results from the active learning algorithm with self-averaging. Arrow mark the $a_s \Lambda$ where the training processes are demonstrated Fig. 3. The inset plots the relative error $\epsilon\%$ around the first atom-dimer resonant point.

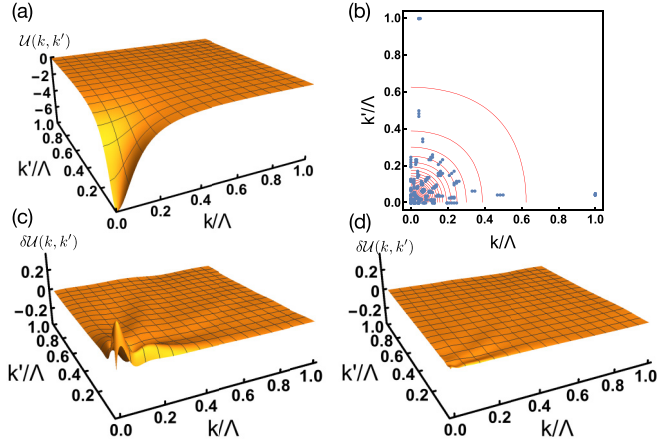


FIG. 5. (a) Profile of the two-dimensional function $\mathcal{U}(k, k')$ in the whole momentum space, plotted with all uniformly sampled 10^4 data points. (b) Visualization of the data set queried during the active learning iterations. The blue dots denote the queried data set with $100 < S \leq 300$. The red lines are the contour plots of (b). [(c) and (d)] Difference between function $\mathcal{U}(k, k')$ generated by the NNs and $\mathcal{U}(k, k')$ generated by uniform sampling. Here we have averaged over all five different NNs and twenty different initializations. The labeled data $S = 120$ for (c) and $S = 320$ for (d).

5(d), we compare the function $\mathcal{U}(k, k')$ generated by NNs and the function generated by uniform sampling. Here the NN results are averaged over five different NNs and over different initializations of different NNs. It shows that the fitting is perfect when the calculation of a_{ad} converges.

IV. ANOMALOUS HALL CONDUCTIVITY PROBLEM

In the second example, we consider the anomalous Hall conductivity of a magnetic Weyl semimetal Mn_3Ge [30], which has been extensively studied recently [31,32]. The anomalous Hall conductivity (σ_{Hall}) is an intrinsic quantity induced by the Berry curvature of the band structure. We have computed $\sigma_{\text{H}} = \sum_{\mathbf{k}} \sigma(k_x, k_y, k_z)$ based on the *ab initio* band structure calculations. The model of this material and the details of computing the Berry curvature are shown in Appendix B. For each k_x and k_z point, we can obtain a value $\sigma(k_x, k_z)$ by performing an integration over k_y . Here we discretize both k_x and k_z into 100 points, and thus one needs to compute 10^4 data points in total without the active learning method. The total Hall conductivity is obtained by summing over k_x and k_z as $\sigma_{\text{H}} = \sum_{k_x, k_z} \sigma(k_x, k_z)$.

As one can see from Fig. 6(a), the function $\sigma(k_x, k_z)$ is much more singular than $\mathcal{U}(k, k')$ in the previous case. There are four broad peaks and four narrow peaks located around $(0.5 \pm 0.1, 0.5 \pm 0.4)$ and $(0.5 \pm 0.4, 0.5 \pm 0.15)$. We follow the same active learning procedure discussed above. In this case, we take five different initializations for each NN. Compared with the first example, computing physical observable is easier because we need only to average over all five different initializations for each NN. Similarly, to suppress the fluctuation at each iteration, we also average over three different predictions of σ_{H} by discarding the largest and the smallest results. We also take a self-average over five

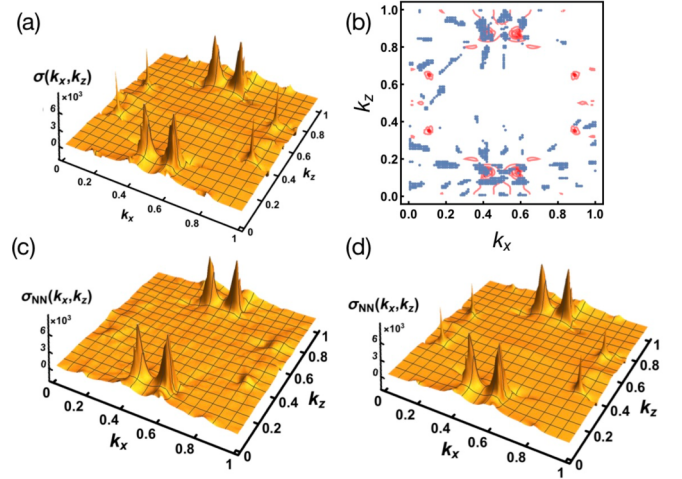


FIG. 6. (a) Profile of the two-dimensional function $\sigma(k_x, k_z)$ in the whole momentum space, plotted with all uniformly sampled 10^4 data points. (b) Visualization of the data set queried during the active learning iterations. The blue dots denote the queried data set with $100 < S \leq 900$. The red lines are the contour-plotting of (b). σ_{H} is in unit of S/cm. [(c) and (d)] The function $\sigma(k_x, k_z)$ generated by the NNs, averaged over all five different NNs and five different initializations for each NN. The labeled data $S = 1200$ for (c) and $S = 4000$ for (d).

successive iterations. As we show in Fig. 7, the prediction of the active learning method approaches the exact value when $S_{\text{max}} > 700$, which is less than 10% of the total number of uniformly sampled data, and the relative error is about 0.8%. As we shown in Fig. 6(b), the “query by committee” strategy distributes most of the queried data in the areas of four broad peaks. In Figs. 6(c) and 6(d), we show the function $\sigma(k_x, k_z)$ generated by NNs. For $S = 1200$ shown in Fig. 6(c), the

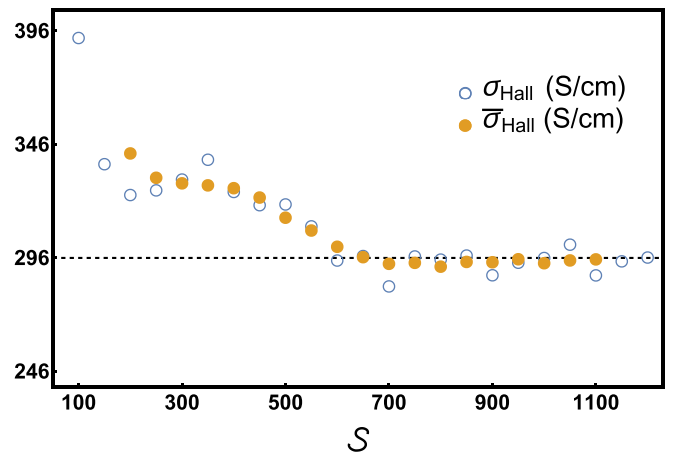


FIG. 7. The Hall conductivity σ_{H} calculated with the active learning method. σ_{H} converges with the increasing number of queried data set S . The blue empty circles are the results averaged over different NNs at each step, and the yellow solid dots are results averaged over the adjacent five steps. The dashed line denotes the exact results obtained with uniformly sampled all 10^4 data points. Here we take the number of initial data set $S_0 = 100$ and at each step $S_i = 50$ data points are added.

fitting has already captured the four main peaks, and since the contribution to the Hall conductance mainly comes from the four main peaks, the results of σ_H already converges very well. Nevertheless, when we continue to add labeled data until $S = 4000$, as shown in Fig. 6(d), these extra data are mainly added in the four narrow peaks such that these four small peaks can also be generated very well.

V. CONCLUSION AND OUTLOOK

In summary, we have developed a neural network based machine learning method to determine multidimensional functions efficiently. The method combines the great expressiveness of NN and the advantage of the active learning scheme to minimally reduce the demand for labeled data. There are two key ingredients to this method. First, we make use of the variances between NNs to steer the sampling of the queried data to those regions where the function varies rapidly, which makes the calculation more efficient. Second, we need to properly average out these variances when calculating physical observables using the fitted results. With two examples, we show our method can work remarkably well. Compared to uniform sampling, our method can achieve an accuracy of less than 1% error with only fewer than 10% of total data points, even in situations where the function has multiple sharp peaks. We believe our method can find broad applications in many areas of computational physics.

ACKNOWLEDGMENTS

We thank colleagues in Microsoft Research Asia for discussing active learning. This work is supported Beijing Outstanding Young Scientist Program (H.Z.), MOST under Grant No. 2016YFA0301600 (H.Z.) and NSFC Grant No. 11734010 (H.Z.), the Willner Family Leadership Institute for the Weizmann Institute of Science (B.Y.), the Benozio Endowment Fund for the Advancement of Science (B.Y.), and Ruth and Herman Albert Scholars Program for New Scientists (B.Y.).

APPENDIX A: THREE-BOSON PROBLEM

The Hamiltonian of the three-boson system is given by

$$\hat{H} = \sum_{i=1}^3 -\frac{\nabla_i^2}{2m} + \sum_{i<j=1}^3 V(|\mathbf{r}_i - \mathbf{r}_j|), \quad (\text{A1})$$

where \mathbf{r}_i ($i = 1, 2, 3$) is the spatial coordinate of the i th particle, and $V(r)$ is an isotropic short-range potential. Normally we focus on the s -wave scattering for ultracold low-energy atoms, and $V(r)$ can be described by the s -wave scattering length a_s and a high-momentum cutoff Λ . Most features of this quantum three-body problem only depend on the dimensionless parameter $a_s\Lambda$ [29]. We consider the case that a_s is large and positive, where $V(r)$ supports a low-energy bound state called a dimer. A key process to this three-body problem is the atom-dimer scattering illustrated in Fig. 8(a), and the quantity to describe this process is a scattering kernel, which is usually denoted by $\mathcal{U}(\mathbf{k}, \mathbf{k}')$. The scattering kernel can be computed diagrammatically with the Feynman diagram shown in Fig. 8(b) [29]. When focusing on the s -wave

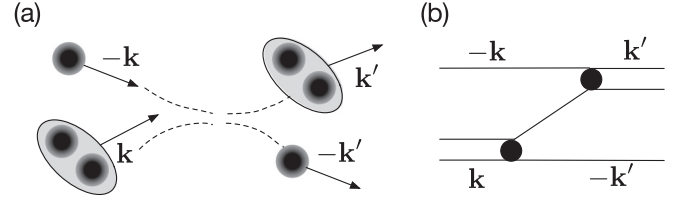


FIG. 8. (a) Schematic of the atom-dimer scattering process. (b) Diagrammatic description of the scattering kernel $\mathcal{U}(\mathbf{k}, \mathbf{k}')$. The single line stands for the free single-particle Green's function $G_0^A(\mathbf{k}, E)$. The double line is the free dimer Green's function $D_s(\mathbf{k}, E)$. The black dot is the s -wave interacting vertex.

scattering only, \mathcal{U} depends solely on the amplitude of $k = |\mathbf{k}|$ and $k' = |\mathbf{k}'|$, and we can simplify \mathcal{U} as $\mathcal{U}(k, k')$.

Knowing $\mathcal{U}(k, k')$, one can calculate the few-body quantity, such as atom-dimer scattering length a_{ad} , through the Skorniakov-Ter-Martirosian (STM) equation, which can be written as

$$\int \frac{dk' k'^2}{2\pi^2} \mathcal{U}(k, k') D_s(k', E) \mathcal{A}(k') - \mathcal{A}(k) = \mathcal{U}(k, 0), \quad (\text{A2})$$

where

$$D_s^{-1}(k', E) = \frac{m}{2\pi a_s} - \frac{m}{2\pi} \sqrt{\frac{3k'^2}{4} - mE}, \quad (\text{A3})$$

is the full propagator for s -wave dimer. Solving Eq. (A2) with E fixed at the dimer energy $-\frac{1}{ma_s^2}$, we can obtain atom-dimer scattering amplitude $\mathcal{A}(k)$. Then the atom-dimer scattering length a_{ad} is given by [29]

$$a_{ad} = -\frac{8}{3ma_s} \mathcal{A}(k=0). \quad (\text{A4})$$

In practice, supposing that we discretize both k and k' into $L = 100$ points between zero and Λ , Eq. (A2) becomes a matrix equation after the discretization, which can be solved by inverting the matrix. Let \mathcal{M} denotes the set of total data points, and the number of data points in \mathcal{M} is 10^4 . With uniform sampling, we need to compute \mathcal{U} for all 10^4 points, with which we solve Eq. (A2) to obtain a_{ad}^{exact} . This is referred as the exact results in the main text.

APPENDIX B: ANOMALOUS HALL CONDUCTIVITY

The band structure of Mn_3Ge calculated with the density-functional theory from Vienna *ab-initio* simulation package [33] in the framework of the generalized-gradient approximation. The calculated plane wave basis result were projected to atomic-orbital-like Wannier functions [34] to get the tight-binding parameters (t_{ij}). Based on the tight-binding Hamiltonian,

$$\hat{H} = \sum_{i,j} t_{ij} c_i^\dagger c_j, \quad (\text{B1})$$

we calculated the Berry curvature and the anomalous Hall conductivity in the clean limit. For the sufficient data set, uniform K point grid as $100 \times 100 \times 100$ employed to produce 10^6 numbers of data points. We have evaluated the AHC (σ_{zx}) and Berry curvature (Ω_{zx}) by the Kubo-formula approach in

the linear response scheme [35],

$$\sigma_{xz}(\mu) = -\frac{e^2}{\hbar} \int_{BZ} \frac{d\mathbf{k}}{(2\pi)^3} \sum_{\epsilon_n < \mu} \Omega_{xz}^n(\mathbf{k}), \quad (\text{B2})$$

$$\Omega_{xz}^n(k) = i \sum_{m \neq n} \frac{\langle n | \hat{v}_x | m \rangle \langle n | \hat{v}_z | m \rangle - (x \leftrightarrow z)}{(\epsilon_n(\mathbf{k}) - \epsilon_m(\mathbf{k}))^2}. \quad (\text{B3})$$

Here, n and m are the band index, ϵ_n is the eigenvalue of the $|n\rangle$ eigenstate of H_k (Fourier transform of \hat{H}), and $\hat{v}_i = dH_k/(\hbar dk_i)$ ($i = x, z$) is the velocity operator, μ is the Fermi level of the system. The calculated anomalous Hall conductivity value is 296 S/cm, where S denoting unit siemens.

-
- [1] P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, *Phys. Rep.* **810**, 1 (2019).
- [2] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [3] D.-L. Deng, X. Li, and S. Das Sarma, *Phys. Rev. B* **96**, 195145 (2017).
- [4] Y. Zhang and E.-A. Kim, *Phys. Rev. Lett.* **118**, 216401 (2017).
- [5] J. Carrasquilla and R. G. Melko, *Nat. Phys.* **13**, 431 (2017).
- [6] E. P. L. Van Nieuwenburg, Y. H. Liu, and S. D. Huber, *Nat. Phys.* **13**, 435 (2017).
- [7] P. Zhang, H. Shen, and H. Zhai, *Phys. Rev. Lett.* **120**, 066401 (2018).
- [8] Y.-H. Liu and E. P. L. van Nieuwenburg, *Phys. Rev. Lett.* **120**, 176401 (2018).
- [9] X.-Y. Dong, F. Pollmann, and X.-F. Zhang, *Phys. Rev. B* **99**, 121104(R) (2019).
- [10] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [11] X. Gao and L. M. Duan, *Nat. Commun.* **8**, 662 (2017).
- [12] Z. Cai and J. Liu, *Phys. Rev. B* **97**, 035116 (2018).
- [13] H. Saito, *J. Phys. Soc. Jpn.* **87**, 074002 (2018).
- [14] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Nat. Phys.* **14**, 447 (2018).
- [15] H. Shen, J. Liu, and L. Fu, *Phys. Rev. B* **97**, 205140 (2018).
- [16] T. Song and H. Lee, *Phys. Rev. B* **100**, 045153 (2019).
- [17] G. Torlai, B. Timar, E. P. L. van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, *Phys. Rev. Lett.* **123**, 230504 (2019).
- [18] A. Macarone-Palmier, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. Biamonte, and S. Kulik, *arXiv:1904.05902*.
- [19] Y. Zhang, A. Mesaros, K. Fujita, S. D. Edkins, M. H. Hamidian, K. Ch'ng, H. Eisaki, S. Uchida, J. C. S. Davis, E. Khatami, and E.-A. Kim, *Nature (London)* **570**, 484 (2019).
- [20] B. S. Rem, N. Käming, M. Tarnowski, L. Asteria, N. Fläschner, C. Becker, K. Sengstock, and C. Weitenberg, *Nat. Phys.* **15**, 917 (2019).
- [21] A. Bohrdt, C. S. Chiu, G. Ji, M. Xu, D. Greif, M. Greiner, E. Demler, F. Grusdt, and M. Knap, *Nat. Phys.* **15**, 921 (2019).
- [22] B. Settles, *Active learning: Synthesis Lectures on Artificial Intelligence and Machine Learning* (Morgan & Claypool, 2012).
- [23] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and W. E, *Phys. Rev. Mater.* **3**, 023804 (2019).
- [24] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, *J. Chem. Phys.* **148**, 241733 (2018).
- [25] K. Gubaev, E. V. Podryabinkin, G. L. W. Hart, and A. V. Shapeev, *Comput. Mater. Sci.* **156**, 148 (2019).
- [26] P. B. Wigley, P. J. Everitt, A. Van Den Hengel, J. W. Bastian, M. A. Sooriyabandara, G. D. McDonald, K. S. Hardman, C. D. Quinlivan, P. Manju, C. C. N. Kuhn, I. R. Petersen, A. N. Luiten, J. J. Hope, N. P. Robins, and M. R. Hush, *Sci. Rep.* **6**, 25890 (2016).
- [27] B. M. Henson, D. K. Shin, K. F. Thomas, J. A. Ross, M. R. Hush, S. S. Hodgman, and A. G. Truscott, *Proc. Natl. Acad. Sci. USA* **115**, 13216 (2018).
- [28] H. S. Seung, M. Pper, and H. Sompolinsky, in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* (Association for Computing Machinery, New York, NY, 1992), p. 287.
- [29] E. Braaten and H.-W. Hammer, *Phys. Rep.* **428**, 259 (2006).
- [30] H. Yang, Y. Sun, Y. Zhang, W.-J. Shi, S. S. P. Parkin, and B. Yan, *New J. Phys.* **19**, 015008 (2017).
- [31] S. Nakatsuji, N. Kiyohara, and T. Higo, *Nature (London)* **527**, 212 (2015).
- [32] A. K. Nayak, J. E. Fischer, Y. Sun, B. Yan, J. Karel, A. C. Komarek, C. Shekhar, N. Kumar, W. Schnelle, J. Kübler, C. Felser, and S. S. P. Parkin, *Sci. Adv.* **2**, e1501870 (2016).
- [33] G. Kresse and J. Furthmüller, *Phys. Rev. B* **54**, 11169 (1996).
- [34] A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, *Comput. Phys. Commun.* **178**, 685 (2008).
- [35] D. Xiao, M.-C. Chang, and Q. Niu, *Rev. Mod. Phys.* **82**, 1959 (2010).