This is the accepted manuscript made available via CHORUS. The article has been published as:

Efficient continuous-time quantum Monte Carlo algorithm for fermionic lattice models
Mauro Iazzi and Matthias Troyer
Phys. Rev. B 91, 241118 — Published 30 June 2015
DOI: 10.1103/PhysRevB.91.241118
Efficient continuous-time quantum Monte Carlo algorithm for fermionic lattice models

Mauro Iazzi and Matthias Troyer

Theoretische Physik, ETH Zurich, 8093 Zurich, Switzerland

Efficient continuous time quantum Monte Carlo (CT-QMC) algorithms that do not suffer from time discretization errors have become the state-of-the-art for most discrete quantum models. They have only been widely used yet for fermionic quantum lattice models, such as the Hubbard model, or other fermionic lattice systems due to a suboptimal scaling of \( O(\beta^3) \) with inverse temperature \( \beta \), compared to the linear scaling of discrete time algorithms. Here we present a CT-QMC algorithms for fermionic lattice systems that matches the scaling of discrete-time methods but is more efficient and free of time discretization errors. This provides an efficient simulation scheme that is free from the systematic errors opening an avenue to more precise studies of large systems at low and zero temperature.

PACS numbers:

Monte Carlo simulations of quantum systems are often performed using an imaginary time path integral formulation \([1]\) to map the partition function of the quantum system to that of an equivalent classical one \([2, 3]\). These imaginary time paths, whose extent corresponds to the inverse temperature \( \beta = 1/k_B T \) are then sampled using Monte Carlo methods. Path integrals are usually formulated on a discrete imaginary time mesh with nonzero time step \( \Delta \tau \) in order to regularize the generally fractal paths. An extrapolation of the measured observables to \( \Delta \tau \to 0 \) is then required to obtain accurate results corresponding to those of the original quantum system.

For discrete quantum lattice models, some quantum Monte Carlo (QMC) algorithms exist that are free from time discretization errors, such as Handscomb’s method for Heisenberg spin models \([4]\) or its generalization, the stochastic series expansion (SSE) algorithm \([5]\). They avoid an explicit introduction of time discretization by working with a Taylor expansion. For these and other discrete quantum lattice models one can also avoid time discretization errors in a path-integral formulation by realizing that the lattice structure already provides a regularization of the path integral.

Over the last two decades a new category of path-integral quantum Monte Carlo algorithms has thus been developed that work directly in the continuous time limit \( \Delta \tau \to 0 \), removing the need for an extrapolation and often significantly speeding up the simulations. The first of these continuous-time quantum Monte Carlo (CT-QMC) algorithms have been for quantum spin systems and boson systems \([6, 7]\). Combined with efficient non-local update algorithms, such as cluster updates \([8]\) or the worm algorithm \([7]\) the gains in efficiency resulting from continuous time schemes are such that the simulation of unfrustrated spins and bosons is now considered a solved problem.

The generalization of CT-QMC to fermionic systems has been less straightforward, but has finally been achieved over the past decade by using time-dependent perturbation theory formulations of continuous time path integrals \([9]\). The first fermionic CT-QMC algorithm for lattice models \([10]\) was followed by a number of algorithms for fermionic quantum impurity problems \([9, 11–13]\). These algorithms have been widely employed as quantum impurity solvers \([9]\) i.e. for simulating an open system embedded in a non-interacting bath. They have replaced discrete time algorithms as the state of the art method by being significantly more efficient, avoiding the need to extrapolate in \( \Delta \tau \), and allowing the simulation of a much wider class of models. In particular, they have revolutionized the solution of the quantum impurity problem arising from self-consistent dynamical mean field (DMFT) theories \([14–16]\) and their cluster extensions \([17]\). They allow the accurate simulation of much larger systems for Hubbard-type problems \([18–20]\) and enable to go beyond density-density interactions by allowing the full Coulomb interaction to be included \([21]\) thus opening the way to realistic materials simulations by multi-orbital DMFT \([22]\).

Despite their enormous success for quantum impurity models, CT-QMC methods are rarely used for fermionic lattice models \([18]\). There, discrete time methods \([23]\) are still the method of choice because of better scaling behavior. Existing fermionic CT-QMC algorithms all scale as \( O(\beta^3 V^2) \) with the inverse temperature \( \beta \) and the lattice size \( V \), since these algorithms require operations to be performed on square matrices with dimension \( O(\beta V) \). For quantum impurity problems, which are described by time-dependent actions after integrating out the bath, also discrete time algorithms have the same scaling \([24]\).

However, for quantum lattice models, discrete time algorithms exist that operate on \( \beta/\Delta \tau \) matrices of dimension \( O(V) \) and the effort thus scales only as \( O(\beta V^3) \). The substantially reduced scaling ensured a significant competitive advantage of the discrete time approach.

Several (unpublished) attempts have been made to develop efficient CT-QMC methods for quantum lattice models. Naive approaches have failed, giving either a worse sign problem or a \( O(\beta V^4) \), erasing the advantage from the better scaling in temperature already for
medium size systems. In this Letter we show how to overcome the issues and present a CT-QMC algorithm that has linear scaling in $\beta$ while retaining the cubic $O(V^3)$ complexity with respect to volume.

While our method is more general, we will – for the sake of simplicity – focus our presentation on the Hubbard model. The Hubbard model is the prototypical example of a strongly correlated fermionic system. It consists of spin-$\frac{1}{2}$ fermionic particles that can hop between neighbouring sites of a lattice and repel via a contact interaction. The full Hamiltonian $H$ is given by a sum of the noninteracting and interacting parts $H = H_0 + H_I$

\[
H_0 = -\sum_{x,y,\sigma} t_{xy} c_{x\sigma}^\dagger c_{y\sigma}
\]

\[
H_I = U \sum_x \hat{n}_x \equiv U \sum_x \left( \hat{n}_x^+ - \frac{1}{2} \right) \left( \hat{n}_x - \frac{1}{2} \right).
\]

Here the creation operators $c_{x\sigma}^\dagger$ introduce a new fermion at site $x$ with spin $\sigma$ and the annihilation operator $c_{x\sigma}$ likewise removes one particle, subject to the canonical anti-commutation relations $\{c_{x\sigma}^\dagger, c_{y\sigma'}\} = \delta_{xy} \delta_{\sigma\sigma'}$. The occupation number for site $x$ is given by the density operator $\hat{n}_x = c_{x\sigma}^\dagger c_{x\sigma}$. The tunneling matrix $t_{xy} = t_{yx}$ gives the energy associated with the hopping from site $y$ to site $x$. When two particles of opposite spin are present in the same site, they repel each other with energy $U$.

To perform the CT-QMC we perform a time-dependent perturbation expansion in $H_I$ obtaining \[9, 25\]

\[
Z = \text{tr} e^{-\beta H} = \text{tr} \left[ e^{-\beta H_0} T e^{-\int_0^\beta d\tau H_I(\tau)} \right] = \sum_{k=0}^\infty \frac{1}{k!} \text{tr} \left[ T e^{-\beta H_0} \int_0^\beta d\tau_1 ... \int_0^\beta d\tau_k \prod_{i=1}^k (-H_I(\tau_i)) \right]
\]

where $H_I(\tau) = e^{\tau H_0} H_I e^{-\tau H_0}$ is the perturbation term $H_I$ in the interaction representation. We then further expand the partition function as

\[
Z = \sum_k \int_0^\beta d\tau_1 ... \int_{\tau_{k-1}}^\beta d\tau_k \sum_{x_1,...,x_k} w(c)
\]

where the factor $\frac{1}{k!}$ is taken care of by time ordering $\tau_1 < \tau_2 < ... < \tau_k$. $c = \{(x_1, \tau_1), \ldots, (x_k, \tau_k)\}$ denotes a continuous time path integral configuration with $k$ vertices and weight

\[
w(c) = \text{tr} \left[ e^{-\beta H_0} \prod_{i=1}^k (-U \hat{n}_{x_i}(\tau_i)) \right]
\]

CT-QMC now proceeds by sampling from all possible configurations $c$ according to their weight $w(c)$.

The structure of the factors in Eq. (5) allows the weight to be rewritten as \[26\]

\[
w(c) = (-U)^k \det(1 + B(c, \beta)),
\]

where single particle propagator matrix $B$ is given by

\[
B(c, \beta) = e^{-\beta H_0} \prod_i \hat{h}(\tau_i, x_i) = e^{-(\beta - \tau_i)H_0} \hat{h}(x_i) \cdots e^{-(\tau_2 - \tau_1)H_0} \hat{h}(x_k) e^{-\tau_1 H_0}
\]

where $H_0$ is a $2V \times 2V$ matrix with elements $[H_0]_{x\sigma,y\sigma'} = t_{xy} \delta_{\sigma\sigma'}$ and the matrix $\hat{h}(x_i)$ is given by

\[
\hat{h}(x_i)|_{x\sigma,y\sigma'} = \delta_{xx_i} \delta_{x_i y} \delta_{\sigma\sigma'} + \delta_{xx_i} \delta_{x y} \delta_{\sigma'\sigma} + \frac{1}{2} \delta_{xy} \delta_{\sigma\sigma'},
\]

and $\hat{h}(\tau_i, x_i) = e^{\tau_i \hat{H}_0} \hat{h}(x_i) e^{-\tau_i \hat{H}_0}$ is its time-displaced counterpart. In the case of the Hubbard the matrix $B$ decomposes into two block matrices for each spin species, giving

\[
w(c) = (-U)^k \det(1 + B^+(c, \beta)) \det(1 + B^-(c, \beta))
\]

The factor $(-U)^k$ introduces a sign problem for positive $U$, since any configuration with an odd number of vertices will have negative weight. On a bipartite lattice with nearest neighbour hoppings this trivial minus sign problem can be removed by mapping the repulsive model into an attractive one with interaction via a particle-hole transformation of the spin-down fermions $c_{x\downarrow} \rightarrow (-1)^k c_{x\downarrow}$ This transformation changes the sign of $U$ and thus removes this trivial sign problem, while the sublattice dependent sign avoids changing the sign of the kinetic energy. However a sign problems can still appear from the determinant.

This formulation of our algorithm is similar in spirit to the interaction-representation (CT-INT) algorithm for quantum impurity problems \[9, 11\] and can be considered a lattice CT-INT (or LCT-INT). Although not required, it has been found to be advantageous to instead use an auxiliary field decomposition \[27\] to remove the trivial sign for fermionic CT-QMC algorithms, leading the continuous time auxiliary field (CT-AUX) algorithm in the case of quantum impurity models \[9, 13\]. The CT expansion is applied with $\hat{h}(x) = (\hat{n}_{x\uparrow} - 1)$. In our algorithm we can introduce an auxiliary field $\rho$ giving an LCT-AUX representation

\[
(1 - \hat{n}_{x\uparrow} \hat{n}_{x\downarrow}) = \frac{1}{2} \sum_{\rho = \pm 1} (1 + \rho \hat{n}_{x\uparrow}) (1 - \rho \hat{n}_{x\downarrow})
\]

so that every vertex now also carries a spin degree of freedom and configurations are of the form $\{(x_i, \tau_i, \rho_i)\}_{i=1,...k}$. The auxiliary field dependent matrix $\hat{h}(x_i, \rho_i)$ now becomes

\[
\hat{h}(x_i, \rho_i)|_{x\sigma,y\sigma'} = \delta_{xx_i} \delta_{x_i y} \delta_{\sigma\sigma'} + \rho_i \delta_{xx_i} \delta_{x y} \delta_{\sigma'\sigma} - \rho_i \delta_{xx_i} \delta_{x y} \delta_{\sigma'\sigma},
\]
and we end up with a weight similar to Eq. (6), but using $h(x_i, \rho_i)$ instead of $h(x_i)$.

To ergodically sample all possible terms in the series (3) it is sufficient to implement two types of Monte Carlo updates: insertion and removal of a vertex. They change the order $k$ by $\pm 1$ and are illustrated in Fig. 1. Starting from a configuration $c$ with $k$ vertices one proposes to randomly insert a new vertex with auxiliary field $\rho$ at position $x$ and time $\tau$. The probability of accepting the new configuration $c'$ is given, using the Metropolis algorithm [28], as $\min(1, R)$ with [34]

$$R = \frac{\beta V[U]}{k+1} \frac{\det[1 + B(c', \beta)]}{\det[1 + B(c, \beta)]}.$$ (12)

Conversely, the probability of removing a vertex is $\min(1, 1/R)$. The same acceptance ratio is derived for the LCT-INT version of the algorithm.

For any observable $O$ one can write an estimator $O(c)$ which must be averaged to obtain an estimate of the quantum expectation value $\langle O \rangle$. For equal-time observables such as densities, kinetic, and interaction energy, these are simple functions of the matrix $B$. The single particle density matrix – or equivalently equal time Green function – $G(x, \sigma; y, \sigma') = \langle c_{x\sigma} c_{y\sigma'}^\dagger \rangle$ is estimated by measuring the matrix elements $G_{x\sigma, y\sigma'}$ of the matrix $G = \frac{B(c, \beta)}{1 + B(c, \beta)}$. (13)

The kinetic energy estimator is simply $E_0 = \text{tr}(H_0 G)$, while the interaction energy is given by $E_I = U \sum_{x,y} G_{x\uparrow, x\uparrow} G_{x\downarrow, x\downarrow}$. 

To demonstrate the reliability and performance of our algorithm we compared it to the discrete time BSS algorithm [23], which has so far been state of the art. Instead of starting from a continuous time representation (3), this algorithm is based on the Suzuki Trotter formula

$$e^{-\beta H} = (e^{-\Delta \tau H_0} e^{-\Delta \tau H_1})^N + O(\beta \Delta^2),$$ (14)

which entails a so-called Trotter error due to time discretization that is quadratic in the time step $\Delta \tau = \beta/N$. Figure 2 shows that the results from our CT-QMC algorithm agree perfectly with those obtained by extrapolating the finite-$\Delta \tau$ results obtained with the BSS algorithm to $\Delta \tau \to 0$. The advantage of our algorithm is that it does not require this extrapolation in $\Delta \tau$. This is particularly important for quantities such as the specific heat or correlation times for the observables being very similar. This is easily understood as both the present method
and the state-of-the-art BSS scheme only employ local updates (vertex insertion/removal and single spin flip respectively).

The main computational effort in the algorithm is calculating the matrix $B(\beta)$ and its changes when performing updates. Naive multiplications of $k$ matrices of dimension $O(V)$ would result in an effort of $kO(V^3)$. Considering that the number of vertices $k$ grows with $\beta$, $U$ and $V$ we obtain a scaling of $O(\beta V^4)$, which is worse than the of the discrete time BSS algorithm. To achieve an overall $O(\beta V^3)$ scaling our algorithm works in the eigenbasis of $H_0$. Since the weight $w(c)$ is a determinant, it is unaffected by a basis change. Basis dependent quantities, such as observables, can be obtained via a rotation of the density matrix $G$.

Diagonalizing $H_0 = U E U^\dagger$, where $E$ is diagonal, the factors $e^{-(\tau_{i+1} - \tau_i)H}$ also become diagonal matrices $e^{-(\tau_{i+1} - \tau_i)E}$. The other factors are of the form $\delta_{kk'} + \gamma U_{kx}^\dagger U_{x'k'}$, which is an identity matrix and an outer product of two vectors. Given this decomposition into sparse matrices and an outer product, the matrix multiplications can be performed with an effort $O(V^2)$, thus recovering the $O(\beta V^3)$ scaling of the BSS algorithm.

The product of matrices is in general an ill-conditioned matrix. To prevent numerical errors from creeping into the simulation, a stabilization procedure must be used, as explained in the Supplementary Material.

Common optimization techniques that are employed in other fermionic QMC algorithms can be applied here as well. Fast updates can be performed due to the fact that $B(c)$ before the insertion (or removal) of a vertex, and $B(c')$ after it, differ only by a single factor $h(x, \tau)$, which is a diagonal matrix having all coefficients equal to 1 except a single one [23]. Several updates can also be combined in a delayed update scheme [19, 29]. We implemented delayed updates for both our discrete time and continuous time codes.

Our performance measurements in Fig. 3 confirm the linear scaling in $\beta$. We have compared LCT-AUX with a BSS code using the same set of optimizations. In these conditions LCT-AUX performs as well as the BSS algorithm with a commonly used time step of $\Delta = 1/8$. Since the BSS simulations have to be repeated for several values of $\Delta$, LCT-AUX extrapolated, our unoptimized CT-QMC is already faster than a full discrete time calculation.

Using projections from a trial wave function our algorithm can be used for ground state simulations, similar to other fermionic QMC algorithms. As for the discrete time algorithms [30] (See Supplementary Material). It can also be used as a a quantum impurity solver and used for DMFT calculations [14–17, 22, 31] by adding $V_b$ non-interacting bath sites. The complexity of such an algorithm is $O(\beta UV(V + V_b)^2)$, which for low temperatures can be better than the $O(\beta^3 V^3)$ scaling of other CT-QMC algorithms [9, 11–13] for Hubbard-type models. The time-dependent Green functions $G(\tau; x, y) = \langle T[c_x(\tau)c_y^\dagger(0)] \rangle$ required for DMFT can be measured using partial propagators $B(c, \tau)$:

$$G(\tau; \{x_i, \tau_i\}) = \frac{B(c, \tau)}{1 + B(c, \beta)}$$

with

$$B(c, \tau) = e^{-\tau H_0} \prod_{\tau_i < \tau} h(\tau_i, x_i).$$

In general it is more stable to measure its Fourier transform, using non-uniform fast Fourier transformations [32].

The algorithm presented here is not specific to the Hubbard model. The only requirements are that the non-interacting Hamiltonian can be diagonalized once at the start of the algorithm to obtain the eigenvector matrix $U$ and that the interacting Hamiltonian can be decomposed into exponentials of quadratic operators, i.e. $e^{A_{xy}c_x^\dagger c_y}$ for some matrix $A$. This is in general possible for any four-fermions interaction of the type $H_1 = \sum_{kk'pp'} c_{k}\dagger c_{k'} c_{p} c_{p'}$ including the Coulomb interaction. As for the discrete time scheme, local interactions will retain the $O(V^3)$ scaling, but this might change in the general case (see Supplementary Material for more discussion).

In summary, we have presented a continuous time QMC algorithm for fermionic lattice models that has the same scaling as discrete time methods. This closes the
last prominent gap in the portfolio of CT-QMC algorithms, which have otherwise become the state of the art for bosons, quantum spins, and fermionic impurity problems. The main advantage of our algorithm is the absence of any time discretization error. This eliminates the need to either guess a small enough time step \( \Delta_t \) or extrapolate from multiple simulations at different \( \Delta_t \) to \( \Delta_t \to 0 \) and leads to shorter simulation times. Our algorithm can profit from the same numerical optimizations previously developed for other fermionic QMC algorithms [19, 23, 29] and can be used for finite temperature simulations, ground state calculations and quantum chemistry simulations and as a quantum impurity solver.

We acknowledge discussions with F.F. Assad and Lei Wang. This work was supported by the ERC Advanced Grant SIMCOFE, Microsoft Research, and the Swiss National Science Foundation through the National Competence Centers in Research NCCR QSIT and MARVEL. MT acknowledges hospitality of the Aspen Center for Physics, supported by NSF grant # 1066293.


[33] Or more generally $t_{xy} = t_{yx}^*$ in the presence of a gauge field. Since gauge fields generally introduce a phase problem we restrict ourselves to the case of real $t_{xy}$

[34] The probability of inserting at a uniformly chosen time and location is $d\tau/(V\beta)$ while the probability of removing one of $k + 1$ vertices is $1/(k + 1)$. The infinitesimal $d\tau$ cancels with infinitesimals in the weights, the a-priori probability $1/2$ of choosing one of the two values of the auxiliary field cancels with a factor $1/2$ in Eq. (10) and the factor $\beta V/(k + 1)$ enters the acceptance ratio.