Elective-momentum ultrasize quantum Monte Carlo method

Zi Hong Liu,[1,2,*] Xiao Yan Xu,[3,*] Yang Qi,[4,5,6,†] Kai Sun,[7] and Zi Yang Meng[1,8,9,‡]

[1]*Beijing National Laboratory for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*
[2]*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100190, China*
[3]*Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China*
[4]*Center for Field Theory and Particle Physics, Department of Physics, Fudan University, Shanghai 200433, China*
[5]*State Key Laboratory of Surface Physics, Fudan University, Shanghai 200433, China*
[6]*Collaborative Innovation Center of Advanced Microstructures, Nanjing 210093, China*
[7]*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*
[8]*CAS Center of Excellence in Topological Quantum Computation and School of Physical Sciences,*
University of Chinese Academy of Sciences, Beijing 100190, China
[9]*Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China*

One bottleneck of quantum Monte Carlo (QMC) simulation of strongly correlated electron systems lies at the scaling relation of computational complexity with respect to the system sizes. For generic lattice models of interacting fermions, the best methodology at hand still scales with βN^3 where β is the inverse temperature and N is the system size. Such scaling behavior has greatly hampered the accessibility of the universal infrared (IR) physics of many interesting correlated electron models at (2+1)D, let alone (3+1)D. To reduce the computational complexity, we develop a new QMC method with inhomogeneous momentum-space mesh, dubbed elective-momentum ultrasize quantum Monte Carlo (EQMC) method. Instead of treating all fermionic excitations on an equal footing as in conventional QMC methods, by converting the fermion determinant into the momentum space, our method focuses on fermion modes that are directly associated with low-energy (IR) physics in the vicinity of the so-called hot spots, while other fermion modes irrelevant for universal properties are ignored. As shown in the paper, for any cutoff-independent quantities, e.g., scaling exponents, this method can achieve the same level of accuracy with orders of magnitude increase in computational efficiency. We demonstrate this method with a model of antiferromagnetic itinerant quantum critical point, realized via coupling itinerant fermions with a frustrated transverse-field Ising model on a triangle lattice. The system size of $48 \times 48 \times 32$ ($L \times L \times \beta$, almost 3 times of previous investigations) are comfortably accessed with EQMC. With much larger system sizes, the scaling exponents are unveiled with unprecedentedly high accuracy, and this result sheds new light on the open debate about the nature and the universality class of itinerant quantum critical points.

I. INTRODUCTION

As an unbiased numerical method, determinantal quantum Monte Carlo (DQMC) is widely used to study sign-problem-free interacting-fermion systems [1–9]. Despite its great success, for systems with large correlation lengths, such as quantum critical systems, it is still highly challenging for this method to accurately reveal the fate of the system at the thermodynamic limit, because of the sharp rise in computational complexity for systems with large sizes, which scales as $O(\beta N^3)$ with β being the inverse temperature and N being the volume of the lattice [3,6]. Although polynomial, this complexity becomes prohibitive for accessing low-temperature physics in large systems, which are nevertheless necessary for the endeavor such as itinerant quantum criticality in correlated electron systems [10–13].

To address this challenge, we develop a new QMC method dubbed elective-momentum ultrasize quantum Monte Carlo (EMUS-QMC, or EQMC in short), which applies generically to any itinerant quantum critical systems, although it is optimized for systems with finite wavelength instabilities (e.g., charge density waves or spin density waves). This method drastically reduces the computational costs and thus can be used to simulate larger systems in order to reveal long-distance physics and universal properties near a quantum critical point. The key to this method lies in the concept of renormalization group (RG) and the following observation. In DQMC, computational costs mainly come from handling the fermionic modes. However, in itinerant fermionic systems, fermionic excitations associated with low-energy excitations, which dominate the universal behaviors at infrared (IR), only inhabit a small part of the Brillouin zone (BZ). All other fermionic modes, which are associated with high-energy physics, are only needed for ultraviolet (UV) completion and are irrelevant as far as universal quantum-critical phenomena are concerned. In conventional DQMC simulations, all fermionic modes are treated on an equal footing, and thus for

---

*These authors have contributed equally to this work.

†qiyang@fudan.edu.cn

‡zymeng@iphy.ac.cn

universal properties at IR, e.g., scaling laws, a large portion of the computational resource is "wasted" on the high-energy UV modes, irrelevant in the sense of RG [14].

In the EMUS-QMC method that we developed, the fermion determinant is computed in the momentum space. Instead of treating all fermionic modes equivalently, we select momentum points within certain patches of the BZ, in which fermionic excitation directly contributes to low-energy (IR) physics. When we compute the fermion determinant, only fermionic modes with momenta inside these patches are taken into account, while all other fermionic degrees of freedom are ignored. As mentioned above, these ignored fermion modes are irrelevant for universal behaviors at infrared (IR), and thus our method will produce the same value for any physics observables independent of UV cutoffs, e.g., scaling exponents. Because this procedure reduces the number of fermionic degrees of freedom, the computational complexity can be greatly reduced from $O(\beta N^3)$ down to $O(\beta N_f^3)$, where N is the volume of the whole BZ and N_f is the volume of the momentum patch that are kept in the simulation (measured in the number of fermion modes). Speedup, to the level of 10^3, can be easily achieved in this method as it is easy to reach $\frac{N}{N_f} \sim 10$. In fact, for the model studied in Sec. III, we showed that reliable and accurate results can be obtained even for $\frac{N}{N_f} = 36$, which implies an even more drastic reduction in computational cost.

Same as many other RG-based techniques, our method only focuses on cut-off independent universal properties. For quantities that are sensitive to UV physics, e.g., the value of the critical temperature or critical coupling strength, our technique is not expected to generate the same value in comparison with the full-momentum-space DQMC simulations. Examples of this type are also shown below.

Our EQMC method is particularly suitable for studying quantum criticality in correlated itinerant electron systems, which is a subject with great theoretical and experimental significance [10–13,15–18], and plays a vital role in the understanding of anomalous transport, strange metal, and non-Fermi-liquid behaviors [19–23] in heavy-fermion materials [24,25], Cu- and Fe-based high-temperature superconductors [26–28], as well as the recently discovered pressure-driven quantum critical point (QCP) between magnetic order and superconductivity in transition-metal monopnictides, CrAs [29], MnP [30], CrAs$_{1-x}$P$_x$ [31], and other Cr/Mn-3d electron systems [32]. Despite the extensive theoretical efforts in the past decades [10–12,33–37], the problem of itinerant quantum criticality is still open and among the most difficult ones in strongly correlated electron systems, due to its nonperturbative nature.

Admittedly, the recent development in sign-problem-free DQMC methods paves the way to sharpen our understanding on this open question. Coupling a Fermi surface (FS) with various bosonic critical fluctuations, many itinerant QCPs, including Ising nematic [38,39], ferromagnetic [23], charge density wave [40,41], spin density wave [42–47], and interaction-driven topological phase transitions [48–50] have been studied. However, the true scaling behaviors of the itinerant QCPs are yet to be explored, because the aforementioned βN^3 complexity prohibits the simulations to reach the true

thermodynamic limit, even with help of the very recent advances such as self-learning Monte Carlo method [41,51–54]. In this regard, the EQMC method, developed in this work, provides a systematic way to reach larger system sizes, and consequently allowing us to access the genuine scaling behaviors in the IR limit for itinerant quantum criticality.

The rest of the paper is organized as follows: the general formulation of EQMC method is laid out in Sec. II, and in Sec. III it is applied to a previously studied model [46] for benchmarking purpose, and with much larger system sizes simulated with EQMC, our previous conclusions are more firmly established. Finally, remarks on future applications of EQMC are provided in Sec. IV.

II. EMUS QUANTUM MONTE CARLO METHOD

A. General ideas

In general, DQMC simulations study the following fermion-boson models:

$$H = H_f + H_{fb} + H_b, \qquad (1)$$

where H_f contains a quadratic free-fermion model,

$$H_f = -\sum_{ija} t_{ij}(c_{ia}^\dagger c_{ja} + \text{H.c.}), \qquad (2)$$

H_{fb} describes the coupling between the fermions and bosonic modes, which is local and also quadratic in terms of fermion operator c and $c^\dagger$s:

$$H_{fb} = \lambda \sum_i c_{ia}^\dagger M_{ab} c_{ib} \phi_i, \qquad (3)$$

H_b describes a Hamiltonian of the bosonic modes ϕ_i, which can take any generic form. Here a, b denotes a combination of quantum numbers, including orbits and spins, that labels different fermion species. The summation over repeated indices of fermion species are taken implicitly if not specified explicitly. Such a model can either be obtained from an interacting fermion model, through introducing auxiliary fields using a Hubbard-Stratonovish transformation [1,2,6], or be constructed directly as a "designer" Hamiltonian [23,46].

In DQMC simulations, configurations of the bosonic modes ϕ_i are stochastically sampled, with weights obtained through integrating out the fermion modes. Traditionally, real-space fermion modes c_{ia} are chosen as eigenstates for this computation. In EQMC, we instead treat fermion modes in the momentum space. To this end, we rewrite Eqs. (2) and (3) in momentum space,

$$H_f = \sum_k [\epsilon(k) - \mu] c_{ka}^\dagger c_{ka} \qquad (4)$$

and

$$H_{fb} = \lambda \sum_{kk'} c_{ka}^\dagger M_{ab} c_{k'b} \phi_{k-k'}. \qquad (5)$$

Here ϕ_k denotes the k component of the Fourier transform of the bosonic field ϕ_i:

$$\phi_k = \frac{1}{N} \sum_i \phi_i e^{-i\mathbf{k}\cdot\mathbf{r}_i}. \qquad (6)$$
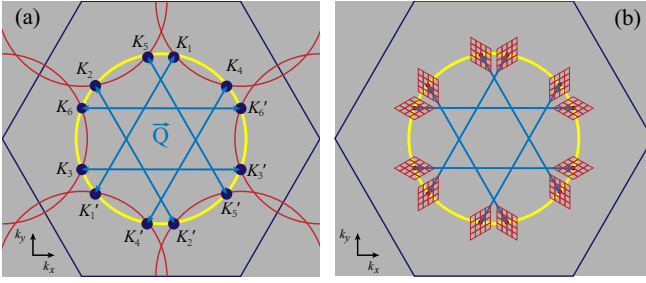
FIG. 1.   (a) The bare Fermi surface of the H_f, here as in Sec. III A, the FS (yellow circle) is for a triangle lattice model. The folded FS (red circles), coming from translating the bare FS by momentum **Q** (blue arrows), which is the wave vector for the antiferromagnetic fluctuations. The folded FS contains Fermi pockets and hot spots (black dots). Two hot spots connected by momentum **Q** forms a hot spot pair and we label it by $\{\mathbf{K}_l, \mathbf{K}'_l\}$, with $l = 1, 2, \ldots, 6$ in case of a triangle lattice model. (b) The inhomogeneous **k** mesh (red rhombus mesh) build around hot spots, the number of momentum points inside each mesh is denoted as N_f.

Rewriting the same problem in the momentum space gives us the freedom of choosing arbitrary k points in the momentum summation in Eqs. (4) and (5). When studying the low-energy and long-range physics, we choose IR fermion modes that are particularly relevant for this physics, and throw away other fermion modes without worrying about a proper UV completion of a lattice model. In particular, for studying fermionic QCPs, only fermion modes near the so-called "hot spots," where two patches of FSs are connected by the ordering wave vector **Q** [see Fig. 1(a)], are relevant to the universality class of the QCPs [34,35]. Therefore, we keep modes in patches around these hot spots, and neglect other modes in the BZ, as shown in Fig. 1(b). In this way, the number of fermion modes used in computing the effective weights is greatly reduced from the total size (or total volume) N to the patch size (or patch volume) N_f, thus drastically leave the computational burden, while retaining the same IR physics of what a lattice DQMC simulation of system size N can achieve.

When working with fermionic QCPs, the ratio N_f/N can be chosen with the following principle. Naively, in a lattice model, the lattice constant a provides a natural UV cutoff scale $\Lambda \sim \frac{2\pi}{a}$, which is also the size of the BZ. However, in practice, in a lattice model with the dispersions shown in Fig. 1(a), the cutoff Λ can be reduced. Due to the complicated shape of the Fermi surface, the universal scaling behavior is achieved assuming a linearized Fermi surface around the hot spots. Therefore, the size of the approximated linear portion of the Fermi surface provides the real UV cutoff Λ, which is much smaller than the size of the BZ. Λ also defines an "optimal" patch around each hot spot: when the patch size is larger than Λ, the cutoff scale is still Λ instead of the patch size. Therefore, the computational cost is reduced, while retaining the same universal physics. In other words, fermion modes outside the optimal patch do not contribute to the universal IR physics, and can be safely removed. However, reducing the patch size beyond Λ does not offer more acceleration: although the computational cost is reduced, the cutoff

is also reduced proportionally, since the cutoff is now given by the patch size. Hence, to reach the same ratio between UV/IR scales, which is needed for studying universal scaling behaviors, a finer mesh of momentum points in the patch must be used, which contains the same amount of fermion modes and thus requires no less amount of computational time. Therefore, the ratio N/N_f should be chosen such that the patch has the optimal size.

We emphasize that by removing fermion modes outside of IR patches, the nonuniversal behavior of the model is quantitatively varied. In other words, the EQMC method is simulating a *different* model, which shares the same universal IR physics, including the scaling behaviors at the QCP, as the original lattice model. This is clearly demonstrated in the example studied in Sec. III.

In EQMC, working in the momentum space means that we can no longer take advantage of the locality in Eq. (3), and therefore cannot efficiently perform local updates in the conventional DQMC [6]. In lattice-based DQMC, the local update [1,2,6,55], which tries to flip the bosonic spins $s_{i,\tau}$ one by one through the space-time lattice of volume βN, can benefit from the so-called "fast update." The acceptance ratio of such a flip involves a ratio of two determinants before and after the flip. The local nature of the update enables one to perform a fast update with a complexity $O(1)$ to compute this ratio, and a complexity $O(N^2)$ to update the Green's function if the flip is accepted [1], although generally the computational complexity for evaluating a determinant is $O(N^3)$. [Since flipping each spin costs $O(N^2)$, a full update that flips spins on the order of space time volume βN costs $O(\beta N^3)$, which is the well-known complexity of DQMC.] In EQMC, due to the dense nature of the coupling in Eq. (5), one can no longer use the fast update algorithm, and a local update would cost $\beta N \cdot O(\beta N_f^3)$. Fortunately, there are global-update algorithms one can choose, including the cumulative update in self-learning Monte Carlo method developed recently by some of us [52,53]. This gives rise to the complexity $O(\beta N_f^3)$ for computing the fermion determinant in a full update, which dominates the total computational cost of EQMC (see the detailed discussion in Sec. II C). Since N_f can be much smaller than N, speedup of $(\frac{N}{N_f})^3 \sim 10^3$ with $\frac{N}{N_f} \sim 10$ can be easily foreseen.

B. Steps of the algorithm

We now describe the details of the EQMC algorithm. Like all DQMC simulations, it samples through bosonic-field configurations $\{\phi_{i,\tau}\}$, with the following weights:

$$W[\phi] = W_b[\phi] \det[\mathbf{I} + \mathbf{B}(\beta, 0; \phi)]. \tag{7}$$

The weight has two parts: $W_b[\phi]$ denotes a bosonic weight determined by the Hamiltonian H_b in Eq. (1), and $\det[\mathbf{I} + \mathbf{B}(\beta, 0; \phi)]$ is a fermion determinant obtained by integrating out the fermion modes in the patches. Here $\mathbf{I}$ denotes the identity matrix; and $\mathbf{B}(\beta, 0; \phi)$ is a short form for the matrix product of $\mathbf{B}^M \mathbf{B}^{M-1} \cdots \mathbf{B}^1$, where the matrix at time slice τ is $\mathbf{B}^\tau = \exp(\Delta\tau \mathbf{K}) \exp(\mathbf{V}[\phi])$, with $\mathbf{K}$ the kinetic-energy matrix of the bare system in H_f in Eq. (4): $K_{ka,k'b} = [\epsilon(\boldsymbol{k}) - \mu]\delta_{kk'}\delta_{ab}$, and $\mathbf{V}[\phi]$ the fermion-boson coupling in H_{fb} in Eq. (5): $V_{ka,k'b} = \lambda M_{ab}\phi_{k-k'}$.

The bosonic weight $W_b[\phi]$ can be computed in standard ways, as in bosonic QMC simulations [56]. Computing the new fermionic determinant, on the other hand, takes three steps: (i) At each time slice, the configuration $\phi_{i,\tau}$ is Fourier transformed into momentum-space components $\phi_{k,\tau}$. (ii) The interaction matrix $\mathbf{V}[\phi]$ is constructed using $\phi_{k,\tau}$. (iii) The determinant $\det[\mathbf{I} + \mathbf{B}(\beta, 0; \phi)]$ is computed using the matrix $\mathbf{K}$, with all its elements fixed, and the matrix $\mathbf{V}[\phi]$, which is varying at each time slice.

The fermion matrices $\mathbf{K}$ and $\mathbf{V}[\phi]$ are generally $N_f \times N_f$ matrices. For the problem of fermion QCP with multiple pairs of hot-spot patches, an approximation can be made such that the matrices have a block-diagonalized structure, which further reduces the computational cost. In such a problem illustrated in Fig. 1, there are 12 hot spots, and they are divided into six pairs, each connected by the ordering wave vector $\mathbf{Q}$. Taking two momenta $\mathbf{k}$ and $\mathbf{k}'$ from the patches around one pair of hot spots, the momentum transfer $\mathbf{k}' - \mathbf{k} \simeq \mathbf{Q}$ corresponds to a low-energy bosonic mode. On the other hand, the momentum transfer $\mathbf{k}' - \mathbf{k}$ between two other momenta, which is not close to $\mathbf{Q}$, corresponds to a high-energy bosonic mode. Therefore, we make an approximation to only include in $\mathbf{V}$ the scattering within a pair of hot-spot patches connected by $\mathbf{Q}$, and ignore other scattering processes. Under this approximation, the matrix $\mathbf{V}$ becomes block diagonalized, with $2N_f \times 2N_f$ blocks, where $2N_f$ is the total number of modes in the pair of patches. Since the matrix $\mathbf{K}$ is already diagonal in momentum space, the resulting matrices $\mathbf{B}^\tau$ is also block diagonalized. This block-diagonalized structure allows the determinant of different blocks to be computed individually, thus saving both computational time and memory.

C. Complexity of the algorithm

The computational complexity of the EQMC algorithm can be determined following the steps outlined in Sec. II B. We first consider the complexity of computing the weight $W[\phi]$ in Eq. (7) for a given configuration $\{\phi_i\}$. The complexity of computing the bosonic weight $W_b[\phi]$ depends on the form of H_b. For a local Hamiltonian with finite-ranged interaction, this can be done with complexity $O(\beta N)$. Next, the Fourier transform from ϕ_i to ϕ_k is done using the fast Fourier transform (FFT) algorithm, which costs $O(N \log N)$ for each time slice, and $O(\beta N \log N)$ in total. Compared to this, the cost of computing $W_b[\phi]$ is much smaller and thus can be neglected. Then, constructing the matrices V costs $O(N_f^2)$ for each time slice, and $O(\beta N_f^2)$ in total. However, this is negligible compared to the cost of computing the determinant $\det[\mathbf{I} + \mathbf{B}(0, \beta; \phi)]$, which costs $O(\beta N_f^3)$. Therefore, the total computational cost for computing the weight of a given bosonic-field configuration $\{\phi_i\}$ is $O(\beta N \log N) + O(\beta N_f^3)$.

In typical models, the optimal patch occupies about 1%–20% of the total BZ. Therefore, in practice, we usually have $N/N_f < 100$. With this parameter range, the second term $O(\beta N_f^3)$ is much larger than the first term $O(\beta N \log N)$ for a large system. Therefore, we can safely ignore the first term, and use $O(\beta N_f^3)$ as the complexity of computing the weight $W[\phi]$.

The complexity of generating an uncorrelated configuration in EQMC is also $O(\beta N_f^3)$, assuming that we have

an effective global-update algorithm satisfying the following assumptions: (i) The computational cost of generating a new configuration $\{\phi_i\}$ is small compared to $O(\beta N_f^3)$. (ii) The autocorrelation time, measuring the number of steps in the generated Markov chain between two uncorrelated configurations, is of order one. Examples of such update algorithms, in the form of the cumulative update scheme in the recently developed self-learning Monte Carlo method [52,53], will be given in Sec. III C, in the context of a concrete model. Compared to the cost of lattice-based DQMC, $O(\beta N^3)$, the EQMC thus offers a very promising acceleration factor of $(\frac{N}{N_f})^3$.

III. DEMONSTRATION OF EQMC

A. Model

In this section we use the model studied in Ref. [46] to benchmark and demonstrate the power of the EQMC method. The Hamiltonian of the model can be organized into the form of Eq. (1), with each part given by

$$H_f = -t \sum_{\langle ij \rangle, \lambda, \sigma} (c_{i,\lambda,\sigma}^\dagger c_{j,\lambda,\sigma} + \text{H.c.}) - \mu \sum_i n_i, \quad (8)$$

$$H_b = J \sum_{\langle ij \rangle} s_i^z s_j^z - h \sum_i s_i^x, \quad (9)$$

$$H_{fb} = -\xi \sum_i s_i^z (\sigma_{i,1}^z + \sigma_{i,2}^z). \quad (10)$$

As shown in Fig. 2(a), fermions, subject to intralayer nearest-neighbor hopping t and chemical potential μ, reside on two of the layers $\lambda = 1, 2$. The middle layer is composed of Ising spins s_i^z with frustrated antiferromagnetic Ising coupling $J > 0$ and a transverse magnetic field h along s^x. Fermions and Ising spins are coupled together via an interlayer onsite Ising coupling H_{fb}, where $\sigma_{i,\lambda}^z = \frac{1}{2}(c_{i,\lambda,\uparrow}^\dagger c_{i,\lambda,\uparrow} - c_{i,\lambda,\downarrow}^\dagger c_{i,\lambda,\downarrow})$ is the fermion spin along z. We set $t = 1, J = 1, \mu = -0.5$ (electron density $\langle n_{i,\lambda} \rangle \sim 0.8$) and leave h and ξ as control parameters.

It is well known that H_b, describing a frustrated triangular-lattice transverse-field Ising model, has extensive ground state degeneracy at $h = 0$. At finite h, this degeneracy is lifted by the quantum order-by-disorder effect, resulting in an ordered ground state with a clock pattern [58], as shown in Fig. 2(b). The clock phase spontaneously breaks the translational symmetry [56,57] and thus has an enlarged unit cell. This phase is characterized by a complex order parameter $me^{i\theta} = m_1 + m_2 e^{i4\pi/3} + m_3 e^{-i4\pi/3}$, where $m_\alpha = \frac{3}{N} \sum_{i=1}^{N/3} s_{i,\alpha}^z$ with $\alpha = 1, 2, 3$ representing magnetization of the three sub-lattices of the $\sqrt{3} \times \sqrt{3}$ enlarged superlattice. In the momentum space, this order parameter has a finite wave-vector $\mathbf{Q} = (\frac{2\pi}{3}, \frac{2\pi}{\sqrt{3}})$, i.e., the corner of the hexagonal Brillouin zone as shown in Fig. 1(a). Upon introducing quantum/classical fluctuations via increasing h or T, the ordered phase can melt. The quantum melting is through a second-order quantum phase transition at $h_c = 1.63(1)$ with an emergent $U(1)$ symmetry [56]. Because of this emergent continuous symmetry, despite that H_b describes an Ising model, this quantum critical point belongs to the $(2 + 1)$D XY universality class and the thermal melting of the clock phase involves an intermediate BKT phase [56].
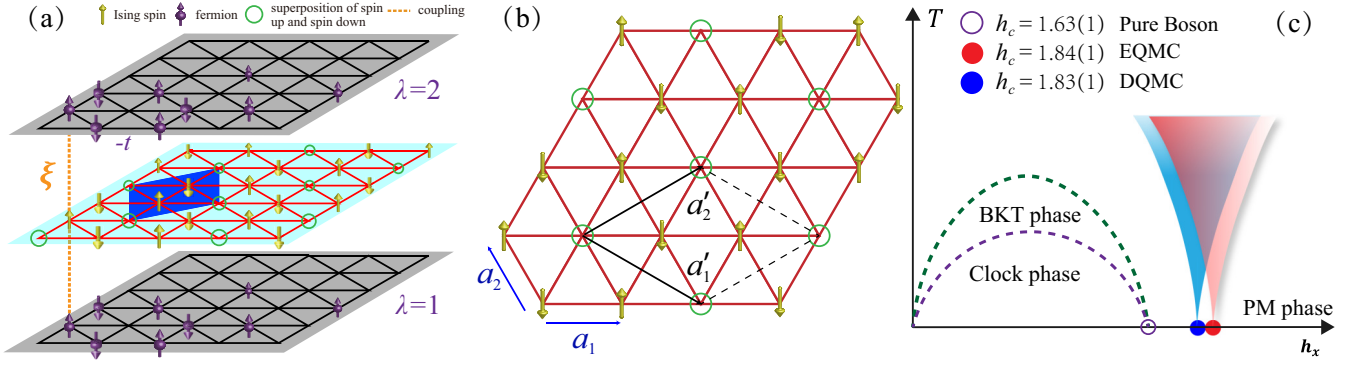
FIG. 2. (a) Illustration of the triangle lattice model in Sec. III A. Fermions reside on two of the layers ($\lambda = 1, 2$) with intralayer nearest-neighbor hopping t. The middle layer is composed of Ising spins s_i^z, subject to nearest-neighbor antiferromagnetic Ising coupling J and a transverse magnetic field h. Between the layers, an onsite Ising coupling is introduced between fermion and Ising spins (ξ). The shaded area in the middle layer stands for the enlarged unit cell of the clock phase before the Ising spins are polarized to the s^x direction. (b) Illustration of the spin arrangement of the clock phase when $h < h_c$. The $\mathbf{a}_1$, $\mathbf{a}_2$ are the lattice vectors of the original triangle lattice and $\mathbf{a}'_1$, $\mathbf{a}'_2$ are the lattice vectors of the clock phase with an enlarged unit cell. (c) Semiquantitative phase diagram. The dashed lines mark the phase boundaries of the pure bosonic model H_b, with a QCP (open magenta dot) at $h_c = 1.63(1)$ [56,57]. After coupling with fermions, the QCP shifts to higher values. The blue solid dot is the QCP of the original model in Ref. [46] with an homogeneous grid [$h_c = 1.83(1)$], and the blue area is the quantum critical region where Hertz-Mills-Moriya scaling has been observed. The red solid dot is the QCP of this study [$h_c = 1.84(1)$], The position of the QCP obtained from DQMC and EQMC is very close, and also the scaling behavior inside the quantum critical region (the red shaded area) is consistent with that observed in the blue shaded area. The EQMC scheme can comfortably capture the IR physics of such antiferromagnetic itinerant QCP, with much larger system sizes, 48×48, compared with previous study [46] 30×30 where great computational efforts have been spent.

In the presence of the fermion-spin coupling, which is relevant in the RG sense, the QCP moves to a higher value of h, as shown in Fig. 2(c). And compared with the QCP of DQMC, $h_c = 1.83(1)$, the QCP of EQMC is a slight different value, $h_c = 1.84(1)$, since the position of the critical is a nonuniversal quantity sensitive to UV cutoffs. But the universal part of the quantum critical point, as will be discussed later, remains the same between DQMC and EQMC.

Furthermore, because fermions and Ising spins are coupled together, the Ising-spin clock phase immediately generates a spin density wave ordering in the fermionic sector with finite ordering wave vector **Q**, which folds the Brillouin zone and renders a new FS with pockets as schematically shown in Fig. 1(a). Near the QCP, as shown in our previous work [46], the quasiparticle at the tip of the FS pockets lose their coherence, forming the so-called hot spots [35,37,59].

As shown in Ref. [46], with the help of cluster updates (Wolff, Swendsen-Wang, and geometric cluster) [60–62] and cumulative updates in the self-learning Monte Carlo method [51–54], we were able to simulate systems as large as $L = 30$ and temperature as low as $\beta = 30$, and overcome the critical slowing down in the vicinity of QCP, to some extent. Up to such system sizes, non-Fermi-liquid around the hot spots can be clearly seen, moreover, the quantum critical scaling in the critical region can be investigated via dynamic susceptibility of the Ising spins, we found that the dynamic susceptibility scales as

$$\chi(T, h, \mathbf{q}, \omega_n)$$
$$= \frac{1}{(c_t T + c'_t T^2) + c_h |h - h_c|^\gamma + c_q |\mathbf{q}|^2 + (c_\omega \omega + c'_\omega \omega^2)},$$
(11)

where c_t, c'_t, c_h, γ, c_q, c_ω, and c'_ω are determined from fitting the QMC data up to $L = \beta = 30$. It is important to note that, at low temperature and frequency, the system encounters crossover behavior from T^2 to T and from ω^2 to ω. This is consistent with the expectation of Hertz-Millis-Moriya theory [10–12]. However, we also note that to reveal the true IR physics of this itinerant antiferromagnetic QCP, $L = 30$ and $\beta = 30$ are clearly not large enough, as there are higher order perturbative RG calculations [19,34,35,37] suggesting the existence of anomalous dimension, i.e., the momentum dependence of χ^{-1} is not $|\mathbf{q}|^2$ but $|\mathbf{q}|^{2-\eta}$, with finite anomalous dimension η. To address such question, we will have to go to even large system sizes and lower temperatures, and that is partially our motivation to develop the EQMC method in this work.

B. Implementation of EQMC

The general practice and step of EQMC have been outlined in Secs. II A and II B. Here we will discuss the model-specific implementation of the algorithm.

After the Fourier transformation, the kinetic energy in Eq. (8) becomes

$$H_f = \sum_{\mathbf{k}, \lambda, \sigma} [\epsilon(\mathbf{k}) - \mu] c^\dagger_{\mathbf{k}, \lambda, \sigma} c_{\mathbf{k}, \lambda, \sigma},$$
(12)

with $\epsilon(\mathbf{k}) = -2t \cos(k_x) - 4t \cos(\frac{\sqrt{3}}{2} k_y) \cos(\frac{1}{2} k_x)$. The bosonic part H_b will still be kept in the real space with size $N \times N$. The coupling term H_{fb} is then transformed as

$$H_{fb} = -\xi \sum_{\mathbf{k}, \mathbf{k}', \lambda} s^z(\mathbf{k} - \mathbf{k}')(c^\dagger_{\mathbf{k}, \lambda, \uparrow} c_{\mathbf{k}', \lambda, \uparrow} - c^\dagger_{\mathbf{k}, \lambda, \downarrow} c_{\mathbf{k}', \lambda, \downarrow}),$$
(13)

where $s^z(\mathbf{q}) = \frac{1}{N} \sum_i e^{-i\mathbf{q}\cdot\mathbf{r}_i} s_i^z$ plays the role of the the bosonic field ϕ_k in Eq. (6).

In the triangular lattice model, the ordering wave vectors $\pm\mathbf{Q}$ connect six pairs of hot spots (12 in total) on the FS. We denote the six pairs of hot spots as $\{\mathbf{K}_l, \mathbf{K}'_l\}$, with $l = 1, 2, \ldots, 6$, as shown in Fig. 1(a). The distance between each two of the pair is one of the two ordering wave vectors $\mathbf{K}'_l - \mathbf{K}_l = \mathbf{Q}_l = \pm\mathbf{Q}$. In the IR limit, only the fluctuations within a pair of hot spots $\{\mathbf{K}_l, \mathbf{K}'_l\}$ are important to the universal scaling behavior in the vicinity of QCP [13,19,33–35,59,63]. Hence, to study this universal behavior, we draw one patch around each $\mathbf{K}_l$ and keep fermion modes therein, while neglect other parts of the BZ. In this way, instead of the original $N = L \times L$ total points one needs to keep, we will only keep the $N_f = L_f \times L_f$ points inside each patch, as marked by the 12 leaflike polygons in Fig. 1(b) that covering the 12 hot spots. Here L and L_f denote the linear size of the original lattice and the size of the patch, respectively.

For the sake of bookkeeping, we label momenta in each patch using the relative difference $\mathbf{q}$ measured from the center of the patch: the real momentum is $\mathbf{k} = \mathbf{K}_l + \mathbf{q}$ or $\mathbf{k}' = \mathbf{K}'_l + \mathbf{q}$, respectively. Correspondingly, we label the fermion mode in each pair of patches as $c_{\mathbf{q}l\lambda\alpha} = c_{\mathbf{K}_l+\mathbf{q},\lambda\alpha}$ and $c'_{\mathbf{q}l\lambda\alpha} = c_{\mathbf{K}'_l+\mathbf{q},\lambda\alpha}$, respectively. In this way, $\mathbf{q}$ is a small momentum around zero. Using this notation, the Hamiltonian in Eqs. (12) and (13) becomes

$$
\begin{aligned}
H_f = \sum_{\mathbf{q}l\lambda\sigma} \{ &[\epsilon(\mathbf{q} + \mathbf{K}_l) - \mu] c^\dagger_{\mathbf{q}l\lambda\sigma} c_{\mathbf{q}l\lambda\sigma} \\
+ &[\epsilon(\mathbf{q} + \mathbf{K}'_l) - \mu] c'^\dagger_{\mathbf{q}l\lambda\sigma} c'_{\mathbf{q}l\lambda\sigma} \},
\end{aligned}
\tag{14}
$$

$$
\begin{aligned}
H_{fb} = -\xi \sum_{\mathbf{q},\mathbf{q}',l,\lambda} [&s^z(\mathbf{q} - \mathbf{q}' + \mathbf{Q}_l) c^\dagger_{\mathbf{q}l\lambda\alpha} \sigma^z_{\alpha\beta} c'_{\mathbf{q}'l\lambda\beta} \\
+ &s^z(\mathbf{q}' - \mathbf{q} - \mathbf{Q}_l) c'^\dagger_{\mathbf{q}'l\lambda\alpha} \sigma^z_{\alpha\beta} c_{\mathbf{q}l\lambda\beta}],
\end{aligned}
\tag{15}
$$

where the sum over $\mathbf{q}$ are in the patches around the hot spots.

We notice that, compared to Eq. (13), only scattering between two modes in a pair of patches is kept in Eq. (15). This is because the distance between two momentum points in a pair of patches can be expressed as

$$
\begin{aligned}
\mathbf{k} - \mathbf{k}' &= \mathbf{q} - \mathbf{q}' + (\mathbf{K}_l - \mathbf{K}'_l) \\
&= \mathbf{q} - \mathbf{q}' + \mathbf{Q}_l,
\end{aligned}
\tag{16}
$$

which is approximately equal to $\mathbf{Q}_l = \pm\mathbf{Q}$. Hence, the momentum transfers in these scattering correspond to low-energy spin fluctuations, and such scattering processes are relevant to the IR universal behavior. On the contrary, other scattering processes have momentum transfers different from $\pm\mathbf{Q}$, and thus are not relevant to the IR universal behavior. This approximation is discussed in the general setting of EQMC in Sec. II B.

We now outline the construction of the $\mathbf{K}$ and $\mathbf{V}$ matrices (defined in Sec. II B) using Eqs. (14) and (15), which are needed in computing the fermion determinant. In total, we keep $12 \times L_f \times L_f$ fermion modes. As discussed in Sec. II B, the matrices are block diagonalized into six blocks, corresponding to the six pairs of hot spots.

First, the fermion dispersion in Eq. (14) is converted into the following $\mathbf{K}$ matrix:

$$
\mathbf{K}_\sigma = \begin{bmatrix} t_{1\sigma} & & & \\ & t_{2\sigma} & & \\ & & \ddots & \\ & & & t_{l\sigma} \end{bmatrix},
\tag{17}
$$

with

$$
t_{l\sigma} = \begin{bmatrix} \epsilon(\mathbf{q} + \mathbf{K}_l) - \mu & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \epsilon(\mathbf{q}' + \mathbf{K}'_l) - \mu & 0 \\ 0 & 0 & 0 & \ddots \end{bmatrix}.
\tag{18}
$$

Here $\mathbf{K}$ is divided into six blocks, each corresponds to one pair of patches. Each block is further divided into two halves, representing patches around $\mathbf{K}_l$ and $\mathbf{K}'_l$, respectively. Within each half, each entry corresponds to one momentum point $\mathbf{q}$ in the patch.

Second, the fermion-spin coupling in Eq. (15) is converted into the following $\mathbf{V}$ matrix:

$$
V_\sigma = \begin{bmatrix} v_{1\sigma} & & & \\ & v_{2\sigma} & & \\ & & \ddots & \\ & & & v_{l\sigma} \end{bmatrix},
\tag{19}
$$

where v_l is an off-diagonal block matrix. The matrix element in v_l represents the interaction within one hot-spot pair $\{\mathbf{K}_l, \mathbf{K}'_l\}$, whereas the other matrix element in V matrix, representing the momentum modes interaction between different hot-spots pair $\{\mathbf{K}_l, \mathbf{K}'_{l'}\}$, with $l \neq l'$, are set to zero under our IR limit approximation.

The structure of each v_l is

$$
v_{l\sigma} = \begin{bmatrix} 0 & 0 & s(\mathbf{q} - \mathbf{q}' + \mathbf{Q}_l) & \cdots \\ 0 & 0 & \vdots & \ddots \\ s(\mathbf{q}' - \mathbf{q} - \mathbf{Q}_l) & \cdots & 0 & 0 \\ \vdots & \ddots & 0 & 0 \end{bmatrix},
\tag{20}
$$

with matrix elements in the two diagonal blocks are zero since there is no scattering within the leaflike polygon of one hot spot $\mathbf{K}_l$ or $\mathbf{K}'_l$, however, the matrix element in the two off-diagonal blocks are nonzero since it is the scatterings between $\mathbf{K}_l$ and $\mathbf{K}'_l$, i.e., $s(\mathbf{q} - \mathbf{q}' + \mathbf{Q}_l)$ in Eq. (20). Since s_i^z are real, $s(\mathbf{q} - \mathbf{q}' + \mathbf{Q}_l) = s^*(\mathbf{q}' - \mathbf{q} - \mathbf{Q}_l)$ and $v_l = v_l^\dagger$, such that one only needs to handle the matrix elements coming from $s(\mathbf{q} - \mathbf{q}' + \mathbf{Q}_l)$.

Once having both block-diagonal matrices of $\mathbf{K}_\sigma$ and $\mathbf{V}_\sigma$ in the fermion determinant in the configurational weight in Eq. (7), the Fermion Green's function matrice also becomes

block diagonal,

$$
G_\sigma = \begin{bmatrix} G_{1\sigma} & & & \\ & G_{2\sigma} & & \\ & & \ddots & \\ & & & G_{l\sigma} \end{bmatrix}, \tag{21}
$$

this means we can update different pair of hot spots, namely, different $G_{l\sigma}$, independently. As claimed in Sec. II C, this block structure not only reduces memory usages, but also reduces computational cost and makes it easier to parallelize the simulation.

Finally, we describe the update scheme used in our simulation. As outlined in Sec. II A, the local-update algorithm, which is widely used in conventional DQMC simulations [1–3,6,55], becomes inefficient in EQMC simulations. This is a result of the fact that the form of Eq. (15) is nonlocal in the momentum space. Consequently, flipping one spin s_i^z in a spin configuration changes all rows and columns in the **V** matrix, and all the Fourier components $s(\boldsymbol{q})$ are affected. As a result, the fast-update method, which computes the ratio between two determinants differing only one row and one column with complexity $O(N)$ (where N is the size of the matrix), no longer applies.

Instead, we use global-update algorithms. In general, these algorithms stochastically propose a completely new spin configuration with a relatively low cost. The new configuration is then either accepted or rejected, following the detailed balance principle. Assuming that starting from a configuration $\{s_i^z\}$, the probability of constructing the configuration $\{\tilde{s}_i^z\}$ is $S(s^z \to \tilde{s}^z)$, then the detailed balance principle requires that probability of accepting this new configuration is

$$
\alpha(s^z \to \tilde{s}^z) = \min \left\{ \frac{S(\tilde{s}^z \to s^z)}{S(s^z \to \tilde{s}^z)} \frac{W[\tilde{s}^z]}{W[s^z]}, 1 \right\}. \tag{22}
$$

A good global update satisfies two conditions: (i) it generates a new configuration sufficiently different from the original one, such that it has almost no correlation with the previous one. (ii) The acceptance ratio computed from Eq. (22) is close to one. These two conditions guarantee that the autocorrelation time in the generated Markov chain is on the order of one.

In our simulations we use two types of global updates: First, we use cluster updates guided by the bosonic part of the Hamiltonian H_b only. In this way, a new configuration is constructed using well-known algorithms, including Wolff [61], Swendsen-Wang [60], and geometric cluster-update schemes [62]. These cluster-update algorithms imply that the probability $S(s^z \to \tilde{s}^z)$ is determined by H_b,

$$
\frac{S(s^z \to \tilde{s}^z)}{S(\tilde{s}^z \to s^z)} = \frac{W_b[\tilde{s}^z]}{W_b[s^z]}. \tag{23}
$$

The acception probability can then be computed using Eq. (22).

Second, we make use of the cumulative update in the self-learning Monte Carlo method [51–53]. To use this method, one first learns an effective bosonic model H_{eff} of the total system from the training steps with the configurational weight generated with other update methods. The effective model is in general nonlocal both in space and time, but since it is a

bosonic model, we can update the entire space-time configuration through it locally with relatively low computational cost (since this step is done without evoking the calculation of Fermi determinant). This step is done repeatedly, until a completely new bosonic configuration with little correlation with the previous one is generated. The collection of all these local steps is called the "cumulative update," and serves as the construction process of a global update. Similar to Eq. (23), the selection probability is given by the learned effective model,

$$
\frac{S(s^z \to \tilde{s}^z)}{S(\tilde{s}^z \to s^z)} = \frac{W_{\text{eff}}[\tilde{s}^z]}{W_{\text{eff}}[s^z]}. \tag{24}
$$

Both types of global updates construct a new configuration with little correlation with the previous one. Furthermore, plugging Eqs. (23) and (24) into Eq. (22), we can see that the resulting acceptance ratio α is close to one, if the bosonic models guiding the construction with weight W_b and W_{eff}, respectively, give good approximations to the full weight W. Hence, these update algorithms satisfy the conditions listed above, and thus can generate statistically independent samples with only a few [$O(1)$] update steps.

In these global-update algorithms, the step dominating the computational cost is computing the true weight $W[s^z]$, in order to determine α in Eq. (22). As discussed in Sec. II C, this in turn is dominated by computing the fermion determinant, which costs $O(\beta N_f^3)$. It is easy to check that constructing a new configuration, using either W_b or W_{eff}, is much faster in comparison. From the above analysis, we can conclude that the computational complexity for EQMC is $O(\beta N_f^3)$, at least $(N/N_f)^3$ times faster than that of the typical DQMCs $O(\beta N^3)$. And because we can easily have $N/N_f \sim 10$ (in the results in Sec. III C, we have $\frac{N}{N_f} = 36$), thousands times of speedup of EQMC over DQMC can be readily accessed, as demonstrated in the next section.

C. Results

To demonstrate the power and efficiency of EQMC, we investigate the triangle lattice model described above, and in particular, pay attention to the scaling behavior in the vicinity of the antiferromagnetic quantum critical point. We chose to fix the ratio of $L/L_f = 6$ such that $N/N_f = 36$, and the size of the leaf-patches $L_f = 4, 6, 8$, and fixed $\beta = 4 \times L_f$. This means that we can simulate the original model in Fig. 2(a) up to size $48 \times 48 \times 32$ ($L \times L \times \beta$), almost 3 times of the largest size ever simulated in the DQMC [46].

In our previous work [46] the bosonic susceptibilities $\chi(T, h, \boldsymbol{q}, \omega)$ close to the QCP, revealed with $L = \beta = 30$, is fitted to the form of Eq. (11). In particular, at low ω, $\chi^{-1}(0, 0, 0, \omega)$ exhibits crossover behavior from ω^2 to ω. It also scales with $\boldsymbol{q}$ as $\chi^{-1}(0, 0, \boldsymbol{q}, 0) \propto |\boldsymbol{q}|^2$.

As we discussed in Sec. II A, the model we simulated with EQMC [Eqs. (9), (14), and (15)], which includes only the k points inside the hot-spot patches, is different from the original model in their nonuniversal properties, due to approximation made in EQMC. This is shown in Fig. 2(c), where the location of the QCP can shift from $h_c = 1.83(1)$ obtained from DQMC in Ref. [46] to $h_c = 1.84(1)$ obtained from EQMC. To determine the position of the new h_c, we
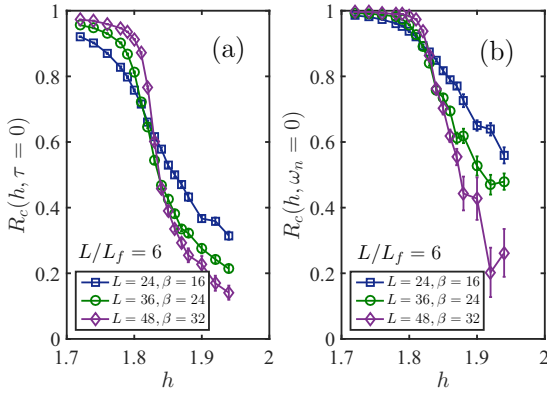
FIG. 3. (a) Correlation ratio $R_c(h, \tau = 0)$ and (b) $R_c(h, \omega = 0)$ of the Ising magnetic order in clock phase, as a function of h, with $L_f = 4, 6, 8$ (corresponding to $L = 24, 36, 48$) and $\beta = 4 \times L_f$, obtained from EQMC simulation. As L increases, the crossing point is found to be converged at $h_c = 1.84(1)$.

measure the correlation ratio [23,64] of the Ising spins as a function of h. We first calculate the magnetic susceptibility in the EQMC simulation

$$\chi(T, h, \mathbf{q}, \omega_n) = \frac{1}{L^2} \sum_{ij} \int_0^\beta d\tau \, \exp^{i(\omega_n \tau - \mathbf{q} \cdot \mathbf{r}_{ij})} \langle s_i^z(\tau) s_j^z(0) \rangle,$$
(25)

then construct the correlation ratio both from the equal-time susceptibility (magnetic structure factor)

$$R_c(h, \tau = 0) = 1 - \frac{\chi(T, h, \mathbf{Q} + d\mathbf{k}, \tau = 0)}{\chi(T, h, \mathbf{Q}, \tau = 0)},$$
(26)

and from the zero frequency susceptibility

$$R_c(h, \omega = 0) = 1 - \frac{\chi(T, h, \mathbf{Q} + d\mathbf{k}, \omega_n = 0)}{\chi(T, h, \mathbf{Q}, \omega_n = 0)},$$
(27)

where $d\mathbf{k}$ can be chosen as $\mathbf{b}_1 = \frac{2\pi}{L}(1, \frac{1}{\sqrt{3}})$ or $\mathbf{b}_2 = \frac{2\pi}{L}(0, 2\sqrt{3})$, which corresponded to the minimum distance in momentum space.

The results are shown in Figs. 3(a) and 3(b). It is clear that with $L_f = 4, 6, 8$, which means $L = 24, 36, 48$, the crossing point of the correlation ratio is converged to $h_c = 1.84(1)$ from EQMC simulations.

What is more important, is that the universal properties of the QCP shall remain intact with EQMC. Therefore, we further explore this assessment by means of analyzing the various divergences of $\chi(T, h, \mathbf{q}, \omega_n)$ in the form of Eq. (11) in the quantum critical region. The results are shown in Fig. 4.

We first look at the $\mathbf{q}$ dependence, as shown in Fig. 4(a), the momentum $|\mathbf{q}|$ is measured with respect to the hot-spot position $\mathbf{K}$. We plot the susceptibility data by subtraction of the finite temperature background as $\chi^{-1}(T, h_c, |\mathbf{q}|, 0) - \chi^{-1}(T, h_c, 0, 0) = c_q |\mathbf{q}|^{a_q}$, and fit the curve to obtain the coefficient c_q and the power a_q, as shown in the red solid
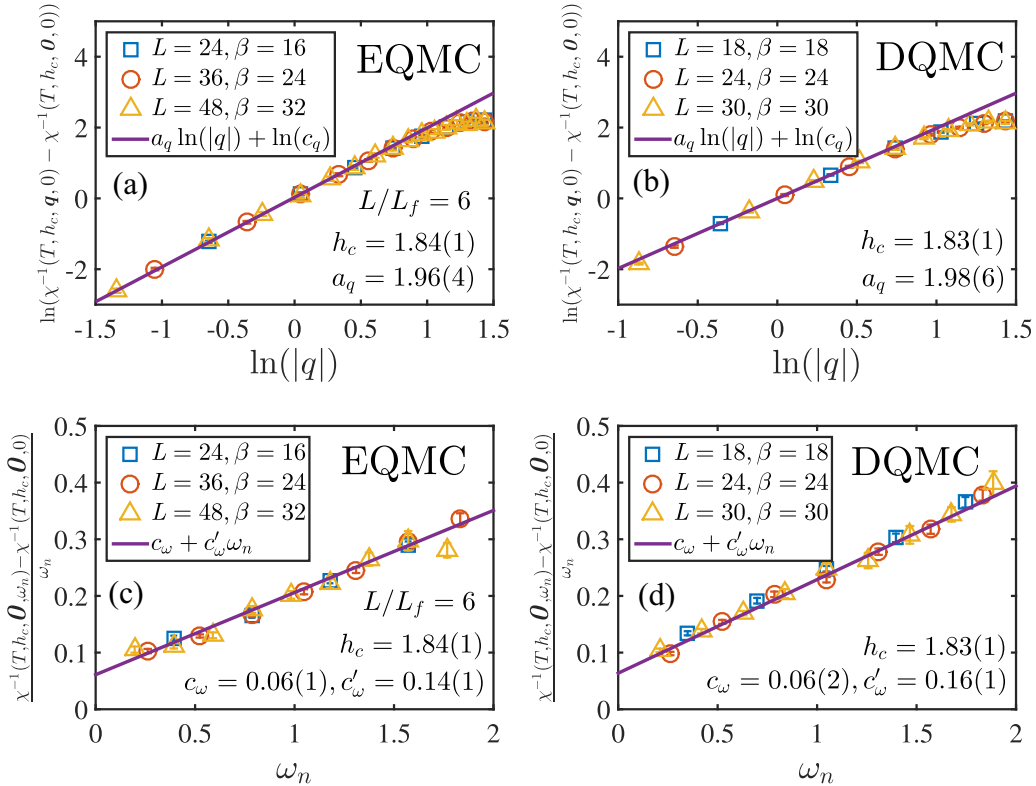

FIG. 4. $|\mathbf{q}|$, ω, and T dependence of the bosonic susceptibilities $\chi(T, h, \mathbf{q}, \omega)$ at the itinerant QCP $h = h_c$. Comparison of the EQMC results with L up to 48 and β up to 32 (a) and (c), with the previous DQMC results with L and β up to 30 (b) and (d). The scaling behaviors with system sizes as large as $48 \times 48 \times 32$ ($L \times L \times \beta$) in EQMC are fully consistent with the form in Eq. (11). The universal quantum critical scaling has been successfully captured by EQMC.

line. With the system size as large as $L = 48$, the $\chi^{-1} \sim |\mathbf{q}|^2$ behavior, clearly manifests, with $a_q = 1.96(4)$. These results are consistent with that obtained in our previous work [46], where the anomalous dimension η of this critical point, in the form of $\chi^{-1} \sim |\mathbf{q}|^{2-\eta}$, is zero within the error bar. For the sake of completeness, the DQMC results from Ref. [46] are also shown in Fig. 4(b).

As for the frequency dependence in χ, it is also consistent with our previous conclusion in Ref. [46]. As shown in Fig. 4(c), we analyze the $[\chi^{-1}(T, h_c, 0, \omega) - \chi^{-1}(T, h_c, 0, 0)]/\omega$, to subtract the finite temperature background. From the plot, it is clear that there is a finite intercept at $\omega = 0$, and from the fit, the form of $\chi^{-1} \sim c_\omega \omega + c'_\omega \omega^2$ can be observed, with $c_\omega = 0.06(1)$ and $c'_\omega = 0.14(1)$, fully consistent with our previously determined the coefficients with DQMC, as also shown in Fig. 4(d).

The results in Fig. 4 conclude that the form in Eq. (11) accurately describe this itinerant antiferromagnetic QCP. This form is different from the bare $(2 + 1)$D $O(2)$ universality which is the description of the bare bosonic problem [56,57], and is consistent with the Hertz-Millis-Moriya description [10–12]. However, the proposed anomalous dimension of this antiferromagnetic itinerant QCP with higher-order perturbative RG calculations [34] has not been observed with system size as large as $L = 48$.

IV. DISCUSSIONS

The elective-momentum ultrasize quantum Monte Carlo method (EMUS-QMC or EQMC), developed in this work, paves the way of performing quantum Monte Carlo simulations for larger system sizes and lower temperature, such that the genuine IR physics of many interesting yet difficult strongly correlated electron problems, exemplified with the itinerant antiferromagnetic quantum critical point here, are now ready to be explored. EQMC manages to reduce the notorious $O(\beta N^3)$ computational complexity of the conventional DQMC down to $O(\beta N_f^3)$, where N_f is the important momentum points inside the patch around hot spots. In practice, one can choose the size of the patches such that they cover the approximately linear-dispersing regions near the hot spots, which contribute to the universal scaling behaviors of the quantum critical point. Since $\frac{N}{N_f} \sim 10$ is easy to design, as shown in this work (here we have $\frac{N}{N_f} = 36$), a speedup to the order of 10^3 is achieved. In the benchmark example we demonstrated, the system sizes of $48 \times 48 \times 32$ ($L \times L \times \beta$) can be comfortably simulated without the great computational effort spent in previous work that only $30 \times 30 \times 30$ can be accessed.

Moreover, the speedup and model flexibility offered by EQMC opens up opportunities to study other interesting universality classes of fermion QCPs, in particular, those that are hard to be realized in a lattice model because the UV completion of the desired IR fermion mode would require many fermion modes in the BZ. For example, a similar antiferromagnetic square-lattice spin-fermion model can be used to study an itinerant QCP with a $\mathbb{Z}_2$ symmetry, where the EQMC method offers a speedup similar to what is demonstrated in this work, and existence or absence of the anomalous dimension in that case can be verified [47]. EQMC can also be applied to models where Dirac fermions interact with bosonic modes [48,50]. Compared to DQMC simulations using a honeycomb-lattice model or a π-flux square-lattice model, EQMC offers a speedup as only fermion modes with the linear dispersion of Dirac fermions are included [65], and therefore much larger system sizes and lower temperatures can be accessed, to resolve the present difference between numerical simulation on finite lattice and analytical field-theory calculations at the thermodynamic limit on the critical exponents of these models [66–69]. Furthermore, EQMC can be applied to investigate interaction effects on surface states of three-dimensional (3D) topological insulators and topological superconductors, which cannot be realized in a two-dimensional (2D) lattice model. With EQMC, we can simulate the 2D surface modes without the corresponding 3D bulk, thus greatly reduce the computational cost.

[1] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, Phys. Rev. D **24**, 2278 (1981).

[2] J. E. Hirsch, Phys. Rev. B **28**, 4059 (1983).

[3] S. R. White, D. J. Scalapino, R. L. Sugar, E. Y. Loh, J. E. Gubernatis, and R. T. Scalettar, Phys. Rev. B **40**, 506 (1989).

[4] F. F. Assaad and M. Imada, J. Phys. Soc. Jpn. **65**, 189 (1996).

[5] F. F. Assaad, Phys. Rev. Lett. **83**, 796 (1999).

[6] F. Assaad and H. Evertz, in *Computational Many-Particle Physics*, Lecture Notes in Physics, edited by H. Fehske, R. Schneider, and A. Weiße (Springer, Berlin, 2008), Vol. 739, pp. 277–356.

[7] Z. Y. Meng, T. C. Lang, S. Wessel, F. F. Assaad, and A. Muramatsu, Nature (London) **464**, 847 (2010).

[8] Y.-Y. He, H.-Q. Wu, Y.-Z. You, C. Xu, Z. Y. Meng, and Z.-Y. Lu, Phys. Rev. B **93**, 115150 (2016).

[9] Y.-Y. He, H.-Q. Wu, Y.-Z. You, C. Xu, Z. Y. Meng, and Z.-Y. Lu, Phys. Rev. B **94**, 241111 (2016).

[10] J. A. Hertz, Phys. Rev. B **14**, 1165 (1976).

[11] A. J. Millis, Phys. Rev. B **48**, 7183 (1993).

[12] T. Moriya, *Spin Fluctuations in Itinerant Electron Magnetism* (Springer, Berlin, 1985).

[13] H. v. Löhneysen, A. Rosch, M. Vojta, and P. Wölfle, Rev. Mod. Phys. **79**, 1015 (2007).

[14] Similar understanding has been realized in the recent work by Lang and Laeuchli [65], where by employing long-range hopping, the fermionic modes at IR can be extended to most of the BZ, thereby minimizes the finite size effect at the quantum criticality of DQMC simulations.

[15] G. R. Stewart, Rev. Mod. Phys. **73**, 797 (2001).

[16] A. V. Chubukov, C. Pépin, and J. Rech, Phys. Rev. Lett. **92**, 147003 (2004).

[17] D. Belitz, T. R. Kirkpatrick, and T. Vojta, Rev. Mod. Phys. **77**, 579 (2005).

[18] A. V. Chubukov and D. L. Maslov, Phys. Rev. Lett. **103**, 216401 (2009).

[19] W. Metzner, D. Rohe, and S. Andergassen, Phys. Rev. Lett. **91**, 066402 (2003).

[20] T. Senthil, Phys. Rev. B **78**, 035103 (2008).

[21] T. Holder and W. Metzner, Phys. Rev. B **92**, 041112 (2015).

[22] M. A. Metlitski, D. F. Mross, S. Sachdev, and T. Senthil, Phys. Rev. B **91**, 115111 (2015).

[23] X. Y. Xu, K. Sun, Y. Schattner, E. Berg, and Z. Y. Meng, Phys. Rev. X **7**, 031058 (2017).

[24] J. Custers, P. Gegenwart, H. Wilhelm, K. Neumaier, Y. Tokiwa, O. Trovarelli, C. Geibel, F. Steglich, C. Pepin, and P. Coleman, Nature (London) **424**, 524 (2003).

[25] A. Steppke, R. Küchler, S. Lausberg, E. Lengyel, L. Steinke, R. Borth, T. Lühmann, C. Krellner, M. Nicklas, C. Geibel, F. Steglich, and M. Brando, Science **339**, 933 (2013).

[26] W. Zhang, J. T. Park, X. Lu, Y. Wei, X. Ma, L. Hao, P. Dai, Z. Y. Meng, Y.-F. Yang, H. Luo, and S. Li, Phys. Rev. Lett. **117**, 227003 (2016).

[27] Z. Liu, Y. Gu, W. Zhang, D. Gong, W. Zhang, T. Xie, X. Lu, X. Ma, X. Zhang, R. Zhang, J. Zhu, C. Ren, L. Shan, X. Qiu, P. Dai, Y.-F. Yang, H. Luo, and S. Li, Phys. Rev. Lett. **117**, 157002 (2016).

[28] Y. Gu, Z. Liu, T. Xie, W. Zhang, D. Gong, D. Hu, X. Ma, C. Li, L. Zhao, L. Lin, Z. Xu, G. Tan, G. Chen, Z. Y. Meng, Y.-F. Yang, H. Luo, and S. Li, Phys. Rev. Lett. **119**, 157001 (2017).

[29] W. Wu, J. Cheng, K. Matsubayashi, P. Kong, F. Lin, C. Jin, N. Wang, Y. Uwatoko, and J. Luo, Nat. Commun. **5**, 5508 (2014).

[30] J.-G. Cheng, K. Matsubayashi, W. Wu, J. P. Sun, F. K. Lin, J. L. Luo, and Y. Uwatoko, Phys. Rev. Lett. **114**, 117001 (2015).

[31] M. Matsuda, F. K. Lin, R. Yu, J.-G. Cheng, W. Wu, J. P. Sun, J. H. Zhang, P. J. Sun, K. Matsubayashi, T. Miyake, T. Kato, J.-Q. Yan, M. B. Stone, Q. Si, J. L. Luo, and Y. Uwatoko, Phys. Rev. X **8**, 031017 (2018).

[32] J. Cheng and J. Luo, J. Phys.: Condens. Matter **29**, 383003 (2017).

[33] A. Abanov, A. V. Chubukov, and J. Schmalian, Adv. Phys. **52**, 119 (2003).

[34] A. Abanov and A. Chubukov, Phys. Rev. Lett. **93**, 255702 (2004).

[35] M. A. Metlitski and S. Sachdev, Phys. Rev. B **82**, 075128 (2010).

[36] S. Sur and S.-S. Lee, Phys. Rev. B **94**, 195135 (2016).

[37] A. Schlief, P. Lunts, and S.-S. Lee, Phys. Rev. X **7**, 021010 (2017).

[38] Y. Schattner, S. Lederer, S. A. Kivelson, and E. Berg, Phys. Rev. X **6**, 031028 (2016).

[39] S. Lederer, Y. Schattner, E. Berg, and S. A. Kivelson, Proc. Natl. Acad. Sci. USA **114**, 4905 (2017).

[40] Z.-X. Li, F. Wang, H. Yao, and D.-H. Lee, Phys. Rev. B **95**, 214505 (2017).

[41] C. Chen, X. Y. Xu, J. Liu, G. Batrouni, R. Scalettar, and Z. Y. Meng, Phys. Rev. B **98**, 041102 (2018).

[42] E. Berg, M. A. Metlitski, and S. Sachdev, Science **338**, 1606 (2012).

[43] Z.-X. Li, F. Wang, H. Yao, and D.-H. Lee, Sci. Bull. **61**, 925 (2016).

[44] Y. Schattner, M. H. Gerlach, S. Trebst, and E. Berg, Phys. Rev. Lett. **117**, 097002 (2016).

[45] M. H. Gerlach, Y. Schattner, E. Berg, and S. Trebst, Phys. Rev. B **95**, 035124 (2017).

[46] Z. H. Liu, X. Y. Xu, Y. Qi, K. Sun, and Z. Y. Meng, Phys. Rev. B **98**, 045116 (2018).

[47] Z. H. Liu, G. Pan, X. Y. Xu, K. Sun, and Z. Y. Meng, arXiv:1808.08878.

[48] X. Y. Xu, K. S. D. Beach, K. Sun, F. F. Assaad, and Z. Y. Meng, Phys. Rev. B **95**, 085110 (2017).

[49] F. F. Assaad and T. Grover, Phys. Rev. X **6**, 041049 (2016).

[50] Y.-Y. He, X. Y. Xu, K. Sun, F. F. Assaad, Z. Y. Meng, and Z.-Y. Lu, Phys. Rev. B **97**, 081110 (2018).

[51] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, Phys. Rev. B **95**, 041101 (2017).

[52] J. Liu, H. Shen, Y. Qi, Z. Y. Meng, and L. Fu, Phys. Rev. B **95**, 241104 (2017).

[53] X. Y. Xu, Y. Qi, J. Liu, L. Fu, and Z. Y. Meng, Phys. Rev. B **96**, 041119 (2017).

[54] Y. Nagai, H. Shen, Y. Qi, J. Liu, and L. Fu, Phys. Rev. B **96**, 161102 (2017).

[55] J. E. Hirsch, Phys. Rev. B **31**, 4403 (1985).

[56] Y.-C. Wang, Y. Qi, S. Chen, and Z. Y. Meng, Phys. Rev. B **96**, 115160 (2017).

[57] S. V. Isakov and R. Moessner, Phys. Rev. B **68**, 104409 (2003).

[58] R. Moessner and S. L. Sondhi, Phys. Rev. B **63**, 224401 (2001).

[59] M. A. Metlitski and S. Sachdev, Phys. Rev. B **82**, 075127 (2010).

[60] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

[61] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).

[62] J. R. Heringa and H. W. J. Blöte, Phys. Rev. E **57**, 4976 (1998).

[63] M. A. Metlitski and S. Sachdev, New J. Phys. **12**, 105007 (2010).

[64] S. Pujari, T. C. Lang, G. Murthy, and R. K. Kaul, Phys. Rev. Lett. **117**, 086404 (2016).

[65] T. C. Lang and A. M. Läuchli, arXiv:1808.01230.

[66] L. Iliesiu, F. Kos, D. Poland, S. S. Pufu, D. Simmons-Duffin, and R. Yacoby, J. High Energy Phys. 01 (2016) 110.

[67] L. Iliesiu, F. Kos, D. Poland, S. S. Pufu, and D. Simmons-Duffin, J. High Energ. Phys. 01 (2018) 0360.

[68] N. Zerf, L. N. Mihaila, P. Marquard, I. F. Herbut, and M. M. Scherer, Phys. Rev. D **96**, 096010 (2017).

[69] B. Ihrig, L. N. Mihaila, and M. M. Scherer, Phys. Rev. B **98**, 125109 (2018).