Symmetry-enforced self-learning Monte Carlo method applied to the Holstein model

Chuang Chen,1,2 Xiao Yan Xu,3,* Junwei Liu,3 George Batrouni,4,5,6,7 Richard Scalettar,8 and Zi Yang Meng1,9,†

1Beijing National Laboratory for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China
2School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100190, China
3Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China
4Université Côte d’Azur, INPHYNI, CNRS, 0600 Nice, France
5Beijing Computational Science Research Center, Beijing 100193, China
6MajuLab, CNRS-UNS-NUS-NTU International Joint Research Unit UMI 3654, Singapore
7Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, 117542 Singapore
8Physics Department, University of California, Davis, California 95616, USA
9CAS Center of Excellence in Topological Quantum Computation and School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100190, China

(Received 5 March 2018; published 11 July 2018)

The self-learning Monte Carlo method (SLMC), using a trained effective model to guide Monte Carlo sampling processes, is a powerful general-purpose numerical method recently introduced to speed up simulations in (quantum) many-body systems. In this Rapid Communication, we further improve the efficiency of SLMC by enforcing physical symmetries on the effective model. We demonstrate its effectiveness in the Holstein Hamiltonian, one of the most fundamental many-body descriptions of electron-phonon coupling. Simulations of the Holstein model are notoriously difficult due to a combination of the typical cubic scaling of fermionic determinant which enters the weight of the configuration, and nonzero autocorrelation times even away from critical points. We argue that our work opens a research area of SLMC, providing a general procedure to deal with ergodicity in situations involving Hamiltonians with multiple, distinct low-energy states.

DOI: 10.1103/PhysRevB.98.041102

Introduction. Electron-phonon coupling is ubiquitously present in condensed matter materials, responsible not only for the nature of basic, single-particle properties such as the resistance and renormalized quasiparticle mass [1], but also for more exotic collective phenomena such as metal-insulator transitions [2], charge density wave (CDW) phases [3,4], and superconductivity (SC) [5,6]. Electron-phonon coupling also has a rich interplay with electron-electron interactions [7,8]. The Holstein Hamiltonian [9], which describes spinful electrons hopping on a lattice and interacting locally with a phonon degree of freedom, is one of the simplest models of this rich physics, incorporating both polaron formation [10–14] in the dilute limit, and collective insulating CDW and SC transitions. Despite its simplicity, an investigation of the Holstein Hamiltonian is extremely challenging, especially within an exact treatment of both its bosonic (phonon) and electronic degrees of freedom.

With the help of the Lang-Firsov transformation, quantum Monte Carlo (QMC) studies of Holstein (and related) models in one dimension (1D) can be performed for large systems and interesting parameter regimes [15]. The case of a retarded interaction in 1D is addressed with directed-loop QMC [16]. There have also been many attempts to explore the phase diagram of the 2D Holstein and Holstein-Hubbard models [17–34]. Recent studies of the metal-to-CDW phase transition in the weak-coupling regime [32] and the competition between CDW and superconductivity at intermediate-coupling strength with phonon dispersion [34] and different Fermi surface topologies [33] have broadened the understanding of the model and its relevance to the microscopic mechanism of superconductivity.

However, for 2D and 3D, these QMC methods are limited by the necessity for an expensive evaluation of a fermion determinant which enters the weight of the configuration, and also long autocorrelation times even away from critical points [35]. Thus, even though a sign problem [36,37] is absent for the Holstein model, reliable results for the complete phase diagram, in particular, for the most interesting intermediate-coupling strength of the model, are still missing. The cost of treating fermion determinants is largely unavoidable. However, the ergodicity problem has been successfully solved, on an instance-by-instance basis, in a number of classical and QMC approaches [38–40]. It remains the key bottleneck of many others [41] (apart from the sign problem). Indeed, the failure of ergodicity challenges determinant and constrained path QMC [42,43], lattice gauge theory simulations [44,45], impurity solvers [46,47] in dynamical mean-field theory, and configuration interaction methods in quantum chemistry [48,49].

A recently developed self-learning Monte Carlo (SLMC) method [50–54], based on a trained effective model to guide
the Monte Carlo simulation \[55,56\], shows substantial improvements over traditional Monte Carlo methods. The central idea of SLMC is to make use of learning algorithms to construct an approximate effective action which can be very rapidly calculated. An exact simulation is recovered by an evaluation of the full determinant which, however, can be done relatively infrequently, owing to the accuracy of the learned effective action. In 2D problems in which fermions coupled to bosonic fluctuations exhibit itinerant quantum critical points, \(L \times L\) spatial lattices with \(L \) up to 100 can be investigated at high temperature \[52\], and \(L\) up to 48 has been achieved at low temperatures with \(\beta \sim L\) scaling \[57,58\].

In this Rapid Communication, we show how SLMC can be applied to the Holstein Hamiltonian. Our key results are the following: (1) Long autocorrelation times can be greatly reduced by designing an effective bosonic Hamiltonian for the phonon fields which incorporates a global \(Z_2\) symmetry in the original model; (2) computational complexity is reduced from roughly \(O(L^{11})\) to \(O(L^3)\), i.e., a speedup of \(O(L^8)\) in SLMC over traditional MC methods; and (3) with such improvements, simulations of lattice sizes up to \(L = 20\) are possible, allowing the evaluation of the metal-to-CW transition temperature to an order of magnitude higher accuracy than previously available. These advantages open a research area of QMC for strongly correlated systems where SLMC provides a powerful and general procedure to improve ergodicity.

**Model.** We study the Holstein Hamiltonian,

\[
    H = H_{el} + H_{lat} + H_{int},
\]

with

\[
    H_{el} = -t \sum_{\langle ij \rangle \sigma} c_{i \sigma}^\dagger c_{j \sigma} - \mu \sum_{i \sigma} n_{i \sigma},
\]

\[
    H_{lat} = \sum_i \left( \frac{M \Omega^2}{2} X_i^2 + \frac{1}{2M} p_i^2 \right),
\]

\[
    H_{int} = g \sum_{i \sigma} n_{i \sigma} X_i.
\]

\(H_{el}\) describes spinful electrons hopping on a 2D square lattice of linear size \(L\), \(H_{lat}\) is the free phonon Hamiltonian, and \(H_{int}\) describes a local coupling between the electron density and the phonon displacement at site \(i\). \(\Omega\) is the phonon frequency, and \(g\) is the electron-phonon coupling. We set \(M = t = 1\) as the units of mass and energy, and focus on half filling (\(\mu = \frac{g^2}{4}\), \(\langle n_i \rangle = 1\)). On a square lattice, \(W = 8\) is the noninteracting bandwidth and \(\lambda = \frac{g^2}{\Omega B W} = \frac{g^2}{\Omega B^2}\) provides a dimensionless measure of the electron-phonon coupling. In this work, we focus on the intermediate-coupling strength \(\lambda = 0.5\) (\(g = 1\), \(\Omega = 0.5\)), a parameter regime which is more challenging than that explored in several recent works \[32,34\].

One can see some of the fundamental physics of the Holstein model by considering the atomic limit (\(t = 0\)). Completing the square in \(\frac{g^2}{4} X_i^2 + g n_i X_i\), and integrating out the phonon degrees of freedom, leads to an effective attraction \(\frac{g^2}{4\Omega^2}\) between spin-up and spin-down electrons, and to pair formation \[29\].

At low temperatures, when \(t\) is made nonzero, these pairs can either organize into an insulating CDW pattern (which tends to happen at half filling), or condense into a SC phase (at incommensurate density). On the other hand, integrating out the fermions leads to a potential energy surface for the phonons with two minima, at \(X_i = 0\) and \(X_i = \frac{2\pi}{L}\), corresponding to empty (\(n_i = 0\)) and double (\(n_i = 2\)) occupation, respectively. The large barrier at single occupation \(n_i = 1\) between these minima is the fundamental cause of long autocorrelation times in QMC simulations.

Several QMC methods have been used to simulate the Holstein model \[10,11,13,17,18,22,32,59\]. Here, we use the determinant quantum Monte Carlo (DQMC) \[42,63,64\], which is especially effective in dimension \(D > 1\) and in the large-coupling regime. In this approach, the inverse temperature \(\beta = L / \Delta\) is discretized (we use \(\Delta = 0.1\) in this work), and a path integral expression for the partition function is constructed in terms of the quantum coordinates in space \(i \in L^2\) and imaginary time index \(l = 1, 2, \ldots, L\). The fermions are integrated out, resulting in a weight \(\omega[X]\) for the phonon fields \(X_{i,l}\) which consists of a product of a bosonic piece \(e^{-S_{\text{bose}} \Delta t}\) with \(S_{\text{bose}} = \frac{\Omega^2}{M} \sum_l \sum_i X_{i,l}^2 + \sum_i \sum_l (X_{i,l+1} - X_{i,l})^2\) and a fermion contribution \([\det M(X_{i,l})]^2\). The square comes from the fact that the two spin species couple to the phonon field in the same way, giving rise to identical determinants. Here, \(M\) is a matrix of dimension \(N = L^2\).

Local updates of a single phonon coordinate \(X_{i,l}\) can be done with a computation cost \(O(N^2)\), so that a sweep through all \(N L\) components scales as \(N^2 L = L^4 L\) in \(D = 2\). The first computational bottleneck, associated with the fermionic degrees of freedom, is immediately evident: Doubling the linear lattice size \(L\) results in a 64-fold increase in computation time in \(D = 2\). Part of the origin of the second bottleneck is also clear from the form of \(S_{\text{bose}}\). The phonon degrees of freedom on adjacent imaginary time slices \(l\) are tightly coupled, especially so as \(\Delta l\) becomes small. Moves of a single coordinate are thus energetically unfavorable. A block update of all the imaginary time phonon coordinates \(l\) of a single spatial lattice site \(i\) helps surmount this problem. It is also important to tune the value of the change in the block update \(\Delta X = \frac{g^2}{4\Omega}\), to shift the phonon fields between the two minima. More details can be found in Sec. I of the Supplemental Material (SM) \[65\].

However, even with the local and block updates, the autocorrelation time \(\tau_e\) of DQMC for the Holstein model is still found to increase rapidly with system size, as shown in Fig. 2(a). (Further aspects of Fig. 2 will be discussed later.) We find \(\tau_e \sim L^{5.1}\), much worse than the dynamic critical exponent \(\tau_e \sim L^z\) with \(z = 2\) associated with classical Monte Carlo simulations with a local update, e.g., of the Ising model near its critical point \[38\]. Such autocorrelation times lead to a situation where in 2D, \(L \sim 10–14\) is at the limit of DQMC simulations. It is important to note that large \(\tau_e\) occurs even away from any critical point, but in this work we focus on the most difficult situation, a critical slowing down near \(T_c\).

**SLMC.** To overcome these problems, we apply SLMC to the Holstein model. The first step \[50–54\] is to obtain an effective model by self-learning on configurations generated with DQMC updates according to Eq. (1). Here, at each temperature studied, we use 80 000 configurations obtained for \(L = 6\) systems to train the effective model \(H^{\text{eff}}\), which we choose to have a polynomial form,

\[
    H^{\text{eff}}[X] = E_0 + J_i X_i + J_{ij} X_i X_j + \cdots ,
\]
where $E_0$ is the zeroth-order background, $J_i$ are the first-order terms, $J_{ij}$ are second-order terms, ..., and indices $i$ and $j$ are now combined space-imaginary time coordinates. Such a form is very natural, as after tracing out the fermions, the bosonic fields acquire long-range interactions in space-time beyond the bare level. One can make use of the symmetry of the original model to further reduce the number of parameters in the effective model, consequently reducing the effort and uncertainty in the fitting step.

The two potential minima of the Holstein model are symmetric with respect to $X = -\frac{\alpha}{\beta} \equiv \alpha$. We build this into the effective model by representing the potential by functions with these two minima (Fig. 1 and see Sec. III of the SM for details). We find that for the phonon fields in the Holstein model, two functions are sufficient to fit an appropriate barrier width and height. Our effective model thus has the following form (up to a constant),

$$-\beta H^{\text{eff}} = J_i \sum_{i\tau} (X_{i\tau+1} - X_{i\tau})^2$$

$$+ J_p \sum_{i\tau} \left( \frac{1}{4} (X_{i\tau} - \alpha)^2 - \frac{\alpha^2}{2} (X_{i\tau} - \alpha)^2 \right)$$

$$+ J'_p \sum_{i\tau} \left( \frac{1}{6} (X_{i\tau} - \alpha)^6 - \frac{\alpha^2}{4} (X_{i\tau} - \alpha)^4 \right)$$

$$+ J_{nn} \sum_{(ij)\tau} (X_{i\tau} - \alpha)(X_{j\tau} - \alpha)$$

$$+ J'_{nn} \sum_{(i\tau)(j\tau')} (X_{i\tau} - \alpha)(X_{j\tau'} - \alpha),$$

where the $J_k$ term comes from the phonon kinetic energy, $J_p$ and $J'_p$ terms are functions which produce the two global minima of Fig. 1, and $J_{nn}$ and $J'_{nn}$ are the nearest-neighbor interactions in the spatial and temporal directions, respectively. Long-range interactions are found to contribute little to the weight and are thus omitted (we have tried spatial interactions to $L/2$ and temporal interactions encompassing ten time slices). One key remark here is this effective model captures a global $Z_2$ symmetry, namely, a global mirror operation on $X$ with axis $\alpha$ leaving $H^{\text{eff}}$ invariant.

With the effective model in the form of Eq. (4), the training procedure is straightforward. Given a configuration $\mathcal{X}$ of the phonon field and its corresponding weight $\omega[\mathcal{X}]$, generated in DQMC, we have

$$-\beta H^{\text{eff}}[\mathcal{X}] = \ln (\omega[\mathcal{X}]).$$

Combining Eqs. (4) and (5), optimized values of $J_i$, $J_p$, $J'_p$, $J_{nn}$, and $J'_{nn}$ (shown in Table S1 in the SM) can be readily obtained through a multilinear regression [50–52] using all the configurations prepared with DQMC. Note, for each temperature, we only train $H^{\text{eff}}$ from a small system size ($L = 6$), but use it with larger systems (up to $L = 20$) in SLMC.

We use the effective model to guide the Monte Carlo simulation of the original model, namely, propose many updates of the phonon fields according to Eq. (4)—this is the so-called cumulative update in SLMC [51,52]. We then calculate the acceptance ratio of the final phonon field configuration via the expensive fermion determinant only rarely. There are two advantages of SLMC over DQMC. First, the effective model is purely bosonic and its local update is $O(1)$ since it bypasses the calculation of fermion determinants. Second, since the effective model is bosonic, global updates, such as Wolff and other cluster update schemes [38,39], are easy to implement. This is crucial since cluster updates in conventional DQMC actually worsen the scaling from $O(N^3 L^z)$ to $O(N^3 L^z)$. The Holstein model exhibits a finite-temperature metal-to-CDW insulator phase transition at half filling belonging to the 2D Ising universality class [32,34]. As discussed in detail in Sec. IV of the SM, we designed a modified Wolff-cluster update on the effective model, by building the cluster in space and including all sites of temporal columns which addresses additional long autocorrelation times associated with the proximity to the critical point. This modified Wolff update successfully reduces the autocorrelation time of Monte Carlo simulations from $L^{5.1}$ to $L^{2.9}$ as shown in Fig. 2(a). So the dynamical exponent is reduced by $\Delta z \geq 2$, an equivalent improvement to that provided by cluster moves in the classical Ising model [38,39].

Using this combination of updates on the effective model, we propose a cumulative move [51,52] of the phonon field for the original model, combined with a final acceptance ratio,

$$A(\mathcal{X} \rightarrow \mathcal{X}') = \min \left\{ 1, \frac{\exp(-\beta H[\mathcal{X}']) \exp(-\beta H^{\text{eff}}[\mathcal{X}])}{\exp(-\beta H[\mathcal{X}]) \exp(-\beta H^{\text{eff}}[\mathcal{X}'])} \right\},$$

which ensures detailed balance and hence simulation of the original Holstein $H[\mathcal{X}]$.

**Results.** To compare SLMC and DQMC, Fig. 2(a) depicts the autocorrelation time of the CDW structure factor $S_{\text{CDW}} = (X^2 \sum_{i\tau} (-1)^{i+1} (n_i n_{\tau+i} - n_{\tau+i} n_i))$ for DQMC and SLMC. We have chosen the most challenging criterion, both by analyzing a long-range quantity associated with the order parameter (which has much longer correlations than simple local quantities such as the energy) and also by tuning the temperature to $T = T_c$. To compare on an equal footing, a MC step in DQMC is defined as a sweep with local update plus four block updates, whereas a MC step in SLMC is defined as local plus Wolff–cluster updates. As can be seen from the fitting of the $\tau_L \sim L^z$, a
severe critical slowing down is observed in DQMC, with a dynamical exponent ($z \sim 5.1$); on the other hand, $z \sim 2.9$ in SLMC. A reduction of $Δ z \geq 2$ is achieved, equivalent to the improvement of cluster over local moves in a classical Ising model [38,39].

In Fig. 2(b), we show how much CPU time is needed to obtain a statistically independent phonon configuration. SLMC provides a $\times 50$ speedup for $L = 12$ and a more than $\times 300$ speedup for $L = 20$. With this dramatic improvement, we are able to simulate the Holstein model in the difficult parameter regime ($λ \sim 0.5$) and determine the critical point with high precision. At the critical point, the finite-size scaling behavior \( S_{\text{CDW}} / L^2 \sim L^{-2β/ν} f\left( L^{1/ν} (T − T_c) / T_c \right) \) is expected, where $β = \frac{1}{8}$, $ν = 1$ are the 2D Ising critical exponents. Figure 3(a) shows $L^{-7/4} S_{\text{CDW}}$ vs $T$ for $L = 10−20$. Their crossing point yields the critical temperature $T_c = 0.244(3)$. In Fig. 3(b), we further rescale the horizontal axis with $L (T − T_c)$, giving an excellent data collapse. Our value of $T_c$ represents a substantial improvement over existing work, which typically reports maximal lattice sizes of $L = 10−12$ [34].

Conclusions. In this Rapid Communication, we applied SLMC to simulations of the electron-phonon interaction in the Holstein model, where QMC is very difficult due to long autocorrelation times and the expense of the fermion determinantal evaluation. By imposing a global $Z_2$ symmetry in the effective model of SLMC, we have successfully captured the global minima of the phonon potential. In addition, we designed a Wolff-cluster update in the effective model which greatly reduces the autocorrelation time challenge that has hampered simulation of the Holstein model for several decades.

With these improvements on the effective model, SLMC simulation of the Holstein Hamiltonian can be pushed to larger system sizes, comparable with other interacting fermion systems [66–69] or itinerant quantum criticality models [57,58,70]). This allows for a much more reliable determination of critical properties.

The idea of imposing symmetry in the effective model can be generalized to other situations. For example, in the Hubbard and Hubbard-like models, one can introduce a continuous auxiliary field $φ$ to decouple the interaction in the spin channel, $αφ(n_{↑} − n_{↓})$. High barriers interfere with movement between minima at $n_{↑} − n_{↓} = ± 1$. Block updates in DQMC can be used [71] but are very time consuming, especially as the lattice size increases. Our work suggests that SLMC can improve this situation by proposing global updates based on effective models such as Eq. (4) with appropriate symmetry to capture the minima and lead to a big reduction in autocorrelation time. Construction of large-scale moves is not only computationally much less expensive within the context of effective models such as Eq. (4), but it is also much easier to incorporate intuitive physical pictures of the low-energy configurations. This suggests SLMC will provide a general framework for improving ergodicity beyond the Holstein model illustrated here.

If successful, outstanding problems such as the spectral properties in the Mott insulator, or the recently discovered phenomena of symmetric mass generation [72,73], could be explored more efficiently. Potential applications extend outside condensed matter physics to, for example, QMC simulations of the shell model Monte Carlo [74] in high-energy physics. The key requirement for our application of SLMC is only that the interactions can be decoupled to continuous bosonic fields associated with fermion bilinears, where several minima are present.

Acknowledgments. C.C. and Z.Y.M. acknowledge valuable discussions with Martin Hohenadler and Fakher As-saad on the Holstein model, and thankful for the support from the Ministry of Science and Technology of China.
under Grant No. 2016YFA0300502, the Key Research Program of the Chinese Academy of Sciences under Grant No. XDPB0803, the National Natural Science Foundation of China under Grants No. 11421092, No. 11574359, and No. 11674370, and the National Thousand-Young Talents Program of China. X.Y.X. acknowledges the hospitality of Institute of Physics, Chinese Academy of Sciences. We thank the following institutions for allocation of CPU time: the Center for Quantum Simulation Sciences in the Institute of Physics, Chinese Academy of Sciences; the Tianhe-1 platform in the National Supercomputer Center in Tianjin. J.L. acknowledges financial support from the Hong Kong Research Grants Council (Project No. ECS26302118). G.G.B. acknowledges support from the Université Côte d’Azur IDEX Jedi. The work of R.T.S. is supported by the U.S. Department of Energy under Grant No. DE-SC0014671.

[65] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.98.041102 for details on the implementation of the DQMC and SLMC algorithms of the Holstein model, which includes Ref. [75].