Accelerated, energy-conserving Born–Oppenheimer molecular dynamics *via* Fock matrix extrapolation

John M. Herbert* and Martin Head-Gordon

*Department of Chemistry, University of California, Berkeley, CA 94720, USA
E-mail: jherbert@calmail.berkeley.edu; mhg@cchem.berkeley.edu*

Born–Oppenheimer molecular dynamics calculations, especially those that exploit information retained from previous time steps in order to accelerate convergence of the electronic structure calculations, can suffer from systematic error in the energy gradient that manifests as a drift in the microcanonical energy. Here, we demonstrate that this is only the case when the self-consistent field (SCF) convergence criterion is set too low; using only a marginally tighter threshold (still two orders of magnitude lower than what is standard for geometry optimizations), the drift disappears completely, for a time scale of several picoseconds. Using a Fock matrix extrapolation technique, SCF convergence is achieved in as few as three iterations per time step, without sacrificing energy conservation. In test calculations for C_2F_4, $(H_2O)_4$, $(H_2O)_6$, and $[Fe(H_2O)_6]^{2+}$, we demonstrate energy-conserving Fock matrix extrapolation that reduces the number of SCF cycles by up to 70% and reduces the computer time per molecular dynamics step by 45–55%, relative to simulations performed without extrapolation.

1. Introduction

In Born–Oppenheimer molecular dynamics (BOMD) calculations, a set of nuclei are propagated according to classical equations of motion, on a potential energy surface obtained by "on-the-fly" solution of the quantum-mechanical electronic structure problem. Fundamental to this approach is the fact that the electronic wave function or density is converged following each nuclear step, in contrast to extended-Lagrangian methods[1] such as Car–Parrinello molecular dynamics[2] and its Gaussian-orbital-based congeners.[3–5] In extended-Lagrangian molecular dynamics (ELMD), the electronic degrees of freedom are *propagated* rather than *optimized*, as in BOMD. Because ELMD algorithms do not iterate the wave function to convergence, these methods are inherently more efficient than BOMD, at least in the sense that simulations consume less computer time per unit of simulated time. For example, in recent studies of liquid water and water clusters, it was concluded that ELMD is three to four times more efficient than BOMD, according to this criterion.[6,7] On the other hand, ELMD represents an approximation to true classical dynamics on the Born–Oppenheimer potential surface; it is known, for example, that errors in ELMD vibrational frequencies can be quite large.[4,8–10] The presence and importance of other artifacts in ELMD continues to be debated.[7,10–12]

BOMD is, by definition, the true classical dynamics on the Born–Oppenheimer potential energy surface, assuming that the electronic problem is solved exactly (within a given model chemistry) at each time step, hence improving the efficiency of this method remains an active area of research. A major concern is reduction of the number of self-consistent field (SCF) iterations required to convergence the electronic density matrix at each nuclear configuration, for this is the only portion of the calculation for which ELMD is computationally more efficient than BOMD. (We limit our discussion to SCF electronic structure theory, as the SCF step is necessary anyway for more accurate treatments.)

Since energy conservation requires small time steps, an obvious means to accelerate BOMD simulations is to use the converged molecular orbitals (MOs) or density matrix at time t to generate an initial guess at time $t + \delta t$, or better still, to extrapolate a guess using converged MOs from several previous time steps.[13] Alternatively, one might extrapolate a sequence of converged Fock matrices instead, as suggested recently by Pulay and Fogarasi.[14] The latter approach has two advantages: first, the Fock matrix can be extrapolated without orthogonality or idempotency constraints; and second, Fock matrix extrapolation obviates the need for a "Fock build" in the first SCF cycle. In Gaussian basis sets (which are used exclusively in this work), the cost of constructing the Fock matrix exceeds the cost of diagonalizing it by an overwhelming factor, except for very large systems that are presently beyond the reach of *ab initio* molecular dynamics. For these reasons, we focus on Fock matrix extrapolation rather than MO or density matrix extrapolation.

It is already known that extrapolation can significantly reduce the number of SCF iterations necessary to reach convergence in BOMD calculations. Given that larger time steps are possible in BOMD than in ELMD, Marx and Hutter[1] go so far as to conclude that BOMD "can be made as fast as (or even faster than) Car–Parrinello molecular dynamics... at the expense of sacrificing accuracy in terms of energy conservation." That last caveat portends a serious problem with BOMD: the resulting trajectories can suffer from systematic error in the energy gradient that is exacerbated by extrapolation,[14] and is manifested as a drift (secular error) in the microcanonical energy.[1,13,14]

Energy conservation is clearly a desirable feature, so the relevant question is how efficient the BOMD algorithm can be made without introducing any energy drift, at least not on time scales that are accessible to *ab initio* molecular dynamics (currently $\lesssim 50$ ps). Using a plane-wave code to simulate 10 ps of dynamics for 64 water molecules in a periodic box, VandeVondele *et al.*[13] recently demonstrated that, given a sufficiently tight SCF convergence threshold, energy drift can be made small enough (2×10^{-7} E_h/ps per atom, where E_h is the atomic unit of energy, the hartree) to make extrapolation useful. These authors used MO extrapolation, though no data

were provided regarding the performance of the extrapolation. Conversely, Pulay and Fogarasi[14] used Fock matrix extrapolation in a Gaussian-orbital-based code and reported convergence in as few as two iterations per time step, but did not quantify the error in energy conservation caused induced by the extrapolation procedure.

In this work, we systematically explore the limits of Pulay and Fogarasi's Fock matrix extrapolation technique,[14] in terms of energy conservation and convergence acceleration. We demonstrate that energy drift is only a problem when the SCF convergence threshold is reduced to a very low value $(10^{-4} E_h)$. For tighter thresholds $(10^{-6} E_h)$, the energy drift per picosecond is smaller than the normal energy fluctuations arising from the finite time step, which is to say that the former is effectively zero. Under demands of high accuracy, we find that Fock matrix extrapolation reduces the cpu time required for SCF convergence by about 50% in most cases.

2. Extrapolation procedure

Following Pulay and Fogarasi,[14] we assume that a Fock matrix element $F_{\mu\nu}$ in the atomic orbital (AO) basis can be represented as a low-order polynomial in the time t or, more conveniently, in the molecular dynamics step number s,

$$F_{\mu\nu}(s) = \sum_{m=0}^{M} c_m^{(\mu\nu)} s^m. \qquad (2.1)$$

We seek a vector $c^{(\mu\nu)}$ of best-fit coefficients for the *ansatz* in eqn (2.1), using the previous N values of $F_{\mu\nu}$ as data points. Let these values be collected into a vector $f^{(\mu\nu)}$. Taking s to have values $-(N-1)$, $-(N-2)$,…,0, the extrapolation coefficients $c^{(\mu\nu)}$ for $s=1$ are obtained by solution of the linear system

$$Ac^{(\mu\nu)} = f^{(\mu\nu)}, \qquad (2.2)$$

in which A is the $N \times (M+1)$ matrix with elements

$$A_{nm} = [-(N-n)]^{m-1}. \qquad (2.3)$$

This system must be solved for each μ and ν, but since A is independent of the AO indices, its generalized inverse need be constructed only once, by singular value decomposition, at the outset of the calculation. Subsequently, the elements $F_{\mu\nu}$ are determined by the action of this inverse on the vectors $f^{(\mu\nu)}$; the operation count for this procedure is negligible in comparison to a proper Fock build.

At $t=0$ and for each of the first N time steps, we use a normal SCF guess for the density matrix, from which a Fock matrix is constructed in the first SCF cycle. For this purpose, all of the calculations reported here use the superposition of atomic densities (SAD) guess, a block-diagonal density matrix whose blocks are obtained from a library of atomic SCF calculations. Beginning on the $(N+1)$st time step, the Fock matrix in the first SCF cycle is obtained by extrapolation and then diagonalized to obtain an initial density matrix. We refer to this procedure as (N, M) extrapolation, where N is the number of saved Fock matrices and M specifies the degree of the extrapolation polynomial. General (N, M) extrapolation has been implemented in the Q-Chem electronic structure package,[15] which was used for all of the calculations in this work.

Fock matrix extrapolation, as described above, embodies a point of view in which each $F_{\mu\nu}$ is simply an oscillatory function in time whose undulations may be fit to a low-order polynomial. Considering the broader point of view, that F is the representation of an operator in a basis that is finite and time-dependent (since the basis functions are tied to atoms), it is tempting to account for the changing nature of the AO basis by projecting all of the saved Fock matrices into the current AO basis at each time step, prior to extrapolation. If F_1 is a

Fock matrix in some AO basis with overlap matrix S_1, then its representation F_2 in a different AO basis is given by

$$F_2 = S_{21}S_1^{-1}F_1S_1^{-1}S_{12}, \qquad (2.4)$$

in which $S_{12} = S_{21}^\dagger$ is the overlap matrix of the two AO basis sets. Upon implementing this procedure, however, we discovered that projection significantly degrades the effectiveness of the extrapolation procedure, and therefore projection is not considered further.

In this work, we report results using the (4,2), (6,3), (8,4), (12,6) and (16,8) extrapolation schemes [Although the accuracy of the extrapolation typically increases with N and M, there are no "magic" combinations; exploratory calculations with other N and M values ($M \leq 8$ and $N \leq 2M$) afforded similar behavior.] We also report results using converged MOs from the previous time step as an initial guess (from which a Fock matrix is constructed in the first SCF cycle) and also using the converged Fock matrix from the previous time step, *i.e.*, (1,0) extrapolation. As a control, we also discuss simulations without extrapolation, in which a SAD guess is used at each time step. (For most applications, the SAD guess is the most accurate non-extrapolated initial guess available in Q-Chem.) For the systems and basis sets examined herein, diagonalization of the Fock matrix represents a negligible amount of overhead, so we characterize the efficiency of various SCF guesses in terms of the average number of Fock builds per time step, N_{FB}.

The performance of these methods is evaluated as a function of the SCF convergence threshold. We use the notation $\varepsilon_{SCF} = n$ to indicate that convergence is achieved when all elements of the occupied-virtual component of the Fock matrix are smaller than $10^{-n} E_h$ in magnitude. (All calculations utilize Pulay's direct inversion in the iterative subspace algorithm[16] to converge the SCF calculations.) $\varepsilon_{SCF} = n$ also implies that the threshold for neglecting shell pairs is $10^{-(n+3)}$ and that the threshold for neglecting AO integrals is $10^{-(n+3)} E_h$. (In cases of near-linear dependencies in the AO basis, these shell pair and integral thresholds would need to be tightened, but we do not anticipate that linear dependencies will affect the behavior of the algorithm with respect to the SCF convergence threshold, which is the main focus of the present study.)

3. Numerical tests

3.1. A toy system

In order to compare with the work of Pulay and Fogarasi,[14] who first introduced the Fock matrix extrapolation procedure described in the previous section, we focus first on one of the test systems used by those authors, namely, the Hartree–Fock (HF)/3-21G dynamics of C_2F_4 at $T = 500$ K. Our simulations are run in the microcanonical ensemble, so "temperature" in this work specifies a Maxwell–Boltzmann distribution from which initial velocities are sampled. When comparing different extrapolation methods, all methods use the same initial velocities and coordinates, the latter corresponding to the minimum-energy molecular geometry. The results discussed below for C_2F_4 are averages over five different trajectories, though the performance statistics (number of Fock builds and energy fluctuations) do not depend strongly on the particular initial conditions. Each C_2F_4 trajectory was propagated for 2.0 ps using the velocity Verlet algorithm, with a time step $\delta t = 20$ au (≈ 0.484 fs). The corresponding calculations by Pulay and Fogarasi[14] were propagated for 0.5 ps with $\delta t = 10$ au and $\varepsilon_{SCF} = 4$.

Table 1 characterizes the results of our simulations in terms of energy fluctuations and the average number of Fock builds per time step. Consider first the $\varepsilon_{SCF} = 4$ results. Using low-order extrapolation schemes like (4,2) and (6,3) we recover the spectacular result demonstrated by Pulay and Fogarasi,[14]

Table 1 Energy fluctuations and average number of Fock builds, N_{FB}, for five 2.0 ps HF/3-21G trajectories for C_2F_4 at $T = 500$ K ($\delta t = 20$ au)

| | $\varepsilon_{SCF} = 4$ | | | $\varepsilon_{SCF} = 6$ | | | $\varepsilon_{SCF} = 8$ | | |
| | | $\delta E/\mu E_h$ | | | $\delta E/\mu E_h$ | | | $\delta E/\mu E_h$ | |
SCF guess	N_{FB}	Noise	Drift/ps^{-1}	N_{FB}	Noise	Drift/ps^{-1}	N_{FB}	Noise	Drift/ps^{-1}
SAD	6.1	6.5	8.4	9.3	3.5	0.1	13.2	3.5	0.1
Previous MOs	2.1	753.1	5044.9	6.5	3.5	1.7	10.3	3.5	0.1
(1,0) extrap.	4.1	53.2	1519.4	7.2	2.8	8.9	10.7	3.5	0.1
(4,2) extrap.	1.6	23.0	795.4	3.8	3.7	38.1	7.5	3.5	0.1
(6,3) extrap.	1.9	9.7	43.3	2.8	3.5	0.8	6.4	3.5	0.1
(8,4) extrap.	2.5	10.5	6.9	2.8	3.5	3.0	6.3	3.5	0.2
(12,6) extrap.	2.9	12.5	13.7	2.9	3.5	1.1	4.5	3.5	0.1
(16,8) extrap.	3.2	11.8	21.3	3.0	3.5	0.4	3.9	3.5	0.1

namely, SCF convergence using fewer than two Fock builds per time step. As noted by those authors, this rapid convergence comes at the cost of serious systematic error in the gradient, which manifests as a rapid drift in the energy. Perhaps paradoxically, higher-order extrapolations (not considered by Pulay and Fogarasi) require *more* SCF cycles per time step than lower-order ones, although these higher-order extrapolations substantially reduce the energy drift.

Fig. 1 typifies the fluctuations in the energy that are observed with $\varepsilon_{SCF} = 4$. Consistent with earlier results regarding energy drift in BOMD[1,14] and in ordinary (analytic potential) molecular dynamics,[17] the drift is linear in time, so we quantify it by fitting $E(t)$ to a linear function of time, the slope of which we report in Table 1. (Energy drifts are reported in units of μE_h ps^{-1}, but the fits utilize the entire 2.0 ps of simulated time). We also tabulate the root mean square fluctuations in the energy *after* removal of the linear drift term. This quantity, termed *energy noise*,[18] provides a measure of the energy fluctuations that arise solely from the finite time step. The energy drift should be interpreted as effectively zero unless the product of

the drift rate and the total simulation time is significantly larger than the energy noise.

The appearance of any energy drift at all may seem curious, given that the velocity Verlet integrator is usually understood to be symplectic,[17] time-reversible,[19] and remarkably stable against roundoff errors,[20,21] properties that are intimately connected to the existence of stable, conservative trajectories at long times.[17,22–25] Strict time reversibility, however, requires that the energy gradient be evaluated in a time-reversible manner. This is trivial if an analytic potential is available, but not in cases where the energy is evaluated *via* self-consistent iteration and converged only to finite precision. In practice convergence is always incomplete, and as such no BOMD algorithm can be strictly time-reversible unless the initial guess for the SCF solution is independent of the previous time steps. SCF guess procedures such as the SAD guess that depend only on the locations of the nuclei (along with fixed parameters such as the one-electron basis set) thus satisfy time-reversibility, but extrapolations do not. Hence the velocity-Verlet/SAD algorithm affords time-reversible, symplectic dynamics (up to errors introduced by the finite convergence threshold), and indeed the "drift" reported for the SAD guess at $\varepsilon_{SCF} = 4$ (Table 1) is only slightly larger than the noise and probably does not represent a true secular error. In contrast, several of the extrapolations schemes afford energy drifts that are one or more orders of magnitude larger than the noise, although the drift decreases significantly when higher-order extrapolations are employed.

In their work on Fock matrix extrapolation, Pulay and Fogarasi[14] propose that energy drift stems from systematic error in the gradient, which in turn arises from overdamped convergence when the SCF guess is very close to the converged solution. In other words, they propose that when using extrapolation, the HF wave function converges from the "same side" in Hilbert space at each time step. To test this idea, we set the SCF convergence threshold to 10^{-10} E_h and recomputed the energy gradient at each molecular geometry for several of the C_2F_4 trajectories obtained with $\varepsilon_{SCF} = 4$. AO integral thresholds were set to the same value in both calculations, in order to isolate the error due to incomplete SCF convergence. Fig. 2 illustrates the difference between the "exact" (10^{-10} E_h) force between the two carbon atoms and that obtained using $\varepsilon_{SCF} = 4$ and $\varepsilon_{SCF} = 6$ with either of two different SCF guesses. It is apparent that the SAD guess provides an unbiased error in the force, whereas the mean error using (4,2) extrapolation is clearly different from zero. Still, the force error is highly oscillatory and may take either sign, which means that the picture is somewhat more complicated than the simple explanation offeered by Pulay and Fogarasi.[14] It is also interesting that the overall magnitude of the force error is smaller when using (4,2) extrapolation than using the SAD guess, despite the systematic error in the former.

Paradoxically, while high-order extrapolations significantly reduce the energy drift at $\varepsilon_{SCF} = 4$, they do not always reduce
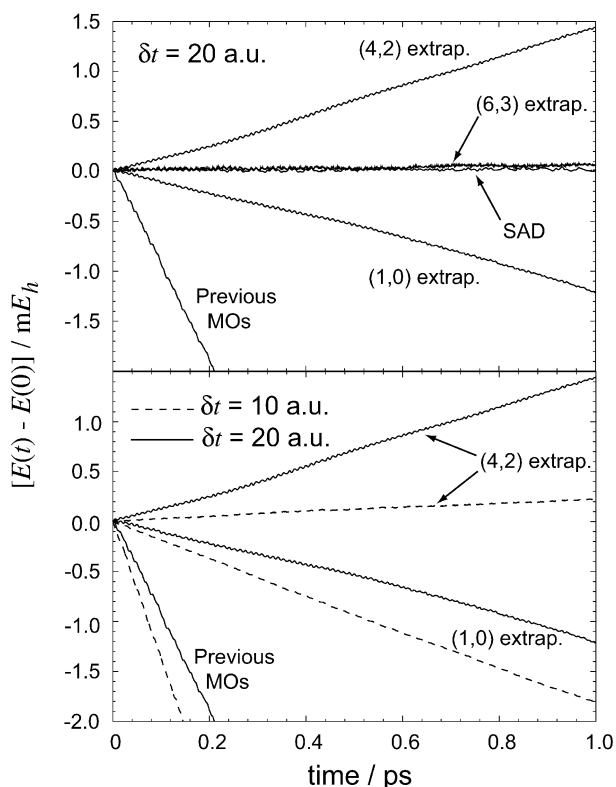
Fig. 1 Energy fluctuations for one C_2F_4 trajectory ($\varepsilon_{SCF} = 4$), obtained using different SCF guesses and time steps.

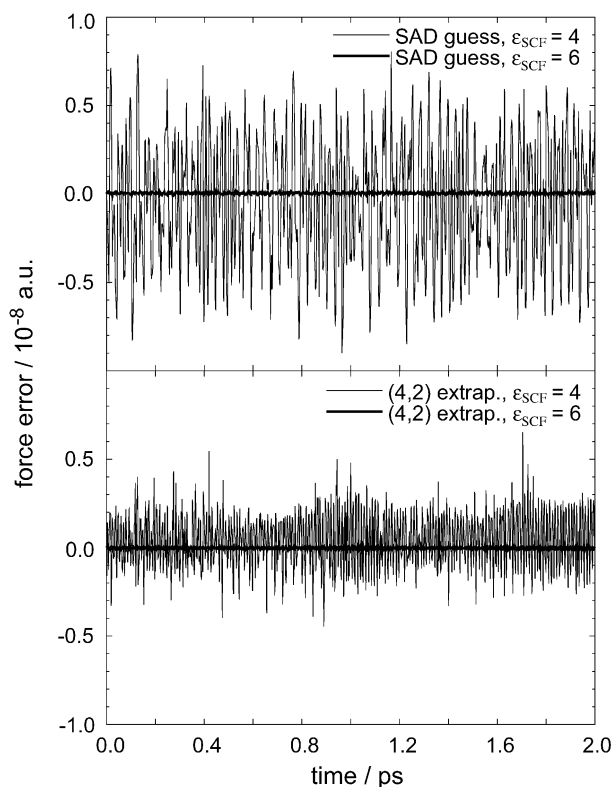**Fig. 2** Errors in the C–C force in C_2F_4.

Fig. 3 Errors in the C–C force in C_2F_4. Each panel uses a different vertical scale.

the overall magnitude of the force errors, as shown in Fig. 3. Only by increasing ε_{SCF} is the force error systematically reduced, and by the time $\varepsilon_{SCF} = 6$ there is little difference, overall, in the force errors for high- *versus* low-order extrapolations. The data listed in Table 1 for $\varepsilon_{SCF} = 6$ and $\varepsilon_{SCF} = 8$ testify that this is generally true: for $\varepsilon_{SCF} \geq 6$, all cubic and higher-order extrapolations afford precisely the same energy noise and energy drift as measured for the SAD guess, and this drift is smaller, per picosecond, than the energy noise. We conclude that the behavior of Fock extrapolation at $\varepsilon_{SCF} = 4$ is an artifact of the erratic behavior induced by a loose convergence threshold. (Q-Chem's default is $\varepsilon_{SCF} = 8$ for calculations that require the energy gradient.) This erratic behavior can be seen, for example, in the fact that for $\varepsilon_{SCF} \geq 6$ the average number of Fock builds decreases steadily as the extrapolation is taken to higher orders (though ultimately reaching some limit), whereas for $\varepsilon_{SCF} = 4$, the smallest value of N_{FB} is obtained using a quadratic extrapolation.

Although Pulay and Fogarasi[14] considered only $\varepsilon_{SCF} = 4$, other work[1] also indicates that BOMD energy drifts are reduced significantly by tightening the SCF convergence threshold, though at a cost of additional SCF cycles at each time step. In the present case, use of $\varepsilon_{SCF} = 6$ affords energy-conserving dynamics and requires essentially just one Fock build beyond what is required for the most efficient (non-energy-conserving) extrapolations at $\varepsilon_{SCF} = 4$. This is a small price to pay for preserving a fundamental invariant of the dynamics.

In Gaussian basis sets, the SCF gradient computation is 2–5 times more expensive than a single Fock build, and intuition therefore suggests that it is disadvantageous to decrease δt, even though this increases the accuracy of the extrapolation. To confirm this intuition, we simulated the same HF/3-21G trajectories for C_2F_2 using a time step of 10 au rather than 20 au; the results are summarized in Table 2. For high-order extrapolations, the smaller time step reduces the average number of Fock builds by one build or fewer, but requires twice as many gradient calculations. Moreover, at $\varepsilon_{SCF} = 4$, energy conservation is not necessarily any better than it is at
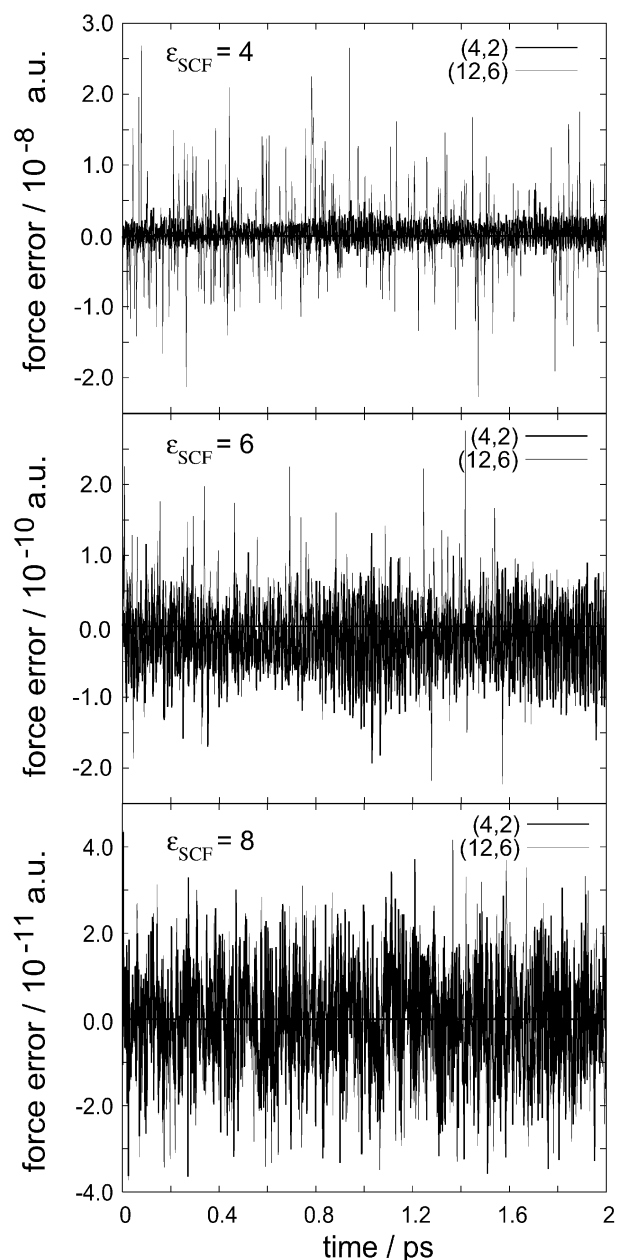
the smaller value of δt; Fig. 1 illustrates that the rate of energy drift may either increase or decrease with the time step, for the same initial conditions. Thus, even when using Fock matrix extrapolation, one should make δt as large as possible, with the upper limit dictated by acceptable energy noise.

In fact, even the larger time step explored above is quite conservative, since the shortest normal mode vibrational period in C_2F_4 is 15 fs. Exploratory calculations with $\delta t = 100$ au (≈ 2.42 fs), in conjunction with (12,6) extrapolation, afforded an energy noise of only 4×10^{-5} E_h over 1.0 ps of simulation time, with essentially zero drift. We will continue to utilize rather conservative (though not unreasonable) time steps in this work, in order to demontrate that energy drift can be eliminated under demands of high accuracy. In chemical applications, it may be possible to increase the time step considerably.

3.2. Realistic applications

The HF/3-21G model is really only a toy, useful for testing a broad swath of extrapolation schemes but not representative of

Table 2 Energy fluctuations and average number of Fock builds, N_{FB}, for five 2.0 ps HF/3-21G trajectories for C_2F_4 at $T = 500$ K ($\delta t = 10$ au)

| | $\varepsilon_{SCF} = 4$ | | | $\varepsilon_{SCF} = 6$ | | | $\varepsilon_{SCF} = 8$ | | |
| | | $\delta E/\mu E_h$ | | | $\delta E/\mu E_h$ | | | $\delta E/\mu E_h$ | |
SCF guess	N_{FB}	Noise	Drift/ps^{-1}	N_{FB}	Noise	Drift/ps^{-1}	N_{FB}	Noise	Drift/ps^{-1}
SAD	6.1	11.2	11.6	9.3	1.0	0.2	13.2	1.0	0.0
Previous MOs	2.0	364.0	6123.6	5.9	1.0	2.0	9.7	1.0	0.0
(1,0) extrap.	4.0	9.2	1136.2	6.6	0.9	2.6	10.0	1.0	0.0
(4,2) extrap.	1.2	4.6	114.5	2.3	1.1	31.6	5.8	1.0	0.1
(6,3) extrap.	2.0	8.0	33.9	2.1	1.0	1.0	4.3	1.0	0.0
(8,4) extrap.	2.5	8.0	20.8	2.4	1.0	0.3	3.0	1.0	0.1
(12,6) extrap.	3.0	7.6	16.5	2.7	0.9	0.3	2.9	1.0	0.0

realistic BOMD applications. In this section we explore several more realistic problems, starting with C_2F_4 in a larger basis set, 6-311G*, and also using potentials from density functional theory (DFT). A statistical summary of HF/6-311G* calculations for C_2F_4 appears in Table 3, while Table 4 gives the same information for BLYP[26,27] and B3LYP[28,29] simulations of this molecule. The trends are mostly the same as those observed at the HF/3-21G level: large energy drifts are observed with $\varepsilon_{SCF} = 4$, but these disappear for $\varepsilon_{SCF} \geq 6$, whereupon all methods yield comparable, δt-limited energy noise, although in the DFT case this requires somewhat higher-order extrapolations than were necessary in HF calculations. For the HF and B3LYP models, we achieve energy-conserving dynamics with an average of as few as three Fock builds per time step. For BLYP, an average of four builds are required per time step, consistent with our experience that BLYP is typically more difficult to converge than B3LYP.

The electronic structure of the C_2F_4 molecule is admittedly very simple, so in the remainder of this work we examine more challenging electronic structure problems, including a hydrogen-bonded cluster, a radical anion, and a transition metal complex. For the first of these, we selected the "book" isomer[30] of $(H_2O)_6$. One might expect a degradation in the performance of AO-based extrapolation procedures when applied to clusters (as opposed to purely covalent systems), since intermolecular

interactions can lead to significant recoupling of the AOs as the monomers reorient relative to one another. Table 5 summarizes the results of three 1.0 ps simulations of $(H_2O)_6$ at $T = 300$ K, using the HF, BLYP, and B3LYP potentials in the 6–31G* basis, with $\varepsilon_{SCF} = 6$. As in the C_2F_4 simulations, high-order extrapolations like (12,6) and (16,8) afford energy-conserving dynamics at a cost of three Fock builds per time step, for both DFT and HF models; lower-order extrapolations like (6,3) occasionally engender a slight energy drift. Even though our simulations extend only to $t = 1.0$ ps (4134 time steps of width $\delta t = 10$ au), we do not expect the rate of energy drift to increase substantially over longer time scales, since BOMD simulations out to 8 ps exhibit energy drifts that are still linear in time,[1] while large-scale molecular dynamics simulations on analytic potentials exhibit energy drifts that are linear in time out to at least 1 ns.[17]

For the $(H_2O)_6$ system, extrapolation reduces the number of Fock builds by 65–70% relative to using the SAD guess at each time step, though the reduction in cpu time is smaller than this factor, because in Gaussian basis sets the cost of an SCF gradient calculation is 2–5 times that of a single Fock build. Specifically, for $(H_2O)_6$ in the 6-31G* basis, the gradient computation is 4.9 times more expensive than a single Fock build, at the HF level; 3.6 times more expensive at the B3LYP level; and 2.1 times more expensive at the BLYP level. (These

Table 3 Energy fluctuations and average number of Fock builds, N_{FB}, for five 2.0 ps HF/6-311G* trajectories for C_2F_4 at $T = 500$ K ($\delta t = 20$ au)

| | $\varepsilon_{SCF} = 4$ | | | $\varepsilon_{SCF} = 6$ | | | $\varepsilon_{SCF} = 8$ | | |
| | | $\delta E/\mu E_h$ | | | $\delta E/\mu E_h$ | | | $\delta E/\mu E_h$ | |
SCF guess	N_{FB}	Noise	Drift/ps^{-1}	N_{FB}	Noise	Drift/ps^{-1}	N_{FB}	Noise	Drift/ps^{-1}
SAD	6.0	11.2	16.9	9.6	4.1	2.5	13.3	3.5	0.3
Previous MOs	2.0	887.7	4997.0	6.0	3.7	16.0	10.2	3.6	0.4
(6,3) extrap.	2.0	24.5	65.8	2.8	3.6	8.9	6.0	3.6	0.2
(12,6) extrap.	3.1	25.2	48.3	3.1	3.6	0.4	4.1	3.6	0.2
(16,8) extrap.	3.5	26.1	53.9	3.1	3.6	0.3	3.9	3.6	0.2

Table 4 Energy fluctuations and average number of Fock builds, N_{FB}, for five 1.0 ps DFT/6-311G* trajectories for C_2F_4 at $T = 500$ K ($\delta t = 20$ au)

| | $\varepsilon_{SCF} = 6$ | | | $\varepsilon_{SCF} = 8$ | | |
| | | $\delta E/\mu E_h$ | | | $\delta E/\mu E_h$ | |
Functional SCF guess	N_{FB}	Noise	Drift/ps^{-1}	N_{FB}	Noise	Drift/ps^{-1}
BLYP SAD	10.8	2.6	0.6	13.4	2.6	0.3
BLYP (6,3)	3.3	2.6	15.9	6.8	2.6	0.3
BLYP (12,6)	4.0	2.7	1.4	4.5	2.6	0.3
B3LYP SAD	9.6	3.4	2.4	13.0	3.4	0.5
B3LYP (6,3)	3.8	3.4	12.2	6.4	3.4	0.5
B3LYP (12,6)	3.0	3.4	0.9	4.1	3.4	0.5

Table 5 Statistical summary of three 1.0 ps trajectories for $(H_2O)_6$ at $T = 300$ K in the 6-31G* basis set ($\varepsilon_{SCF} = 6$, $\delta t = 10$ au). Also listed is the relative cpu time for each SCF guess method, as well as the ratio of time spent in SCF convergence to time spent in gradient computation

Functional	SCF guess	N_{FB}	Relative cpu time	SCF time/grad. time	$\delta E/\mu E_h$ Noise	Drift/ps^{-1}
HF	SAD	9.0	1.00	1.85	8.7	1.8
HF	(6,3)	3.1	0.58	0.64	8.8	92.9
HF	(12,6)	3.2	0.56	0.61	8.7	0.5
HF	(16,8)	3.2	0.57	0.61	8.7	0.9
BLYP	SAD	10.5	1.00	4.92	8.1	5.5
BLYP	(6,3)	4.0	0.50	2.02	7.0	10.2
BLYP	(12,6)	3.4	0.46	1.76	7.0	2.9
BLYP	(16,8)	3.0	0.44	1.61	5.0	2.8
B3LYP	SAD	9.0	1.00	2.56	6.9	1.9
B3LYP	(6,3)	3.9	0.60	1.15	6.9	5.3
B3LYP	(12,6)	3.2	0.55	0.97	6.9	1.7
B3LYP	(16,8)	3.0	0.54	0.91	6.9	2.0

values are averages over all time steps, and incremental Fock builds[31,32] have been used to accelerate computation of the Fock matrix.) The ratio of SCF convergence time to SCF gradient time for each $(H_2O)_6$ calculations is listed in Table 5. Absent Fock matrix extrapolation, SCF convergence is the overwhelming bottleneck at each time step, but with high-order extrapolations this step is comparable to, or even faster than, computation of the SCF energy gradient. Finally, we list in Table 5 the total cpu time that is required for dynamics with Fock matrix extrapolation, as a fraction of the time required for dynamics using only the SAD guess. Due to the relatively high cost of gradient calculations, the aforementioned 65–70% reduction in the number of Fock builds translates into about a 45–55% reduction in overall cpu time.

In SCF electronic structure calculations, radicals and transition metal complexes typically require a larger number of SCF iterations than closed-shell, main-group molecules. In recent work,[33] we have simulated temperature-dependent photoelectron spectra for the radical anion $(H_2O)_4^-$, and here we take advantage of the rather large number of trajectories required for that study in order to demonstrate the robustness of the Fock matrix extrapolation procedure. The $(H_2O)_4^-$ system presents a more challenging test of extrapolation than either C_2F_4 or $(H_2O)_6$ because the excess electron in $(H_2O)_4^-$ is bound only by the dipole moment of the underlying water cluster, so highly diffuse basis sets are required in order to obtain accurate electron detachment energies. Our simulations of $(H_2O)_4^-$ utilize the B3LYP/6-31(1+,3+)G* model,[34] where the indicated basis set contains one set of diffuse sp functions on each oxygen atom and three sets of diffuse s functions on each hydrogen atom, with orbital exponents that are decre-

mented by successive factors of 3.32. The most diffuse of these Gaussian basis functions has a full width at half maximum of 15.4 Å, hence the diffuse basis functions centered on different atoms overlap significantly and thus the singly-occupied molecular orbital can fluctuate markedly, coupling together different AOs as the dipole moment vector changes in time.

A statistical summary of $(H_2O)_4^-$ trajectories appears in Table 6, where the averages involve three different structural isomers and 50–100 half-picosecond trajectories at each of six different temperatures. All simulations used (12,6) extrapolation with $\varepsilon_{SCF} = 8$. Whereas the SAD guess consistently requires 16 Fock builds per time step, extrapolation reduces N_{FB} to 7.5–8.5 (depending somewhat on the temperature). In hindsight, the use of $\varepsilon_{SCF} = 8$ seems overly conservative; exploratory calculations for $T = 50$ K and $T = 300$ K using $\varepsilon_{SCF} = 6$ show a reduction in N_{FB} to 5–7 builds per time step using (12,6) extrapolation, whereas the SAD guess consistently requires 11 Fock builds with this threshold. In either case, Fock matrix extrapolation reduces N_{FB} by 45–50% relative to the SAD guess. This is not quit as dramatic as the 65–70% reduction obtained for $(H_2O)_6$, reflecting the more complicated electronic structure of $(H_2O)_4^-$.

As a final example, we have simulated the dynamics of the octahedrally-coordinated $[Fe(H_2O)_6]^{2+}$ complex at $T = 150$ K using the B3LYP/6-31G* potential. The results (Table 7) are averaged over three 0.5 ps trajectories using $\varepsilon_{SCF} = 6$. Relative to B3LYP/6-31G* simulations of $(H_2O)_6$ using the same convergence threshold (Table 5), this complex requires approximately four more Fock builds per time step, absent extrapolation. Using high-order extrapolation, however, only about two additional builds are required relative to extrapolated simulations for the electronically simpler $(H_2O)_6$ complex. Just as we saw for $(H_2O)_6$, in the case of $[Fe(H_2O)_6]^{2+}$ Fock matrix extrapolation reduces the total cpu time for SCF convergence by about 45%, meaning that this step consumes essentially the same amount of time as calculation of the SCF energy gradient.

Table 6 Energy fluctuations and average number of Fock builds, N_{FB}, for simulations of $(H_2O)_4^-$ at various temperatures, using (12,6) extrapolation and $\varepsilon_{SCF} = 8$. (Standard deviations are given in parentheses.)

T/K	N_{FB}	$\delta E/\mu E_h$ Noise	Drift/ps^{-1}
50[a]	7.3(0.9)	2.7(1.4)	3.1 (4.8)
100[a]	7.8(0.8)	5.4(2.9)	2.9 (2.5)
150[b]	8.0(0.8)	7.7(3.5)	2.7 (2.9)
200[b]	8.2(0.9)	10.4(5.4)	4.0 (3.9)
250[c]	8.4(1.0)	16.1(8.5)	5.9 (6.7)
300[c]	8.5(1.1)	16.5(8.6)	9.5 (11.5)

[a] Average over 50 trajectories. [b] Average over 100 trajectories. [c] Average over 60 trajectories.

Table 7 Statistical summary of three 0.5 ps B3LYP/6-31G* trajectories for $[Fe(H_2O)_6]^{2+}$ at $T = 150$ K ($\varepsilon_{SCF} = 6$, $\delta t = 10$ au)

SCF guess	N_{FB}	Relative cpu time	SCF time/ grad. time	$\delta E/\mu E_h$ Noise	Drift/ps^{-1}
SAD	13.2	1.00	2.71	8.1	24.4
(6,3)	4.3	0.50	0.86	7.1	18.4
(12,6)	5.0	0.54	0.98	7.0	1.9
(16,8)	5.3	0.55	1.03	7.1	2.3

4. Summary

Catastrophic energy drift in BOMD is an artifact of excessively loose SCF convergence thresholds; using even a modest threshold (10^{-6} E_h in the occupied-virtual Fock matrix elements), systematic energy drift is reduced to a few microhartree ($\sim 10^{-3}$ kcal mol^{-1}) per picosecond of simulated time, a rate that appears to be stable over time scales currently accessible to *ab initio* molecular dynamics. For all practical purposes, this means that energy drift is eliminated entirely, since typical molecular dynamics time steps introduce unbiased energy fluctuations on the order of a few microhartree.

Extrapolation of Fock matrices retained from previous time steps reduces the overall cost of BOMD simulations by 45–55%. For the convergence threshold mentioned above, extrapolation using quartic or higher polynomials is recommended, as opposed to the quadratic scheme originally recommended by Pulay and Fogarasi.[14] In particular, the "(12,6)" extrapolation scheme (12 saved Fock matrices and a sixth-degree polynomial extrapolation), in conjuction with an SCF convergence threshold of 10^{-6} E_h, provides energy-conserving dynamics for a variety of nontrivial systems, including radicals and transition metal complexes, using both Hartree–Fock and density functional potentials in a variety of common Gaussian basis sets. With the aforementioned threshold, SCF convergence is consistently achieved in 3–5 Fock builds, which means that SCF convergence is no longer the overwhelming bottleneck in BOMD calculations, but instead is comparable in cost to calculation of the energy gradient. Fock matrix extrapolation considerably narrows the gap in efficiency between BOMD and extended-Lagrangian molecular dynamics methods, without introducing any additional approximations.

Acknowledgements

References

1 D. Marx and J. Hutter, in *Modern Methods and Algorithms of Quantum Chemistry*, ed. J. Grotendorst, John von Neumann Institute for Computing, Jülich, Germany, 2nd edn, 2000, p. 329.
2 R. Car and M. Parrinello, *Phys. Rev. Lett.*, 1985, **55**, 2471.
3 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.*, 2002, **114**, 9758.
4 J. M. Herbert and M. Head-Gordon, *J. Chem. Phys.*, 2004, **121**, 11542.
5 C. Raynaud, L. Maron, J.-P. Daudey and F. Jolibois, *Phys. Chem. Chem. Phys.*, 2004, **6**, 4226.
6 H. B. Schlegel, S. S. Iyengar, X. Li, J. M. Millam, G. A. Voth, G. E. Scuseria and M. J. Frisch, *J. Chem. Phys.*, 2002, **117**, 8694.
7 I.-F. W. Kuo, C. J. Mundy, M. J. McGrath, J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein, F. Mohamed, M. Krack and M. Parrinello, *J. Phys. Chem. B*, 2004, **108**, 12990.
8 V. Wathelet, B. Champagne, D. H. Mosley, J.-M. André and S. Massidda, *Chem. Phys. Lett.*, 1997, **275**, 506.
9 V. Wathelet, B. Champagne, D. H. Mosley, É. A. Perpète and J.-M. André, *J. Mol. Struct. (THEOCHEM)*, 1998, **425**, 95.
10 P. Tangney and S. Scandolo, *J. Chem. Phys.*, 2002, **116**, 14.
11 J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi and G. Galli, *J. Chem. Phys.*, 2004, **120**, 300.
12 E. Schwegler, J. C. Grossman, F. Gygi and G. Galli, *J. Chem. Phys.*, 2004, **121**, 5400.
13 J. VandeVondele, M. Krack, F. Mohammed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103.
14 P. Pulay and G. Fogarasi, *Chem. Phys. Lett.*, 2004, **386**, 272.
15 J. Kong, C. A. White, A. I. Krylov, C. D. Sherrill, R. D. Adamson, T. R. Furlani, M. S. Lee, A. M. Lee, S. R. Gwaltney, T. R. Adams, C. Ochsenfeld, A. T. B. Gilbert, G. S. Kedziora, V. A. Rassolov, D. R. Maurice, N. Nair, Y. Shao, N. A. Besley, P. E. Maslen, J. P. Dombroski, H. Daschel, W. Zhang, P. P. Korambath, J. Baker, E. F. C. Byrd, T. Van Voorhis, M. Oumi, S. Hirata, C. Hsu, N. Ishikawa, J. Florian, A. Warshel, B. G. Johnson, P. M. W. Gill, M. Head-Gordon and J. A. Pople, *J. Comput. Chem.*, 2000, **21**, 1532.
16 P. Pulay, *J. Comput. Chem.*, 1982, **3**, 556.
17 S. K. Gray, D. W. Noid and B. G. Sumpter, *J. Chem. Phys.*, 1994, **101**, 4062.
18 H. J. C. Berendsen, W. F. van Gunsteren, in *Molecular-Dynamics Simulation of Statistical-Mechanical Systems*, ed. G. Ciccotti and W. G. Hoover, North Holland, New York, 1986, vol. 97 of Proceedings of the International School of Physics 'Enrico Fermi', p. 43.
19 G. J. Martyna, M. E. Tuckerman, D. J. Tobias and M. L. Klein, *Mol. Phys.*, 1996, **87**, 1117.
20 S. Toxvaerd, *Phys. Rev. E*, 1993, **47**, 343.
21 A. K. Mazur, *J. Comput. Phys.*, 1997, **136**, 354.
22 P. J. Channell and C. Scovel, *Nonlinearity*, 1990, **3**, 231.
23 J. Candy and W. Rozmus, *J. Comput. Phys.*, 1991, **92**, 230.
24 J. J. Biesiadecki and R. D. Skeel, *J. Comput. Phys.*, 1993, **109**, 318.
25 M. Shimada and H. Yoshida, *Publ. Astron. Soc. Jpn.*, 1996, **48**, 147.
26 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098.
27 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.
28 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.
29 P. J. Stephens, J. F. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623.
30 J. K. Gregory and D. C. Clary, *J. Phys. Chem.*, 1996, **100**, 18014.
31 J. Almlöf, K. Faegri and K. Korsell, *J. Comput. Chem.*, 1982, **3**, 385.
32 E. Schwegler, M. Challacombe and M. Head-Gordon, *J. Chem. Phys.*, 1997, **106**, 9708.
33 J. M. Herbert and M. Head-Gordon, manuscript in preparation.
34 J. M. Herbert and M. Head-Gordon, *J. Phys. Chem. A*, 2005, **109**, 5217.