Dynamical tuning of the chemical potential to achieve a target particle number in grand canonical Monte Carlo simulations

Cole Miles,1,* Benjamin Cohen-Stead,2 Owen Bradley,2 Steven Johnston,3,4 Richard Scalettar,2 and Kipton Barros5,†

1Department of Physics, Cornell University, Ithaca, New York 14850, USA
2Department of Physics, University of California, Davis, California 95616, USA
3Department of Physics and Astronomy, The University of Tennessee, Knoxville, Tennessee 37996, USA
4Institute for Advanced Materials and Manufacturing, University of Tennessee, Knoxville, Tennessee 37996, USA
5Theoretical Division and CNLS, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

(Received 13 January 2022; accepted 24 March 2022; published 18 April 2022)

We present a method to facilitate Monte Carlo simulations in the grand canonical ensemble given a target mean particle number. The method imposes a fictitious dynamics on the chemical potential, to be run concurrently with the Monte Carlo sampling of the physical system. Corrections to the chemical potential are made according to time-averaged estimates of the mean and variance of the particle number, with the latter being proportional to thermodynamic compressibility. We perform a variety of tests, and in all cases find rapid convergence of the chemical potential—inexactness of the tuning algorithm contributes only a minor part of the total measurement error for realistic simulations.

DOI: 10.1103/PhysRevE.105.045311

I. INTRODUCTION

A fundamental attribute of statistical mechanics is the equivalence of thermodynamic ensembles in the limit of large system size. In particular, the canonical ensemble, with fixed particle number, should be equivalent to a grand canonical ensemble in which the chemical potential \( \mu \) is suitably selected to fix the average particle number. However, there may be practical reasons to prefer working in the grand canonical ensemble, particularly in the context of Monte Carlo (MC) simulations. In classical MC simulations, for example, moves that modify the particle number can be useful for reducing decorrelation times or for studying coexistence between phases [1–5]. Similarly, the starting point of many [6–12] finite temperature quantum Monte Carlo (QMC) simulations is the grand canonical partition function \( Z = \text{Tr} \exp(-\beta (H + \mu N)) \), where \( \beta \) is the inverse temperature, \( H \) is the Hamiltonian, and \( N \) is the number operator. The trace above runs over all quantum wave functions, not just those constrained to a fixed particle number.

In MC and QMC simulations, we often wish to specify the average particle number \( \langle N \rangle \) directly, e.g., to fix the electron filling fraction. Determining the \( \mu \) value which satisfies this condition has traditionally required a tedious manual search, with additional searches necessary after every update to the model parameters.

Here we present a method to efficiently converge the chemical potential \( \mu_t \) to a solution value that produces the desired mean particle number within the same MC or QMC simulation where measurements are performed. We take the chemical potential \( \mu_t \) to be continually evolving in sampling time \( t \). Corrections to \( \mu_t \) are performed whenever new measurements \( N_t / \langle N \rangle \) are collected.

A central challenge is that, in certain cases, it can be difficult to collect good statistical samples for systems with long autocorrelation times. That is, modifications to \( \mu_t \) may not fully impact the samples \( N_t \) until quite some time later. We address this problem by employing increasingly long-time averages, which incorporate a fixed fraction of the entire history of MC data.

Roughly speaking, our proposed strategy for tuning \( \mu_t \) is as follows: Given continually improving approximate measurements of the particle number \( \langle N \rangle \) and compressibility \( \kappa = d \langle N \rangle / d \mu \), we update the value of the chemical potential under the assumption of linear response. Our method shares some conceptual similarities with proportional-integral-derivative (PID) controllers, which have previously been applied to \( \mu \) tuning [17,18]. A disadvantage of PID controllers is that they introduce several parameters that must be carefully selected for each new problem. In contrast, the method we introduce here is simple and works robustly across a wide range of problems using a single default set of algorithm parameters.

We benchmark the new method on two problems: (1) Tuning the applied field in the classical two-dimensional ferromagnetic Ising model to achieve a target magnetization, and (2) tuning the chemical potential in simulations of the quantum Holstein model to achieve a target electron filling fraction. For the Ising model, simulation temperatures approaching \( T_c \) give rise to long autocorrelation times, which can make it difficult to achieve good statistical sampling. For QMC simulations of the Holstein model, phonons mediate an effective attractive electron-electron interaction, which, in turn, gives rise to challenging metastability effects.

*cmms572@cornell.edu
†kbarros@lanl.gov
The benchmarks indicate that accurate measurements of a system with a specified mean particle number can be acquired from a single MC simulation run, with \( \mu \) tuning enabled throughout. This works because \( \mu \), mostly converges within the burn-in period of the MC simulation. After burn-in there remains a small (and steadily decreasing) error in \( \mu \), but it does not seem to contribute significantly to the overall statistical error in measurements.

II. METHOD

We present a method to tune any thermodynamic field according to its conjugate observable. For concreteness, consider the task of tuning the chemical potential \( \mu \) to produce a target mean particle number, \( \langle N \rangle = N^* \). The same method could be applied to a magnetic system, in which case one would replace \( \mu \) with the applied magnetic field and \( \langle N \rangle \) with the total magnetization (an example is considered in Sec. IIIA).

At fixed \( \mu \), one can estimate the mean particle number \( \langle N \rangle \) using MC sampling. The compressibility \( \kappa \) will play an essential role in our tuning scheme. A fundamental result from thermodynamics states

\[
\kappa = \frac{d \langle N \rangle}{d \mu} = \beta \text{Var}[N],
\]

where \( \text{Var}[N] = \langle N^2 \rangle - \langle N \rangle^2 \). Thus, \( \kappa \) can be estimated using the observed variance of \( N \).

A. Prior work with iterated simulation

Previous work proposed the following \( \mu \)-tuning strategy [17]: At fixed \( \mu \), run MC over some time window to collect statistical estimates \( \bar{N} \) and \( \bar{\kappa} \) of the mean particle number and compressibility. To find the chemical potential that will approximately achieve a target particle number \( N^* \), solve for the \( \mu' \) value that satisfies \( \bar{\kappa} = (N^* - \bar{N})/(\mu' - \mu) \). Assign \( \mu \leftarrow \mu' \) and repeat.

A practical challenge with this iterated update scheme is that it may be difficult to acquire sufficiently good estimates \( \bar{N} \) and \( \bar{\kappa} \). It is hard to know \textit{a priori} how much sampling time should be devoted to any particular \( \mu \) value. Statistical estimates of \( \kappa \), associated with fluctuations in \( N \), are particularly error prone. Also, if many iterative updates to \( \mu \) are required, it would seem advantageous to incorporate information from all previous MC runs, not just from the most recent sampling window.

B. Dynamical \( \mu \) tuning

In contrast to the iterated simulation scheme, here we explore an approach where \( \mu_t \) evolves dynamically in the context of a single simulation. At each iteration \( t \), a MC update step or sweep is performed using the instantaneous value \( \mu_t \) of the chemical potential. Next, the chemical potential is updated using the rule

\[
\mu_{t+1} = \frac{\bar{N}_t + (N^* - \bar{N}_t)/\bar{\kappa}_t}{\bar{\kappa}_t},
\]

We use the notation \( \langle \cdot \rangle_t \) to signify an appropriate time average over a subset of the sampled data up to time \( t \). Note that the effective window size is continually increasing with sampling time. Many types of time-averaging are possible; for simplicity, we average over the most recent half of all collected data, weighting each sample equally. Averages of \( \mu \) and \( N \) up to time \( t \) are defined as

\[
\bar{\mu}_t = \frac{1}{L_t} \sum_{t'=\lceil t/2 \rceil}^t \mu_{t'}, \quad \bar{N}_t = \frac{1}{L_t} \sum_{t'=\lceil t/2 \rceil}^t N_{t'}. \tag{3}
\]

The ceiling function \( \lceil \cdot \rceil \) rounds up to the nearest integer and \( L_t = t - \lceil t/2 \rceil + 1 \) is the number of samples in the average. We select this form for the running-time averages partly for simplicity, and partly because it allows updates in constant time, as described in Appendix.

An important aspect of Eq. (2) is that it defines \( \mu_{t+1} \) as a correction to the time-averaged chemical potential \( \bar{\mu}_t \), and not as a correction to the previous instantaneous chemical potential \( \mu_t \). In this way, the estimator \( \mu_t \) captures important information from the entire sampling history, and evolves on the same timescale as \( \bar{N}_t \).

1. Estimating compressibility

The success of Eq. (2) depends crucially on the definition of the time-averaged compressibility \( \bar{\kappa}_t \). Equation (1) suggests that we can estimate \( \kappa \) using the time-averaged variance,

\[
\bar{\kappa}_t = \beta \text{Var}[N]. \tag{5}
\]

For classical systems, \( \text{Var}_t[N] \) is defined as the sample variance for the data \( \{N_{t_1}, \ldots, N_{t_n}\} \). For quantum systems, a slightly modified definition is given in Eq. (17). Generally, the fluctuation-based estimator \( \kappa_t^{\text{fluc}} \) becomes valid at late times, once \( \mu_t \) settles to a near constant value. At early times, however, \( \mu_t \) is evolving rapidly, and the estimator \( \kappa_t^{\text{fluc}} \) is error prone. To ensure that the update rule of Eq. (2) is reasonable at all times, we impose carefully defined lower and upper bounds for our final compressibility estimator,

\[
\bar{\kappa}_t = \max \left[ \kappa_t^{\text{min}}, \min (\kappa_t^{\text{max}}, \kappa_t^{\text{fluc}}) \right]. \tag{6}
\]

This ordering of the max and min operators ensures that \( \bar{\kappa}_t \) never vanishes.

Imposing the lower bound \( \kappa_t^{\text{min}} \) protects against the case where early-time fluctuations of \( N_t \) are artificially small. This could happen in a QMC simulation, for example, if the initial guess \( \mu_{t=0} \) is in one of the system’s band gaps (\( N_t \) associated with fully occupied bands), leading to a divergence in \( 1/\kappa_t^{\text{fluc}} \). Since the error in statistical observables decays like the inverse square root of the number of samples, a reasonable lower bound on \( \bar{\kappa}_t \) is given by

\[
\kappa_t^{\text{min}} = \frac{\alpha}{\sqrt{t + 1}}, \tag{7}
\]

for some appropriately defined \( \alpha \). Referring to Eq. (1), we see that \( \kappa \) should scale like the system volume \( V \) divided by an intensive energy. For the benchmarks in Sec. III, we will select \( \alpha = N_{\text{sites}}/u_0 \) where \( N_{\text{sites}} \) is the number of lattice sites, and \( u_0 \) is a characteristic energy scale. Our tests indicate that algorithm performance is largely insensitive to the precise choice of \( \alpha \). By design, \( \kappa_t^{\text{min}} \) decays to zero at large times \( t \).
Imposing the upper bound $\kappa_{t}^{\text{max}}$ protects against the case where early-time fluctuations of $N$ are artificially large. Although Eq. (5) is correct in thermodynamic equilibrium, it produces a very poor estimator in the out-of-equilibrium context of a dynamically evolving $\mu_t$. Changes to $\mu_t$ will, by design, generate a strong response in $N_t$. Drift in $N_t$ will cause a large overestimate of the compressibility, $\kappa_{t}^{\text{max}} \sim \text{Var}[N] \sim V^2$. Recall that physical compressibility must scale like system size, $\kappa \sim V$, which is apparent from the definition $\kappa = d \langle N(t) \rangle / d \mu$. To get a compressibility estimator with the correct scaling, we can compare the typical variations in $N_t$ and $\mu_t$. Specifically, we define the upper bound on $\kappa$, to be

$$\kappa_{t}^{\text{max}} = \sqrt{\text{Var}[N]/\text{Var}[\mu]},$$

where $\text{Var}[\mu]$ is the sample variance for $\{\mu_t/2, \ldots, \mu_t\}$. At early times, when both $\mu_t$ and $N_t$ are varying significantly, it is assured that $\kappa_{t}^{\text{max}} \sim N \sim V$. At late times, $\mu_t$ should settle to the target chemical potential, whereas $N_t$ will continue to exhibit equilibrium fluctuations. Then $\kappa_{t}^{\text{max}}$ grows very large, and we expect to recover the fluctuation-based estimator, $\kappa_t = \kappa_{t}^{\text{fluc}}$.

### 2. Method summary

Pseudocode for the full $\mu$-tuning algorithm is listed in Algorithm 1. The mean and variance estimators can be updated in constant time using the methods described in the Appendix. The user must provide an initial guess $\mu_{t=0}$ for the chemical potential. Also required is a parameter $\alpha$ that sets an appropriate scale for the compressibility.

We make two final remarks regarding the algorithm. Note, first, that convergence, $\mu_{t+1} = \Pi_t$, implies that the target condition is satisfied, $N_t = N^*$. Second, the dynamical update rules are inherently self-stabilizing. Suppose that changes to $\mu_t$ are having no major effect on $N_t$. Then the sample variance $\text{Var}[N]$ will decrease, leading to smaller $\kappa_t$. This, in turn, will drive larger updates to the chemical potential. Eventually the magnitude of these updates will be enough to produce the necessary changes in $N_t$. Conversely, if changes to $\mu_t$ are having too large of an effect on $N_t$, then compressibility estimator $\kappa_t$ will also grow large, and this will dampen the magnitude of updates to $\mu_{t+1}$. These self-stabilizing mechanisms share conceptual similarities to those in standard PID controllers [18,19], though all parameters in our algorithm are physically motivated and work robustly across a range of systems.

### III. RESULTS

#### A. Ising model

We begin by demonstrating our method on a well-understood test case, the ferromagnetic Ising model on a two-dimensional square lattice. The Hamiltonian is

$$H = -J \sum_{\langle ij \rangle} s_i s_j - B \sum_i s_i,$$

where the sum over $\langle ij \rangle$ is over all nearest neighbor sites, and $s_i = \pm 1$. For simplicity, all energies will be measured in units of $J = 1$ and we likewise set the Boltzmann constant $k_B = 1$. The tunable quantity here is the magnetic field $B$, which couples linearly to the total magnetization $M = \sum_i s_i$. In this case, the magnetic susceptibility $\chi = d \langle M \rangle / dB = \beta(\langle M^2 \rangle - \langle M \rangle^2)$ plays the role of the compressibility $\kappa$. Algorithm 1 applies upon making the following substitutions:

$$\mu \mapsto B, \quad N \mapsto M, \quad \kappa \mapsto \chi.$$

It will also be convenient to refer to the magnetization per site, $m = M/N_{\text{sites}}$.

In zero field, this system undergoes a second-order phase transition between the paramagnetic and ferromagnetic phases at a critical temperature $T_c \approx 2.27$ [20]. Approaching $T_c$ from above causes both the susceptibility $\chi$ and the autocorrelation time to diverge. The distance $T - T_c$ from the critical point offers an excellent means to scale the “tuning difficulty.”

To begin, we investigate an example tuning run on a $100 \times 100$ Ising system at a temperature of $T = 2.5$, starting with a uniformly random initial state. By symmetry, zero magnetization is achieved at zero field. To make the tuning task more interesting, we aim to find the magnetic field $B = B^*$ that produces a nonzero target magnetization per site of $m^* = 1/2$. Despite the presence of a small but finite field $B^* > 0$, we still observe very large autocorrelations times when $T$ approaches $T_c$. To explore this effective critical slowing down, we opt to use the standard single spin-flip Metropolis-Hastings algorithm [21], though more advanced cluster updates [22,23] would also be compatible with the tuning algorithm. The time index $t$ is incremented once per MC sweep, at which point a new measurement of the total magnetization $M_t$ is taken, and the field $B_{t+1}$ is computed according to Algorithm 1 under the substitutions of Eq. (10). The initial field is $B_{t=0} = 0$. To set a scale for $\kappa_{t}^{\text{min}}$, we select $\alpha = N_{\text{sites}}/J$.

In Fig. 1, we plot the result of dynamically tuning the field $B_t$. Figures 1(a) and 1(b) show the time evolution of $B_t$ and $m_t$, and their time averages. Figure 1(c) shows estimators for the susceptibility $\chi$ (compressibility $\kappa$ in lattice gas language). The dynamics undergoes three different regimes, corresponding to the three branches in the expression

Algorithm 1 Chemical potential tuning. In the language of magnetism, we would make the substitutions of Eq. (10).

| Input: | Target particle number $N^*$ |
|-------|-----------------------------|
| Input: | Initial guess $\mu_{t=0}$ for the chemical potential |
| Input: | Characteristic compressibility scale $\alpha$ |
| for $i = 0, 1, \ldots$ do |
| Perform MC sampling with chemical potential $\mu$ |
| Collect samples for $\langle N \rangle$ and (in the QMC context) for $\langle N^2 \rangle$ |
| Update time averages $\overline{\Pi}_t$ and $\overline{N}_t$ |
| Update variance estimators $\text{Var}[\mu]$ and $\text{Var}[N]$ |
| Calculate $\kappa$ estimate $\kappa_{t}^{\text{est}} = \beta \overline{\text{Var}}[N]$ |
| Calculate lower bound $\kappa_{t}^{\text{min}} = \alpha/(t + 1)^{1/2}$ |
| Calculate upper bound $\kappa_{t}^{\text{max}} = \sqrt{\text{Var}[N]/\text{Var}[\mu]}$ |
| Calculate bounded $\kappa$ estimate $\overline{\kappa}_t = \max(\kappa_{t}^{\text{max}}, \min(\kappa_{t}^{\text{min}}, \kappa_{t}^{\text{est}}))$ |
| Update $\mu_{t+1} = \overline{\Pi}_t + (N^* - \overline{N}_t)/\overline{\kappa}_t$ |
| end |
FIG. 1. Tuning the magnetic field $B$ of the classical Ising model to achieve a target magnetization per site, $m^* = 1/2$. We employ a 100 × 100 square lattice, and temperature $T = 2.5$. (a) The dynamically evolving field $B_t$ and its running average $\overline{B}_t$ both eventually converge to $B^* \approx 3.096(2) \times 10^{-2}$. (b) Similar plots for $m_t = M_t/N_{\text{sites}}$ and its running average $\overline{m}_t$; significant equilibrium fluctuations in $m_t$ are observed. (c) The susceptibility $\chi$ (analogous to $\kappa$) is reasonably estimated by $\overline{\chi}$, with the lower and upper bounds, $\chi_{\text{min}}$ and $\chi_{\text{max}}$, playing important roles at early times. At long times the susceptibility converges to 7.76(4) × 10^4. Note that panel (c) is a log-log plot and extends over longer times than panels (a) and (b).

$\chi_t = \max[\chi_{t,\text{min}}, \min(\chi_{t,\text{max}}, \chi_{t,\text{fluc}})]$. For the first few MC sweeps ($t \leq 3$) there is essentially no good susceptibility data. Here, the lower bound $\chi_{t,\text{min}}$ of Eq. (7) controls the estimator $\overline{\chi}$, and prevents the algorithm from making overly large corrections to $B_t$. In the intermediate time regime of $3 < t \lesssim 10^3$, the field $B_t$ is evolving significantly, and driving large changes to $m_t$. Here, the upper bound $\chi_{t,\text{max}}$ of Eq. (8) controls the estimator $\overline{\chi}$, and correctly captures the approximate sensitivity of magnetism to changes in the applied field. By the end of this regime, fluctuations in $B_t$ decrease significantly. Finally, at times $t \gtrsim 10^3$, the fluctuation-based estimator $\chi_{t,\text{fluc}}$ becomes accurate, and $B_t$ converges precisely toward the solution $B^* \approx 3.096(2) \times 10^{-2}$.

In Fig. 2, we show the evolution of $B_t$ for the same Ising model, but now over a variety of temperatures. At high $T$, the tuner converges to the correct value $B^*$ very quickly. At temperatures approaching $T_c \approx 2.27$, both the autocorrelation time and the magnetic susceptibility grow large, leading to difficulty in collecting accurate statistics, and very high sensitivity to small changes in the applied field. Despite the slower convergence near $T_c$, the tuner appears to be working correctly in all cases tested.

Figure 3 shows the error of the instantaneous field $|B_t - B^*|$ and the running-mean magnetization $|\overline{m}_t - m^*|$, throughout the tuning process. Here we take $T = 2.5$ and $m^* = 1/2$ as in Fig. 1, and allow system size to vary. Our best estimates for $B^*$ are obtained by averaging over ten independent runs, extending to $t = 5 \times 10^6$ Monte Carlo sweeps. The results are $B^* = 3.113(3) \times 10^{-2}$ and $B^* = 3.096(1) \times 10^{-2}$, for lattice sizes $N_{\text{sites}} = 25^2$ and $100^2$, respectively. The system size dependence is relatively minor, given our choice of $T = 2.5$. At temperatures nearer to $T_c \approx 2.27$, the $B^*$ would become smaller, and their relative variation would depend more strongly on system size.

The early-time dynamics (up to $\sim 10^2$ sweeps) of both errors are seen to be system-size independent due to the dynamics being dominantly controlled by our $\chi_{\text{min}}$ and $\chi_{\text{max}}$ bounding scheme. Once $\chi_{\text{fluc}}$ gains control, we can see a separation emerge as larger systems enjoy improved statistics from increased self-averaging, resulting in faster tuning. At these large times, the errors decay as $t^{-3/2}$, and the limiting factor in tuning becomes the statistical error in the Monte Carlo estimates. In this regime, note that the error in the average magnetization, $|\overline{m}_t - m^*|$, is roughly independent of whether $B_t$ is being tuned dynamically, or fixed to the correct value $B = B^*$ throughout the simulation [shown as the red curve in Fig. 3(b)].

**B. Holstein model**

We now demonstrate our method in a QMC setting. As a test case we consider the Holstein model, one of the simplest models describing interactions between electrons and
where $B$ describes hopping between nearest-neighbor sites sizes, with $T$ function $Z$ of methods begins by expressing the many-body partition phonons on a lattice [24]. The Hamiltonian is are estimated, per system size, by averaging DYNAMICAL TUNING OF THE CHEMICAL POTENTIAL TO … PHYSICAL REVIEW E and the trace runs over the entire Fock space.

where $\beta$ is understood to mean $\langle \hat{O} \rangle = Z^{-1} \text{Tr} e^{-\beta H} \hat{O}$, (12)

and the trace runs over the entire Fock space.

1. Overview of quantum Monte Carlo

There are many forms of QMC. One widely used set of methods begins by expressing the many-body partition function $Z = \text{Tr} e^{-\beta H}$ as a path integral involving fields that fluctuate in imaginary time. The aim is then to perform ordinary Monte Carlo sampling of these fluctuating variables according to some appropriate probability distribution. In the determinant-QMC approach [6], which is our focus here, one uses a Suzuki-Trotter expansion inside the trace

$$Z = \text{Tr} e^{-\beta \hat{H}} = \text{det} M = \text{det} \left[ \sum_{\tau} e^{-\Delta \tau \hat{H}} \right] + O(\Delta \tau^2),$$

with a carefully selected decomposition $\hat{H} = \hat{H}_0 + \hat{H}_1$. Inserting a complete set of states at each discrete imaginary time slice, $0 \leq \tau < \beta$, introduces an effectively classical field $x_{\tau,i}$ which, when sampled, allows one to estimate observables $\langle \hat{O} \rangle$. Each sample $x_{\tau,i}$ is typically weighted according to a fermionic determinant, $P[x] \propto \left| \det M \right|$, for an appropriate matrix function $M_x[x]$.

When this procedure is applied to the Holstein model, the field $x_{\tau,i}$ can be interpreted as “imaginary time fluctuations” of the phonons. An analogous formalism is used for sampling the gluon field in lattice quantum chromodynamics, and we can borrow techniques from that community. In particular, Langevin [7] and hybrid Monte Carlo (HMC) sampling [26,27] have both proven effective for simulating electron-phonon models [28,29], and make it possible to update the entire field $x_{\tau,i}$ at a cost that scales near linearly with system size. Here we employ HMC. A complete account of our QMC methodology is presented in [30].

2. Chemical potential tuning for quantum models

Algorithm 1 remains valid in the QMC context provided that we are careful in estimating the compressibility $\kappa$. The thermodynamic relationship of Eq. (1) continues to hold,

$$\kappa = \frac{d \langle \hat{N} \rangle}{d \mu} = \beta \text{Var}[\hat{N}],$$

where $\text{Var}[\hat{N}] = \langle \hat{N}^2 \rangle - \langle \hat{N} \rangle^2$, and each expectation value on the right-hand side is to be interpreted in the sense of Eq. (12). An interesting feature of QMC, however, is that unbiased
samples $N_t$ of $\langle \hat{N} \rangle$ do not generally contain sufficient information to estimate the variance of $\hat{N}$ due to neglecting within-sample fluctuations [31]. We can still define a fluctuation-based estimator in the form of Eq. (5),

$$k_t^{\text{fluc}} = \beta \text{Var}_t[\hat{N}],$$

but now we must use

$$\text{Var}_t[\hat{N}] = N_t^{(2)} - (\bar{N}_t)^2,$$

where $N_t^{(2)}$ denotes a statistical sample of the expectation value $\langle \hat{N}^2 \rangle$ defined in Eq. (12). Time averages have the same form as in Eqs. (3) and (4), but now we must also track

$$\bar{N}_t^{(2)} = \frac{1}{L} \sum_{i=1/2}^{i+L/2} N_t^{(2)}.$$

The quantity $k_t^{\text{max}}$ remains as in Eq. (8), but using the sample variance of Eq. (17). With these refinements to $k_t^{\text{fluc}}$ and $k_t^{\text{max}}$, we can directly apply Algorithm 1 to tune the chemical potential.

### 3. Single-site limit

As an initial demonstration, we examine the behavior of Algorithm 1 in the single-site limit of the Holstein model ($t_s = 0$). This model admits an analytic solution by the Lang-Firsov transformation [32,33], but nonetheless serves as a challenging test case for $\mu$ tuning. The effective action $S[x]$ resulting from Eq. (11) possesses two deep local minima (associated with electron number 0 or 2) due to an effective electron-electron attraction that is mediated by the phonons. MC sampling of the phonon field $x_t$ is characterized by long periods of trapping within one minima, punctuated by rare hops across the barrier into the other. Strong hysteresis creates a challenge for $\mu$ tuning; adjusting the chemical potential $\mu$, may affect measurements of electron number $N_t$ only after a very large amount of simulation time, $t' \gg t$.

We consider a model with $\omega = 1$, $\lambda = \sqrt{2}$ at $\beta = 4$. These values are commonly used in studies of the charge-density-wave transition in the Holstein model, which occurs at $\beta_c \sim 6$ when $t_s = 1$ [34]. We aim for a target electron number of $\langle \hat{N} \rangle = 1$, i.e., half-filling. This corresponds to finding the chemical potential at which the average occupancies of the two metastable states are equal. In the single-site limit at half-filling, the exact chemical potential and compressibility are $\mu^* = -\lambda^2 / \omega^2$ and $\kappa^* = \beta / \sqrt{1 + \beta^2 \mu^*}$.

Figure 4 illustrates a representative $\mu$-tuning run. Each increment in time $t$ corresponds to a single HMC trial update. At early times, Fig. 4(b) shows sharp transitions ("jumps") in the measurements $N_t$, which are largely driven by changes in $\mu_t$. These jumps correspond to transitions between the two metastable wells, and in the absence of $\mu$ tuning, would occur on the timescale of $10^5$ HMC trial updates.

Figure 4(c) shows that each of the early-time jumps between metastable wells is accompanied by a large spike in $\bar{N}_t$, which reflects the large change in $N_t$. After each jump, there is a significant period of time where $N_t$ is roughly constant, which causes $\bar{N}_t$ to drop. At times $t \lesssim 100$, the upper bound $\kappa_t^{\text{max}}$ is instrumental in allowing the tuning dynamics to make significant corrections to $\mu$, which drive the density back and forth between metastable wells on an exponentially growing timescale. At around $t \approx 100$, we reach the timescale required for natural (equilibrium) jumps between the two metastable wells. At this point, the algorithm switches over to the fluctuation-based compressibility estimator $\bar{N}_t = k_t^{\text{fluc}}$, as observed in Fig. 4(c). At times $t \gtrsim 100$, the errors in statistical estimators decay like $t^{-1/2}$ in a controlled fashion.

### 4. Full 2D system

Finally, to benchmark our algorithm in a more realistic setting, we consider a square lattice Holstein model with phonon frequency $\omega = 1$ and coupling strength $\lambda = \sqrt{2}$. We consider a square lattice of size $L = 10$, with $N_{\text{sites}} = L^2$ total sites. At half-filling the ground state is characterized by a finite temperature phase transition to charge-density-wave (CDW) order, where the electrons localize onto one of the two sublattices, spontaneously breaking a $Z_2$ symmetry. For our chosen parameters the critical inverse temperature is approximately $\beta \sim 6$. 

![Graphical representation](image-url)
susceptibility and at sufficiently low temperature, the system is expected to CDW phase. When doped sufficiently away from half-filling, DYNAMICAL TUNING OF THE CHEMICAL POTENTIAL TO … PHYSICAL REVIEW E 105

We measure the CDW order using the staggered charge susceptibility

\[ \chi_{\text{CDW}} = \int_0^\beta \sum_r (-1)^{r_+ r} C(r, \tau) d\tau, \quad (19) \]

defined in terms of the real-space density-density correlation function

\[ C(r, \tau) = \frac{1}{N} \sum_i \langle \hat{n}_{i+r} \hat{n}_i(0) \rangle, \quad (20) \]

where \( \hat{n}_i(\tau) = \hat{n}_{i,\uparrow}(\tau) + \hat{n}_{i,\downarrow}(\tau) \) denotes the total electron number on site \( i \) at imaginary time \( \tau \). A signature for superconducting order is given by the pair susceptibility

\[ P_\sigma = \frac{1}{N_{\text{sites}}} \int_0^\beta \langle \hat{\Delta}(\tau) \hat{\Delta}^\dagger(0) \rangle d\tau, \quad (21) \]

where \( \hat{\Delta}(\tau) = \sum_i \hat{c}_{i,\uparrow}(\tau) \hat{c}_{i,\downarrow}(\tau) \).

In Fig. 5 we compare two sets of simulation results, one where \( \mu \) is held fixed, and the other where we tune \( \mu \) to a target density that was measured in the first set of results. All simulations employed \( t = 5 \times 10^3 \) thermalization HMC steps. Following this, we performed \( 5 \times 10^4 \) steps, with measurements taken at every step. For the dynamical \( \mu \) simulations we initialized the chemical potential to \( \mu_{\tau=0} = 0 \) and set \( \alpha = N_{\text{sites}}/\omega \). The chemical potential was continually updated throughout the simulation, but had largely converged already by the end of the thermalization process. There is very good agreement between the two sets of data; the error bars with \( \mu \) tuning enabled are not discernibly larger than with \( \mu \) fixed to its target value.

Figure 5(a) shows the density as a function of the chemical potential. The plateau at half-filling (\( \mu = -2 \)) illustrates the gapped CDW phase. The relatively large horizontal error bars in the tuned value of \( \mu \) near half-filling are associated with a vanishing compressibility \( \kappa \), Fig. 5(b). In other words, a fairly wide range of chemical potentials give rise to (approximately) half-filling. Observe that the plateau in \( \langle n \rangle = \langle N \rangle / N_{\text{sites}} \) corresponds to a strong enhancement of \( \chi_{\text{CDW}} \) in Fig. 5(c) and a suppression of \( P_\sigma \) in Fig. 5(d). With enough doping, at approximately \( \mu = -2.3 \) and \( \mu = -1.7 \), the marker \( \chi_{\text{CDW}} \) for CDW order rapidly vanishes. Simultaneously, the density rapidly shifts away from half-filling, as reflected by the two peaks in the compressibility \( \kappa \). We emphasize that the \( \mu \)-tuning algorithm performs well throughout the diverse range of behaviors exhibited in this model.

IV. DISCUSSION

The dynamical \( \mu \)-tuning algorithm presented in Algorithm 1 enables simulation in the grand canonical ensemble while targeting a fixed mean particle number. The algorithm is straightforward to implement, and imposes negligible computational overhead. Note that all running time averages can be updated in constant time using the formulas of Appendix.

Under the proposed scheme, the chemical potential \( \mu \) is adjusted concurrently with the MC sampling. Although we do not provide formal convergence guarantees, the method works well in practice. For long-running simulations, most statistical samples will be collected after \( \mu \) has approximately converged to its target value, and our benchmarks show that errors are well controlled.

Models with long autocorrelation times present a practical challenge, in that it becomes difficult to assess the impact of a modified \( \mu \) value on the resulting mean particle number \( \langle N \rangle \). Our solution is to effectively collect statistics over increasingly large time windows such that, eventually, both \( \langle N \rangle \) and its sensitivity \( \kappa = d \langle N \rangle / d \mu \) can be accurately measured. We demonstrated that our scheme works well even in very challenging cases, such as the Ising model approaching criticality, and the single-site Holstein model with strong metastability due to phonon-mediated electron binding and associated large energy barriers. The method also works well for larger-size quantum Monte Carlo simulations of the Holstein model on the square lattice, including at filling fractions coinciding with a charge density wave gap, where \( \kappa \) approximately vanishes.

Many variations of the \( \mu \)-tuning algorithm could be studied. For example, one could modify the definition of the running time averages in Eqs. (3) and (4) to “smoothly forget” past data, with the goal of reducing underdamped oscilla-
tions in the early-time dynamics of the tuned field, e.g., in Figs. 1(a) and 4(a). One might also explore whether ideas for accelerating fixed point solvers (e.g., Anderson mixing) could somehow be incorporated into Eq. (2), which updates \( \mu \) as a correction to the time average \( \overline{\mu} \). In our preliminary tests, however, we could not find any modifications to the algorithm that significantly improved accuracy over large simulation times. Indeed, Figs. 3 and 5 suggest that the \( \mu \)-tuning algorithm is already close to optimal; errors in long-time statistical measurements are observed to be about the same, whether \( \mu \) is dynamically tuned or statically fixed to the exactly correct value.

Code and an interactive interface to explore tuning for the Ising system is available at [37]. Holstein simulations, with \( \mu \) tuning built in, are available at [38]. A general purpose implementation of the algorithm is provided by the Julia package MuTuner.jl [39].

ACKNOWLEDGMENTS

K.B. acknowledges support from the Center of Materials Theory as a part of the Computational Materials Science (CMS) program, funded by the U.S. Department of Energy, Office of Basic Energy Sciences. S.J., O.B., and R.T.S. acknowledge support from the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award No. DE-SC0022311. C.M. acknowledges support by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Graduate Fellowship under Award No. DE-SC0020347. B.C.-S. was funded by a U.C. National Laboratory In-Residence Graduate Fellowship through the U.C. National Laboratory Fees Research Program.

This article was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

The authors declare no conflicts of interest.

APPENDIX: UPDATING RUNNING TIME AVERAGES

The \( \mu \)-tuning algorithm requires one to maintain running averages of the form

\[
\overline{x}_t = \frac{1}{L_t} \sum_{r=\lceil ct \rceil}^{t} x_r, \tag{A1}
\]

After a somewhat lengthy derivation, one finds the recursion relation,

\[
M_{t+1} = \begin{cases} 
M_t + (x_{t+1} - \overline{x}_t)(x_{t+1} - \overline{x}_{t+1}) & \text{case A} \\
M_t + (x_{t+1} - x_{\lceil ct \rceil})(x_{t+1} - \overline{x}_{t+1} + x_{\lceil ct \rceil} - \overline{x}_t) & \text{case B},
\end{cases} \tag{A7}
\]

which is numerically stable and easy to implement given that we are already maintaining the running average \( \overline{x}_t \). Note that cases A and B coincide when \( x_{\lceil ct \rceil} = \overline{x}_t \).

where \( L_t = t - \lceil ct \rceil + 1 \) is the count of samples in the average, and \( \lceil \cdot \rceil \) denotes the ceiling function. We selected \( c = 1/2 \) for our study, but other values \( 0 < c < 1 \) are possible. The \( x_t \) data could be one of the following: the instantaneous chemical potential \( \mu_t \), a sample \( N_t \) for the particle number, or (in the QMC context) a sample \( N_t^{(2)} \) for the particle number squared.

After each MC time step, we wish to update the running average from \( \overline{x}_t \) to \( \overline{x}_{t+1} \). It is helpful to distinguish between two cases:

\[
\begin{align*}
\text{case A:} & \quad \lceil ct + 1 \rceil = \lceil ct \rceil, \\
\text{case B:} & \quad \lceil ct + 1 \rceil = \lceil ct \rceil + 1.
\end{align*}
\]

These are the only two possibilities given our assumptions that \( 0 < c < 1 \) and \( t \) is integer.

In case A we must add the new datapoint \( x_{t+1} \) to the running average. In case B we must additionally remove the datapoint \( x_{\lceil ct \rceil} \) from the running average. The update rule is then

\[
\overline{x}_{t+1} = \begin{cases} 
(L_t \overline{x}_t + x_{t+1})/L_{t+1} & \text{case A} \\
(L_t \overline{x}_t + x_{t+1} - x_{\lceil ct \rceil})/L_{t+1} & \text{case B}.
\end{cases}
\] (A2)

In case A the number of datapoints increases by one, \( L_{t+1} = L_t + 1 \), whereas in case B, \( L_{t+1} = L_t \). We can therefore rearrange as

\[
\overline{x}_{t+1} = \begin{cases} 
\overline{x}_t + (x_{t+1} - \overline{x}_t)/L_t & \text{case A} \\
\overline{x}_t + (x_{t+1} - x_{\lceil ct \rceil})/L_t & \text{case B},
\end{cases}
\] (A3)

which improves numerical accuracy.

We are also interested in keeping a running estimate of the sample variance,

\[
\overline{\text{var}}_t[x] = \overline{x}_t^2 - (\overline{x}_t)^2, \] (A4)

or, for quantum observables, the closely related Eq. (17). The formula suggests that we maintain a running average \( \overline{x}_t^2 \), through which the variance follows immediately. To estimate \( \text{Var}[\hat{N}] \) in QMC simulations, such a strategy may be necessary. Whenever possible, however, direct numerical implementation of Eq. (A4) should be avoided due to potentially large floating point round-off error. A much improved algorithm was proposed by Welford [40], which we here adapt.

The sample variance can be equivalently written,

\[
\overline{\text{var}}_t[x] = M_t/L_t, \] (A5)

where

\[
M_t = \sum_{r=\lceil ct \rceil}^{t} (x_r - \overline{x}_t)^2. \] (A6)
[1] J. Yao, R. A. Greenkorn, and K. C. Chao, Monte Carlo simulation of the grand canonical ensemble, Mol. Phys. 46, 587 (1982).
[2] F. A. Escobedo and J. J. de Pablo, Expanded grand canonical and Gibbs ensemble Monte Carlo simulation of polymers, J. Chem. Phys. 105, 4391 (1996).
[3] P. Kowalczyk, H. Tanaka, R. Holyst, K. Kaneko, T. Ohmori, and J. Miyamoto, Storage of hydrogen at 303 K in graphite slitlike pores from grand canonical Monte Carlo simulation, J. Phys. Chem. B 109, 17174 (2005).
[4] M. Clark, F. Guarnieri, I. Shkurko, and J. Wiseman, Grand canonical Monte Carlo simulation of ligand-protein binding, J. Chem. Inf. Model. 46, 231 (2006).
[5] H. Eslami and F. Müller-Plathe, Molecular dynamics simulation in the grand canonical ensemble, J. Comput. Chem. 28, 1763 (2007).
[6] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, Monte Carlo calculations of coupled boson-fermion systems, I, Phys. Rev. D 24, 2278 (1981).
[7] G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Langevin simulations of lattice field theories, Phys. Rev. D 32, 2736 (1985).
[8] R. T. Scalettar, D. J. Scalapino, R. L. Sugar, and D. Toussaint, Hybrid molecular-dynamics algorithm for the numerical simulation of many-electron systems, Phys. Rev. B 36, 8632 (1987).
[9] S. R. White, D. J. Scalapino, R. L. Sugar, E. Y. Loh, J. E. Gubernatis, and R. T. Scalettar, Numerical study of the two-dimensional Hubbard model, Phys. Rev. B 40, 506 (1989).
[10] N. Kawashima, J. E. Gubernatis, and H. G. Evertz, Loop algorithms for quantum simulations of fermion models on lattices, Phys. Rev. B 50, 136 (1994).
[11] S. Zhang, Finite-Temperature Monte Carlo Calculations for Systems with Fermions, Phys. Rev. Lett. 83, 2777 (1999).
[12] Y.-Y. He, M. Qin, H. Shi, Z.-Y. Lu, and S. Zhang, Finite-temperature auxiliary-field quantum Monte Carlo: Self-consistent constraint and systematic approach to low temperatures, Phys. Rev. B 99, 045108 (2019).
[13] J. E. Hirsch, R. L. Sugar, D. J. Scalapino, and R. Blankenbecler, Monte Carlo simulations of one-dimensional fermion systems, Phys. Rev. B 26, 5033 (1982).
[14] A. W. Sandvik, Finite-size scaling of the ground-state parameters of the two-dimensional Heisenberg model, Phys. Rev. B 56, 11678 (1997).
[15] S. Zhang, J. Carlson, and J. E. Gubernatis, Constrained Path Quantum Monte Carlo Method for Fermion Ground States, Phys. Rev. Lett. 74, 3652 (1995).
[16] S. Zhang, J. Carlson, and J. E. Gubernatis, Constrained path Monte Carlo method for fermion ground states, Phys. Rev. B 55, 7464 (1997).
[17] J. A. Speidel, J. R. Banfielder, and M. Mezei, Automatic control of solvent density in grand canonical ensemble Monte Carlo simulations, J. Chem. Theory Comput. 2, 1429 (2006).
[18] K. G. Kleiner, Implementing a Self-Corrected Chemical Potential Scheme in Determinant Quantum Monte Carlo Simulations, Chancellor’s Honors Project (University of Tennessee, Knoxville, 2019), https://trace.tennessee.edu/utk_chanhonproj/2225.
[19] J. Bechhoefer, Feedback for physicists: A tutorial essay on control, Rev. Mod. Phys. 77, 783 (2005).
[20] L. Onsager, Crystal statistics. I. A two-dimensional model with an order-disorder transition, Phys. Rev. 65, 117 (1944).
[21] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57, 97 (1970).
[22] J.-S. Wang and R. H. Swendsen, Cluster Monte Carlo algorithms, Phys. A (Amsterdam, Neth.) 167, 565 (1990).
[23] J. Kent-Dobias and J. P. Sethna, Cluster representations and the Wolff algorithm in arbitrary external fields, Phys. Rev. E 98, 063306 (2018).
[24] T. Holstein, Studies of polaron motion: Part I. The molecular-crystal model, Ann. Phys. 8, 325 (1959).
[25] S. Johnston, E. A. Nowadnick, Y. F. Kung, B. Moritz, R. T. Scalettar, and T. P. Devereaux, Determinant quantum Monte Carlo study of the two-dimensional single-band Hubbard-Holstein model, Phys. Rev. B 87, 235133 (2013).
[26] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid Monte Carlo, Phys. Lett. B 195, 216 (1987).
[27] R. Neal, MCMC Using Hamiltonian Dynamics, in Handbook of Markov Chain Monte Carlo (Chapman and Hall/London/CRC, Boca Raton, FL, 2011).
[28] G. G. Batrouni and R. T. Scalettar, Langevin simulations of a long-range electron-phonon model, Phys. Rev. B 99, 035114 (2019).
[29] S. Beyl, F. Goth, and F. F. Assaad, Revisiting the hybrid quantum Monte Carlo method for Hubbard and electron-phonon models, Phys. Rev. B 97, 085144 (2018).
[30] B. Cohen-Stead, O. Bradley, C. Miles, G. Batrouni, R. Scalettar, and K. Barros, Fast and scalable quantum Monte Carlo simulations of electron-phonon models, arXiv:2203.01291 (2022).
[31] Classical microstates sampled from the grand canonical equilibrium distribution have a well-defined particle number. Our QMC simulations, in contrast, sample a phonon field that fluctuates in imaginary time, and each phonon configuration is associated with an entire statistical distribution of electron numbers.
[32] I. G. Lang and Y. A. Firsov, Zh. Eksp. Teor. Fiz. 43, 1843 (1962) [Sov. Phys. JETP 16, 1301 (1963)].
[33] G. Mahan, Many-Particle Physics, 3rd ed. (Springer US, Boston, MA, 2000).
[34] C. Feng and R. T. Scalettar, Interplay of flat electronic bands with Holstein phonons, Phys. Rev. B 102, 235152 (2020).
[35] O. Bradley, G. G. Batrouni, and R. T. Scalettar, Superconductivity and charge density wave order in the two-dimensional Holstein model, Phys. Rev. B 103, 235104 (2021).
[36] B. Nosarzewski, E. W. Huang, P. M. Dee, I. Esterlis, B. Moritz, S. A. Kivelson, S. Johnston, and T. P. Devereaux, Superconductivity, charge density waves, and bipolarons in the Holstein model, Phys. Rev. B 103, 235156 (2021).
[37] https://colemiles.github.io/isng-tuner
[38] https://github.com/cohensbw/ElPhDynamics
[39] https://github.com/cohensbw/MuTuner.jl
[40] B. P. Welford, Note on a method for calculating corrected sums of squares and products, Technometrics 4, 419 (1962).