

# Machine Learning Phases of Strongly Correlated Fermions

Kelvin Ch'ng,<sup>1</sup> Juan Carrasquilla,<sup>2</sup> Roger G. Melko,<sup>2,3</sup> and Ehsan Khatami<sup>1</sup>

<sup>1</sup>*Department of Physics and Astronomy, San José State University, San José, California 95192, USA*

<sup>2</sup>*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*

<sup>3</sup>*Department of Physics and Astronomy, University of Waterloo, Ontario N2L 3G1, Canada*

(Received 14 June 2017; published 30 August 2017)

Machine learning offers an unprecedented perspective for the problem of classifying phases in condensed matter physics. We employ neural-network machine learning techniques to distinguish finite-temperature phases of the strongly correlated fermions on cubic lattices. We show that a three-dimensional convolutional network trained on auxiliary field configurations produced by quantum Monte Carlo simulations of the Hubbard model can correctly predict the magnetic phase diagram of the model at the average density of one (half filling). We then use the network, trained at half filling, to explore the trend in the transition temperature as the system is doped away from half filling. This transfer learning approach predicts that the instability to the magnetic phase extends to at least 5% doping in this region. Our results pave the way for other machine learning applications in correlated quantum many-body systems.

DOI: 10.1103/PhysRevX.7.031038

Subject Areas: Computational Physics,  
Condensed Matter Physics,  
Strongly Correlated Materials

## I. INTRODUCTION

The various modern architectures of neural networks consisting of multiple layers and neuron types (see Fig. 1 for an example) can be *trained* to classify, with a high degree of accuracy, intricate sets of labeled data [1]. The data, e.g., a series of handwritten digits, are fed to the network input layer, and the outcome, read at the output layer, are neuron activations corresponding to the different digits. Common to most algorithms involving neural networks is the training procedure, which is an optimization problem where the free parameters associated with connections between neurons in adjacent layers and their biases (additive constants) are slowly adjusted until a high classification accuracy is attained. Embodied in the study of quantum and classical statistical mechanics are the many-body states, which can be understood as immense data sets associated with the equilibrium state of the system, and over which machine learning techniques can be naturally applied. Early applications of machine learning ideas in condensed matter physics focused on their connection to renormalization group methods [2], obtaining the Green's function of the Anderson impurity model [3], categorizing real materials [4–7], or learning ground states and thermodynamics of many-body systems [8,9]. Tensor-network

representations of quantum states have also been proposed recently as a powerful tool for supervised learning [10].

Recently, neural-network machine learning algorithms have been successfully adopted to distinguish phases of

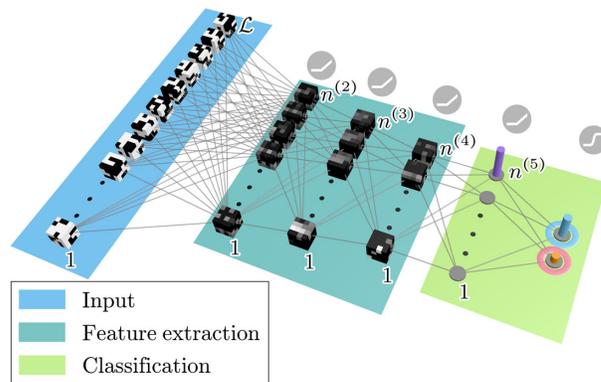


FIG. 1. Architecture of the 3D convolutional neural network used to obtain  $T_N$  for the 3D Hubbard model. For input, we use the auxiliary field configurations in a four-dimensional grid, three spatial dimensions of size 4 (total of  $N = 4^3$  sites), and one imaginary time dimension of size  $\mathcal{L} = 200$ . Numbers of volumetric feature maps in the hidden-feature extraction layers are  $n^{(2)} = 32$ ,  $n^{(3)} = 16$ , and  $n^{(4)} = 8$ . Here,  $n^{(5)} = 8$  in the fully connected layer. During training, dropout regularization with a rate of 0.5 was used to mitigate overtraining. To classify the input system as ordered or unordered, each of the eight fully connected neurons is connected to each of the two readout neurons using the softmax function as a neural activation function. The output neuron with the highest probability represents the activated neuron.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

matter in classical Ising-type models, effectively locating critical temperatures at which transitions between phases take place [11,12]. Two of the authors found that by using a simple network consisting of only one hidden layer, one can predict the transition temperature with up to 99% accuracy for two-dimensional (2D) Ising models, solely based on spin configurations generated by Monte Carlo simulations and in the absence of any information about the underlying lattice or the order parameter [11].

The extension of the technique to quantum mechanical systems is less straightforward, as the quantum Monte Carlo simulations of interacting particles involve an additional dimension associated with imaginary time in the path integral formalism at finite temperatures, or the projection parameter for ground-state calculations; quantum fluctuations can distort the easily recognizable picture of spin configurations in the ordered phase of the classical system and therefore significantly affect the training process of the neural network. Here, using quantum Monte Carlo simulations of the Hubbard model of strongly correlated fermions on cubic lattices and convolutional neural networks (CNNs), we show that one can successfully classify finite-temperature phases of quantum systems and estimate transition temperatures with a reasonable degree of accuracy on relatively small lattice sizes.

## II. MODEL

The Fermi-Hubbard Hamiltonian [13,14] in the particle-hole invariant form is expressed as

$$H = -t \sum_{\langle ij \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i \left( n_{i\uparrow} - \frac{1}{2} \right) \left( n_{i\downarrow} - \frac{1}{2} \right) - \mu \sum_{i\sigma} n_{i\sigma}, \quad (1)$$

where  $c_{i\sigma}$  ( $c_{i\sigma}^\dagger$ ) annihilates (creates) a fermion with spin  $\sigma$  on site  $i$ ,  $n_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$  is the number operator,  $U$  is the onsite Coulomb interaction,  $\langle \cdot \rangle$  denotes nearest neighbors,  $t$  is the corresponding hopping integral, and  $\mu$  is the chemical potential. Here,  $\mu = 0$  corresponds to the half-filled model (average density of one fermion per site,  $n = 1$ ). We set  $t = 1$  as the unit of energy and consider the model on three-dimensional (3D) cubic lattices.

The 3D model at half filling realizes a finite-temperature transition to the antiferromagnetic Néel phase for any  $U > 0$ , analogous to the magnetic ordering in the 2D classical Ising model. The transition temperature  $T_N$ , which is relatively well known from the analysis of the staggered spin structure factor, or the staggered susceptibility [15–21], is a non-monotonic function of the interaction strength; it increases rapidly with increasing  $U$  in the weak-coupling regime ( $U \lesssim 8$ ), a result that can be captured using the random phase approximation [15], and decreases at large  $U$ . In the strong-coupling regime ( $U \gtrsim 12$ ), the half-filled model can be effectively described by the antiferromagnetic (AFM)

Heisenberg model, whose exchange constant, and hence, Néel temperature, is proportional to  $1/U$  [22].

## III. METHOD

Our goal here is to train a CNN to identify finite-temperature phase boundaries of the Hubbard model. We utilize the determinantal quantum Monte Carlo (DQMC) [23], which reduces the numerical evaluation of the observables of the Fermi-Hubbard model to a stochastic averaging over a set of discrete auxiliary fields extending in space and along an imaginary time dimension. The spin correlations of the model can be written directly in terms of the correlations in our particularly chosen auxiliary field (see Appendix A), rendering it an obvious choice to be used in the identification of magnetic phases through machine learning, although a previous attempt by including two of the authors has not been successful [24]. The training is performed using the field configurations generated during DQMC simulations in a range of temperatures around one or two critical points. The objective is to use the trained network to map out the entire phase boundary associated with the same critical phenomenon by varying the parameters driving the transition and generating *test* data sets of the field configurations. In this work, we focus on the magnetic properties of the Hubbard model.

We use a 3D CNN, originally developed for human action recognition in videos [25], implemented in Tensorflow [26]. Convolutions are designed to return information about spatial dimension and locality to the simpler idea of a fully connected feed-forward neural network. In our case, the three spatial dimensions of the cubic lattice are treated with the convolution, while slices in the fourth imaginary time axis are used as different filter channels [1]. The network architecture for  $N = 4^3$  is shown in Fig. 1. We use three or four hidden layers, depending on the spatial size of the system, for feature extraction, followed by a fully connected layer before the output layer. The optimal number of neurons in each layer (resulting in the largest accuracy) for  $N = 8^3$  is found using a Monte Carlo optimization procedure (see Appendix B).

## IV. RESULTS

To benchmark our results and validate our approach, we start with the 3D Hubbard model at half filling and explore the accuracy with which we can predict the Néel phase boundary in the temperature-interaction space. We train the network to distinguish (by activating the corresponding output neuron) spin configurations belonging to the ordered phase ( $T < T_N$ ) from those of the unordered high-temperature phase (see Appendix C). The approximately 80 000 labeled configurations at various temperatures around  $T_N$  are generated through DQMC simulations for two interaction strengths,  $U = 5$  and 16, one in the weak-coupling and one in the strong-coupling regime, and shuffled before they are used in the training. The trained network is then

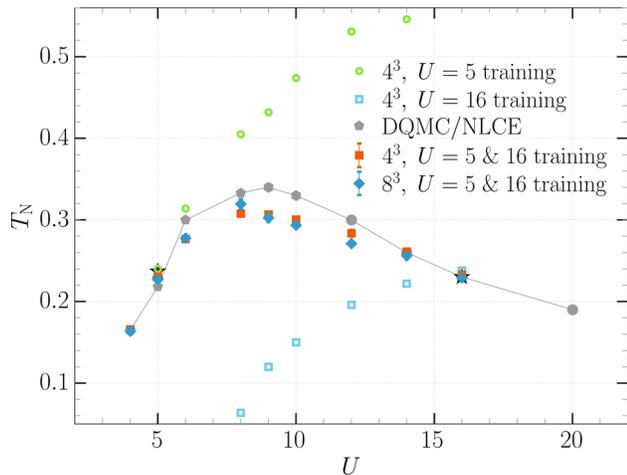


FIG. 2. Prediction of the Néel transition temperature by the neural network. Using the auxiliary spin configurations, the network is trained separately at  $U = 5$  and  $U = 16$  for  $N = 4^3$ , and simultaneously at  $U = 5$  and  $16$  for  $N = 4^3$  and  $N = 8^3$ . The error bars are the standard error of the mean of six different classifications using CNNs that were trained starting from different random weights and biases. The critical temperatures used for the training of the network with  $N = 4^3$  are shown as stars (see text). Grey filled symbols are the estimates for  $T_N$  in the thermodynamic limit from DQMC and NLCE simulations. Grey pentagons, hexagons, and circles for weak-, intermediate-, and strong-coupling regimes are taken from Refs. [18,19,21], respectively. The solid line is a guide to the eye.

used to classify other configurations as the temperature is varied across the estimated critical values for other values of  $U$  between 4 and 16.

The results for two system sizes  $N = 4^3$  and  $N = 8^3$  are shown in Fig. 2 as full squares and diamonds. Figure 2 also summarizes  $T_N$  in the thermodynamic limit from recent unbiased studies [18,19,21]. One can see that, despite notable disagreements at intermediate values of  $U$ , remarkably, the network can predict the nontrivial shape of the phase boundary with a reasonable degree of accuracy even with a system size of  $N = 4^3$ . The results for  $N = 8^3$  show the same trend. However, we find that in the strong-coupling regime, the latter are smaller than  $T_N$  obtained with  $N = 4^3$ . This counterintuitive behavior is likely rooted in the lack of precise knowledge of the critical temperature at  $U = 16$ , and the sensitivity of CNNs' predictions to our choice of  $T_N$  during training as explained below. For each system size, we train six different CNNs with the same architecture but with different initial random weights and biases (see Appendix C for training details), and we show the average  $T_N$  over different CNNs to make sure results are not biased towards any particular training. The error bars are slightly larger for  $N = 8^3$  as the neural network contains a significantly larger number of parameters in this case.

We note that  $T_N$  is not well defined for finite clusters, which can lead to uncertainties in the labeling of the

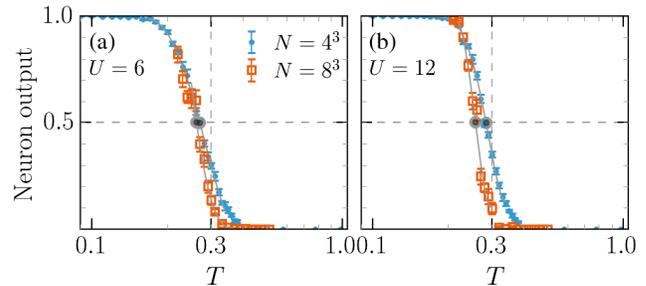


FIG. 3. Average output of the neuron that is trained to be activated (return 1) in the ordered phase as a function of temperature at half filling. The network is maximally confused (average is 0.5) at the transition temperatures. We use 90 000 configurations for  $N = 4^3$  and 10 000 configurations for  $N = 8^3$ . The grey solid lines are fits to data in the ranges shown. The vertical dashed lines indicate the estimates for  $T_N$  in the thermodynamic limit (about 0.30 for both  $U$  values) from other studies (see caption of Fig. 2).

configurations in the training process. However, the *exact* value of  $T_N$  is, in principle, not required for the training. One can omit configurations from the training data for an arbitrary temperature range in the proximity of the estimated critical value, e.g., the temperature at which the correlations reach the linear size of the cluster, at the expense of losing some predicting accuracy by the network after the training. Here, we take advantage of the fact that finite-size errors are small in the strong-coupling regime and use  $T_N = 0.23$  for  $U = 16$ , obtained for the thermodynamic limit [21], for both cluster sizes. For  $U = 5$ , however, we use the approximate critical temperatures  $T_N = 0.24$  and  $0.25$  for  $N = 4^3$  and  $8^3$ , respectively, from the analysis of the Binder ratio for the order parameter [19].

The predicted critical temperatures for other values of  $U$  during the classification are taken as temperatures at which the network is maximally “confused,” i.e., when the average output crosses 0.5. Figure 3 shows the average output of the neuron, in the classification, that is trained to be activated, i.e., to return 1, in the ordered. Results are shown for the two system sizes and for  $U = 6$  and  $12$ . We perform the classifications on 90 000 (10 000) configurations per temperature for  $N = 4^3$  ( $N = 8^3$ ). The data become noisier with a smaller number of configurations for the larger system and by increasing  $U$ , and so we use fits to a third-degree polynomial using data near the 0.5 crossing to better estimate  $T_N$ .

The simultaneous use of configurations for  $U = 5$  and  $16$  is crucial in obtaining the results in Fig. 2. We find that a network that is trained to distinguish phases only in the weak-coupling regime, e.g., for  $U = 5$ , correctly predicts the trend in  $T_N$  vs  $U$  for systems in the weak-coupling regime but grossly overestimates it for systems in the intermediate- and strong-coupling regimes. Similarly, a network that is trained with configurations only for  $U = 16$  underestimates  $T_N$  for any  $U < 16$  (see the open symbols in Fig. 2). Physically, this can be understood as

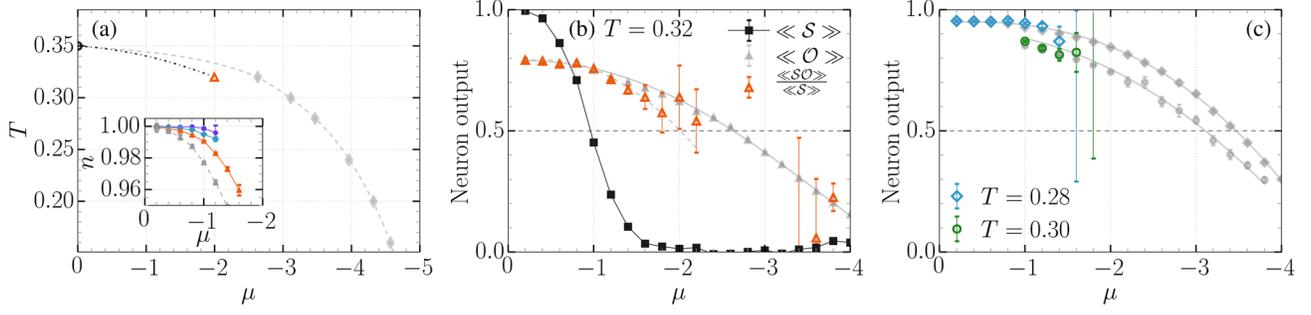


FIG. 4. Phase transitions away from half filling. (a) Magnetic phase diagram of the 3D Hubbard model away from half filling for  $U = 9$ . Empty red (filled grey) symbol(s) are estimates for the Néel temperatures, taking (not taking) the minus sign problem into account, obtained from a neural network that is trained to identify the phase at half filling. Lines are guides to the eye. Inset: Equations of state (density vs  $\mu$ ) for  $U = 9$  at  $T = 0.24$ ,  $0.28$ , and  $0.32$  (solid lines from top to bottom, respectively). The grey dashed line is calculated without taking the sign problem into account at  $T = 0.32$ . (b) Average neuron output  $\mathcal{O}$  calculated taking (without taking) the sign into account at  $T = 0.32$  ( $\langle\langle SO \rangle\rangle/\langle\langle S \rangle\rangle$  and  $\langle\langle O \rangle\rangle$ , respectively). See Appendix A for details. The grey dotted line is a fit to the data near the 0.5 crossing point. (c) Same as in panel (b) but for  $T = 0.30$  and  $0.28$ .

follows: The transition to the AFM ordered phase is less sharp in the weak-coupling regime, with substantial double occupancy still present in the system above the transition. As a result, the network that is trained only in this regime tracks the onset of Mott physics, which moves to higher temperatures as the interaction strength increases, not the Néel transition. On the other hand, the network that is trained only in the strong-coupling regime tracks the onset of the region with significantly reduced double occupancy and stronger AFM correlations, which takes place at lower temperatures for smaller values of  $U$ . The competition between these two scenarios when training with both  $U = 5$  and  $U = 16$  data is encoded in our CNN and is key in obtaining reasonable  $T_N$  for other  $U$ .

Does the magnetically ordered phase survive if we dope the system away from half filling, and if so, what is the dependence of  $T_N$  on doping? We try to answer these questions for  $U = 9$ , where the transition temperature is highest at half filling. We train the network using spin configurations obtained for  $U = 9$  at half filling and then classify other configurations generated at fixed temperatures less than the half-filled value of  $T_N$ , but across the chemical potential axis. As soon as  $\mu$  deviates from zero (the value corresponding to half filling), the weight of configurations in the DQMC can become negative, leading to the so-called “sign problem” [27,28]. We treat the neuron output as a (binary) physical observable, which, like spin correlations, can be written as a nonlinear function of the auxiliary spins (see Appendix A). Therefore, the expectation value is calculated in the conventional way by including the sign in the averaging and dividing by the average sign, typical for DQMC (see Appendix A). This procedure is valid as long as the average sign does not vanish. For completeness, in the following, we also show results in cases where we have ignored the sign problem and performed neuron output averages simply by using the absolute value of weights of the configurations.

The results are summarized in Fig. 4(a), where we show the location of the critical  $\mu$ , the onset of dominant AFM correlations, at  $T = 0.32$  (red triangle). Similarly to  $T_N$  at half filling, the critical  $\mu$  is estimated as the chemical potential where the expectation value of the neuron output crosses 0.5. To note the effect of the negative sign problem on our results, the grey filled points show the results if the sign is ignored during the calculation of the expectation value. The transition temperature remains nonzero at  $n \neq 1$  but is expected to rapidly decrease by decreasing the density. The flat region at  $\mu > -1.0$  is a direct consequence of the Mott physics setting in near half filling. At the temperatures we have access to, the Mott gap is not fully developed yet, and the density starts deviating from unity around  $\mu = -1.0$ . This is more clearly seen in the equations of state shown in the inset of Fig. 4(a) at  $T = 0.24$ ,  $0.28$ , and  $0.32$ .

In Figs. 4(b) and 4(c), we show the average neuron output taking and without taking the sign problem into account vs  $\mu$  at  $T = 0.32$ ,  $0.30$ , and  $0.28$ . In Fig. 4(b), we also show the average sign in the DQMC at  $T = 0.32$ . As the latter approaches zero around  $\mu = -2$ , the accuracy in the expectation value of the neuron output is largely compromised. Although it appears that the sign problem does not have a considerable effect on the value of our observable for this particular magnetic phase transition, we find significant differences between the results obtained taking and without taking the sign into account at  $T = 0.32$  starting at  $\mu = -1.4$ . The latter can call into question a recent suggestion that the sign problem can be circumvented using neural networks by essentially ignoring the sign in training or classifications [24]. We fit the data with error bars to a third-degree polynomial for  $\mu \geq -2.2$  and deduce a critical  $\mu$  of  $-2.0$  for the 0.5 crossing, which is larger than  $-2.6$  obtained by ignoring the sign problem. Note that here we are “transfer learning” by employing a network that has been trained in the sign-problem-free parameter region and using it in the sign

problematic region, an approach that can be used in other machine learning applications where training in the sign-problematic region cannot be justified. At  $T < 0.32$ , the sign problem becomes more severe and prevents us from obtaining estimates for  $T_N$  [see Fig. 4(c)].

The inset of Fig. 4(a) shows that ignoring the sign problem also leads to a smaller average density at a given  $\mu$  (see the grey dashed line in the inset corresponding to  $T = 0.32$ ). This leads to a magnetic phase diagram in the more physical temperature-density space that very much depends on whether or not the sign has been properly taken into account. Our results suggest that, at  $T = 0.32$  and the critical  $\mu = -2$ , the density is less than 0.95, consistent with recent results from a more conventional method [21].

## V. CONCLUSION

We have utilized neural-network machine learning techniques to predict the onset of the finite-temperature magnetically ordered phases of the 3D Fermi-Hubbard model. We train a 3D CNN using auxiliary spin configurations for a range of temperatures around the transition temperature, sampled over during DQMC simulations of two systems in the weak- and strong-coupling regimes. We show that the trend in the Néel temperature of the half-filled model as a function of the interaction strength can be captured by using the trained network to classify configurations generated for other interaction strengths. We then train a network at half filling for  $U = 9$  and use it to predict the fate of the ordered phase as the system is doped away from half filling. We find that the instability persists in the latter region in close proximity to the commensurate filling and that not including the sign in the expectation value of the neuron outputs leads to different results.

## ACKNOWLEDGMENTS

We acknowledge useful conversations with Rajiv R. P. Singh, Andy Millis, Simon Trebst, and Peter Broecker. K. C. and E. K. acknowledge support from the U.S. National Science Foundation under Grant No. DMR-1609560. R. M. acknowledges support from NSERC and the Canada Research Chair program. Additional support was provided by the Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science.

## APPENDIX A: DETERMINANT QUANTUM MONTE CARLO

In the DQMC, the partition function  $Z = \text{Tr}e^{-\beta H}$  is expressed as a path integral by discretizing the inverse temperature  $\beta$  into  $\mathcal{L}$  slices of length  $\Delta\tau$ . The one-body (kinetic) and two-body (interaction) terms of the Hubbard

Hamiltonian are then separated in each time slice, leading to a product of two exponentials:  $Z = \text{Tr}(e^{-\Delta\tau K} e^{-\Delta\tau V})^{\mathcal{L}} + \mathcal{O}(\Delta\tau^2)$ , where  $K$  and  $V$  are the kinetic and interaction parts of the Hamiltonian, respectively. Since the two terms do not commute, this process introduces a small controlled error of the order of  $\Delta\tau^2$ . The interaction terms at different times can then be written in terms of one-body (quadratic) fermion operators using the Hubbard-Stratonovich (HS) transformation

$$e^{-U\Delta\tau n_{\uparrow}n_{\downarrow}} = \frac{1}{2} e^{-U\Delta\tau(n_{\uparrow}+n_{\downarrow})/2} \sum_{s=\pm 1} e^{-s\lambda(n_{\uparrow}-n_{\downarrow})}, \quad (\text{A1})$$

where  $\cosh \lambda = e^{U\Delta\tau/2}$ , at the expense of introducing a sum over the field of auxiliary (spin) variables like  $s$  in a  $D + 1$ -dimensional space ( $D$  spatial and 1 imaginary time). The fermionic degrees of freedom in the quadratic form can be integrated out analytically, resulting in the product of two determinants, one for each fermion species, and leaving behind the sum over the  $2^{N\mathcal{L}}$  configurations of the auxiliary spins, where  $N$  is the spatial size of the system. Hence, the partition function can be expressed as

$$Z \propto \sum_{\{s_{i,l}\}} \det M_{\uparrow}(\{s_{i,l}\}) \det M_{\downarrow}(\{s_{i,l}\}), \quad (\text{A2})$$

where  $M_{\sigma}$  is an  $N \times N$  matrix that depends on the spin configuration  $\{s_{i,l}\}$ , and  $i$  and  $l$  represent the spatial and time indices.

Observables are estimated by important sampling over the configurations, which is performed, e.g., using the Metropolis algorithm, with the product of determinants as the probability, to accept or reject proposed local changes to the field. However, the determinants are costly to update and, other than in a few special cases, can take different signs at low temperatures, leading to the infamous ‘‘sign problem’’ [27,28]. At half filling, the two determinants have the same sign by symmetry, and thus, there is no sign problem. Away from half filling, one can take the absolute value of the product of determinants as the new probability in the Metropolis algorithm, but we have to treat the sign with the observable explicitly. The expectation value of a physical observable,  $O$ , can be calculated using

$$\langle O \rangle = \frac{\langle\langle SO \rangle\rangle}{\langle\langle S \rangle\rangle}, \quad (\text{A3})$$

where  $S$  denotes the sign of the product of determinants and  $\langle\langle \dots \rangle\rangle$  represents the Monte Carlo average using the absolute value of the product of determinants as the probability. Here, we take the neuron outputs as observables. We use jackknife resampling to estimate the error bars.

The particular decoupling scheme we have chosen is especially useful for inferring magnetic ordering by the machine purely based on the auxiliary fields. This is because the correlations between fermionic spins  $S$  can be written in terms of auxiliary spins [29,30]:

$$\langle S_i(\tau)S_j(0) \rangle = (1 - e^{-\Delta\tau U})^{-1} \langle s_i(\tau)s_j(0) \rangle, \quad (\text{A4})$$

(not valid when  $\tau = 0$  and  $i = j$ ). Therefore, the auxiliary field variables are effectively representing the fermion spins with the same correlations in space and time. The decoupling can be done using other variables that couple to density or pairing operators, which can be more helpful in detecting other phases, including charge density wave or superconductivity. We do not explore those possibilities here.

We use the quantum electron simulation toolbox (QUEST) [31] for our DQMC simulations. The toolbox, with minimal modifications, allows for outputting of the HS field configurations at arbitrary intervals during the simulations. We use 1000 (500) warmup sweeps and between 10 000 and 100 000 (1000 and 10 000) measurement sweeps for  $N = 4^3$  ( $N = 8^3$ ) in each run. We repeat runs with different random number seeds to obtain the desired number of configurations. In most cases, the HS field configuration is written into a file after every 10 sweeps, the same frequency typically used to perform physical measurements. For training, we choose the temperature grid for every value of  $U$  so that we have the same number of temperature points above and below the expected  $T_N$ . The temperatures are chosen in a nonuniform grid that can extend from 0.03 to about 3.0 depending on  $U$ .

The discretization of the imaginary time interval  $[0 - \beta]$  ( $\beta = 1/T$ ) in the DQMC introduces a systematic Trotter error, which scales like  $\Delta\tau^2$ , where  $\Delta\tau = \beta/\mathcal{L}$  and  $\mathcal{L}$  is the number of time slices. Thus, in principle, one has to choose a variable  $\mathcal{L}$  at different temperatures so that the error in physical observables is of the same order. However, such configurations, with variable size of the fourth dimension, are not useful in the training of a network whose size of the input layer has to be kept fixed. Hence, we choose a large fixed  $\mathcal{L} = 200$  for all  $\beta$  such that  $\Delta\tau$  remains below the threshold value of  $(8U)^{-1/2}$ , where the Trotter error can be neglected, for the largest  $U$  values and the lowest temperatures considered here. Moreover, we are not interested in calculating physical observables; rather, we are interested in locating the critical points at which the correlations of the system, both in the spatial and in the imaginary time dimensions, diverge; thus, the discretization errors can essentially be ignored.

## APPENDIX B: CONVOLUTIONAL NEURAL NETWORK

### 1. Case of $N = 4^3$

Each auxiliary field configuration is fed into a 3D CNN, with each imaginary time slice encoded as a different filter channel (input block with  $4^3$  neurons—see the input layer in Fig. 1). Each block from the  $\mathcal{L}$  imaginary time slices is immediately connected to 32 hidden volumetric feature maps (blocks in the first hidden layer) for feature

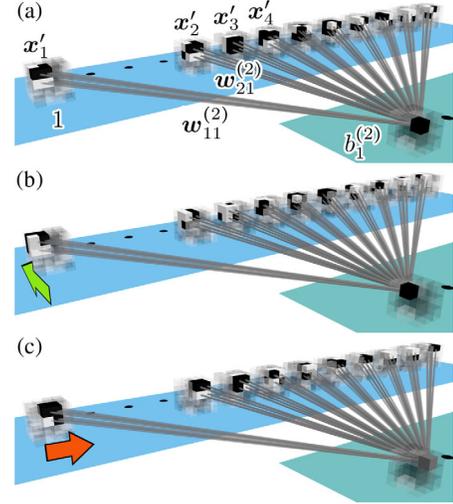


FIG. 5. Construction of a volumetric feature map. To produce the highlighted neuron in the volumetric feature map, from the input, a shared filter  $w_{11}^{(2)}$  is convolved with a  $2 \times 2 \times 2$  local receptive cube  $x'_i$  across all the inputs in the imaginary time slices, adding a bias offset  $b_1^{(2)}$ , before being activated using ReLU. Sliding the  $2 \times 2 \times 2$  cube in each of the spatial dimension while repeating the convolutional procedure completes a volumetric feature map. The spatial size of the volumetric feature map can be preserved by using zero padding.

extraction, for the network to detect a collection of unique features.

To produce the volumetric feature maps in the first hidden layer, we convolve a shared filter (also known as a kernel)  $w_{lm}^{(2)}$  with a  $2 \times 2 \times 2$  local receptive cube that sweeps each input block, and we offset the result with a shared bias  $b_m^{(2)}$ . The kernel  $w_{lm}^{(2)}$  can be thought of as a  $2^3$  tensor, where the superscript denotes the layer it corresponds to,  $l \in [1, \mathcal{L}]$  is the input block label, which is the same as the imaginary time index, and  $m \in [1, 32]$  labels the blocks in the first hidden layer. The convolution can be written as a tensor operation  $z_{lmn} = w_{lm}^{(2)} x_{ln} + b_m^{(2)}$ , where  $x_{ln}$  represents the  $n$ th input as picked up by the local receptive cube as it sweeps the  $l$ th input block. The process is depicted in Fig. 5. Before  $z$  is written in the corresponding element in the hidden layer, we pass it through a rectified linear unit (ReLU)  $f(z) = \max(0, z)$ , serving as the neural activation function.

Sliding the local receptive cube with a stride of one in each spatial dimension on the  $\mathcal{L}$  input cubes completes the construction of a volumetric feature map. To ensure that the information around the borders does not deplete too quickly, especially with smaller input size, as more convolutions are performed in the next layers, we use zero padding to preserve the input size after each convolution; we pad each of the input cubes with zeros, increasing their size to  $5 \times 5 \times 5$  so that the output volume has the same size as the original data, i.e.,  $4 \times 4 \times 4$ .

As deeper networks are typically more powerful, we use two more convolutional layers with ReLUs. We employ 16 and 8 volumetric feature maps for the second and third hidden convolutional layers, respectively. We repeat the procedure described above to construct the feature maps in these layers.

The final feature extraction layer is connected to the first classification layer, where some judgement on the input is made (see Fig. 1). There are eight neurons in this layer. Each of the  $8 \times 4^3$  individual neurons from the third feature extraction layer is connected to each of the fully connected neurons in the first classification layer. Finally, each of the eight hidden neurons is connected to two output neurons, where the final judgement is made, using softmax as their activation function. One of the output neurons represents the ordered state ( $T < T_N$ ), and the other represents the unordered state ( $T > T_N$ ). The output of the softmax functions represents the likelihood of the input configuration belonging to each of the two categories. As such, the sum of outputs from the two neurons is always 1. The activated output neuron is taken to be the one with the highest probability.

During training, we use dropout regularization with a dropout rate of 0.5 on the eight fully connected neurons in order to mitigate overfitting. Namely, half of the eight fully connected neurons are chosen at random and temporarily deactivated. This forces the neurons to adapt to more robust features.

An exponentially decaying learning rate is also used to ensure that the network starts out learning rapidly and slows down the learning after the model is close to convergence. The learning rate is given by

$$\eta = \eta_0 \lambda^{\text{training epoch}}, \quad (\text{B1})$$

where the initial learning rate  $\eta_0 = 10^{-3}$ , and the decay rate  $\lambda = 0.925$ . A complete training epoch is defined as when the network has stepped through the whole set of training data once.

We use a cross-entropy cost function given by [1]

$$C = -\frac{1}{n_{\text{td}}} \sum_x \sum_{i=1}^2 [y_i \ln a_i + (1 - y_i) \ln(1 - a_i)], \quad (\text{B2})$$

where  $x$  is the input data,  $n_{\text{td}}$  is the number of training data,  $y_i$  is the desired output, and  $a_i$  is the network output.

## 2. Case of $N = 8^3$

The same procedure described above for the  $N = 4^3$  system is used for  $N = 8^3$ . However, to keep the computational cost of the optimizations manageable, for  $N = 8^3$ , we perform convolutions without zero padding. We also use four hidden-feature extraction layers in this case. To prevent

overshooting the optimal model, we use a more aggressive decay rate  $\lambda = 0.875$ .

We find that handpicking the number of feature maps and fully connected neurons, similar to what we do for the case of  $N = 4^3$ , yields no learning. Thus, to avoid a costly hyperparameter grid search, where hyperparameter refers to the number of feature maps or number of neurons in the classification layer, we perform a Monte Carlo sampling for 20 search iterations in the hyperparameter space. We constrain the number of feature maps and fully connected neurons to the interval [8, 128]. The optimized hyperparameters we found (the set that yields the largest training accuracy) are  $n^{(2)} = 54$ ,  $n^{(3)} = 26$ ,  $n^{(4)} = 14$ ,  $n^{(5)} = 18$ , and  $n^{(6)} = 64$ , where  $n^{(2)}$  through  $n^{(5)}$  are the number of feature maps in the feature extraction layers and  $n^{(6)}$  is the number of neurons in the first classification layer.

## APPENDIX C: TRAINING THE CNN

Figure 6 shows the training progress of our networks and the evolution of the cost function in different cases vs training epochs. In a training epoch, we go through the entire set of shuffled training data once through iterations that involve batch sizes of 200. Here, 85% of the input data was used for training and 15% for testing. Test data are used as a gauge for overfitting (overtraining). The network overfits to the training data when the training accuracy continues to improve, but the testing accuracy stalls. We stop the training once we detect overfitting. The criterion is that the training accuracy and testing accuracy should not deviate from each other for more than 10 successive training evaluations.

We save the model (weights and biases) every time (i) the difference between the training accuracy and testing accuracy is less than 2.5%, (ii) the training accuracy is greater than the testing accuracy, and (iii) the current testing

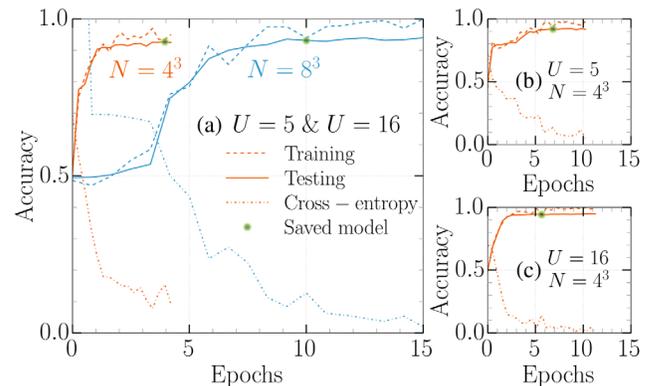


FIG. 6. Progress of the training process. Training accuracy for (a) simultaneous training with  $U = 5$  and 16, (b)  $U = 5$  alone, and (c)  $U = 16$  alone vs epochs. The location of last saved model is shown as a green circle in each panel. The testing accuracies for saved models are 92.7% (93.2%) for  $N = 4^3$  ( $N = 8^3$ ) in (a), 91.9% in (b), and 94.3% in (c).

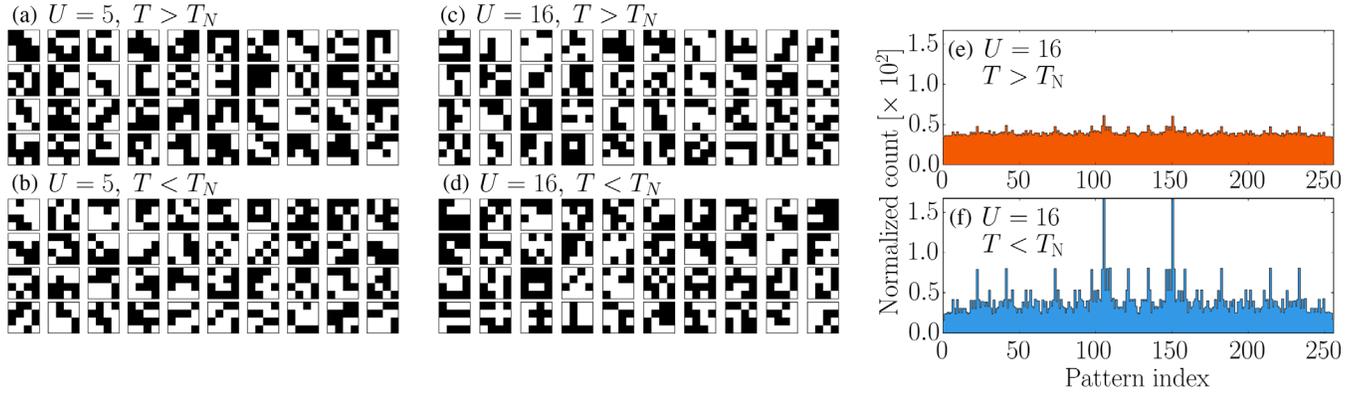


FIG. 7. (a)–(d) Typical auxiliary field patterns above and below  $T_N$  used as input to the CNN for the  $N = 4^3$  system. In each panel, the four rows correspond to the four layers in the  $z$  direction. We show the field only in the first 10 imaginary time slices (left to right) for each case. (e,f) Normalized count of the 256 patterns within  $2 \times 2 \times 2$  cubes as detected in the raw configurations for  $U = 16$  below and above  $T_N$ . The two largest peaks in (f) correspond to the two AFM patterns.

accuracy is better than the last recorded testing accuracy. The locations of the last saved models are shown as green circles for each case in Fig. 6.

#### APPENDIX D: WHAT DOES THE MACHINE LEARN?

One may wonder what features are learned by the CNN, as encoded in the weights and biases. Before addressing this question, we make one observation. Unlike in cases of image or sound recognition, the information put into our machine (auxiliary spin configurations) consist of binary numbers; namely, each input neuron takes 1 or  $-1$  as the value. Therefore, the “features” extracted by the eight-site receptive cube will be no more than  $2^8 = 256$  distinct patterns of 1’s and  $-1$ ’s on a  $2 \times 2 \times 2$  cube. In Figs. 7(a)–7(d), we show a typical auxiliary field for the first 10 time slices from DQMC simulations of the system with  $U = 5$  and  $U = 16$  both below and above  $T_N$ . One can see that, due to quantum fluctuations, a clear AFM pattern that can be discerned does not emerge in the 2D images. However, plotting the histogram of all 256 patterns in a sample set for  $U = 16$  in Figs. 7(e) and 7(f) reveals that the AFM pattern is in fact the dominant one at  $T < T_N$ .

We focus on the first feature extraction layer immediately after the input layer. What has been encoded in the weights and biases in that layer should be a good indication of what the CNN is looking for in all the feature extraction layers. Specifically, we would like to know to what extent the saved  $w_{lm}^{(2)}$  correlate with each of the 256 ordering patterns. To find out, we convolve each of the ordering patterns with each  $w_{lm}^{(2)}$ . We introduce the following overlap function:

$$\gamma_x = \sum_{lm} |w_{lm}^{(2)} x|, \quad (\text{D1})$$

where each  $x$  is a  $2 \times 2 \times 2$  tensor with 1’s and  $-1$ ’s as elements corresponding to one of the 256 ordering patterns on the eight-site cube. We have safely ignored the biases as they introduce a uniform shift and do not pertain to local correlations. We take the absolute value of the tensor product before performing the sums since we expect the spin inversion symmetry to have been encoded in the CNN to a good extent during the training; i.e.,  $\sum_{lm} w_{lm}^{(2)} x$  is orders of magnitude smaller than  $\gamma_x$  for any of the ordering patterns, something we confirm with our network. The dominant features, as seen by the CNN, are those with the largest  $\gamma_x$ . Here,  $\gamma_x$  plays a role analogous to the order parameter for each of the ordering patterns. For example,  $w x^{\text{AFM}}$ , where  $x^{\text{AFM}}$  corresponds to the perfect AFM pattern, would be the staggered “magnetization” of  $w$ .

In Fig. 8, we show histograms of  $\gamma_x$  for the 256 possibilities for  $x$  at  $U = 5$  and  $U = 16$ , respectively.

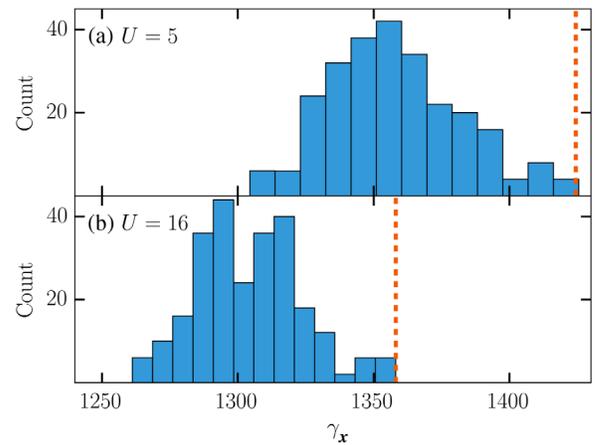


FIG. 8. Histogram of the overlap between the learned features and all the possible ordering patterns picked up by the  $2 \times 2 \times 2$  receptive cube. Top and bottom panels correspond to cases with  $U = 5$  and  $U = 16$ , respectively. The vertical dashed (red) line marks the value for the perfect AFM ordering pattern.

The ones in the large- $\gamma$  tail of the distribution are the dominant patterns. The red vertical line denotes the magnitude of the convolution (overlap) with the two degenerate AFM patterns. We find that for both a CNN trained by the  $U = 5$  data and one trained by the  $U = 16$  data, the AFM pattern is clearly the dominant pattern learned by the network.

- 
- [1] M. A. Nielsen, *Neural Networks and Deep Learning* (Determination Press, 2015).
- [2] P. Mehta and D. J. Schwab, *An Exact Mapping Between the Variational Renormalization Group and Deep Learning*, arXiv:1410.3831.
- [3] L.-F. Arsenault, A. Lopez-Bezanilla, O. Anatole von Lilienfeld, and A. J. Millis, *Machine Learning for Many-Body Physics: The Case of the Anderson Impurity Model*, *Phys. Rev. B* **90**, 155136 (2014).
- [4] A. G. Kusne, T. Gao, A. Mehta, L. Ke, M. C. Nguyen, K.-M. Ho, V. Antropov, C.-Z. Wang, M. J. Kramer, C. Long, and I. Takeuchi, *On-the-Fly Machine-Learning for High-Throughput Experiments: Search for Rare-Earth-Free Permanent Magnets*, *Sci. Rep.* **4**, 6367 (2014).
- [5] S. V. Kalinin, B. G. Sumpter, and R. K. Archibald, *Big-Deep-Smart Data in Imaging for Guiding Materials Design*, *Nat. Mater.* **14**, 973 (2015).
- [6] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, *Big Data of Materials Science: Critical Role of the Descriptor*, *Phys. Rev. Lett.* **114**, 105503 (2015).
- [7] S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, *A Structural Approach to Relaxation in Glassy Liquids*, *Nat. Phys.* **12**, 469 (2016).
- [8] G. Carleo and M. Troyer, *Solving the Quantum Many-Body Problem with Artificial Neural Networks*, *Science* **355**, 602 (2017).
- [9] G. Torlai and R. G. Melko, *Learning Thermodynamics with Boltzmann Machines*, *Phys. Rev. B* **94**, 165134 (2016).
- [10] E. M. Stoudenmire and D. J. Schwab, *Supervised Learning with Quantum-Inspired Tensor Networks*, *Adv. Neural Inf. Process. Syst.* **29**, 4799 (2016).
- [11] J. Carrasquilla and R. G. Melko, *Machine Learning Phases of Matter*, *Nat. Phys.* **13**, 431 (2017).
- [12] L. Wang, *Discovering Phase Transitions with Unsupervised Learning*, *Phys. Rev. B* **94**, 195105 (2016).
- [13] N. F. Mott, *The Basis of the Electron Theory of Metals and with Special Reference to the Transition Metals*, *Proc. Phys. Soc. London Sect. A* **62**, 416 (1949).
- [14] J. Hubbard, *Electron Correlations in Narrow Energy Bands*, *Proc. R. Soc. A* **276**, 238 (1963).
- [15] R. T. Scalettar, D. J. Scalapino, R. L. Sugar, and D. Toussaint, *Phase Diagram of the Half-Filled 3D Hubbard Model*, *Phys. Rev. B* **39**, 4711 (1989).
- [16] R. Staudt, M. Dzierzawa, and A. Muramatsu, *Phase Diagram of the Three-Dimensional Hubbard Model at Half Filling*, *Eur. Phys. J. B* **17**, 411 (2000).
- [17] P. R. C. Kent, M. Jarrell, T. A. Maier, and Th. Pruschke, *Efficient Calculation of the Antiferromagnetic Phase Diagram of the Three-Dimensional Hubbard Model*, *Phys. Rev. B* **72**, 060411 (2005).
- [18] T. Paiva, Y. L. Loh, M. Randeria, R. T. Scalettar, and N. Trivedi, *Fermions in 3D Optical Lattices: Cooling Protocol to Obtain Antiferromagnetism*, *Phys. Rev. Lett.* **107**, 086401 (2011).
- [19] E. Kozik, E. Burovski, V. W. Scarola, and M. Troyer, *Néel Temperature and Thermodynamics of the Half-Filled Three-Dimensional Hubbard Model by Diagrammatic Determinant Monte Carlo*, *Phys. Rev. B* **87**, 205102 (2013).
- [20] D. Hirschmeier, H. Hafermann, E. Gull, A. I. Lichtenstein, and A. E. Antipov, *Mechanisms of Finite-Temperature Magnetism in the Three-Dimensional Hubbard Model*, *Phys. Rev. B* **92**, 144409 (2015).
- [21] E. Khatami, *Three-Dimensional Hubbard Model in the Thermodynamic Limit*, *Phys. Rev. B* **94**, 125114 (2016).
- [22] A. W. Sandvik, *Critical Temperature and the Transition from Quantum to Classical Order Parameter Fluctuations in the Three-Dimensional Heisenberg Antiferromagnet*, *Phys. Rev. Lett.* **80**, 5196 (1998).
- [23] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, *Monte Carlo Calculations of Coupled Boson-Fermion Systems. I*, *Phys. Rev. D* **24**, 2278 (1981).
- [24] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, *Machine Learning Quantum Phases of Matter Beyond the Fermion Sign Problem*, arXiv:1608.07848.
- [25] S. Ji, W. Xu, M. Yang, and K. Yu, *3D Convolutional Neural Networks for Human Action Recognition*, *IEEE Trans. Pattern Analysis Machine Intelligence* **35**, 221 (2013).
- [26] M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems* (2015); software available from <http://www.tensorflow.org>.
- [27] E. Y. Loh, J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino, and R. L. Sugar, *Sign Problem in the Numerical Simulation of Many-Electron Systems*, *Phys. Rev. B* **41**, 9301 (1990).
- [28] V. I. Iglovikov, E. Khatami, and R. T. Scalettar, *Geometry Dependence of the Sign Problem in Quantum Monte Carlo Simulations*, *Phys. Rev. B* **92**, 045110 (2015).
- [29] J. E. Hirsch, *Discrete Hubbard-Stratonovich Transformation for Fermion Lattice Models*, *Phys. Rev. B* **28**, 4059 (1983).
- [30] J. E. Hirsch, *Connection Between World-Line and Determinantal Functional-Integral Formulations of the Hubbard Model*, *Phys. Rev. B* **34**, 3216 (1986).
- [31] Available at <http://quest.ucdavis.edu/> and <https://code.google.com/p/quest-qmc>.