PHYSICAL REVIEW B **73**, 041105(R) (2006)

Linear scaling electronic structure calculations and accurate statistical mechanics sampling with noisy forces

Florian R. Krajewski and Michele Parrinello
Computational Science, Department of Chemistry and Applied Biosciences, ETH Zurich, USI Campus,
Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland

(Received 5 December 2005; published 25 January 2006)

Numerical simulations based on electronic structure calculations are finding ever growing applications in many areas of physics. A major limiting factor, however, is the cubic scaling of the algorithms used. Building on previous work [Phys. Rev. B **71**, 233105 (2005)] we introduce a statistical method for evaluating interatomic forces, which scales linearly with system size and is applicable also to metals. The method is based on exact decomposition of the fermionic determinant and on a mapping onto a field theoretical expression. We solve the problem of an accurate sampling of the Boltzmann distribution with noisy forces. This novel approach can be used in such diverse fields as quantum chromodynamics, quantum Monte Carlo, or colloidal physics.

DOI: 10.1103/PhysRevB.73.041105 PACS number(s): 71.15.−m, 31.15.−p

Atomistic simulations in which interactions are computed on the fly from electronic structure calculation play an important role in modern science and have proven their relevance in many fields. However, their computational cost is a severe limitation. In particular, simulating large systems has proven challenging due to the cubic dependence of the computation time on the number of electrons. This has long since been recognized[1,2] and a number of algorithms have been suggested that in principle lead to linear scaling.[3–10] Most are based on the possibility in semiconductors or insulators of localizing the electronic orbitals. Linear scaling is then achieved by neglecting interactions between faraway atoms. This approach, however, suffers from poor convergence and leads to errors that are not easy to control. The wavefunctions cannot be localized in metals and only very few alternative methods have been suggested.[9,10] All in all it can be stated that in spite of considerable progress performing linear scaling *ab initio* simulations is still very challenging.

Very recently we have proposed a new algorithm that scales linearly in all physical dimensions for semiconductors and metals alike.[11] Here we reformulate the algorithm as a field theory and sample the resulting action stochastically. We show that in spite of the statistical noise present in the evaluation of the forces accurate sampling can be performed. Our way of solving this problem is general and can also solve problems of similar nature that are encountered in quantum chromodynamics,[12,13] quantum Monte Carlo,[14] and colloidal physics,[15] where the interaction is also determined stochastically.

Let us start with the generic expression for the total energy in theories that can be formulated in an effective single particle form:

$$E = 2\sum_{i=1}^{N} \epsilon_i + V_r. \tag{1}$$

The first term is the so-called band-structure term given by the sum of the lowest N doubly occupied eigenvalues ϵ_i of a Hamiltonian $\mathbf{H}$. For instance, in density functional theory $\mathbf{H}$ is the Kohn and Sham Hamiltonian and V_r corrects for double counting and accounts for the direct nuclear nuclear interaction, while in tight binding and other semiempirical approaches $\mathbf{H}$ is a Hamiltonian that depends parametrically on the atomic positions, and V_r is a pairwise additive energy. In either case, V_r can be calculated in $\mathcal{O}(N)$ operations, while the calculation of the band-structure term has an apparent $\mathcal{O}(N^3)$ complexity and has been the limiting factor that has so far prevented simulating very large systems.

Following Refs. 16 and 17 we write the band-structure term as the low-temperature limit of the grand canonical potential for independent fermions:

$$\Omega = -\frac{2}{\beta} \ln \det(1 + e^{\beta(\mu - \mathbf{H})}). \tag{2}$$

In Eq. (2) μ is the electron chemical potential and it easy to see that $\lim_{\beta\to\infty}\Omega = 2\sum_{i=1}^{N}\epsilon_i - \mu N_e$, where $N_e = 2N$ is the total number of electrons. We now factorize the operator in Eq. (2) as

$$1 + e^{\beta(\mu - \mathbf{H})} = \prod_{l=1}^{P/2} (\mathbf{M}_l^\dagger \mathbf{M}_l), \tag{3}$$

where P is an even integer and $\mathbf{M}_l = 1 + e^{i\pi(2l-1)/P} e^{\beta/P(\mu - \mathbf{H})}$. Here we depart from Ref. 11 and since $\mathbf{M}_l^\dagger \mathbf{M}_l$ is a positive definite operator we can follow the well-known practice in lattice gauge simulations[12] of writing its inverse determinant as an integral over a field $\mathbf{\Phi}_l$ that has the dimension of the full Hilbert space in the form[18]

$$\det(\mathbf{M}_l^\dagger \mathbf{M}_l)^{-1/2} = \frac{\int \mathcal{D}[\mathbf{\Phi}_l] e^{-\mathbf{\Phi}_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \mathbf{\Phi}_l}}{\int \mathcal{D}[\mathbf{\Phi}_l] e^{-\mathbf{\Phi}_l^\dagger \mathbf{\Phi}_l}}. \tag{4}$$

The appearance of determinants is ubiquitous in fermionic theories and is another way of formulating the minus sign problem. The use of Eq. (4) that implies sampling a positive definite distribution circumvents this problem. Inserting relation (3) into Eq. (2) after having used Eq. (4) we end up with the following expression for the grand canonical potential:

$$\Omega = \frac{4}{\beta}\sum_{l=1}^{P/2} \ln \int \mathcal{D}[\mathbf{\Phi}_l] e^{-\mathbf{\Phi}_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \mathbf{\Phi}_l} + \text{const}, \qquad (5)$$

which is the promised field theoretical formulation. The quantities of physical interest like energy or force can all be calculated as derivatives of Ω relative to an appropriate external parameter. For instance $N_e = -(\partial/\partial\mu)\Omega$, and assuming that β^{-1} is so low that temperature effects on the electrons can be neglected $E^{\text{band}} = 2\sum_{i=1}^{N}\epsilon_i = (\partial/\partial\beta)(\beta\Omega) + \mu N_e$ while the contribution to the force on particle I at position $\mathbf{R}_I$ coming from the band term is given by $\mathbf{F}_I^{\text{band}} = -\mathbf{\nabla}_{\mathbf{R}_I}\Omega$. In taking the derivatives the constant in Eq. (5) vanishes and one is left with calculating expressions of the type:

$$\frac{\partial\Omega}{\partial\lambda} = \frac{4}{\beta}\sum_{l=1}^{P/2} \frac{\int \mathcal{D}[\mathbf{\Phi}_l]\left[\mathbf{\Phi}_l^\dagger\left(\frac{\partial\mathbf{M}_l^\dagger}{\partial\lambda}\mathbf{M}_l + \mathbf{M}_l^\dagger\frac{\partial\mathbf{M}_l}{\partial\lambda}\right)\mathbf{\Phi}_l\right] e^{-\mathbf{\Phi}_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \mathbf{\Phi}_l}}{\int \mathcal{D}[\mathbf{\Phi}_l] e^{-\mathbf{\Phi}_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \mathbf{\Phi}_l}}. \qquad (6)$$

Thus all relevant properties can be evaluated as averages of the $P/2$ distributions $e^{-\mathbf{\Phi}_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \mathbf{\Phi}_l}$.

So far no approximation has been made and no computational advantage has been gained either. In order to make further progress we must use the fact that in $\mathbf{M}_l$ the operator $e^{(\beta/P)(\mu-\mathbf{H})}$ appears and that P can be taken to be sufficiently large for suitable approximations to the exponential operator to be accurate. In Ref. 11 we used as the basis set a grid in real space, and a Trotter decomposition was the natural approximation to use. Here we will apply our method to a tight binding Hamiltonian and simply use a high-temperature expansion

$$\mathbf{M}_l = 1 + e^{\pi(2l-1)/P}\left[1 + \frac{\beta}{P}(\mu - \mathbf{H})\right] + \mathcal{O}\left[\left(\frac{\beta}{P}\right)^2\right]. \qquad (7)$$

In this manner the operator $\mathbf{M}_l$ has the same sparsity of $\mathbf{H}$, a fact which will eventually lead to linear scaling. Note that no assumption has been made on the energy spectrum or on the local character of the wavefunctions and therefore our method will be valid both for metals and nonmetals. It simplifies the calculation of the properties of the system if in Eq. (6) we use the expression

$$\frac{\partial\mathbf{M}_l}{\partial\lambda} = \frac{1}{2P}\{(\mathbf{M}_l - 1)\mathbf{O}_\lambda + \mathbf{O}_\lambda(\mathbf{M}_l - 1)\} + \mathcal{O}\left(\frac{1}{P^3}\right), \qquad (8)$$

which has an accuracy compatible with Eq. (7). For instance in Eq. (8) $\mathbf{O}_\lambda$ can be β, $(\mu - \mathbf{H})$, or $-\beta\mathbf{\nabla}_{\mathbf{R}_I}\mathbf{H}$ for $\lambda = \mu$, β, or $\mathbf{R}_I$ respectively. A standard approach to sampling $e^{-\mathbf{\Phi}_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \mathbf{\Phi}_l}$ is to draw a sequence of normal distributed random numbers $\mathbf{\Psi}_l$ and compute $\mathbf{\Phi}_l$ solving the equations $\mathbf{M}_l\mathbf{\Phi}_l = \mathbf{\Psi}_l$. Since $\mathbf{M}_l$ is sparse, this equation can be solved in $\mathcal{O}(N)$ operations using for instance a biconjugated gradient method.[19] It can be easily shown that this approach is equivalent to the stochastic inversion method advocated in Ref. 11 Here we have decided instead to sample $e^{-\mathbf{\Phi}_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \mathbf{\Phi}_l}$ using Langevin dynamics.

$$m_l\ddot{\mathbf{\Phi}}_l = -\mathbf{M}_l^\dagger\mathbf{M}_l\mathbf{\Phi}_l - \gamma_e\dot{\mathbf{\Phi}}_l + \xi_l \qquad (9)$$

where the components α of the white random noise vector ξ_l obey the relation

$$\langle \xi_l^\alpha(0)\xi_l^\alpha(t)\rangle = 2m_l\gamma_e\delta(t). \qquad (10)$$

In this way we circumvent the need to invert the matrix $\mathbf{M}_l$. This solves the problem that for a tight binding Hamiltonian as opposed to the Hamiltonian used in Ref. 11 we were not able to find good preconditioners. Furthermore, since different $\mathbf{M}_l$'s have different eigenvalue spectra one can choose the m_l so as to achieve the optimal sampling speed in each l channel. This problem is particularly serious for metals where $\mathbf{M}_{l\approx P/2}$ can have eigenvalues close to zero, which leads to a much slower sampling speed.[11] Furthermore, since we will use a Langevin sampling also for the ions it is pleasing to use the same sampling methodology for electronic and ionic degrees of freedom.

Inevitably the interatomic forces that are calculated by this procedure will be affected by a statistical error. This will prevent us from using these forces to perform energy conserving molecular dynamics (MD) calculations. However, we will show that sampling the Boltzmann distribution is still possible. Similar to what was done for the electronic degrees of freedom we sample the ionic configurations with a Langevin equation

$$M\ddot{\mathbf{R}}_I = \mathbf{F}_I - \gamma_I\dot{\mathbf{R}}_I + \mathbf{\Xi}_I \qquad (11)$$

where the random force obeys the relations

$$\langle \mathbf{\Xi}_I(0)\mathbf{\Xi}_I(t)\rangle = 6k_{\text{B}}TM\gamma_I\delta(t) \qquad (12)$$

and

$$\langle \mathbf{F}_I(0)\mathbf{\Xi}_I(t)\rangle = 0. \qquad (13)$$

From the electrons' Langevin dynamics we do not get the exact forces $\mathbf{F_I}$ but an approximation $\mathbf{F_I}^L = \mathbf{F_I} + \mathbf{\Xi}_\mathbf{I}^L$ that is affected by a statistical error $\mathbf{\Xi}_I^L$, and therefore there is in principle no guarantee that correct Boltzmann averages are obtained from the solution of Eq. (11). But let us assume, and we will show later that this assumption is accurate, that $\mathbf{\Xi}_I^L$, is also a white noise obeying

$$\langle \mathbf{\Xi}_I^L(0)\mathbf{\Xi}_I^L(t)\rangle \cong 6k_{\text{B}}TM\gamma_I^L\delta(t),$$

$$\langle \mathbf{F}_I(0)\mathbf{\Xi}_I^L(t)\rangle \cong 0. \qquad (14)$$

In this case the noise $\mathbf{\Xi}_\mathbf{I}^L$ simply adds to $\mathbf{\Xi}_\mathbf{I}$ and if we modify Eq. (11) so as to read

$$M\ddot{\mathbf{R}}_I = \mathbf{F}_I - (\gamma_I + \gamma_I^L)\dot{\mathbf{R}}_I + \mathbf{\Xi}_I + \mathbf{\Xi}_I^L \qquad (15)$$

we recover a Langevin equation whose trajectories can still be used to obtain a Boltzmann sampling. Equation (14) is in

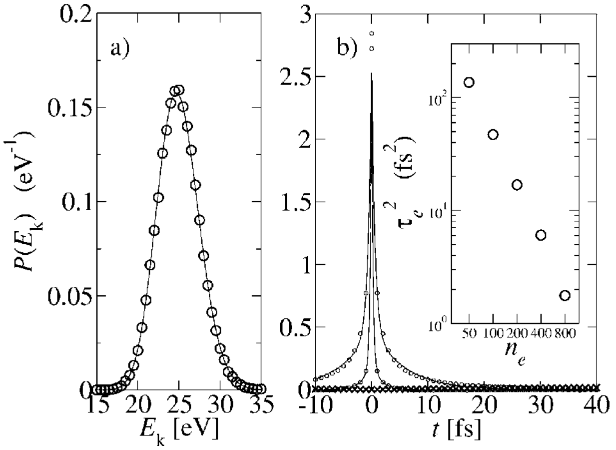

FIG. 1. Statistical properties in a 64 atoms liquid Si simulation at 3000 K. (a) The ionic kinetic energy distribution (circles) is compared with the exact Maxwell distribution (line). (b) Autocorrelation of the noise $\langle \Xi_I^L(0)\Xi_I^L(t)\rangle$ (circles) for $n_e=100$ and $n_e=800$. The broader distribution corresponds to the $n_e=100$ case and the lines are two exponential fits to the data. The crosses correspond to the cross correlation $\langle \mathbf{F}_I(0)\Xi_I^L(t)\rangle$. All correlation functions are normalized to the average square fluctuation of the band energy contribution to the ionic forces. In the inset the variation of τ_e^2 as function of n_e is shown in double logarithmic scale and exhibits approximately a $n_e^{-3/2}$ dependence.

principle an approximation but we will show below how to make it more and more accurate.

At first sight it would appear that we are defeating our object since in general we know $\mathbf{F}_I+\Xi_I^L$ but not each term individually. However, we can determine γ_I^L by varying it until the equipartition theorem $\left\langle \frac{1}{2}M\dot{\mathbf{R}}_I^2\right\rangle=\frac{3}{2}k_{\mathrm{B}}T$ is satisfied. With this choice the sampling will be correct and noisy forces can be used in the sampling. It is important to note that once the value of γ_I^L is determined it has to be kept constant during the simulation. With this procedure one can exactly calculate static observables within the framework of Langevin dynamics without knowing the exact force. This is at variance with Ref. 20 where explicit assumptions on the noise have to be made.

We tested our scheme simulating Si in the solid and the liquid phase using a tight binding model.[21] We checked that a value of $P=200$ is sufficient for the approximation of Eq. (7) to be valid. The Langevin dynamics parameters used were $\delta t_e=1$, $\gamma_e\delta t_e=1/20$ where δt_e is the discretized integration time step and the algorithm of Ricci and Ciccotti[22] has been used throughout. The masses m_l are adjusted such that the average force fluctuations are $(\mathbf{M}_l^\dagger\mathbf{M}_l\boldsymbol{\Phi}_l)^2(\delta t_e^2/2m_l) < 0.025$. In this way we gain at least one order of magnitude in speed. The integration time step for the ion dynamics is $\delta t=1$ fs. After each ionic displacement we let the $\boldsymbol{\Phi}_l$ evolve under the action of the new $\mathbf{M}_l^\dagger\mathbf{M}_l$ until the distribution is equilibrated. The time needed for the equilibration is problem dependent and can be measured by looking at the correlation function $\langle\boldsymbol{\Phi}_l(0)\boldsymbol{\Phi}_l(t)\rangle$. In the present case we make the rather conservative choice of running the electronic Langevin equation for $n_e=100$ time steps. The $\boldsymbol{\Phi}_\mathbf{l}$'s thus obtained are used to calculate the ionic forces for the next integration step. The chemical potential is continuously adjusted such that the number of electrons fluctuates around the desired value.

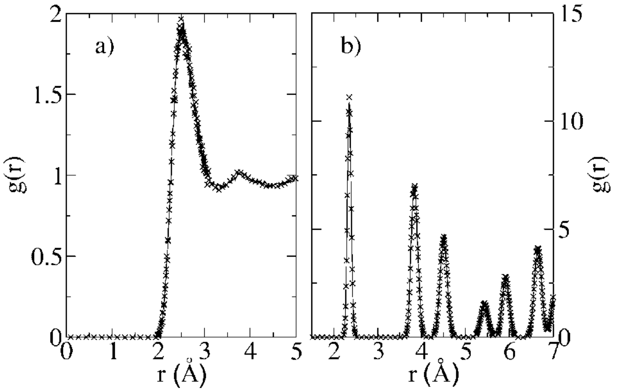

FIG. 2. Pair correlation functions for (a) liquid silicon (3000 K) and (b) crystalline silicon in the diamond phase (300 K). The crosses show the results from our method and the lines are calculated using standard diagonalization of the tight-binding Hamiltonian.

We first consider the case of 64 Si atoms in a periodically repeated cell of length 10.86 Å at a temperature of 3000 K where the system is metallic. Using the procedure described above we find that if we take $\gamma_I=\frac{1}{30}$ fs^{-1} we need to add a correction due to the noise in the forces $\gamma_I^L=\frac{1}{379}$ fs^{-1} in order to fix the average kinetic energy to the desired value. In Fig. 1(a) it is also seen that not only is the average energy correct but also that its fluctuations follow Maxwell distribution.

In this small system it is possible to calculate the correct forces on the ions and check the statistical properties of Ξ_I^L. In Fig. 1(b) it can be seen that $\langle \mathbf{F}_I(0)\Xi_I^L(t)\rangle$ is very close to zero and that $\langle \Xi_I^L(0)\Xi_I^L(t)\rangle$ is very localized in time and can be made sharper at will by increasing n_e. This is illustrated also by the behavior of the noise correlation width

$$\tau_e^2=\frac{\displaystyle\int dt\, t^2\,\langle \Xi_I^L(0)\Xi_I^L(t)\rangle}{\displaystyle\int dt\,\langle \Xi_I^L(0)\Xi_I^L(t)\rangle}$$

as a function of n_e. If we approximate $\langle \Xi_I^L(0)\Xi_I^L(t)\rangle$ as a delta function whose strength is given by its integral we find an estimate for $\gamma_I^L\approx 1/345$ fs^{-1} in good agreement with the empirical determination. In the solid phase the close value of $\gamma_I^L=1/502$ fs^{-1} leads to correct sampling. Therefore we expect that with some adjustment phase transitions can be studied with our method, at least in this case.

In Fig. 2 we compare the pair correlation functions $g(r)$ calculated with noisy forces and those evaluated with a standard approach. We see that the agreement is excellent and that the use of noisy forces does not degrade the quality of the simulation.

The break-even point between standard calculations and the present linear scaling method is not easy to determine since it depends on the accuracy required, which in our case is related to the number P used in the decomposition of the fermion determinant from Eq. (3). For the case of Si we estimate on a single processor the crossing point to be at ‘about 500 atoms. In other cases larger values of P might be necessary, thus shifting the crossing point to larger systems. However, our algorithm is trivially parallelizable and therefore we expect it to be more favorable in terms of wall clock time on massive parallel platforms.

A simple and interesting extension of the present work would be to use it for fully self-consistent density functional theory. In this case one further source of noise would be the fact that only a noisy estimator of the density can be given. Whether this leads to a practically working simulation scheme will be studied elsewhere.

We also stress that the procedure described here for handling noisy forces is fully self contained since the parameter n_e can be varied until convergence on the physical properties of interest is reached. The only requirement is that the stochastic process through which the forces are calculated has a finite correlation time. As such, the method described transcends the field of linear scaling algorithms and offers an alternative to other Monte Carlo (MC) methods with noisy estimators.[13,20]

We thank A. Laio for critically reading the manuscript.

---

[1] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).

[2] G. Galli and M. Parrinello, Phys. Rev. Lett. **69**, 3547 (1992).

[3] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).

[4] A. A. Mostofi, P. D. Haynes, C. Skylaris, and M. C. Payne, J. Chem. Phys. **119**, 8842 (2003).

[5] G. E. Scuseria, J. Phys. Chem. A **103**, 4782 (1999).

[6] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).

[7] E. Hernandez and M. J. Gillan, Phys. Rev. B **51**, 10157 (1995).

[8] X. P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).

[9] S. Goedecker and L. Colombo, Phys. Rev. Lett. **73**, 122 (1994).

[10] R. Baer and M. Head-Gordon, Phys. Rev. Lett. **79**, 3962 (1997).

[11] F. R. Krajewski and M. Parrinello, Phys. Rev. B **71**, 233105 (2005).

[12] I. Montvay and G. Münster, *Quantum Fields on a Lattice* (Cambridge Monographs on Mathematical Physics, 1994).

[13] G. Bhanot and A. D. Kennedy, Phys. Lett. B **157**, 70 (1985).

[14] C. Pierleoni, D. M. Ceperley, and M. Holzmann, Phys. Rev. Lett. **93**, 146402 (2004).

[15] F. Bitzer, T. Palberg, H. Löwen, R. Simon, and P. Leiderer, Phys. Rev. E **50**, 2821 (1994).

[16] A. Alavi and D. Frenkel, J. Chem. Phys. **97**, 9249 (1992).

[17] A. Alavi, J. Kohanoff, M. Parrinello, and D. Frenkel, Phys. Rev. Lett. **73**, 2599 (1994).

[18] An elegant expression for the partition function of the total system (electrons+ions) can be obtained if one uses the alternate expression $\det(\mathbf{M}_l^\dagger \mathbf{M}_l)^{1/2} = \int \mathcal{D}[\Phi_l] e^{-\Phi_l^\dagger (\mathbf{M}_l^\dagger \mathbf{M}_l)^{-1} \Phi_l} / \int \mathcal{D}[\Phi_l] e^{-\Phi_l^\dagger \Phi_l}$, leading to $Z \propto \int dR_I e^{-\beta V_r} \int \prod_{l=1}^{P} \mathcal{D}[\Phi_l] e^{-\Sigma_{l=1}^{P} \Phi_l^\dagger (\mathbf{M}_l^\dagger \mathbf{M}_l)^{-1} \Phi_l}$. Unfortunately this led to an algorithm which was less efficient than that described here. This will be discussed in detail elsewhere.

[19] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77* (Cambridge University Press, Cambridge, 1992).

[20] D. M. Ceperley and M. Dewing, J. Chem. Phys. **110**, 9812 (1999).

[21] L. Goodwin, A. J. Skinner, and D. J. Pettifor, Europhys. Lett. **9**, 701 (1989).

[22] A. Ricci and G. Ciccotti, Mol. Phys. **101**, 1927 (2003).