

Extended Born-Oppenheimer Molecular Dynamics

Anders M. N. Niklasson

*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
and Applied Materials Physics, Department of Materials Science and Engineering, Royal Institute of Technology,
SE-100 44 Stockholm, Sweden*

(Received 4 September 2007; published 25 March 2008)

A Lagrangian generalization of time-reversible Born-Oppenheimer molecular dynamics Niklasson *et al.* [Phys. Rev. Lett. **97**, 123001 (2006)] is proposed. The formulation enables the application of higher-order symplectic or geometric integration schemes that are stable and energy conserving even under incomplete self-consistency convergence. It is demonstrated how the accuracy is improved by over an order of magnitude compared to previous formulations at the same level of computational cost. The proposed Lagrangian includes extended electronic degrees of freedom as auxiliary dynamical variables in addition to the nuclear coordinates and momenta. While the nuclear degrees of freedom propagate on the Born-Oppenheimer potential energy surface, the extended auxiliary electronic degrees of freedom evolve as a harmonic oscillator centered around the adiabatic propagation of the self-consistent ground state.

DOI: [10.1103/PhysRevLett.100.123004](https://doi.org/10.1103/PhysRevLett.100.123004)

PACS numbers: 31.15.xv, 34.10.+x, 71.15.Pd

Born-Oppenheimer molecular dynamics [1–6] based on self-consistent field (SCF) methods, such as Hartree-Fock or density functional theory [7–9], is currently a gold standard in molecular dynamics simulations. It is derived from the well defined adiabatic approximation for the separation of the nuclear and electronic degrees of freedom, where the forces acting on the atoms are calculated at the self-consistent electronic ground state [10]. However, the ability to achieve physically accurate and stable micro-canonical simulations, while keeping the computational cost low, has been limited by the requirement of a high degree of SCF convergence in the nuclear force calculations [11,12]. Only very recently, by restoring the time-reversal symmetry in the underlying adiabatic propagation of the electronic degrees of freedom [13–15], has it been possible to achieve efficient energy conserving simulations also under incomplete SCF convergence. Unfortunately, these techniques are not able to take advantage of powerful symplectic or geometric integration methods developed for celestial and classical molecular dynamics [16–21]. Symplectic integration, which in general requires a Hamiltonian or Lagrangian formulation of the dynamics, enables highly efficient simulations while keeping a rigorous control over physical properties.

The purpose of this Letter is to take advantage of geometric integration methods in Born-Oppenheimer molecular dynamics by introducing a Lagrangian generalization of the recently proposed time-reversible Born-Oppenheimer molecular dynamics [14]. This gives time-reversible Born-Oppenheimer molecular dynamics a theoretically more solid and physically transparent framework and, most significantly, thanks to the Lagrangian formulation, higher-order symplectic integration algorithms can be applied, which provide excellent performance in molecular dynamics simulations. It will be demonstrated how the accuracy is

increased by over an order of magnitude at the same level of computational cost compared to previous formulations.

The conventional Born-Oppenheimer (BO) Lagrangian for *ab initio* molecular dynamics is given by

$$\mathcal{L}^{\text{BO}}(\mathbf{R}, \dot{\mathbf{R}}) = \frac{1}{2} \sum_k M_k \dot{R}_k^2 - U_{\text{SCF}}[\mathbf{R}; D], \quad (1)$$

where the potential U_{SCF} is the total electronic energy in, for example, Hartree-Fock or density functional theory, including the nuclear-nuclear repulsion terms. The potential energy $U_{\text{SCF}}[\mathbf{R}; D]$ is defined at the electronic ground state given by the SCF optimized solution D of the electronic degrees of freedom. D is assumed to be the symmetric single-particle density matrix in an orthogonal basis-set representation, though generalizations to other representations, such as the density, the wave functions, or the Kohn-Sham Hamiltonian, are straightforward. Notice that D is not an independent dynamical variable, since it is determined by the external potential at atomic configuration $\mathbf{R} = \{R_k\}$. It is included to show that the Lagrangian \mathcal{L}^{BO} is calculated at the self-consistent Born-Oppenheimer ground state. The nuclear degrees of freedom are given by the atomic coordinates R_k and velocities \dot{R}_k , with the corresponding masses M_k . The dots denote time derivatives.

The key result of this Letter is the extension of the Born-Oppenheimer Lagrangian \mathcal{L}^{BO} in Eq. (1) by the addition of auxiliary electronic degrees of freedom P and \dot{P} that evolve in a harmonic potential centered around the self-consistent solution D . The extended auxiliary dynamical variables, P and \dot{P} , are assumed to be of the same form as D , i.e., a density matrix and its time derivative. The extended Born-Oppenheimer (XBO) Lagrangian is given by

$$\mathcal{L}^{\text{XBO}}(\mathbf{R}, \dot{\mathbf{R}}, P, \dot{P}) = \mathcal{L}^{\text{BO}} + \frac{\mu}{2} \text{Tr}[\dot{P}^2] - \frac{\mu\omega^2}{2} \text{Tr}[(D - P)^2]. \quad (2)$$

Here μ and ω are fictitious mass and frequency parameters for the auxiliary electronic degrees of freedom. There are no additional terms imposing constraints on the electronic degrees of freedom, i.e., wave function orthogonality or density matrix idempotency [22–24]. These constraints are not necessary, since the potential energy $U_{\text{SCF}}[\mathbf{R}; D]$ and the nuclear forces are calculated at the normalized and idempotent ground state D .

The time evolution of the dynamical system described by the extended Lagrangian \mathcal{L}^{XBO} is determined by Euler-Lagrange equations of motion

$$M_k \ddot{R}_k = - \frac{\partial U_{\text{SCF}}[\mathbf{R}; D]}{\partial R_k} - \mu\omega^2 \text{Tr}[(D - P) \partial D / \partial R_k], \quad (3)$$

$$\mu \ddot{P} = \mu\omega^2 (D - P). \quad (4)$$

In the limit $\mu \rightarrow 0$, i.e., when $\mathcal{L}^{\text{XBO}} \rightarrow \mathcal{L}^{\text{BO}}$, the dynamics is determined by the coupled equations of motion,

$$M_k \ddot{R}_k = - \frac{\partial U_{\text{SCF}}[\mathbf{R}; D]}{\partial R_k}, \quad (5)$$

$$\ddot{P} = \omega^2 (D - P). \quad (6)$$

These two equations of motion reveal two properties that are of fundamental importance. (a) The nuclear forces are calculated at the self-consistent ground state D as with the Born-Oppenheimer Lagrangian. The molecular trajectories therefore evolve on the Born-Oppenheimer potential energy surface with the total Born-Oppenheimer energy, $E^{\text{BO}} = \frac{1}{2} \sum_k M_k \dot{R}_k^2 + U_{\text{SCF}}[\mathbf{R}; D]$, as a constant of motion. (b) The equations of motion do not include the fictitious electron mass parameter μ , which otherwise could cause problems [25].

Since the extended electronic degrees of freedom $P(t)$ evolve in a harmonic potential centered around their own self-consistent solutions $D(t)$ in Eq. (2), the auxiliary density matrix $P(t)$ and its self-consistent solution $D(t)$ will stay close together. We can therefore use $P(t)$ as an efficient initial guess to $D(t)$ in the iterative SCF optimization,

$$D(t) = \text{SCF}[\mathbf{R}(t), P(t)]. \quad (7)$$

This strongly reduces the computational cost to reach the self-consistent ground state $D(t)$ at which the nuclear forces are calculated in Eq. (5).

In conventional Born-Oppenheimer molecular dynamics the initial guess for the iterative SCF optimization is not given by an auxiliary dynamical variable, as in Eq. (7), but by an extrapolation from previous time steps [5,11,12,26–

28]. Unfortunately, because the SCF procedure is irreversible and in practice never complete, this extrapolation breaks the time-reversal symmetry in the underlying propagation of the electronic degrees of freedom, which causes serious stability problems with a systematic drift in the total energy [11,12]. Only by increasing the SCF convergence, at great computational cost, can the energy drift be reduced, though it never fully disappears. The fundamental problem with the broken time-reversal symmetry in the electron propagation was recently solved by the introduction of time-reversible Born-Oppenheimer molecular dynamics based on a lossless dual filter integration scheme [14]. With the extended Lagrangian formulation the time-reversal problem is avoided in a similar way: the auxiliary electronic degrees of freedom $P(t)$, and thus the initial SCF guesses in Eq. (7), occur, not through extrapolation, but as dynamical variables that can be integrated by time-reversible algorithms [16,29,30]. The nuclear forces are then calculated with an underlying electron propagation that is time reversible. In this way a systematic energy drift is avoided even under incomplete SCF convergence.

It is easy to see that the extended Lagrangian formulation is a generalization of time-reversible Born-Oppenheimer molecular dynamics. If we apply the time-reversible Verlet scheme [29] to the integration of the electronic degrees of freedom in Eq. (6) we get

$$P(t + \delta t) = 2P(t) - P(t - \delta t) + \delta t^2 \omega^2 [D(t) - P(t)]. \quad (8)$$

If we choose the dimensionless factor $\kappa = \delta t^2 \omega^2 = 2$ this propagation is identical to the linear integration scheme in time-reversible Born-Oppenheimer molecular dynamics [14]. Thus, the extended Born-Oppenheimer Lagrangian in Eq. (2) forms a natural framework for time-reversible Born-Oppenheimer molecular dynamics, which thereby is given a more rigorous and physically transparent formulation. Instead of propagation through a time-reversible dual filter process, the auxiliary electronic degrees of freedom $P(t)$ occur as dynamical variables that evolve through a time-reversible integration scheme.

Possibly the most important advantage of the extended Born-Oppenheimer molecular dynamics is that it enables the application of higher-order symplectic or geometric integration methods [16–21]. Whereas a conventional integration algorithm can be seen as a numerical approximation for the integration of an underlying exact Hamiltonian dynamics, a symplectic integration can be seen as an exact integration for an underlying approximate Hamiltonian. The conservation of various physical properties of the approximate Hamiltonian can then be rigorously controlled [17,20,31]. For the nuclear coordinates in Eq. (5) a quite general symplectic integration [20,31] over a time length δt is divided in m steps ($i = 1, 2, \dots, m$)

$$\begin{aligned} \dot{R}_k(t_i) &= \dot{R}_k(t_{i-1}) + b_i \delta t \ddot{R}_k(t_{i-1}), \\ R_k(t_i) &= R_k(t_{i-1}) + a_i \delta t \dot{R}_k(t_i). \end{aligned} \quad (9)$$

Here $[R_k(t_0), \dot{R}_k(t_0)] = [R_k(t), \dot{R}_k(t)]$ and $[R_k(t + \delta t), \dot{R}_k(t + \delta t)] = [R_k(t_m), \dot{R}_k(t_m)]$. For the electronic degrees of freedom in Eq. (6), for $i = 1, 2, \dots, m$, and using the variable substitution $\delta t \dot{P}(t) \rightarrow \dot{P}(t)$, the symplectic integration is

$$\begin{aligned} \dot{P}(t_i) &= \dot{P}(t_{i-1}) + b_i \kappa [D(t_{i-1}) - P(t_{i-1})], \\ P(t_i) &= P(t_{i-1}) + a_i \dot{P}(t_i), \end{aligned} \quad (10)$$

where $D(t_i) = \text{SCF}[R(t_i), P(t_i)]$. Here $\kappa = \delta t^2 \omega^2$, $[P(t_0), \dot{P}(t_0)] = [P(t), \dot{P}(t)]$ and $[P(t + \delta t), \dot{P}(t + \delta t)] = [P(t_m), \dot{P}(t_m)]$. Examples of coefficients a_i and b_i for various number of steps m can be found in Ref. [31].

Before applying a symplectic integration algorithm the value of the dimensionless constant $\kappa = \delta t^2 \omega^2$ must be chosen. Since the SCF convergence is always incomplete the self-consistent solutions $D(t_i)$ will be calculated only approximately. The idea is to choose κ such that the integration in Eq. (10) is always stable under approximate SCF convergence. The optimal choice is the largest κ -value that is consistent with stability, since this choice gives the largest value of ω^2 for a given time step δt . A larger ω^2 corresponds to a higher curvature of the harmonic potential in Eq. (2), which keeps the auxiliary dynamical variables closer to the self-consistent ground state. This reduces the error and/or the cost for the SCF optimization. Based on a linearization of the SCF optimization procedure in Eq. (7) around its exact ground state D^* , we can express an approximately SCF optimized density matrix as

$$D \approx D^* + \Gamma(P - D^*). \quad (11)$$

Here Γ corresponds to the SCF response kernel, which is given as a “super matrix” acting on the matrix $(P - D^*)$. Assuming at least some amount of convergence in the SCF procedure the eigenvalue of Γ with the largest magnitude, γ , will be somewhere in the interval $\gamma \in [-1, 1]$. Following the analysis by Arias *et al.* [26], we insert the linearized SCF expression in Eq. (11), with Γ replaced by γ , in the symplectic integration, Eq. (10), and look at the homogeneous part of the equation,

$$\begin{bmatrix} \dot{P}(t) \\ P(t) \end{bmatrix} = \mathbf{T}_m \mathbf{T}_{m-1} \dots \mathbf{T}_1 \begin{bmatrix} \dot{P}(t - \delta t) \\ P(t - \delta t) \end{bmatrix}. \quad (12)$$

Here \mathbf{T}_i ($i = 1, 2, \dots, m$) are the matrices

$$\mathbf{T}_i = \begin{bmatrix} 1 & b_i \kappa (\gamma - 1) \\ a_i & a_i b_i \kappa (\gamma - 1) + 1 \end{bmatrix}. \quad (13)$$

Equation (12) corresponds to a mapping of the phase space from one time step to the next for a linearized test system with the constant solution $D^*(t) \equiv 0$. The optimal choice of κ is the largest value for which the mapping $\mathbf{T}_m \mathbf{T}_{m-1} \dots \mathbf{T}_1$ has all its eigenvalues on the unit circle for all degrees of incomplete SCF convergence, i.e., for $\gamma \in [-1, 1]$. This case avoids exponentially increasing solutions leading to numerical instabilities or

unphysical dissipation. The mapping in Eq. (12) always preserves the phase space, i.e., the “area” spanned by $\dot{P}(t)$ and $P(t)$, since the determinant of the mapping $\det(\mathbf{T}_m \mathbf{T}_{m-1} \dots \mathbf{T}_1) = \prod_{i=1}^m \det(\mathbf{T}_i) = 1$ for all values of κ and γ .

For the optimal 4th order integration scheme by McLachlan and Atela [31] (where $a_1 \approx 0.515352837$, $a_2 \approx -0.0857820194$, $a_3 \approx 0.441583024$, $a_4 \approx 0.128846158$, $b_1 \approx 0.134496199$, $b_2 \approx -0.224819803$, $b_3 \approx 0.756320001$, and $b_4 \approx 0.334003603$) the largest possible κ -value consistent with stability under incomplete SCF convergence is $\kappa = 4.617$. For the conventional leap-frog or velocity-Verlet scheme [31], as well as the time-reversible Verlet integration in Eq. (8), the largest possible value is $\kappa = 2$.

The great advantage with stability under incomplete SCF convergence is that any amount of convergence suffices for stability, which typically means that only one SCF cycle per force calculation is necessary. This is in contrast to possibly all previous higher-order extrapolation schemes (beyond linear order), for example, the Fock-Matrix dynamics schemes by Pulay and Fogarasi [11] and the higher-order time-reversible schemes by Niklasson *et al.* [14], which are unstable under incomplete SCF convergence, i.e., for the full interval of $\gamma \in [-1, 1]$. Simulations based on those methods will thus diverge if not a certain finite degree of SCF convergence can be guaranteed. To take full advantage of the higher-order symplectic integration schemes the accuracy in the electronic force calculations in Eqs. (6) and (10) should preferably match the accuracy in the integration of the nuclear degrees of freedom. This generally motivates an improved SCF convergence, though it is not a requirement for stability.

Figure 1 shows the fluctuations in the total Born-Oppenheimer energy E^{BO} for a C_2F_4 molecule using Hartree-Fock theory with a standard Gaussian basis set

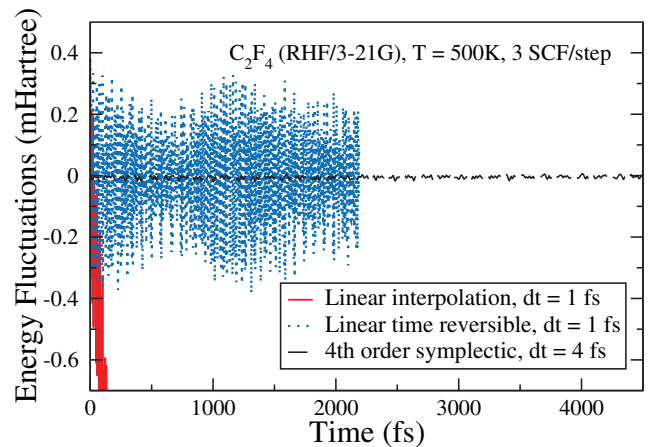


FIG. 1 (color). The fluctuations in total Born-Oppenheimer energy $[E^{\text{BO}}(t) - E_0]$ using three different *ab initio* molecular dynamics approaches described in the text. The integration time length dt is adjusted to allow for a direct comparison between the methods at the same level of computational cost.

TABLE I. Estimated total energy fluctuation amplitudes for the linear time-reversible (A) [14] and the optimized 4th order symplectic (B) integration in Eqs. (9) and (10) [31]. (3 SCF/force calculation).

System	A ($\delta t = 0.5$ fs)	B ($\delta t = 2.0$ fs)
(H ₂ O) ₁₀ (RHF/3-21G)	50 μ hartree	2 μ hartree
C ₂ F ₄ (RHF/3-21G)	80 μ hartree	2 μ hartree
F ₂ (RHF/6-31G)	20 μ hartree	0.07 μ hartree

for three schemes: (i) conventional linear interpolation of the electronic degrees of freedom from two previous time steps, (ii) time-reversible Born-Oppenheimer molecular dynamics, Eq. (8) [14] and (iii) the proposed symplectic integration scheme, Eqs. (9) and (10), with the optimal 4th order coefficients by McLachlan and Atela [31], with $\kappa = 4.617$. In scheme (i) and (ii) the nuclear coordinates are integrated by the velocity-Verlet algorithm [30]. Three SCF fixed point iterations without mixing were used in each force calculation. This example clearly illustrates three levels of performance. The linear interpolation scheme has a rapid decay, whereas no systematic energy drift is seen for the time-reversible and the symplectic integration schemes. Most importantly, the higher-order symplectic approach improves the numerical accuracy, as measured by the amplitude of the energy fluctuations, by over an order of magnitude ($\sim 1/40$) compared to the previous time-reversible scheme, at the same level of computational cost. Similar improvements are found in Table I.

In summary, a Lagrangian formulation of time-reversible Born-Oppenheimer molecular dynamics was proposed, where extended auxiliary electronic degrees of freedom evolve as a harmonic oscillator centered around the adiabatic propagation of the self-consistent electronic ground state. The Lagrangian formulation enables the application of highly efficient and accurate symplectic or geometric integration methods that are shown to be stable under incomplete SCF convergence. For example, using a 4th-order symplectic integration scheme it was demonstrated how the accuracy is improved by over an order of magnitude compared to previous formulations. With a Lagrangian formulation it is also possible to extend micro-canonical simulations to other ensembles, for example, to Nosé-Hoover thermostats or Langevin dynamics. The extended Born-Oppenheimer molecular dynamics presented here, which is an example based on a density matrix representation, may thus open the door to alternative and even more versatile formulations of molecular dynamics.

We gratefully acknowledge the support of the U.S. Department of Energy through the LANL LDRD/ER program for this work. Discussions with N. Bock, M. Challacombe, E. Chisholm, C. Greeff, W. Harrison, E. Holmström, J. Hutter, S. Niklasson, A. Odell,

T. Peery, E. Rubensson, C.J. Tymczak, V. Weber, and J. Wills are gratefully acknowledged.

- [1] C. Leforestier, *J. Chem. Phys.* **68**, 4406 (1978).
- [2] T. Helgaker and E. Uggerud, *Chem. Phys. Lett.* **173**, 145 (1990).
- [3] R. N. Barnett, U. Landman, A. Nitzan, and G. Rajagopal, *J. Chem. Phys.* **94**, 608 (1991).
- [4] R. M. Wentzcovitch and J. L. Martins, *Solid State Commun.* **78**, 831 (1991).
- [5] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, *Rev. Mod. Phys.* **64**, 1045 (1992).
- [6] D. Marx and J. Hutter, *Modern Methods and Algorithms of Quantum Chemistry*, edited by J. Grotendorst (John von Neumann Institute for Computing, Jülich, Germany, 2000), 2nd ed.
- [7] C. C. J. Roothaan, *Rev. Mod. Phys.* **23**, 69 (1951).
- [8] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- [9] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [10] P. Pulay, *Mol. Phys.* **17**, 197 (1969).
- [11] P. Pulay and G. Fogarasi, *Chem. Phys. Lett.* **386**, 272 (2004).
- [12] J. M. Herbert and M. Head-Gordon, *Phys. Chem. Chem. Phys.* **7**, 3269 (2005).
- [13] J. Kolafa, *J. Comput. Chem.* **25**, 335 (2004).
- [14] A. M. N. Niklasson, C. J. Tymczak, and M. Challacombe, *Phys. Rev. Lett.* **97**, 123001 (2006).
- [15] T. D. Kühne, M. Krack, F. R. Mohamed, and M. Parrinello, *Phys. Rev. Lett.* **98**, 066401 (2007).
- [16] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic, New York, 2002), 2nd ed.
- [17] B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics* (Cambridge University Press, Cambridge, England, 2004).
- [18] R. D. Ruth, *IEEE Trans. Nucl. Sci.* **30**, 2669 (1983).
- [19] K. Feng, *J. Comput. Math.* **4**, 279 (1986).
- [20] H. Yoshida, *Phys. Lett. A* **150**, 262 (1990).
- [21] M. Tuckerman, B. J. Berne, and G. J. Martyna, *J. Chem. Phys.* **97**, 1990 (1992).
- [22] R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
- [23] H. B. Schlegel, J. M. Millam, S. S. Iyengar, G. A. Voth, A. D. Daniels, G. Scuseria, and M. J. Frisch, *J. Chem. Phys.* **114**, 9758 (2001).
- [24] J. M. Herbert and M. Head-Gordon, *J. Chem. Phys.* **121**, 11 542 (2004).
- [25] P. Tangney, *J. Chem. Phys.* **124**, 044111 (2006).
- [26] T. Arias, M. Payne, and J. Joannopoulos, *Phys. Rev. Lett.* **69**, 1077 (1992).
- [27] J. Millan, V. Bakken, W. Chen, L. Hase, and H. B. Schlegel, *J. Chem. Phys.* **111**, 3800 (1999).
- [28] C. Raynaud, L. Maron, J.-P. Daudey, and F. Jolibois, *Phys. Chem. Chem. Phys.* **6**, 4226 (2004).
- [29] L. Verlet, *Phys. Rev.* **159**, 98 (1967).
- [30] W. C. Swope, H. C. Andersen, P. H. Berens, and K. Wilson, *J. Chem. Phys.* **76**, 637 (1982).
- [31] R. I. McLachlan and P. Atela, *Nonlinearity* **5**, 541 (1992).