Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. II. Generalizations based on mass-weighting, idempotency, energy conservation and choice of initial conditions

Srinivasan S. Iyengar
Department of Chemistry and Henry Eyring Center for Theoretical Chemistry, University of Utah, Salt Lake City, Utah 84112-0850

H. Bernhard Schlegel and John M. Millam
Department of Chemistry, Wayne State University, Detroit, Michigan 48202-3489

Gregory A. Voth[a)]
Department of Chemistry and Henry Eyring Center for Theoretical Chemistry, University of Utah, Salt Lake City, Utah 84112-0850

Gustavo E. Scuseria
Department of Chemistry and Center for Nanoscale Science and Technology, Rice University, Houston, Texas 77005-1892

Michael J. Frisch
Gaussian, Inc., North Haven, Connecticut 06473-1712

(Received 16 July 2001; accepted 19 September 2001)

A generalization is presented here for a newly developed approach to *ab initio* molecular dynamics, where the density matrix is propagated with Gaussian orbitals. Including a tensorial fictitious mass facilitates the use of larger time steps for the dynamics process. A rigorous analysis of energy conservation is presented and used to control the deviation of the fictitious dynamics trajectory from the corresponding Born–Oppenheimer dynamics trajectory. These generalizations are tested for the case of the $Cl^-(H_2O)_{25}$ cluster. It is found that, even with hydrogen atoms present in the system, no thermostats are necessary to control the exchange of energy between the nuclear and the fictitious electronic degrees of freedom. © *2001 American Institute of Physics.* [DOI: 10.1063/1.1416876]

I. INTRODUCTION

The method of *ab initio* molecular dynamics (AIMD) relies on a calculation of the electronic potential energy surface traversed by the (classical) nuclei ''on-the-fly'' during the dynamics procedure. Both Born–Oppenheimer (BO) molecular dynamics (MD),[1–3] as well as Car–Parrinello (CP) molecular dynamics[4–7] fall into this category. The CP scheme differs from the BO dynamics approach in that the wave functions are propagated together with the classical nuclear degrees of freedom using an extended Lagrangian. This, in turn, relies on an adjustment of the relative nuclear and electronic time scales, which facilitates the adiabatic propagation of the electronic wave function in response to the nuclear motion with suitably large time steps. This adjustment of time scales through the use of a fictitious electronic wave function kinetic energy and mass, enables the CP approach to predict effectively similar nuclear dynamics as MD on the BO surface at significantly reduced cost. In this respect, CP differs from methods which rigorously treat the detailed dynamics (rather than structure) of the electrons (see Ref. 8 and references therein).

It is also interesting to consider molecular dynamics methods from the perspective of multiple-time scale problems which are well-known in statistical mechanics.[9] The slow convergence associated with the multiple-time scales has resulted in recent advances[10] that serve to overcome, partially, the computational problems associated with such systems. In some sense these new methods may be considered to facilitate the CP approach, since both provide suitable computational alternatives to simulate systems with coupled fast–slow subsystems.

We have recently proposed an alternative formalism for AIMD.[11] Our method differs from the standard Car–Parrinello approach,[4–7] as we employ atom-centered Gaussian basis sets and the single-particle density matrix within the extended Lagrangian formalism, whereas the Kohn–Sham molecular orbitals in plane-wave basis sets are chosen as dynamical variables in the CP algorithm. The electronic variables (the density matrix elements in our case) are allowed to have fictitious masses, which leads to a simple adjustment of the relative time scales, thus facilitating the adiabatic propagation of the electronic variables along with the nuclei.

One of the main advantages of using atom-centered Gaussian basis sets in electronic structure calculations is that these are quite effective at describing the wave function for a molecular system due to their compact and localized nature. This allows the use of a smaller number of basis functions to describe the state of a molecular system to within a desired degree of accuracy. Furthermore, it is well established[12–18]

[a)]Electronic mail: voth@chem.utah.edu

that molecular calculations with atom centered functions can be carried out with computational times that scale linearly with system size in the large-system limit.[12–14] This is achieved through the use of Gaussian orbitals in Kohn–Sham density-functional theory (DFT) calculations, fast multipole methods for the Coulomb problem,[15,16] and density matrix search alternatives with sparse matrix multiplication techniques to bypass the $\mathcal{O}(N^3)$ Hamiltonian diagonalization bottleneck.[14,17,18] For similar reasons, our approach using density matrices and Gaussian basis sets for AIMD has the fundamental advantage of leading to $\mathcal{O}(N)$ scaling of computational time with system size. As shown in Ref. 11 and as will be seen in greater detail in this paper, another advantage of using atom-centered Gaussian basis sets and density matrices is the freedom to use smaller values for the fictitious mass of the density matrix. This is crucial to maintaining the desired separation in time scales between the nuclear and electronic motions, *even for hydrogen nuclei*, while maintaining significantly large time steps.

Our method may be contrasted with other approaches to CP that use Gaussian basis sets. Gaussian basis functions within the generalized valence bond (GVB)[19,20] and Hartree–Fock[21] framework, where orbital and wave function coefficients are propagated, have been used; however, the scheme has had some difficulty with energy conservation.[19] In addition floating Gaussian orbitals have also been used[22] in the CP scheme. By contrast, the present approach employs atom-centered Gaussian basis functions.

In this paper we present further generalizations of our Gaussian-basis density matrix AIMD method[11] and also analyze some of its important computational properties. For completeness, in Sec. II, some of the basic concepts from Ref. 11 are briefly reviewed and a new scheme to maintain idempotency of the density matrix is also provided. In Sec. III, the concepts in Sec. II are generalized to include tensorial fictitious mass. This generalization allows larger time steps to be used during the propagation. In Sec. IV the conservation properties of the AIMD method are studied, and an algorithm is provided to help control the difference between a rigorously adiabatic Born–Oppenheimer trajectory and that obtained from the AIMD scheme presented here. In Sec. V some methods are presented for choosing initial conditions for the fictitious density matrix velocity, while in Sec. VI results are given for the energy conservation and electronic adiabaticity corresponding to different choices of the fictitious mass, time step, and numerical integration method for a chloride–water cluster. In Sec. VII the concluding remarks are given.

II. BASIC CONCEPTS OF GAUSSIAN-BASIS DENSITY MATRIX *AB INITIO* MOLECULAR DYNAMICS

An extended Lagrangian describing the combined nuclear-density matrix system, in an orthonormal basis,[11] can be defined as

$$\mathcal{L}= \tfrac{1}{2}Tr(\mathbf{V}^{\mathbf{T}}\mathbf{M}\mathbf{V})+\tfrac{1}{2}\mu\, Tr(\mathbf{W}\mathbf{W})-E(\mathbf{R},\mathbf{P}) - Tr[\boldsymbol{\Lambda}(\mathbf{P}\mathbf{P}-\mathbf{P})], \tag{1}$$

where $\mathbf{M}$, $\mathbf{R}$, and $\mathbf{V}$ are the nuclear masses, positions, and velocities, respectively. The density matrix, density matrix velocity, and the fictitious mass for the density matrix elements are $\mathbf{P}$, $\mathbf{W}$ and μ, respectively. Idempotency and particle number are conserved by the Lagrangian constraint matrix $\boldsymbol{\Lambda}$. Using the principle of stationary action,[23] the Lagrange equations for density matrix and nuclei are obtained as

$$\mu\frac{d^2\mathbf{P}}{dt^2}=-\left[\left.\frac{\partial E(\mathbf{R},\mathbf{P})}{\partial\mathbf{P}}\right|_{\mathbf{R}}+\boldsymbol{\Lambda}\mathbf{P}+\mathbf{P}\boldsymbol{\Lambda}-\boldsymbol{\Lambda}\right], \tag{2}$$

$$\mathbf{M}\frac{d^2\mathbf{R}}{dt^2}=-\left.\frac{\partial E(\mathbf{R},\mathbf{P})}{\partial\mathbf{R}}\right|_{\mathbf{P}}. \tag{3}$$

The term $\partial E(\mathbf{R},\mathbf{P})/\partial\mathbf{P}|_{\mathbf{R}}$, above, implies that the partial derivative is taken under constant $\mathbf{R}$, and similarly for $\partial E(\mathbf{R},\mathbf{P})/\partial\mathbf{R}|_{\mathbf{P}}$. Equations (2) and (3) are used in Ref. 11 to propagate the combined nuclear-density matrix system. Using the velocity Verlet algorithm,[24] the propagation of the density matrix is given by[11]

$$\mathbf{P}_{i+1}=\mathbf{P}_i+\mathbf{W}_i\Delta t-\frac{\Delta t^2}{2\mu}\left[\left.\frac{\partial E(\mathbf{R}_i,\mathbf{P}_i)}{\partial\mathbf{P}}\right|_{\mathbf{R}}+\boldsymbol{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\boldsymbol{\Lambda}_i-\boldsymbol{\Lambda}_i\right], \tag{4}$$

$$\mathbf{W}_{i+1/2}=\mathbf{W}_i-\frac{\Delta t}{2\mu}\left[\left.\frac{\partial E(\mathbf{R}_i,\mathbf{P}_i)}{\partial\mathbf{P}}\right|_{\mathbf{R}}+\boldsymbol{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\boldsymbol{\Lambda}_i-\boldsymbol{\Lambda}_i\right] =\frac{\mathbf{P}_{i+1}-\mathbf{P}_i}{\Delta t}, \tag{5}$$

$$\mathbf{W}_{i+1}=\mathbf{W}_{i+1/2}-\frac{\Delta t}{2\mu}\left[\left.\frac{\partial E(\mathbf{R}_{i+1},\mathbf{P}_{i+1})}{\partial\mathbf{P}}\right|_{\mathbf{R}}+\boldsymbol{\Lambda}_{i+1}\mathbf{P}_{i+1} +\mathbf{P}_{i+1}\boldsymbol{\Lambda}_{i+1}-\boldsymbol{\Lambda}_{i+1}\right]. \tag{6}$$

As in the conjugate gradient density matrix search (CG-DMS) method,[17,18] the energy, $E(\mathbf{R}_i,\mathbf{P}_i)$, is calculated using the McWeeny purification transformation of the density,[25] $\tilde{\mathbf{P}}=3\mathbf{P}^2-2\mathbf{P}^3$. For both DFT and Hartree–Fock, we may write the expression for the energy as

$$E=Tr[\mathbf{h}'\tilde{\mathbf{P}}'+\tfrac{1}{2}\mathbf{G}'(\tilde{\mathbf{P}}')\tilde{\mathbf{P}}']+E_{xc}+V_{NN} =Tr[\mathbf{h}\tilde{\mathbf{P}}+\tfrac{1}{2}\mathbf{G}(\tilde{\mathbf{P}})\tilde{\mathbf{P}}]+E_{xc}+V_{NN}. \tag{7}$$

Here, for both Hartree–Fock and DFT, $\mathbf{h}'$ is the one electron matrix in the nonorthogonal Gaussian basis. The matrix $\mathbf{G}'(\tilde{\mathbf{P}}')$ represents the two electron matrix in the nonorthogonal Gaussian basis for Hartree–Fock calculations, but for DFT it represents the Coulomb potential. The term E_{xc} is the DFT exchange-correlation functional (for Hartree–Fock $E_{xc}=0$), while V_{NN} represents the nuclear repulsion energy. In the orthonormal basis, these matrices are $\mathbf{h}=\mathbf{U}^{-\mathbf{T}}\mathbf{h}'\mathbf{U}^{-\mathbf{1}}$, etc., where the overlap matrix for the nonorthogonal Gaussian basis, $\mathbf{S}'$, is factorized to yield $\mathbf{S}'=\mathbf{U}^{\mathbf{T}}\mathbf{U}$. There are a number of choices for this transformation (e.g., $\mathbf{U}$ can be obtained by Cholesky decomposition,[26] or $\mathbf{U}=\mathbf{S}'^{1/2}$ for Löwdin symmetric orthonormalization). The matrix $\mathbf{U}$ can also

include an additional transformation so that overall rotation of the system is factored out of the propagation of the density. The density matrix in the orthonormal basis, $\mathbf{P}$, is related to the density matrix in the nonorthogonal Gaussian basis, $\mathbf{P}'$, by $\mathbf{P}\equiv\mathbf{U}\mathbf{P}'\mathbf{U}^{\mathbf{T}}$.

The gradient terms involved in the nuclear and density matrix equations of motion, i.e., Eqs. (2) and (3), are given by[11]

$$\left.\frac{\partial E(\mathbf{R},\mathbf{P})}{\partial \mathbf{P}}\right|_{\mathbf{R}} = 3\mathbf{F}\mathbf{P}+3\mathbf{P}\mathbf{F}-2\mathbf{F}\mathbf{P}^2-2\mathbf{P}\mathbf{F}\mathbf{P}-2\mathbf{P}^2\mathbf{F} \tag{8}$$

and

$$\begin{aligned}\left.\frac{\partial E(\mathbf{R},\mathbf{P})}{\partial \mathbf{R}}\right|_{\mathbf{P}} &= Tr\left[\frac{d\mathbf{h}'}{d\mathbf{R}}\tilde{\mathbf{P}}'+\frac{1}{2}\left.\frac{\partial \mathbf{G}'}{\partial \mathbf{R}}\right|_{\mathbf{P}'}\tilde{\mathbf{P}}'\right]\\ &\quad -Tr\left[\mathbf{F}'\mathbf{U}^{-1}\frac{d\mathbf{U}}{d\mathbf{R}}\tilde{\mathbf{P}}'+\tilde{\mathbf{P}}'\frac{d\mathbf{U}^{\mathbf{T}}}{d\mathbf{R}}\mathbf{U}^{-\mathbf{T}}\mathbf{F}'\right]\\ &\quad +\left.\frac{\partial E_{xc}}{\partial \mathbf{R}}\right|_{\mathbf{P}}+\frac{\partial V_{NN}}{\partial \mathbf{R}}\\ &= Tr\left[\mathbf{U}^{-\mathbf{T}}\frac{d\mathbf{h}'}{d\mathbf{R}}\mathbf{U}^{-\mathbf{1}}\tilde{\mathbf{P}}\right.\\ &\quad \left.+\frac{1}{2}\mathbf{U}^{-\mathbf{T}}\left.\frac{\partial \mathbf{G}'(\tilde{\mathbf{P}}')}{\partial \mathbf{R}}\right|_{\mathbf{P}'}\mathbf{U}^{-\mathbf{1}}\tilde{\mathbf{P}}\right]\\ &\quad -Tr\left[\mathbf{F}\frac{d\mathbf{U}}{d\mathbf{R}}\mathbf{U}^{-\mathbf{1}}\tilde{\mathbf{P}}+\tilde{\mathbf{P}}\mathbf{U}^{-\mathbf{T}}\frac{d\mathbf{U}^{\mathbf{T}}}{d\mathbf{R}}\mathbf{F}\right]+\left.\frac{\partial E_{xc}}{\partial \mathbf{R}}\right|_{\mathbf{P}}\\ &\quad +\frac{\partial V_{NN}}{\partial \mathbf{R}}.\end{aligned} \tag{9}$$

Again, the Fock matrix in the nonorthogonal basis-set

$$\mathbf{F}'_{\nu,\sigma}\equiv\mathbf{h}'_{\nu,\sigma}+\mathbf{G}'(\tilde{\mathbf{P}}')_{\nu,\sigma}+\frac{\partial E_{xc}}{\partial \mathbf{P}'_{\nu,\sigma}}, \tag{10}$$

is related to the Fock matrix in the orthonormal basis by $\mathbf{F}=\mathbf{U}^{-\mathbf{T}}\mathbf{F}'\mathbf{U}^{-\mathbf{1}}$.

The Lagrangian constraint matrix for time step i, $\mathbf{\Lambda}_i$, in Eqs. (4)–(6) is chosen to satisfy $Tr[\mathbf{P}_{i+1}]=N_e$ and $\mathbf{P}_{i+1}^2=\mathbf{P}_{i+1}$. One can solve directly for $\mathbf{P}_{i+1}$ by a simple iterative procedure that minimizes $Tr[(\mathbf{P}_{i+1}^2-\mathbf{P}_{i+1})^2]$. Starting with

$$\mathbf{P}_{i+1}=\mathbf{P}_i+\mathbf{W}_i\Delta t-\left[\left.\frac{\partial E(\mathbf{R}_i,\mathbf{P}_i)}{\partial \mathbf{P}}\right|_{\mathbf{R}}\right]\frac{\Delta t^2}{2\mu}, \tag{11}$$

which is obtained by choosing $\mathbf{\Lambda}_i=0$ in Eq. (4), the idempotency of $\mathbf{P}_{i+1}$ is improved iteratively using

$$\mathbf{P}_{i+1}\leftarrow\mathbf{P}_{i+1}+\mathbf{P}_i\mathbf{T}\mathbf{P}_i+\mathbf{Q}_i\mathbf{T}\mathbf{Q}_i, \tag{12}$$

where $\mathbf{T}=\tilde{\mathbf{P}}_{i+1}-\mathbf{P}_{i+1}=3\mathbf{P}_{i+1}^2-2\mathbf{P}_{i+1}^3-\mathbf{P}_{i+1}$. The iteration converges rapidly and is stopped when $\{Tr[(\mathbf{P}_{i+1}^2-\mathbf{P}_{i+1})^2]\}^{1/2}/N<10^{-12}$. The above algorithm comprises an update of only the occupied–occupied and virtual–virtual blocks of $\mathbf{P}_{i+1}$, since the constraint terms satisfy

$$\mathbf{P}_i[\mathbf{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\mathbf{\Lambda}_i-\mathbf{\Lambda}_i]\mathbf{Q}_i=\mathbf{Q}_i[\mathbf{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\mathbf{\Lambda}_i-\mathbf{\Lambda}_i]\mathbf{P}_i=0, \tag{13}$$

where $\mathbf{Q}=\mathbf{I}-\mathbf{P}$ and $\mathbf{I}$ is the identity matrix. The final density matrix velocity needs to satisfy the time derivative of the idempotency condition, given by

$$\mathbf{W}_{i+1}\mathbf{P}_{i+1}+\mathbf{P}_{i+1}\mathbf{W}_{i+1}=\mathbf{W}_{i+1}. \tag{14}$$

This can be solved exactly to yield $\mathbf{\Lambda}_{i+1}=[2\mathbf{P}_{i+1}\mathbf{W}_{i+1}^*\mathbf{P}_{i+1}-\mathbf{W}_{i+1}^*](2\mu/\Delta t)$, where $\mathbf{W}_{i+1}^*=\mathbf{W}_{i+1/2}-[\partial E(\mathbf{R}_{i+1},\mathbf{P}_{i+1})/\partial\mathbf{P}|_{\mathbf{R}}](\Delta t/2\mu)$. With this, Eq. (6) simplifies to

$$\mathbf{W}_{i+1}=\mathbf{W}_{i+1}^*-\mathbf{P}_{i+1}\mathbf{W}_{i+1}^*\mathbf{P}_{i+1}-\mathbf{Q}_{i+1}\mathbf{W}_{i+1}^*\mathbf{Q}_{i+1}. \tag{15}$$

Equation (15) estimates the value of $\mathbf{\Lambda}_{i+1}$, as required by Eq. (6). This estimate for $\mathbf{\Lambda}_{i+1}$ is refined in the next step of the propagation, so that idempotency is satisfied for the density matrix $\mathbf{P}_{i+2}$ using Eqs. (11) and (12).

It should be noted that the iterative scheme in Eqs. (11) and (12) corresponds to starting with $\mathbf{\Lambda}_i=0$ [as can be verified from Eq. (4)]. This, however, is not the only possible choice. A better initial value for $\mathbf{\Lambda}_i$ may be obtained from the second-order Taylor expansion for $\mathbf{P}_{i+1}$

$$\mathbf{P}_{i+1}=\mathbf{P}_i+\mathbf{W}_i\Delta t+\frac{d\mathbf{W}_i}{dt}\frac{\Delta t^2}{2}. \tag{16}$$

As seen from Eq. (2)

$$\frac{d\mathbf{W}_i}{dt}=-\frac{1}{\mu}\left[\left.\frac{\partial E(\mathbf{R}_i,\mathbf{P}_i)}{\partial \mathbf{P}}\right|_{\mathbf{R}}+\mathbf{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\mathbf{\Lambda}_i-\mathbf{\Lambda}_i\right]. \tag{17}$$

To obtain an alternative expression for $(d\mathbf{W}_i/dt)$, we consider the second derivative of the idempotency constraint, i.e.,

$$\frac{d\mathbf{W}_i}{dt}=\mathbf{P}_i\frac{d\mathbf{W}_i}{dt}+2\mathbf{W}_i\mathbf{W}_i+\frac{d\mathbf{W}_i}{dt}\mathbf{P}_i. \tag{18}$$

From Eqs. (8), (17), and (18), it is seen that

$$\mathbf{P}_i\frac{d\mathbf{W}_i}{dt}\mathbf{P}_i=-2\mathbf{P}_i[\mathbf{W}_i\mathbf{W}_i]\mathbf{P}_i=-\frac{1}{\mu}\mathbf{P}_i\mathbf{\Lambda}_i\mathbf{P}_i, \tag{19}$$

$$\mathbf{Q}_i\frac{d\mathbf{W}_i}{dt}\mathbf{Q}_i=2\mathbf{Q}_i[\mathbf{W}_i\mathbf{W}_i]\mathbf{Q}_i=\frac{1}{\mu}\mathbf{Q}_i\mathbf{\Lambda}_i\mathbf{Q}_i \tag{20}$$

and

$$\mathbf{P}_i\frac{d\mathbf{W}_i}{dt}\mathbf{Q}_i=-\frac{1}{\mu}\mathbf{P}_i\left[\left.\frac{\partial E(\mathbf{R}_i,\mathbf{P}_i)}{\partial \mathbf{P}}\right|_{\mathbf{R}}\right]\mathbf{Q}_i=\frac{1}{\mu}\mathbf{P}_i\mathbf{F}_i\mathbf{Q}_i. \tag{21}$$

From Eqs. (19), (20) and (4), one obtains

$$\begin{aligned}\frac{1}{\mu}[\mathbf{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\mathbf{\Lambda}_i-\mathbf{\Lambda}_i] &= \frac{1}{\mu}[\mathbf{P}_i\mathbf{\Lambda}_i\mathbf{P}_i-\mathbf{Q}_i\mathbf{\Lambda}_i\mathbf{Q}_i]\\ &= 2[\mathbf{P}_i(\mathbf{W}_i\mathbf{W}_i)\mathbf{P}_i-\mathbf{Q}_i(\mathbf{W}_i\mathbf{W}_i)\mathbf{Q}_i],\end{aligned} \tag{22}$$

which is accurate to second-order, and can be used as an initial guess [instead of the $\mathbf{\Lambda}_i=0$ guess used in Eq. (11)], for the iterative scheme in Eq. (12).

III. MASS-WEIGHTING AND IDEMPOTENCY

Consider a partitioning of the $\mathbf{P}_{i+1}$ matrix given in Eq. (4). In particular, consider the projected blocks: $[\mathbf{P}_i\mathbf{P}_{i+1}\mathbf{P}_i]$, $[\mathbf{P}_i\mathbf{P}_{i+1}\mathbf{Q}_i]$, $[\mathbf{Q}_i\mathbf{P}_{i+1}\mathbf{P}_i]$, and $[\mathbf{Q}_i\mathbf{P}_{i+1}\mathbf{Q}_i]$. It is clear from Eqs. (4) and (8) that

$$\mathbf{P}_i[\mathbf{P}_{i+1}]\mathbf{P}_i=\mathbf{P}_i+\frac{\Delta t^2}{2\mu}\mathbf{P}_i\mathbf{\Lambda}_i\mathbf{P}_i, \tag{23}$$

$$\mathbf{P}_i[\mathbf{P}_{i+1}]\mathbf{Q}_i=\mathbf{P}_i\{\mathbf{W}_i\Delta t\}\mathbf{Q}_i-\mathbf{P}_i\left\{\frac{\Delta t^2}{2\mu}\frac{\partial E(\mathbf{R}_i,\mathbf{P}_i)}{\partial\mathbf{P}}\bigg|_{\mathbf{R}}\right\}\mathbf{Q}_i, \tag{24}$$

and

$$\mathbf{Q}_i[\mathbf{P}_{i+1}]\mathbf{Q}_i=-\frac{\Delta t^2}{2\mu}\mathbf{Q}_i\mathbf{\Lambda}_i\mathbf{Q}_i. \tag{25}$$

Note that $\mathbf{P}_i[\partial E(\mathbf{R}_i,\mathbf{P}_i)/\partial\mathbf{P}|_{\mathbf{R}}]\mathbf{P}_i=\mathbf{Q}_i[\partial E(\mathbf{R}_i,\mathbf{P}_i)/\partial\mathbf{P}|_{\mathbf{R}}]\mathbf{Q}_i=0$ from Eq. (8), and $\mathbf{P}_i\mathbf{W}_i\mathbf{P}_i=\mathbf{Q}_i\mathbf{W}_i\mathbf{Q}_i=0$ from Eq. (14). Also, the term $[\mathbf{Q}_i\mathbf{P}_{i+1}\mathbf{P}_i]\equiv[\mathbf{P}_i\mathbf{P}_{i+1}\mathbf{Q}_i]^T$, by definition, since all matrices are real-symmetric. Hence, the Lagrangian constraint matrix $\mathbf{\Lambda}_i$ affects the occupied–occupied [see Eq. (23)] and virtual–virtual [see Eq. (25)] blocks of the density matrix (with reference to $\mathbf{P}_i$), but not the occupied–virtual block [Eq. (24)]. The occupied–virtual block of the density matrix is free to respond to the fictitious velocity of the density matrix, i.e., $\mathbf{W}_i$, and the force on the density matrix, i.e., $\{-\partial E(\mathbf{R}_i,\mathbf{P}_i)/\partial\mathbf{P}|_{\mathbf{R}}\}$. Furthermore, the force on the density matrix does not affect the dynamics of the occupied–occupied and virtual–virtual blocks.

To understand the dynamics of the density matrix more clearly, let us consider the propagation of the individual blocks of $\mathbf{P}_{i+1}$ as defined by Eqs. (23)–(25). The occupied–virtual block, denoted by Eq. (24), dictates how the virtual orbitals at time step i couple with the occupied orbitals at time step i to produce the density matrix at $(i+1)$. This coupling is, however, governed by the magnitude of the fictitious density velocity (which may be arbitrary) and the force on the density matrix, as shown in Eq. (24). To control this arbitrariness in the dynamics of the occupied–virtual block of the density matrix, and in particular the core–virtual block, a mass weighting scheme is introduced here by generalizing the fictitious mass to a matrix of masses, μ. (Note that the fictitious mass matrix is represented by the symbol μ, whereas its scalar counterpart, used earlier in this paper and in Ref. 11, is represented simply as μ.) As will be seen in the results section, this has the effect of maintaining good energy conservation even for larger time steps, as compared to those allowed when the fictitious mass is treated as a scalar quantity.

Another rationalization for mass-weighting may be obtained from considering the following fact. For dynamics on the Born–Oppenheimer surface, the density matrix elements for the core orbitals of an atom change more slowly than for the valence orbitals since the core is more tightly bound to the nucleus. Hence, it is useful to have a larger mass for the core orbitals and a smaller mass for the valence orbitals in order to increase integration efficiency.

To incorporate mass-weighting into our formalism, the mass-weighted density velocity is defined as $[\mu^{1/4}\mathbf{W}\mu^{1/4}]$. The Lagrangian in Eq. (1) may then be generalized as

$$\mathcal{L}=\tfrac{1}{2}Tr(\mathbf{V}^{\mathbf{T}}\mathbf{M}\mathbf{V})+\tfrac{1}{2}Tr([\mu^{1/4}\mathbf{W}\mu^{1/4}]^2)-E(\mathbf{R},\mathbf{P})-Tr[\mathbf{\Lambda}(\mathbf{P}\mathbf{P}-\mathbf{P})], \tag{26}$$

which yields the Lagrange equation for the density matrix, given by

$$\mu^{1/2}\frac{d^2\mathbf{P}}{dt^2}\mu^{1/2}=-\left[\frac{\partial E(\mathbf{R},\mathbf{P})}{\partial\mathbf{P}}\bigg|_{\mathbf{R}}+\mathbf{\Lambda}\mathbf{P}+\mathbf{P}\mathbf{\Lambda}-\mathbf{\Lambda}\right], \tag{27}$$

where we have assumed that μ does not depend on t. Eq. (27) can be integrated as usual to obtain

$$\mathbf{P}_{i+1}=\mathbf{P}_i+\mathbf{W}_i\Delta t-\frac{\Delta t^2}{2}\mu^{-1/2}\left[\frac{\partial E(\mathbf{R}_i,\mathbf{P}_i)}{\partial\mathbf{P}}\bigg|_{\mathbf{R}}+\mathbf{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\mathbf{\Lambda}_i-\mathbf{\Lambda}_i\right]\mu^{-1/2}, \tag{28}$$

$$\mathbf{W}_{i+1/2}=\mathbf{W}_i-\frac{\Delta t}{2}\mu^{-1/2}\left[\frac{\partial E(\mathbf{R}_i,\mathbf{P}_i)}{\partial\mathbf{P}}\bigg|_{\mathbf{R}}+\mathbf{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\mathbf{\Lambda}_i-\mathbf{\Lambda}_i\right]\mu^{-1/2}=\frac{\mathbf{P}_{i+1}-\mathbf{P}_i}{\Delta t}, \tag{29}$$

$$\mathbf{W}_{i+1}=\mathbf{W}_{i+1/2}-\frac{\Delta t}{2}\mu^{-1/2}\left[\frac{\partial E(\mathbf{R}_{i+1},\mathbf{P}_{i+1})}{\partial\mathbf{P}}\bigg|_{\mathbf{R}}+\mathbf{\Lambda}_{i+1}\mathbf{P}_{i+1}+\mathbf{P}_{i+1}\mathbf{\Lambda}_{i+1}-\mathbf{\Lambda}_{i+1}\right]\mu^{-1/2}. \tag{30}$$

Unlike the quantity $\mathbf{P}_i\{(1/\mu)[\mathbf{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\mathbf{\Lambda}_i-\mathbf{\Lambda}_i]\}\mathbf{Q}_i$ [see Eq. (13)] for the scalar mass case, the corresponding quantity for the mass-matrix case, i.e.,

$$\mathbf{P}_i\{\mu^{-1/2}[\mathbf{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\mathbf{\Lambda}_i-\mathbf{\Lambda}_i]\mu^{-1/2}\}\mathbf{Q}_i\neq 0, \tag{31}$$

if $[\mu,\mathbf{P}_i]\neq 0$. As a result, mass-weighting allows for the idempotency force, $\{\mu^{-1/2}[\mathbf{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\mathbf{\Lambda}_i-\mathbf{\Lambda}_i]\mu^{-1/2}\}$, to affect the occupied–virtual blocks of the density matrix in (28) [to control the arbitrary effect of $\mathbf{W}_i$ in Eq. (24)]. Hence, by appropriate choice of the μ matrix, a method is obtained to control the occupied–virtual (and in particular the core-virtual) part of the density matrix, since now the total force in Eq. (28) (i.e., including the gradient term as well as the force due to deviation from idempotency) acts on all four projected blocks

$$\mathbf{P}_i[\mathbf{P}_{i+1}]\mathbf{P}_i=\mathbf{P}_i-\frac{\Delta t^2}{2}\mathbf{P}_i\left\{\mu^{-1/2}\left[\frac{\partial E(\mathbf{R}_i,\mathbf{P}_i)}{\partial\mathbf{P}}\bigg|_{\mathbf{R}}+\mathbf{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\mathbf{\Lambda}_i-\mathbf{\Lambda}_i\right]\mu^{-1/2}\right\}\mathbf{P}_i, \tag{32}$$

$$\mathbf{P}_i[\mathbf{P}_{i+1}]\mathbf{Q}_i=\mathbf{P}_i\{\mathbf{W}_i\Delta t\}\mathbf{Q}_i-\frac{\Delta t^2}{2}\mathbf{P}_i\left\{\mu^{-1/2}\left[\frac{\partial E(\mathbf{R}_i,\mathbf{P}_i)}{\partial\mathbf{P}}\bigg|_{\mathbf{R}}+\mathbf{\Lambda}_i\mathbf{P}_i+\mathbf{P}_i\mathbf{\Lambda}_i-\mathbf{\Lambda}_i\right]\mu^{-1/2}\right\}\mathbf{Q}_i, \tag{33}$$

and

$$\mathbf{Q}_i[\mathbf{P}_{i+1}]\mathbf{Q}_i = -\frac{\Delta t^2}{2}\mathbf{Q}_i\left\{\underline{\mu}^{-1/2}\left[\left.\frac{\partial E(\mathbf{R}_i,\mathbf{P}_i)}{\partial \mathbf{P}}\right|_{\mathbf{R}} + \mathbf{\Lambda}_i\mathbf{P}_i + \mathbf{P}_i\mathbf{\Lambda}_i - \mathbf{\Lambda}_i\right]\underline{\mu}^{-1/2}\right\}\mathbf{Q}_i. \tag{34}$$

As a consequence of Eq. (31), an alternate iteration scheme is needed to conserve idempotency for the mass-weighting case, since the scheme presented in Ref. 11 and in Sec. II [see Eq. (12)] only updates the occupied–occupied and virtual–virtual blocks of the density matrix, as is seen from Eq. (13). In this section two new schemes are presented to preserve idempotency for the mass-matrix case.

In the first scheme, an approach similar to that in Ref. 11 is adopted, by starting with

$$\mathbf{P}^*_{i+1} = \mathbf{P}_i + \mathbf{W}_i\Delta t - (\Delta t^2/2) \times\{\underline{\mu}^{-1/2}[\partial E(\mathbf{R}_i,\mathbf{P}_i)/\partial \mathbf{P}|_{\mathbf{R}}]\underline{\mu}^{-1/2}\}. \tag{35}$$

From Eq. (28) it is noted that the correction from the Lagrangian multiplier matrix has the form $[\underline{\mu}^{-1/2}\mathbf{A}\underline{\mu}^{-1/2}]$, where the matrix $\mathbf{A}$ has only occupied–occupied and virtual–virtual blocks. Using Eq. (12) as a guide, one may then iterate

$$\mathbf{P}_{i+1} \leftarrow \mathbf{P}^*_{i+1} + \underline{\mu}^{-1/2}[\mathbf{P}_i\mathbf{T}\mathbf{P}_i + \mathbf{Q}_i\mathbf{T}\mathbf{Q}_i]\underline{\mu}^{-1/2}, \tag{36}$$

where $\mathbf{T} = \underline{\mu}^{1/2}[\tilde{\mathbf{P}}_{i+1} - \mathbf{P}^*_{i+1}]\underline{\mu}^{1/2} = \underline{\mu}^{1/2}[3\mathbf{P}^2_{i+1} - 2\mathbf{P}^3_{i+1} - \mathbf{P}^*_{i+1}]\underline{\mu}^{1/2}$. As in the scalar mass case, the iterative scheme here converges rapidly and is stopped when $\{Tr[(\mathbf{P}^2_{i+1} - \mathbf{P}_{i+1})^2]\}^{1/2}/N < 10^{-12}$, where N is the size of the matrices involved (i.e., N is the number of basis functions).

To obtain $\mathbf{W}_{i+1}$, it is necessary to satisfy the time derivative of the idempotency condition, i.e., Eq. (14). This is solved iteratively, by first choosing

$$\mathbf{W}^*_{i+1} = \mathbf{W}_{i+1/2} - (\Delta t/2) \times\{\underline{\mu}^{-1/2}[\partial E(\mathbf{R}_{i+1},\mathbf{P}_{i+1})/\partial \mathbf{P}|_{\mathbf{R}}]\underline{\mu}^{-1/2}\}, \tag{37}$$

and then iterating

$$\mathbf{W}_{i+1} \leftarrow \mathbf{W}^*_{i+1} + \underline{\mu}^{-1/2}\{\mathbf{P}_{i+1}\tilde{\mathbf{T}}_{i+1}\mathbf{P}_{i+1} + \mathbf{Q}_{i+1}\tilde{\mathbf{T}}_{i+1}\mathbf{Q}_{i+1}\}\underline{\mu}^{-1/2}, \tag{38}$$

where $\tilde{\mathbf{T}}_{i+1} = \underline{\mu}^{1/2}[\tilde{\mathbf{W}}_{i+1} - \mathbf{W}^*_{i+1}]\underline{\mu}^{1/2}$ and $\tilde{\mathbf{W}}_{i+1} = d\tilde{\mathbf{P}}_{i+1}/dt = \mathbf{P}_{i+1}\mathbf{W}_{i+1}\mathbf{Q}_{i+1} + \mathbf{Q}_{i+1}\mathbf{W}_{i+1}\mathbf{P}_{i+1}$. Here again, the iteration converges rapidly and is stopped when $\{Tr[(\mathbf{W}_{i+1}\mathbf{P}_{i+1} + \mathbf{P}_{i+1}\mathbf{W}_{i+1} - \mathbf{W}_{i+1})^2]\}^{1/2}/N < 10^{-12}$.

It should be noted that, as in the scalar mass case, the initial guess for $\mathbf{P}_{i+1}$ and $\mathbf{W}_{i+1}$ above, i.e., the expressions in Eqs. (35) and (37), correspond to choosing $\mathbf{\Lambda}_i = 0$. A higher-order, initial guess for an idempotent $\mathbf{P}_{i+1}$ may be obtained by using the fact that $\mathbf{W}_i\mathbf{P}_i + \mathbf{P}_i\mathbf{W}_i = \mathbf{W}_i$ is already satisfied by the use of Eqs. (37) and (38), for the previous time step i. Since, a solution to $\mathbf{W}_i$, requires having an approximation to $\mathbf{\Lambda}_i$ [see Eq. (30)], this approximation may be used in generating an initial guess for $\mathbf{P}_{i+1}$. Thus, using Eqs. (28)–(30), one obtains

$$\mathbf{P}_{i+1} = \mathbf{P}_i + 2\mathbf{W}_i\Delta t - \mathbf{W}_{i-1/2}\Delta t, \tag{39}$$

which is a higher order initial guess for $\mathbf{P}_{i+1}$ and may be used in conjunction with the iterative scheme described in Eq. (36). The term $\mathbf{W}_{i-1/2}$ in Eq. (39) is given by Eq. (29), for the time step i.

IV. ENERGY CONSERVATION AND ADIABATIC CONTROL

To study the energy conservation properties for the Lagrangian in Eq. (26), one can write down the conjugate Hamiltonian, which is given by the Legendre transform[23] of the Lagrangian

$$\mathcal{H}(\mathbf{P},\mathcal{W},\mathbf{R},\mathcal{V},t) = Tr(\mathcal{W}\mathbf{W}) + Tr(\mathcal{V}^{\mathbf{T}}\mathbf{V}) - \mathcal{L}(\mathbf{P},\mathbf{W},\mathbf{R},\mathbf{V},t), \tag{40}$$

where $\mathcal{W}$ and $\mathcal{V}$ are the conjugate momenta for $\mathbf{P}$ and $\mathbf{R}$, respectively, and are given by

$$\mathcal{W} = \frac{\partial \mathcal{L}}{\partial \mathbf{W}} = \underline{\mu}^{1/2}\mathbf{W}\underline{\mu}^{1/2} \tag{41}$$

and

$$\mathcal{V} = \frac{\partial \mathcal{L}}{\partial \mathbf{V}} = \mathbf{M}\mathbf{V}. \tag{42}$$

Using Eqs. (41) and (42) in Eq. (40), one obtains the conjugate Hamiltonian as

$$\mathcal{H} = \tfrac{1}{2}Tr(\mathcal{V}^{\mathbf{T}}\mathbf{M}^{-1}\mathcal{V}) + \tfrac{1}{2}Tr(\mathcal{W}\underline{\mu}^{-1/2}\mathcal{W}\underline{\mu}^{-1/2}) + E(\mathbf{R},\mathbf{P}) + Tr[\mathbf{\Lambda}(\mathbf{P}\mathbf{P} - \mathbf{P})]. \tag{43}$$

To study the conservation property for the above Hamiltonian, consider the total derivative of $\mathcal{H}$ with respect to t, i.e.,

$$\begin{aligned}\frac{d\mathcal{H}}{dt} &= Tr\left[\frac{\partial \mathcal{H}}{\partial \mathbf{P}}\frac{d\mathbf{P}}{dt} + \frac{\partial \mathcal{H}}{\partial \mathcal{W}}\frac{d\mathcal{W}}{dt} + \frac{\partial \mathcal{H}}{\partial \mathbf{R}}\frac{d\mathbf{R}}{dt} + \frac{\partial \mathcal{H}}{\partial \mathcal{V}}\frac{d\mathcal{V}}{dt}\right] \\ &= Tr\left[\left(\left.\frac{\partial E(\mathbf{R},\mathbf{P})}{\partial \mathbf{P}}\right|_{\mathbf{R}} + \mathbf{\Lambda}\mathbf{P} + \mathbf{P}\mathbf{\Lambda} - \mathbf{\Lambda}\right)\mathbf{W}\right] \\ &\quad + Tr\left[\mathcal{W}\frac{d^2\mathbf{P}}{dt^2}\right] + Tr\left[\left.\frac{\partial E(\mathbf{R},\mathbf{P})}{\partial \mathbf{R}}\right|_{\mathbf{P}}\mathbf{V}\right] \\ &\quad + Tr\left[\mathbf{V}\mathbf{M}\frac{d^2\mathbf{R}}{dt^2}\right],\end{aligned} \tag{44}$$

where the definitions in Eqs. (41) and (42) have been used and it is also assumed that $\underline{\mu}$ is time independent. Now, using the Lagrange equations of motion, i.e., Eqs. (27) and (3), it follows that:

$$\frac{d\mathcal{H}}{dt} = 0. \tag{45}$$

This proves that the Hamiltonian in Eq. (43) [and hence the Lagrangian in Eq. (26) and its special scalar fictitious mass case in Eq. (1)] represents a conservative system, i.e., the total energy associated with this system should always be a constant.

Let us now partition the Hamiltonian such that

$$\mathcal{H} = \mathcal{H}_{\text{real}} + \mathcal{H}_{\text{fict}}, \tag{46}$$

where $\mathcal{H}_{\rm real}$ and $\mathcal{H}_{\rm fict}$ represent the real and fictitious parts of the full Hamiltonian and are defined as

$$\mathcal{H}_{\rm real}=\tfrac{1}{2}Tr(\mathcal{V}^{\mathbf{T}}\mathbf{M}^{-1}\mathcal{V})+E(\mathbf{R},\mathbf{P})+Tr[\mathbf{\Lambda}(\mathbf{PP}-\mathbf{P})] \quad (47)$$

and

$$\mathcal{H}_{\rm fict}=\tfrac{1}{2}Tr[\mathcal{W}\mu^{-1/2}\mathcal{W}\mu^{-1/2}]. \quad (48)$$

The definitions for $\mathcal{H}_{\rm real}$ and $\mathcal{H}_{\rm fict}$, above, are consistent with the usual definitions[5] for the real and fictitious energy in the CP scheme. Using Eq. (27), it then follows that:

$$\begin{aligned}\frac{d\mathcal{H}_{\rm fict}}{dt}&=Tr\left[\frac{\partial\mathcal{H}_{\rm fict}}{\partial\mathcal{W}}\frac{d\mathcal{W}}{dt}\right]\\&=Tr\left[\mu^{-1/2}\mathcal{W}\mu^{-1/2}\frac{d\mathcal{W}}{dt}\right]\\&=Tr\left[\mathbf{W}\mu^{1/2}\frac{d^2\mathbf{P}}{dt^2}\mu^{1/2}\right]\\&=-Tr\left[\mathbf{W}\left(\left.\frac{\partial E(\mathbf{R},\mathbf{P})}{\partial\mathbf{P}}\right|_{\mathbf{R}}+\mathbf{\Lambda P}+\mathbf{P\Lambda}-\mathbf{\Lambda}\right)\right]. \quad (49)\end{aligned}$$

In accordance with Eq. (45), $d\mathcal{H}_{\rm real}/dt$ is simply the negative of $d\mathcal{H}_{\rm fict}/dt$. The value of $d\mathcal{H}_{\rm fict}/dt$ (and hence $d\mathcal{H}_{\rm real}/dt$), for the special case of scalar fictitious masses is obtained by noting that in this case $\mu^{1/2}\equiv\mu^{1/2}\mathbf{I}$, where $\mathbf{I}$ is the identity and $\mu^{1/2}$ is the square root of the scalar fictitious mass seen in Sec. II and Ref. 11.

A few comments regarding Eq. (49) are now in order. (i) Equation (49) represents the rate at which a trajectory generated from the solution to Eqs. (27) and (3) [or the special scalar mass equations (2) and (3)] deviates from a standard Born–Oppenheimer trajectory[1–3] (since for a trajectory that exactly follows the Born–Oppenheimer surface $d\mathcal{H}_{\rm real}/dt=0$). Furthermore, since Eq. (49) exhibits a dependence on $\mu^{1/2}$, the magnitude of the fictitious mass (or in the mass-weighting case the L^2-norm of the fictitious mass matrix) determines the deviations from the Born–Oppenheimer trajectory.[27] This is an important result, since as noted later in Sec. VI and as seen in Ref. 11, the current method indeed allows for relatively small values of the fictitious mass, while maintaining larger time steps. Additionally, in molecular dynamics it is usually defined that the dynamics of nuclei occur on the electronic ground state potential energy surface (in the absence of conical intersections[28–30]), hence it is important to keep the terms in Eq. (49) small. (ii) Since it is also desirable that time averages of various properties derived from the actual Born–Oppenheimer dynamics be identical to those derived from a fictitious dynamics calculation, like that described here, it is necessary that the quantities $d\mathcal{H}_{\rm fict}/dt$ and $d\mathcal{H}_{\rm real}/dt$ oscillate about zero. In the rest of this section, an algorithm will be derived that is useful to control and monitor the magnitude of the quantities in Eq. (49) during the progress of a simulation.

The first condition described above may be achieved by checking the absolute value of the right-hand side in Eq. (49) for every time step. If this absolute value is less than some user-defined threshold, the dynamics is representative of the Born–Oppenheimer trajectory to within that threshold. Note that it is also possible to check the value of the fictitious kinetic energy at the point where $d\mathcal{H}_{\rm fict}/dt\approx 0$, i.e.

$$\mathcal{H}_{\rm fict}|_{d\mathcal{H}_{\rm fict}/dt=0}\equiv\tfrac{1}{2}Tr[\mu^{1/2}\mathbf{W}\mu^{1/2}\mathbf{W}]|_{d\mathcal{H}_{\rm fict}/dt=0}. \quad (50)$$

The value of the fictitious kinetic energy on the right-hand side of Eq. (50) represents the maximum (or minimum) deviation of the fictitious dynamics trajectory from the *true* Born–Oppenheimer energy surface.

The second condition may be checked by evaluating

$$\int_{t=0}^{t=T}dt\,\frac{d\mathcal{H}_{\rm fict}}{dt}\equiv\mathcal{H}_{\rm fict}|_{t=T}-\mathcal{H}_{\rm fict}|_{t=0}, \quad (51)$$

where $t=T$ represents the total propagation time. The value of the right-hand side in Eq. (51), along with a chosen set of stationary points from Eq. (50), will help ascertain the oscillatory nature of $\mathcal{H}_{\rm fict}(t)$.

Equations (49)–(51) together provide a method to gauge the deviation of the fictitious dynamics from the *true* Born–Oppenheimer dynamics trajectory. If any of the conditions above is not satisfied, then it will be necessary to go back to the previous time step and re-propagate, but now with either a smaller time step or a smaller fictitious mass, chosen such that the above conditions are met to within the user-defined threshold. Choosing the smaller fictitious mass will result in a jump or sharp change in the dynamics trajectory, which is akin to a quenching process commonly employed in molecular-dynamics methods. This kind of method leads to an *adaptive* molecular dynamics scheme that always remains within a specified limit from the Born–Oppenheimer trajectory, i.e., the adiabaticity of the trajectory can be *controlled* in this fashion.[31]

The evaluation of all conditions described above involves the trace of a matrix product. The number of operations here scales as N^2 multiplications for full matrices and is $\mathcal{O}(N)$ for sparse matrices, where N is the size of the matrices involved, i.e., the number of basis functions used. For the remaining portion of this section, we will investigate how to obtain good estimates to the right-hand side in Eq. (49) by using computationally faster algorithms. (These will be useful only for the full dense matrix case, since for sparse matrices the methods described above are already $\mathcal{O}(N)$.) Using the concept of L^2-norm, one may write

$$\begin{aligned}\left|\frac{d\mathcal{H}_{\rm fict}}{dt}\right|&\leqslant N\left\|\mathbf{W}\mu^{1/2}\frac{d^2\mathbf{P}}{dt^2}\mu^{1/2}\right\|\\&\leqslant N\|\mathbf{W}\|\left\|\mu^{1/2}\frac{d^2\mathbf{P}}{dt^2}\mu^{1/2}\right\|\\&\equiv N\|\mathbf{W}\|\left\|\left.\frac{\partial E(\mathbf{R},\mathbf{P})}{\partial\mathbf{P}}\right|_{\mathbf{R}}+\mathbf{\Lambda P}+\mathbf{P\Lambda}-\mathbf{\Lambda}\right\|, \quad (52)\end{aligned}$$

where the Schwartz inequality[32] has been used and $\|\mathbf{W}\|$ represents the L^2-norm of the matrix $\mathbf{W}$, i.e., the maximum absolute eigenvalue of $\mathbf{W}$. Since it is possible to obtain an upper bound to $\|\mathbf{W}\|$ (see, for example, Ref. 33), and similarly for $\|\mu^{1/2}(d^2\mathbf{P}/dt^2)\mu^{1/2}\|$, and since all matrices, $\mathbf{P}$, $\mathbf{W}$, $d^2\mathbf{P}/dt^2$, $\mathcal{W}$, etc. are finite dimensional, it is possible to obtain an upper bound to the right-hand side in Eq. (49) quite

easily by using Eq. (52). But, first, it is important to ascertain under what conditions truly useful bounds can be obtained from Eq. (52). Obviously,

$$\lim_{\|\mathbf{W}\underline{\mu}^{1/2}(d^2\mathbf{P}/dt^2)\underline{\mu}^{1/2}\|\to 0} \frac{d\mathcal{H}_{\text{fict}}}{dt}=0, \tag{53}$$

since, as the maximum absolute eigenvalue of a finite-dimensional matrix goes to zero, the trace of the matrix should also go down to zero. Hence, for small values of $\|\mathbf{W}\underline{\mu}^{1/2}(d^2\mathbf{P}/dt^2)\underline{\mu}^{1/2}\|$, it may be expected that Eq. (52) provides truly ''tight'' bounds to $d\mathcal{H}_{\text{fict}}/dt$. However, for large values of $\|\mathbf{W}\underline{\mu}^{1/2}(d^2\mathbf{P}/dt^2)\underline{\mu}^{1/2}\|$, it is conceivable that the bounds obtained from Eq. (52) are not as tight. In such cases, the upper bound to $\|\mathbf{W}\underline{\mu}^{1/2}(d^2\mathbf{P}/dt^2)\underline{\mu}^{1/2}\|$ may directly be used as an estimate for $d\mathcal{H}_{\text{fict}}/dt$, i.e.,

$$\left|\frac{d\mathcal{H}_{\text{fict}}}{dt}\right| \approx \|\mathbf{W}\| \left\| \left.\frac{\partial E(\mathbf{R},\mathbf{P})}{\partial \mathbf{P}}\right|_{\mathbf{R}} + \mathbf{\Lambda P} + \mathbf{P\Lambda} - \mathbf{\Lambda} \right\|. \tag{54}$$

The upper bound in Eq. (52), or the estimate in Eq. (54), may be compared to the user-defined threshold to evaluate the accuracy of the trajectory. Alternately, the Frobenius norm[32] of the matrix product in Eq. (49) may also be used in Eq. (52), since it is an upper bound to the trace. However, the Frobenius norm is not a very tight upper bound[33] to the trace, hence the L^2-norm has been used in Eqs. (52) and (54). Also the evaluation of the L^2-norm scales as N^2 additions, which has a much smaller prefactor as compared to the evaluations of the exact expressions stated above, in Eq. (49).

In this section a methodology has been presented to estimate the error in the fictitious dynamics trajectory with respect to a *true* Born–Oppenheimer dynamics trajectory. The quantities in Eqs. (49)–(51) provide exact deviations, which can be readily evaluated. However, for dense matrix systems, approximations to these results are provided, which may be calculated at lower computational cost. These results can be used to control the deviation of the fictitious dynamics trajectory from the true Born–Oppenheimer dynamics trajectory (and hence the adiabaticity) to within physically enforced limits. For all the techniques described above, the number of operations required scales linearly with system size for sparse matrices.

When dealing with metals and other small band-gap systems, the above formalism will limit the size of the fictitious mass, and hence the time step, in order to achieve proper adiabaticity. This may result in a propagation scheme that is not computationally as efficient. Hence, for small band-gap systems, it will be necessary to use an alternative scheme to enforce adiabaticity.[34]

V. CHOICE OF INITIAL CONDITIONS FOR THE DENSITY MATRIX VELOCITY

The choice of initial velocities for the nuclei are governed by the temperature at which the simulations are to be performed. However, the choice of initial density matrix velocity is not as intuitively obvious. There are restrictions on such a choice. For one it is necessary to maintain the density matrix ''temperature'' to be as low as possible, since a higher temperature would give rise to energy exchange and potentially thermal equilibration between the density matrix and nuclear degrees of freedom. Such a thermal equilibrium could at any time lead to nonphysical electronic surface crossings. By avoiding such a situation it is possible to maintain the dynamics on a given electronic state, i.e., close to the associated Born–Oppenheimer surface.

There are two facets to controlling the fictitious electronic kinetic energy: (i) The choice of initial velocity of the density matrix elements must in some sense be consistent with the choice of initial velocities of the nuclei, and (ii) the density matrix velocities at any given time must be maintained such that the ''temperature'' of the fictitious electronic degrees of freedom are restricted below a certain value. In this section the first problem is addressed. The second problem can be controlled as explained in Sec. IV.

Among the various possible choices for initial density matrix velocities, one of the conceptually simplest choices includes choosing a fixed value for the fictitious kinetic energy at time $t=0$. The individual elements of $\mathbf{W}$, at time $t=0$, may, however, be chosen randomly as long as the total fictitious kinetic energy is constant. In this case, the electronic degrees of freedom will need to ''catch-up'' with the nuclei, if the initial choice is not consistent with the choice of the nuclear kinetic energy, and a certain number of simulation steps will be expended for this equilibration process. Such an equilibration process will also see some exchange in energy between the nuclear and density matrix parts of the system, before the equilibration is complete. As stated above this exchange could give rise to spontaneous excitations depending upon the band-gap of the system. Hence, for large band-gap systems, like insulators and some semiconductors, this may be a viable choice but precious simulation time could be lost in the equilibration process.

A better, but more expensive, choice is to obtain the velocity from converged or approximately converged self-consistent field (SCF) calculations. In this case, the initial nuclear positions are propagated forward in time for exactly one time step using the exact gradients of the converged initial density matrix. At the new nuclear conformation either a full SCF calculation could be performed or a limited number of conjugate gradient density matrix search (CG-DMS)[17,18] steps could be used to obtain a converged or approximately converged density matrix. In a similar fashion, the initial nuclear positions may be propagated backwards in time for again one time step, and a fully converged, or approximately converged, density matrix obtained. Using the three density matrices at times $t=-\Delta t$, $t=0$ and $t=+\Delta t$, one may obtain the density matrix velocity at $t=0$ using a finite difference approximation[35] to the derivative of a function. Higher order, and hence more accurate, approximations could also be used by obtaining more converged or approximate densities at different times such as $t=-2\Delta t$, $t=+2\Delta t$ and so on.

VI. COMPUTATIONAL TESTS

Solvation of ions in aqueous medium is of great practical interest and has recently attracted considerable attention, both from theoretical[36–42] and experimental[43–45] groups. Here, the dynamics of a $Cl^-(H_2O)_{25}$ cluster is considered.

TABLE I. Trajectories for the $Cl^-(H_2O)_{25}$ cluster with no thermostats,[a] using the scalar mass formalism.[a]

$\mathcal{M}_e$[b] (amu bohr^2)	Time step (femtosecond)	Trajectory time (femtosecond)	Conservation of total energy (hartree)[c]	Conservation of real energy (hartree)[d]
0.1	0.01	20	0.0002	0.003
0.1	0.03	30	0.0004	0.002
0.1	0.06	120	0.0024	0.002
0.15	0.06	105	0.0022	0.005
0.2	0.06	100	0.0022	0.005
0.2	0.07	80	0.0017	0.005

[a]All calculations used DFT with the PBE exchange-correlation functional (Ref. 51), while the basis set used was 3-21G* (number of basis function=344). Cholesky orthogonalization procedure was used in all calculations to obtain the orthonormal basis set (see Ref. 11 for details).

[b]1 amu bohr^2=1823 a.u.

[c]Maximum deviation of the total energy of the system (defined as the sum of the total potential energy, E, the nuclear kinetic energy and the fictitious kinetic energy of the density matrix) during the trajectory.

[d]Maximum deviation of the real energy of the system (defined as the sum of the total potential energy, E and the nuclear kinetic energy, i.e., the total energy described above minus the fictitious kinetic energy of the density matrix) after equilibration. The initial equilibration process, discussed in the text, involved exchange of energy between the nuclear and electronic degrees of freedom where the electrons ''catch-up'' with the nuclei.

AIMD trajectories are obtained using three different techniques: The scalar mass equations presented in Ref. 11 and here in Sec. II, the mass-weighting scheme presented in Sec. III, and by using pseudo-potentials[46–48] to replace the core-electrons. The effect of the magnitude of the fictitious mass on energy conservation, adiabaticity, and allowable time steps were studied. The all-electron calculations (for both scalar mass and mass-matrix schemes) were performed using an all electron double zeta basis set with polarization functions on the chloride ion (the 3-21G* basis set). The pseudopotential calculations are performed using the CEP-31G[46–48] basis set, where a valence double zeta basis set with polarization functions on the chloride ion are used along with appropriate pseudopotentials to replace the oxygen $1s$ and the chlorine $1s$, $2s$, and $2p$ core functions.[46–48] Further, we chose the Cholesky decomposition scheme[26] to obtain the orthonormal basis set (see Ref. 11 or Sec. II) in all our calculations. The reason for this choice has to do with the more expensive N^3 computational scaling in obtaining the Löwdin symmetric $\mathbf{S}'^{1/2}$ matrix.[49] The Cholesky decomposition, on the other hand, scales only as N^2 even for full-matrices and is strictly $\mathcal{O}(N)$ for the sparse matrix case. Furthermore, the use of Löwdin symmetric orthogonalization does not provide any noticeably different trajectories from those obtained by Cholesky orthogonalization, as was demonstrated in Ref. 11. Starting nuclear geometries for all trajectories were obtained from an MM3 minimization. All calculations were performed using a development version of the *Gaussian* series of electronic structure codes[50] and with density-functional theory using the PBE exchange-correlation functional.[51] The results are shown in Tables I–III.

For the mass-weighting scheme, a diagonal fictitious mass-matrix, μ was chosen. The diagonal representation implies that the orthonormal basis set vectors (Cholesky or Löwdin, depending upon the choice) are the eigenstates of the mass-matrix.[52] Since these basis set vectors, in general, will never be the eigenstates of the $\mathbf{P}_i$ matrix, this choice trivially yields $[\mu,\mathbf{P}_i]\neq 0$. There are, of course, other possible and more general choices for μ, and these will be investigated in future publications. Here, the value of a particu-

TABLE II. Trajectories for the $Cl^-(H_2O)_{25}$ cluster with no thermostats,[a] using the mass-weighting scheme.[a]

$\mathcal{M}_e$[b] (amu bohr^2)	Time step (femtosecond)	Trajectory time (femtosecond)	Conservation of total energy (hartree)[c]	Conservation of real energy (hartree)[d]
0.05	0.20	320	0.0027	0.005
0.1	0.20	400	0.0027	0.005
0.1	0.25	500	0.0025	0.005
0.15	0.20	400	0.0025	0.005
0.15	0.25	500	0.0025	0.007

[a]All calculations used DFT with the PBE exchange-correlation functional (Ref. 51), while the basis set used is 3-21G* (number of basis function=344). Cholesky orthogonalization procedure was used in all calculations to obtain the orthonormal basis set (see Ref. 11 for details).

[b]1 amu bohr^2=1823 a.u.

[c]Maximum deviation of the total energy of the system (defined as the sum of the total potential energy, E, the nuclear kinetic energy and the fictitious kinetic energy of the density matrix) during the trajectory.

[d]Maximum deviation of the real energy of the system (defined as the sum of the total potential energy, E and the nuclear kinetic energy, i.e., the total energy described above minus the fictitious kinetic energy of the density matrix) after equilibration. The initial equilibration process, discussed in the text, involved exchange of energy between the nuclear and electronic degrees of freedom where the electrons ''catch-up'' with the nuclei.

TABLE III. Trajectories for the $Cl^-(H_2O)_{25}$ cluster with no thermostats,[a] using pseudopotentials.[a]

$\mathcal{M}_e$[b] (amu bohr^2)	Time step (femtosecond)	Trajectory time (femtosecond)	Conservation of total energy (hartree)[c]	Conservation of real energy (hartree)[d]
0.1	0.05	55	0.0001	0.001
0.1	0.10	45	0.0001	0.001
0.1	0.15	230	0.0002	0.002

[a]All calculations uses DFT with the PBE exchange-correlation functional (Ref. 51), while the basis set used is CEP-31G (Refs. 46 to 48) (number of basis function=314). Cholesky orthogonalization procedure was used in all calculations to obtain the orthonormal basis set (see Ref. 11 for details).

[b]1 amu bohr^2=1823 a.u.

[c]Maximum deviation of the total energy of the system (defined as the sum of the total potential energy, E, the nuclear kinetic energy and the fictitious kinetic energy of the density matrix) during the trajectory.

[d]Maximum deviation of the real energy of the system (defined as the sum of the total potential energy, E and the nuclear kinetic energy, i.e., the total energy described above minus the fictitious kinetic energy of the density matrix) after equilibration. The initial equilibration process, discussed in the text, involved exchange of energy between the nuclear and electronic degrees of freedom where the electrons ''catch-up'' with the nuclei.

lar diagonal element of μ was chosen based on the value of the corresponding element in the Fock matrix at the initial time step. Particularly, it was assumed that if the diagonal Fock matrix element $\mathbf{F}_{i,i}<-2$ a.u., at the initial time step, the corresponding orbital is a core orbital, and if $\mathbf{F}_{i,i}\geqslant -2$ a.u., at the initial time step, the corresponding orbital is valence. The mass-matrix chosen has the following constant form:

$$\mu_{i,i}^{1/2}=\sqrt{\mathcal{M}_e}, \quad \mathbf{F}_{i,i}\geqslant -2 \text{ a.u.}, \tag{55}$$

$$\mu_{i,i}^{1/2}=\sqrt{\mathcal{M}_e}[2\sqrt{|\mathbf{F}_{i,i}+2|}+1], \quad \mathbf{F}_{i,i}<-2 \text{ a.u.}, \tag{56}$$

$$\mu_{i,j}^{1/2}=0, \qquad i\neq j, \tag{57}$$

where $\mathcal{M}_e$ is a constant and may be considered as the mass of the valence orbitals, as seen from Eq. (55). The choice of the functional form in Eq. (56) is based on a harmonic-oscillatorlike assumption for the orbital energies. The result of including the mass-weighting factor in Eq. (56) is that the core orbital oscillations get adjusted to roughly the same frequency as that for the valence orbitals. Note that, in general, the above choice for μ may not be the most effective, since it is hard to fix a universal energy criterion for separation between core and valence orbitals for all atoms,[53] and also the harmonic oscillator approximation, while reasonable, may not be the most effective choice in many cases. The current choice of μ is used here only to demonstrate the applicability of our mass-weighting scheme, and as will be seen later in this section, it is useful for the application studied here. More general approaches to the mass-matrix will be considered in future publications.

The most striking aspect seen from the tables is the significantly larger time steps that are accessible due to the use of the mass-weighted propagation scheme presented in this paper. Needless to say, this results in being able to propagate longer and more stable trajectories. Employing pseudopotentials permits time steps larger than those allowed in the scalar mass case, but smaller than allowed by the use of the mass-matrix scheme.

Increasing the fictitious mass is another way to obtain larger time steps. However, as pointed out in Sec. III this is not in general advisable, since larger fictitious masses lead to deterioration in electronic adiabaticity. This aspect is also noted in our calculations, where it was found that, for the mass-weighting case, increasing the value of $\mathcal{M}_e$ in Eqs. (55) and (56) results in a drift in the fictitious kinetic energy. For example, compare the drift in the fictitious kinetic energy in Fig. 1(b), obtained from a value of $\mathcal{M}_e=0.2$ amu bohr^2 ≈ 364 a.u., as opposed to its more steady behavior after equilibration in Fig. 1(a), obtained for $\mathcal{M}_e=0.1$ amu bohr^2 ≈ 182 a.u. Both figures are for a time step of 0.25 fs and total propagation time of 500 fs ($=0.5$ ps). The deviations are even greater for larger values of $\mathcal{M}_e$. A reduction in the size of the fictitious mass ($\mathcal{M}_e=0.05$ amu bohr$^2\approx 91$ a.u.) results in a reduction of allowable time step size (to 0.2 fs).

Another interesting observation from the tables is that the real energy is conserved to a lower (but acceptable) degree in the mass-weighted calculation as compared to the scalar mass and pseudopotential calculations. There are two factors that contribute to this. Firstly, the time steps used in the mass-weighted calculations are higher than those used in the scalar mass and pseudopotential calculations. Secondly, as seen from Eq. (56), the core orbitals in the mass-weighting scheme are ''heavier'' than the valence orbitals. This yields a larger value for the L^2-norm of the $\mu^{1/2}$ matrix, and as is seen from Eqs. (49) and (52), this in turn leads to a larger value of $|(d\mathcal{H}_{\text{fict}}/dt)|$, and hence poorer conservation of the real energy in the mass-weighted propagation scheme. In any case, the conservation of the real energy is found to be acceptable since the deviations seen in Table II are of the order of 0.0002% with respect to the total energy of the system (see Fig. 1 and caption). It should also be noted here that, as stated after Eq. (57), the quantity $\mathcal{M}_e$ has the ''physical'' interpretation of being the fictitious mass of the valence orbitals in the mass-weighting scheme. This should be compared with the notation used in the plane-wave Car–Parrinello method, where fictitious mass stands for the mass of the valence orbitals *only*, since pseudopotentials are used to replace the core.

Apart from the discussion in Sec. III and the results seen in our study, another important reason to keep the value of the fictitious mass of the density matrix elements small, is that the mass of a single hydrogen atom is 1836 a.u.

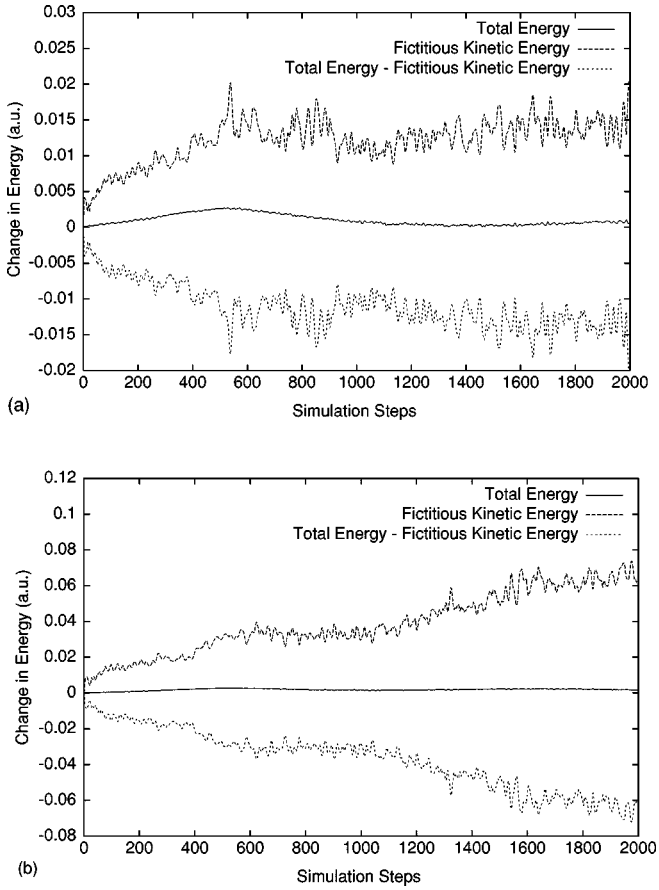

FIG. 1. AIMD trajectories for the $Cl^-(H_2O)_{25}$ cluster dynamics calculation using the mass-weighting scheme: (a) fictitious electronic mass, $\mathcal{M}_e = 0.1$ amu bohr$^2 \approx 182$ a.u., (b) fictitious electronic mass, $\mathcal{M}_e = 0.2$ amu bohr$^2 \approx 364$ a.u. In both figures $\Delta t = 0.25$ fs and total simulation time $= 0.5$ ps. The average electronic energy and average total energy for the full duration of both simulations is of the order of -2355 a.u. Note the gradual drift in the fictitious kinetic energy in (b) (although small compared to the total electronic energy), as opposed to the more steady behavior in (a) after equilibration.

($\equiv 1.008$ amu). Hence, as the value of the fictitious electronic mass gets closer to this number, the separation in time scales between electronic motions and motions of hydrogen and hydrogenlike systems becomes smaller, and the fictitious system deviates further away from the real system due to energy exchange. The value of $\mathcal{M}_e$ used in our calculations is small as compared to the mass of the hydrogen atom, hence a good degree of separation of time scales is maintained through all of the simulations. It is useful to note that, typically,[36–41] in order to utilize time steps of reasonable size, the magnitude of the fictitious mass chosen in plane-wave Car–Parrinello calculations for water is ~ 0.6 amu bohr^2 (≈ 1100 a.u.[36–41]), i.e., six times larger in magnitude as compared to our $\mathcal{M}_e = 0.1$ amu bohr^2 ≈ 182 a.u. calculation. The associated effects of the larger electronic mass in the CP calculations are often countered by using Deuterium nuclei instead of Hydrogen nuclei in water simulations. This, however, is not necessary here, since the fictitious electronic masses used are much smaller.

The time steps allowed in the mass-weighted scheme in this study (0.25 fs for the $\mathcal{M}_e = 0.1$ amu bohr$^2 \approx 182$ a.u. calculation), are nearly twice the time steps commonly used in plane-wave Car–Parrinello studies of solvation.[36–41] The difference between the present AIMD implementation and that in the plane-wave Car–Parrinello scheme is the use of localized Gaussian orbitals and the single particle density matrix here, as opposed to global plane-waves and the occupied Kohn–Sham wave functions in the CP method. The other important difference is the use of pseudopotentials in plane-wave Car–Parrinello scheme, whereas the present method has the freedom to either treat all electrons rigorously (the scalar mass formalism discussed in Ref. 11 and in Sec. II in this paper), weigh the core differently from the valence to reduce core-unoccupied orbital coupling (the mass-weighting scheme discussed in Sec. III), or to use pseudopotentials.

For the scalar mass case, again, it is found that larger values of the fictitious mass resulted in larger fictitious kinetic energies, and hence greater deviations from adiabaticity. A fictitious mass of ≈ 0.1 amu bohr$^2 \approx 182$ a.u. allowed a time step of 0.06 fs within this scheme (cf. Table I). Further increase in the fictitious mass (0.15 and 0.2 amu bohr^2) does not, in this case, provide any reasonable gain in the size of the time step to be chosen. (As may be noted from Table I, a larger fictitious mass of 0.2 amu bohr^2 (≈ 364 a.u.) allows only a slightly larger time step. Increasing the time step any further results in complete loss of idempotency of the density matrix, and hence nonphysical trajectories.)

For all calculations the total angular momentum of the system was conserved to a very high degree ($10^{-11}\hbar$) for the duration of the simulation, since projection methods were used to remove residual angular forces.[54] Idempotency was always conserved to better than 10^{-12} using the iterative scheme presented in Eqs. (36) and (38) (with fewer than 10 iterations necessary in all cases). It was found that the guess in Eq. (39) provided an initial $\mathbf{P}_{i+1}$ for Eq. (36) that was in most cases an order of magnitude or better in idempotency as compared to the guess in Eq. (35).

It was found that a fictitious mass of ≈ 0.1 amu bohr^2 ≈ 182 a.u. is an appropriate choice for the water–cluster system for both mass-weighting (where $\mathcal{M}_e$ is chosen to be 0.1 amu bohr^2) and scalar mass schemes. The mass-weighting case is definitely advantageous since it allows for larger time steps, while providing stable dynamics trajectories.

VII. CONCLUSIONS

In this paper, the approach to *ab initio* molecular dynamics introduced in Ref. 11, has been generalized by considering nonscalar fictitious masses. The conservation properties of this AIMD scheme have been rigorously studied by employing the Hamiltonian formalism. This allows us to establish, on-the-fly, if a given fictitious dynamics trajectory obtained from the present scheme is close enough to the *true* Born–Oppenheimer trajectory (to within user-specified limits). Approximations to the exact expressions have also been derived that may be useful to evaluate the accuracy of the trajectories for dense matrix systems. The exact expressions, in addition to being useful in general, are computationally $\mathcal{O}(N)$ when systems become sparse.

The matrix fictitious mass generalization allows larger time steps to be used in the AIMD propagation algorithm. This is accomplished with moderate values of the fictitious mass, hence keeping the loss of adiabaticity to within controllable limits and thus maintaining reasonable separation between the time scales of motion of the electronic (density matrix) parts and the nuclear parts. For the $Cl^-(H_2O)_{25}$ cluster case studied here, a choice of $\mathcal{M}_e = 0.1\,\text{amu bohr}^2 \approx 182\,\text{a.u.}$, within the mass-weighting formalism, and $\Delta t = 0.25\,\text{fs}$ (or 0.2 fs) was found to be appropriate. This choice leads to a stable AIMD trajectory within a reasonable computational expense.

It was also shown in the discussion that the magnitude of the fictitious mass determines the separations in time scales and hence it is important to keep its value as small as possible. As a result of the combination of smaller fictitious mass and a good separation between the time-scales of electronic and nuclear motions, in the present AIMD method hydrogen atom systems may be routinely studied. It is not necessary to substitute heavier isotopes to maintain electronic adiabaticity, as is frequently done in plane-wave CP calculations.[36–41]

The combination of a Gaussian-basis along with the single particle density matrix representation of the electronic degrees of freedom, when used in conjunction with an extended Lagrangian scheme to enable parallel dynamical propagation of nuclei and the electronic density matrix, yields a novel and robust computational tool to perform *ab initio* molecular-dynamics simulations. The advantages of this method include: (i) the freedom to rigorously treat all electrons in the system or to use pseudopotentials; (ii) the ability to use reasonably large time steps with small values for the fictitious mass, which allows one to retain hydrogen atoms in the system and not substitute heavier isotopes; (iii) the freedom to employ a wide variety of accurate and effective exchange-correlation functionals, including hybrid density functionals; (iv) the ability to treat charged molecular systems which is non-trivial in most implementations of the plane-wave Car–Parrinello method due to the need to treat molecular systems as periodic;[55] (v) rigorous on-the-fly control of the deviation from the Born–Oppenheimer surface and the mixing of fictitious and real kinetic energies; and (vi) good computational efficiency due to the use of fewer basis functions per atom, larger time steps, and asymptotic $\mathcal{O}(N)$ scaling using established techniques.[13,17,18] Applications and extensions of the new AIMD methodology will be the subject of future research.

ACKNOWLEDGMENTS

This work was supported by the Natural Science Foundation (CHE-9982156 and CHE-9874005), the Office of Naval Research (GAV), and Gaussian, Inc. The authors acknowledge Professor Jack Simons for his input on a preliminary version of this manuscript, and Dr. Nelson H. F. Beebe and Dr. James Lewis for fruitful discussions on related issues.

[1] K. Bolton, W. L. Hase, and G. H. Peslherbe, in *Modern Methods for Multidimensional Dynamics Computation in Chemistry*, edited by D. L. Thompson (World Scientific, Singapore, 1998).
[2] J. M. Millam, V. Bakken, W. Chen, W. L. Hase, and H. B. Schlegel, J. Chem. Phys. **111**, 3800 (1999).
[3] X. Li, J. M. Millam, and H. B. Schlegel, J. Chem. Phys. **113**, 10062 (2000).
[4] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[5] D. K. Remler and P. A. Madden, Mol. Phys. **70**, 921 (1990).
[6] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).
[7] M. E. Tuckerman, P. J. Ungar, T. von Rosenvinge, and M. L. Klein, J. Phys. Chem. **100**, 12878 (1996).
[8] E. Deumens, A. Diz, R. Longo, and Y. Öhrn, Rev. Mod. Phys. **66**, 917 (1994).
[9] R. I. Cukier and J. M. Deutch, Phys. Rev. **177**, 240 (1969); N. N. Bogoliubov and Y. A. Mitropolsky, *Asymptotic Theory of Nonlinear Oscillations* (Gordon and Breach Science, New York, 1961); N. Krylov and N. N. Bogoliubov, *Introduction to Nonlinear Mechanics* (Princeton University Press, Princeton, NJ, 1947); C. M. Bender and S. A. Orzag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).
[10] M. E. Tuckerman, B. J. Berne, and A. Rossi, J. Chem. Phys. **94**, 1465 (1991).
[11] H. B. Schlegel, J. M. Millam, S. S. Iyengar, G. A. Voth, A. D. Daniels, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **114**, 9758 (2001).
[12] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).
[13] G. E. Scuseria, J. Phys. Chem. A **103**, 4782 (1999).
[14] X.-P. Li, W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).
[15] C. A. White and M. Head-Gordon, J. Chem. Phys. **101**, 6593 (1994).
[16] M. C. Strain, G. E. Scuseria, and M. J. Frisch, Science **271**, 51 (1996).
[17] J. M. Millam and G. E. Scuseria, J. Chem. Phys. **106**, 5569 (1997).
[18] A. D. Daniels, J. M. Millam, and G. E. Scuseria, J. Chem. Phys. **107**, 425 (1997).
[19] D. A. Gibson, I. V. Ionova, and E. A. Carter, Chem. Phys. Lett. **240**, 261 (1995).
[20] B. Hartke and E. A. Carter, J. Chem. Phys. **97**, 6569 (1992).
[21] B. Hartke and E. A. Carter, Chem. Phys. Lett. **189**, 358 (1992).
[22] G. Martyna, C. Cheng, and M. L. Klein, J. Chem. Phys. **95**, 1318 (1991).
[23] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Cambridge, Mass. 1980).
[24] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, J. Chem. Phys. **76**, 637 (1982); M. E. Tuckerman and M. Parrinello, *ibid.* **101**, 1302 (1994).
[25] R. McWeeny, Rev. Mod. Phys. **32**, 335 (1960).
[26] G. H. Golub and C. F. van Loan, *Matrix Computations* (The Johns Hopkins University Press, Baltimore, 1996).
[27] A similar result has been obtained for the plane-wave Car–Parrinello method. See, F. A. Bornemann and C. Schütte, Numer. Math. **78**, 359 (1998); F. A. Bornemann and C. Schütte, *ibid.* **83**, 179 (1999).
[28] M. V. Berry, Proc. R. Soc. London, Ser. A **392**, 45 (1984).
[29] C. A. Mead and D. G. Truhlar, J. Chem. Phys. **70**, 2284 (1979); **78**, 6344 (1983).
[30] A. Kuppermann, in *Dynamics of Molecules and Chemical Reactions*, edited by R. E. Wyatt and J. Z. H. Zhang (Dekker, New York, 1996).
[31] It must be noted that changing the value of fictitious mass introduces a time-dependent nature to the mass, which is inconsistent with our assumption in Eqs. (27) and (44). However, the time-dependence lasts only for one time step, hence the assumptions are still valid, independently, before and after the reduction (change) in the size of the fictitious mass.
[32] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Dover, New York, 1990).
[33] S. S. Iyengar, D. J. Kouri, G. A. Parker, and D. K. Hoffman, Theor. Chem. Acc. **103**, 507 (2000).
[34] For the case of metals and small band-gap system, it will also be necessary to modify the Lagrangian constraints, since it is only necessary that $\mathbf{P}^2 \leqslant \mathbf{P}$. The concept of Fractional Occupation Numbers, that are currently very popular in DFT, may be useful in this regard. Methods based on Nosé–Hoover thermostats (Refs. 56–59), can also be used.
[35] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, New York, 1992).
[36] C. J. Mundy, J. Hutter, and M. Parrinello, J. Am. Chem. Soc. **122**, 4837 (2000).
[37] F. Bruge, M. Bernasconi, and M. Parrinello, J. Am. Chem. Soc. **121**, 10883 (1999).

[38] L. M. Ramaniah, M. Bernasconi, and M. Parrinello, J. Chem. Phys. **111**, 1587 (1999).
[39] E. J. Meijer and M. Sprik, J. Phys. Chem. A **102**, 2893 (1998).
[40] D. Marx, M. Sprik, and M. Parrinello, Chem. Phys. Lett. **373**, 360 (1997).
[41] M. Tuckerman, K. Laasonen, M. Sprik, and M. Parrinello, J. Phys. Chem. **99**, 5749 (1995).
[42] G. K. Schenter, B. C. Garrett, and G. A. Voth, J. Chem. Phys. **113**, 5171 (2000).
[43] J.-H. Choi, K. T. Kuwata, Y.-B. Cao, and M. Okumura, J. Phys. Chem. A **102**, 503 (1998).
[44] P. Ayyotte, G. H. Weddle, J. Kim, and M. A. Johnson, J. Am. Chem. Soc. **120**, 12361 (1998).
[45] P. Ayyotte, S. B. Nielsen, G. H. Weddle, M. A. Johnson, and S. S. Xantheas, J. Phys. Chem. A **103**, 10665 (1999).
[46] W. Stevens, H. Basch, and J. Krauss, J. Chem. Phys. **81**, 6026 (1984).
[47] W. J. Stevens, M. Krauss, H. Basch, and P. G. Jasien, Can. J. Chem. **70**, 612 (1992).
[48] T. R. Cundari and W. J. Stevens, J. Chem. Phys. **98**, 5555 (1993).
[49] There are faster methods based on the Arnoldi iterative scheme to obtain the Löwdin symmetric $\mathbf{S}'^{1/2}$ matrix. The convergence (and hence the computational complexity) of these methods, however, depends on the average spacing between the eigenvalues of the $\mathbf{S}'$ matrix.
[50] M. J. Frisch, G. W. Trucks, H. B. Schlegel, *et al.* GAUSSIAN 99, Development Version Revision C.01, Gaussian, Inc., Pittsburgh, PA, 1998.
[51] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996); **78**, 1396(E) (1997).
[52] In this study we choose μ to be diagonal in the orthonormal basis instead of the nonorthogonal basis. This avoids the need to explicitly treat velocity dependent forces that arise when the diagonal, μ (in the nonorthogonal basis) is transformed to the orthonormal basis for use in Eqs. (27), (2), and (3). The transformation depends on, **U**, and hence on, **R** and t.
[53] G. A. Petersson (personal communication).
[54] For the semiempirical calculations with converged densities, the angular forces originate primarily from the numerical differentiation of the integrals used in the gradients (as seen in Ref. 11, angular momentum conservation is $\sim 0.3\hbar$ for PM3 without projection). When the electronic structure is propagated, additional angular forces can arise from the fact that the electronic energy is not converged with respect to the density.
[55] Charged systems are generally treated in plane-wave CP calculations by introducing a potential to screen the Coulomb interaction and by increasing the size of the periodic cell. For a detailed account, see, for example, Refs. 60 and 61.
[56] S. Nosé, J. Chem. Phys. **81**, 511 (1984).
[57] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[58] G. J. Martyna, M. L. Klein, and M. Tuckerman, J. Chem. Phys. **97**, 2635 (1992).
[59] M. E. Tuckerman and M. Parrinello, J. Chem. Phys. **101**, 1302 (1994).
[60] D. Marx, and J. Hutter, in *Modern Methods and Algorithms of Quantum Chemistry*, edited by J. Grotendorst (John von Neumann Institute for Computing, Jülich, NIC Series, 2000), Vol. 1, p. 301.
[61] R. N. Barnett and U. Laudman, Phys. Rev. B **48**, 2081 (1993).