Elastic molecular dynamics with self-consistent flexible constraints

Jin Zhou
Laboratory of Biophysical Chemistry, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892-5626

Sebastian Reich
Department of Mathematics and Statistics, University of Surrey, Guildford, Surrey GU2 5XH, England

Bernard R. Brooks
Laboratory of Biophysical Chemistry, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892-5626

(Received 30 April 1999; accepted 11 November 1999)

A new algorithm for constrained molecular dynamics is proposed. In contrast to the standard approach, the constrained bond-length/bond-angle value is adjusted at each time step so that total energy is minimized with respect to the constrained distances. This can be viewed as modifying the equilibrium bond-length/bond-angle according to external and centripetal forces. Two approaches are constructed to implement the algorithm. Method I includes all energy terms, but it is neither holonomic nor symplectic. Method II neglects a rotational kinetic energy term, resulting in a more expensive symplectic integrator. Both integrators are reversible and well conserve total energy. Due to the iterative nature of the optimization, which requires an energy evaluation at each iteration, these methods are intended for use in conjunction with other integration techniques that require an optimization at each time step iteration, such as the self-consistent electronic polarization methods or iterative combined quantum mechanics/molecular mechanics (QM/MM) hybrid simulations. Simulation results for collisions between two diatomic molecules, two water molecules, and for a periodic water box are compared and contrasted with fully constrained and free dynamics. These results demonstrate the utility and efficacy of these new methods. © *2000 American Institute of Physics.* [S0021-9606(00)50906-4]

I. INTRODUCTION

In molecular dynamics, the primary restriction on time step comes from the presence of rapid vibrations in the chemical bonds and angles. There are quite a few techniques reviewed in a special issue of Journal of Computational Physics to achieve computational gains.[1] One of the popular methods to increase time step is to constrain those bonds to their equilibrium length. The corresponding modifications in the Verlet method lead to the SHAKE[2] or RATTLE[3] discretization. Although SHAKE/RATTLE have been applied successfully to a large variety of problems, it has also been observed that constraining bond-angles (and perhaps also bond-lengths) can lead to a too rigid approximation of the free dynamics.[4,5]

In this article we propose a new constrained formulation. In our method the bonds are no longer constrained to their equilibrium length. Instead we define an instantaneous bond-length constraint by demanding that the same total force acting on the bond is equal to zero. The key to this method is to allow molecular flexibility without the necessity of including high frequency vibrations. This allows longer time steps with fully flexible potential energy functions and avoids non-physical heat capacity for high frequency vibrations when constraints are not used.

The derivation of our method is given in Sec. III. In Sec. IV, we discuss various numerical aspects and show how to implement the method within the standard SHAKE/RATTLE routine. Finally, in Sec. V, we present results from simulations of collision between two diatomic molecules, collision between two water molecules, and for a periodic water box.

Since the new methods require an energy optimization at each time step, they are computationally more expensive than the SHAKE method, which in turn is only slightly more expensive than the free dynamics method. While we did not attempt to optimize the efficiency of our method, the methods are more useful when combined with other integration techniques that require an optimization at each time step, such as the self-consistent electronic polarization methods[6–8] or iterative combined QM/MM hybrid simulations.[9]

II. BACKGROUND

For classical molecular dynamics (MD) of macromolecules, atomic trajectories for N atoms obey the Hamiltonian equations of motion[10]

$$\dot{p} = -\nabla_q V(q), \tag{1}$$

$$\dot{q} = M^{-1} p, \tag{2}$$

where $q \in \mathbf{R}^{3N}$ and $p \in \mathbf{R}^{3N}$ are the vectors containing all positions and momenta, M is a diagonal matrix of atomic masses with each mass replicated thrice, and $V(q)$ is the (empirical) potential energy function. Typically this potential energy function might be written[10]

 © 2000 American Institute of Physics

$$V(q)=V_b(q)+V_\theta(q)+V_\phi(q)+\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\psi_{ij}(|q_i-q_j|), \tag{3}$$

where $q_i\in\mathbf{R}^3$ and $q_j\in\mathbf{R}^3$ denote the positions of the ith and jth atoms, ψ represents terms due to nonbonded interactions, and $|\cdots|$ denotes Euclidian distance in $\mathbf{R}^3$. $V_b(q)$ is the component of the potential due to deformations of chemical bonds between atom pairs. A bond α between atoms i and j is usually described by a harmonic term of the form

$$\frac{K_\alpha}{2}(|q_i-q_j|-L_\alpha)^2, \tag{4}$$

where L_α is the equilibrium length of the bond, and K_α is the force constant of the bond. Terms V_θ and V_ϕ for bond-angles and dihedrals are also present. Often the rapid vibrations in the chemical bonds are removed by imposing the algebraic constraints

$$g_\alpha(q):=|q_i-q_j|-L_\alpha=0. \tag{5}$$

This leads to Hamiltonian equations of motion subject to m holonomic constraints

$$\dot{p}=-\nabla_q V(q)-G(q)^t\lambda, \tag{6}$$

$$\dot{q}=M^{-1}p, \tag{7}$$

$$0=g(q). \tag{8}$$

Here m is the number of constrained bond-vibrations, $g:\mathbf{R}^{3N}\to\mathbf{R}^m$, $g=\{g_\alpha\}_{\alpha=1,\ldots,m}$, $G(q)=g_q(q)$ is the matrix of partial derivatives of g with respect to position q, and $\lambda\in\mathbf{R}^m$ is a vector of Lagrange multipliers. The solution manifold underlying Eqs. (6)–(8) is

$$\mathcal{M}=\{(q,p)|g(q)=0,\quad G(q)M^{-1}p=0\}. \tag{9}$$

The *hidden* constraint $G(q)M^{-1}p=0$ is obtained through time differentiation of the position constraint. One further differentiation of the hidden constraint with respect to time yields

$$0=G(q)M^{-1}\dot{p}+g_{qq}(q)(M^{-1}p,\dot{q}), \tag{10}$$

where $g_{qq}(q)$ denotes the second derivative of g with respect to position q. Using now Eqs. (6) and (7), we obtain an equation solvable for the Langrange multipliers λ, i.e.,

$$\lambda(q,p):=[G(q)M^{-1}G(q)^t]^{-1}[g_{qq}(q)(M^{-1}p,M^{-1}p)-G(q)M^{-1}\nabla V_q(q)]. \tag{11}$$

The constrained equations of motion (6)–(8) can conveniently be discretized by the RATTLE modification[3] of the Verlet method, i.e.,

$$q_{n+1}=q_n+\Delta tM^{-1}p_{n+1/2}, \tag{12}$$

$$p_{n+1/2}=p_n-\tfrac{1}{2}\Delta t\nabla_q V(q_n)-\tfrac{1}{2}\Delta tG(q_n)^t\lambda_n, \tag{13}$$

$$0=g(q_{n+1}), \tag{14}$$

$$p_{n+1}=p_{n+1/2}-\tfrac{1}{2}\Delta t\nabla_q V(q_{n+1})-\tfrac{1}{2}\Delta tG(q_{n+1})^t\mu_n, \tag{15}$$

$$0=G(q_{n+1})M^{-1}p_{n+1}, \tag{16}$$

which was shown to be symplectic on the solution manifold $\mathcal{M}$,[11] i.e., the solutions preserve the wedge product $dq\wedge dp$,[12] i.e.,

$$dq_n\wedge dp_n=dq_{n+1}\wedge dp_{n+1}, \tag{17}$$

which implies good long-term energy conservation.

Upon neglecting entropic corrections to the equations of motion,[5,13,14] the constrained system, Eqs. (6)–(8) can be shown to reproduce the dynamics of the free system correctly if the force constants K_α are sufficiently large.[13] In MD, this assumption is not strictly satisfied and one might have to take into account effects due to finite values of K_α. One possibility is to let the actual bond-lengths $r_\alpha=|r_i-r_j|$ adjust according to the external forces acting on them instead of freezing them to their equilibrium values L_α. Two approaches to realize that are described in detail in the next section.

III. THE METHODS

The idea is simple: The Lagrangian multipliers $\lambda(q,p)$ introduce through $-G(q)^t\lambda(q,p)$ the force that is needed to satisfy the constraint $g(q)=0$. The same force is obtained through bond-stretching if it satisfies

$$-G(q)^tKg(q)=-G(q)^t\lambda(q,p), \tag{18}$$

which is equivalent to

$$g(q)=K^{-1}\lambda(q,p), \tag{19}$$

where the $m\times m$ matrix K is the diagonal matrix of force constants K_α. This suggests the modified constraint

$$0=g(q)-K^{-1}\lambda(q,p), \tag{20}$$

which can be understood as adjusting the equilibrium bond-length L_α to the modified value

$$\tilde{L}_\alpha(q,p):=L_\alpha+K_\alpha^{-1}\lambda_\alpha(q,p), \tag{21}$$

where λ_α is the Lagrange multiplier corresponding to the bond $r_\alpha=|r_i-r_j|$. If K_α is large enough, $\tilde{L}_\alpha(q,p)\approx L_\alpha$ and the correction can be neglected. We are interested in the case that K_α takes moderate values, typical of those required to reproduce experimental spectroscopic data for small molecules. The modified constraint Eq. (20) can be imposed in the same way as the original constraint $g(q)=0$ and we obtain

$$\dot{p}=-\nabla_q V(q)-G(q)^t\Lambda, \tag{22}$$

$$\dot{q}=M^{-1}p, \tag{23}$$

$$0=g(q)-K^{-1}\lambda(q,p). \tag{24}$$

The new Langrange multipliers $\Lambda\in\mathbf{R}^m$ are again obtained by twice differentiating the constraint Eq. (24) with respect to time. This time this leads to

$$0=\ddot{g}-K^{-1}\ddot{\lambda}=-G(q)M^{-1}\nabla_q V(q)+g_{qq}(q)\times(M^{-1}p,M^{-1}p)-[G(q)M^{-1}G(q)^t]\Lambda-K^{-1}\ddot{\lambda}, \tag{25}$$

and the force term $-G(q)^t\Lambda$ due to the Langrange multipliers Λ is now only of order K^{-1} since

$$-G(q)M^{-1}\nabla_q V(q)+g_{qq}(q)(M^{-1}p,M^{-1}p)=O(K^{-1}), \tag{26}$$

under the modified constraint Eq. (20). Thus the equations of motion (22)–(24) without the force term $G(q)^t\Lambda$ are almost self-consistent and provide a better approximation to the free dynamics of Eqs. (1)–(2) than the system Eqs. (6)–(8). We like to point out that, upon neglecting entropic effects, Eqs. (22)–(24) can be justified rigorously through averaging over the highly oscillatory bond-vibrations.[14]

The constraint Eq. (24) can also be understood as the minimum of total energy

$$E(q,p)=\frac{p^tM^{-1}p}{2}+V(q), \tag{27}$$

with respect to the bond lengths r_α, i.e., Eq. (24) is, for each bond r_α, equivalent to

$$\nabla_{r_\alpha}E(q,p)=0. \tag{28}$$

We call this approach our Method I.

Let us now discuss conservation properties of Eqs. (22)–(24). First note that the constraint (24) is nonholonomic. Thus the equations of motion Eqs. (22)–(24) are no longer Hamiltonian. However, the equations (22)–(24) are time-reversible. This follows from the invariance of the equations under the transformation $(q,p)\rightarrow(q-p)$ and subsequent time-reversion $t\rightarrow -t$. The overall system is also invariant under translations and rotations which implies exact preservation of linear and angular momentum. Due to the non-Hamiltonian character of the equations, total energy is no longer a first integral. But

$$\dot{E}=p^tM^{-1}\dot{p}+\dot{q}^t\nabla_q V(q)=-p^tM^{-1}G(q)^t\Lambda, \tag{29}$$

and, since $GM^{-1}p=O(K^{-1})$ and $\Lambda=O(K^{-1})$,

$$\dot{E}=-\Lambda^tG(q)M^{-1}p=O(K^{-2}), \tag{30}$$

and total energy is an adiabatic invariant.[12]

If we only minimize the potential energy $V(q)$ instead of the total energy as in Eq. (28), i.e.,

$$\nabla_{r_\alpha}V(q)=0, \tag{31}$$

then the momentum-dependent term in the constraint Eq. (24) vanishes and we obtain the following modified equilibrium bond-lengths

$$\tilde{L}_\alpha(q)=L_\alpha+K_\alpha^{-1}\lambda_\alpha(q,p=0), \tag{32}$$

which leads to the holonomic constraint

$$0=\tilde{g}(q):=g(q)+K^{-1}[G(q)M^{-1}G(q)^t]^{-1}G(q)M^{-1}\nabla V_q(q). \tag{33}$$

These constraints can be imposed on the free dynamics Eqs. (1)–(2) through

$$\dot{p}=-\nabla_q V(q)-\tilde{G}(q)^t\Lambda, \tag{34}$$

$$\dot{q}=M^{-1}p, \tag{35}$$

$$0=\tilde{g}(q), \tag{36}$$

with $\tilde{G}(q)$ the Jacobian of $\tilde{g}(q)$, i.e., $\tilde{G}(q)=\tilde{g}_q(q)$. Now the equations of motion are Hamiltonian and total energy $E(q,p)$ is preserved exactly.[12,15] The system is also time-reversible and preserves linear and angular momentum exactly. This follows from the invariance of the equations of motion under the corresponding transformations as described for Method I.

The price to be paid is a modified Lagrangian force term $-\tilde{G}(q)^t\Lambda$ in Eq. (34) which will become important when discussing the numerical integration of (34)–(36). We term this approach our Method II. A modification to this method leads to a symplectic method that does not require the Hessian of the force field.[1,16]

Note that the alternate simplified formulation

$$\dot{p}=-\nabla_q V(q)-G(q)^t\lambda, \tag{37}$$

$$\dot{q}=M^{-1}p, \tag{38}$$

$$0=\tilde{g}(q), \tag{39}$$

using $-G(q)^t\Lambda$ instead of $-\tilde{G}(q)^t\Lambda$ is not Hamiltonian. Even worse, the time derivative of total energy satisfies only $\dot{E}=O(K^{-1})$. This will become important when considering numerical discretization.

In summary, two methods are presented. Method I includes a kinetic energy contribution (related to centrifugal forces on rotating systems) in the formulation which causes the constraint to be nonholonomic and the system of equations to be non-Hamiltonian. There is, however, an ''energy like'' quantity that is conserved with this system. Method II derives from the same basic equations except that the kinetic energy term is discarded. This provides a Hamiltonian system allowing a symplectic integration scheme. Both methods are time-reversible and both conserve linear and angular momentum.

IV. NUMERICAL ASPECTS

A. Method I

We start with the discretization of our Method I of the constrained formulation, Eqs. (22)–(24). Straightforward generalization of the RATTLE discretization (12)–(16) yields

$$q_{n+1}=q_n+\Delta tM^{-1}p_{n+1/2}, \tag{40}$$

$$p_{n+1/2}=p_n-\tfrac{1}{2}\Delta t\nabla_q V(q_n)-\tfrac{1}{2}\Delta tG(q_n)^t\lambda_n, \tag{41}$$

$$p_{n+1}=p_{n+1/2}-\tfrac{1}{2}\Delta t\nabla_q V(q_{n+1})-\tfrac{1}{2}\Delta tG(q_{n+1})^t\mu_n, \tag{42}$$

$$0=g(q_{n+1})-K^{-1}\lambda(q_{n+1},p_{n+1}), \tag{43}$$

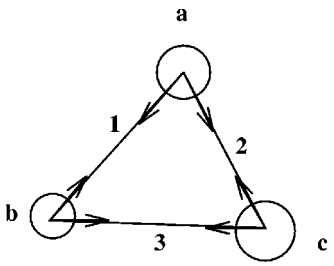

FIG. 1. The centripetal forces in a nonlinear triatomic molecule.

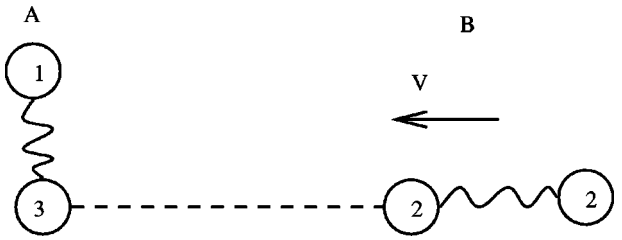

FIG. 2. The collision model for two hypothetical diatomic molecules.

$$0 = G(q_{n+1}) M^{-1} p_{n+1}. \tag{44}$$

The method is time-reversible as can be seen from replacing the momenta p by $-p$ and the step size Δt by $-\Delta t$ which leads back to the given method. The method also exactly preserves linear and angular momentum as follows from the fact that the underlying equations of motion preserve linear and angular momentum and the fact that this property is maintained under a SHAKE/RATTLE-type discretization.[17,18]

Note that, through Eq. (43), the scheme is now implicit in the force field $F_{n+1} = -\nabla V(q_{n+1})$. However, due to the factor K^{-1} in front of $\lambda(q_{n+1}, p_{n+1})$, one might only have to include the strong nearest-neighborhood interactions in the evaluation of $\lambda(q,p)$. The corresponding splitting of the potential energy V into strong and weak contributions can be done as for multiple-time stepping methods.[1] The evaluation of λ is now done by replacing V in Eq. (11) by V_{strong}.

To implement our method within the framework of the standard SHAKE iteration, the following iterative procedure for the solution of the nonlinear equation

$$0 = g(q_{n+1}) - K^{-1} \lambda(q_{n+1}, p_{n+1}), \tag{45}$$

can be used: Compute a first approximation $(q_{n+1}^{[1]}, p_{n+1}^{[1]})$ by solving Eqs. (40)–(44) with the constraint Eq. (43) replaced by the standard constraint

$$0 = g(q_{n+1}^{[1]}), \tag{46}$$

and the variable substitution $q_{n+1} = q_{n+1}^{[1]}$, $p_{n+1} = p_{n+1}^{[1]}$. This can be done by applying the standard SHAKE iteration.[2] Now evaluate $K^{-1}\lambda(q_{n+1}^{[1]}, p_{n+1}^{[1]})$ and compute an approximation $(q_{n+1}^{[2]}, p_{n+1}^{[2]})$ by solving Eqs. (40)–(44) again with Eq. (43) now replaced by

$$0 = g(q_{n+1}^{[2]}) - K^{-1} \lambda(q_{n+1}^{[1]}, p_{n+1}^{[1]}), \tag{47}$$

and the variable substitution $q_{n+1} = q_{n+1}^{[2]}$, $p_{n+1} = p_{n+1}^{[2]}$. Again the nonlinear equations can be solved using SHAKE. Continue this procedure till convergence. Of course, one could think of more efficient techniques for solving the nonlinear Eqs. (40)–(44). This can be achieved by replacing the SHAKE iteration by a more efficient solution method[19] and, more importantly, by iterating g and λ simultaneously in a quasiNewton-type iteration procedure.

We implement an algorithm for Method I by introducing an artificial force, called centrifugal force, due to the rotation of a bond. At each integration time step, a self-consistent iteration method is used to adjust the bond lengths. The algorithm first goes through each constrained bond and calculates the total forces, $\mathbf{F_1}$ and $\mathbf{F_2}$, including centrifugal force on the two atoms that form the bond. It then calculates the new bond length $L^{[j+1]}$ according to

$$L^{[j+1]} = L^{[j]} - \frac{(m_2 \mathbf{F_1} - m_1 \mathbf{F_2}) \cdot \mathbf{L}}{m L^{[j]} k_b}, \tag{48}$$

where m_1 and m_2 are masses of the two atoms, $m = m_1 + m_2$, k_b is the force constant of the bond, and j is the iteration step. After all new bond lengths are calculated, the standard SHAKE routine is invoked to adjust the coordinates to fit the new bond lengths. The procedure is iterated until the new bond lengths, $L^{[j+1]}$, and old bond lengths, $L^{[j]}$ are sufficiently close.

In calculating the centrifugal force, one needs the velocities at the time step t_{n+1} for each atoms. The velocities are not available in the leap-frog algorithm. We used the following velocity predictor to calculate the approximate velocity for each atom,

$$\mathbf{v}_{n+1} = \mathbf{v}_{n+1/2} + \frac{1}{2m} \Delta t \mathbf{F}_{n+1}, \tag{49}$$

where $\mathbf{F}_{n+1}$ is the potential force on the atom at time t_{n+1}, m is the mass of the atom and Δt is the integration time step.

It is somewhat trickier to calculate the centrifugal force for bonds in nonlinear triatomic molecules such as water since the three bonds are correlated. The three centripetal forces (opposite of centrifugal forces) must be solved simultaneously.

Consider a nonlinear triatomic molecule formed by three atoms a, b and c with masses m_a, m_b, and m_c and bonds 1, 2, and 3 (Fig. 1). Let the three centripetal forces along the bonds 1, 2, and 3 be f_1, f_2, and f_3, respectively. For the bond 1, the following equations must be satisfied for atom a and b.

$$\frac{f_1 + f_2 \cos(\angle bac)}{m_a} = \omega_1^2 r_a, \tag{50}$$

$$\frac{f_1 + f_3 \cos(\angle abc)}{m_b} = \omega_1^2 r_b, \tag{51}$$

where

$$r_a = \frac{m_b}{m_a + m_b} r_1, \quad r_b = \frac{m_a}{m_a + m_b} r_1, \tag{52}$$

r_1 is the bond length and ω_1 is the rotational speed of the bond 1 around its center of mass. Add Eqs. (51) and (52), one can get

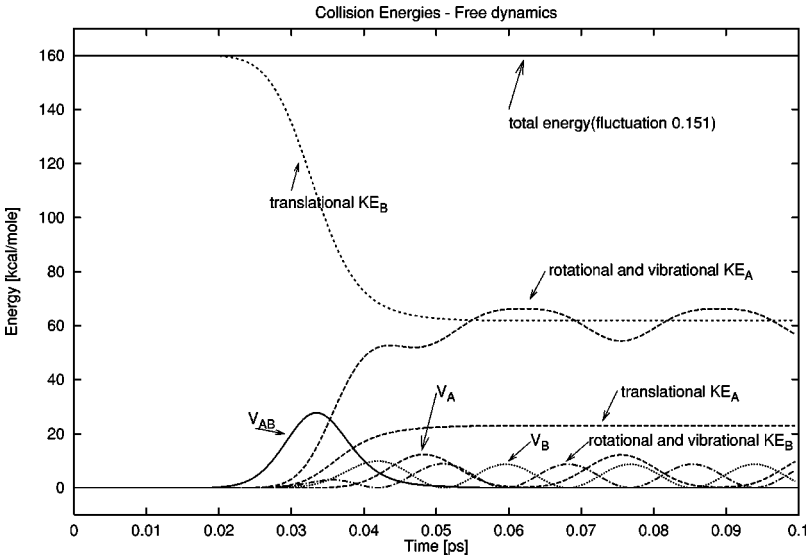

FIG. 3. Free dynamics method energy profile: Total energy and potential energy V_{AB} are in solid line; energies associated with molecule A are in dashed line; and energies associated with molecule B are in dotted line.

$$f_1\frac{m_a+m_b}{m_a m_b}+f_2\frac{\cos(\angle bac)}{m_a}+f_3\frac{\cos(\angle abc)}{m_b}=\omega_1^2 r_1. \tag{53}$$

Similarly for the bonds 2 and 3, one gets

$$f_1\frac{\cos(\angle bac)}{m_a}+f_2\frac{m_a+m_c}{m_a m_c}+f_3\frac{\cos(\angle acb)}{m_c}=\omega_2^2 r_2, \tag{54}$$

$$f_1\frac{\cos(\angle abc)}{m_b}+f_2\frac{\cos(\angle acb)}{m_3}+f_3\frac{m_b+m_c}{m_b m_c}=\omega_3^2 r_3. \tag{55}$$

Now f_1, f_2, and f_3 can be solved simultaneously from Eqs. (53), (54), and (55). This approach can be generalized for larger molecular fragments. It is also possible to recast these equations to allow for presence of massless particles in a flexible system.

B. Method II

While algorithms (40)–(44) is not symplectic, discretization of the Method II of the constrained system (34)–(36) leads to the symplectic method[15]

$$q_{n+1}=q_n+\Delta t M^{-1}p_{n+1/2}, \tag{56}$$

$$p_{n+1/2}=p_n-\frac{1}{2}\Delta t\nabla_q V(q_n)-\frac{1}{2}\Delta t\tilde{G}(q_n)^t\lambda_n, \tag{57}$$

$$0=\tilde{g}(q_{n+1}), \tag{58}$$

$$p_{n+1}=p_{n+1/2}-\frac{1}{2}\Delta t\nabla_q V(q_{n+1})-\frac{1}{2}\Delta t\tilde{G}(q_{n+1})^t\mu_n, \tag{59}$$

$$0=\tilde{G}(q_{n+1})M^{-1}p_{n+1}. \tag{60}$$

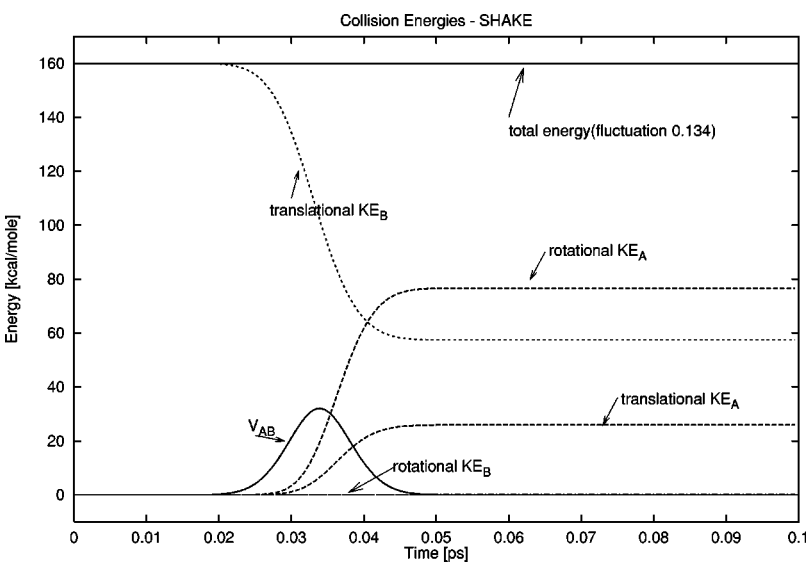

FIG. 4. SHAKE method energy profile: Total energy and potential energy V_{AB} are in solid line; energies associated with molecule A are in dashed line; and energies associated with molecule B are in dotted line.

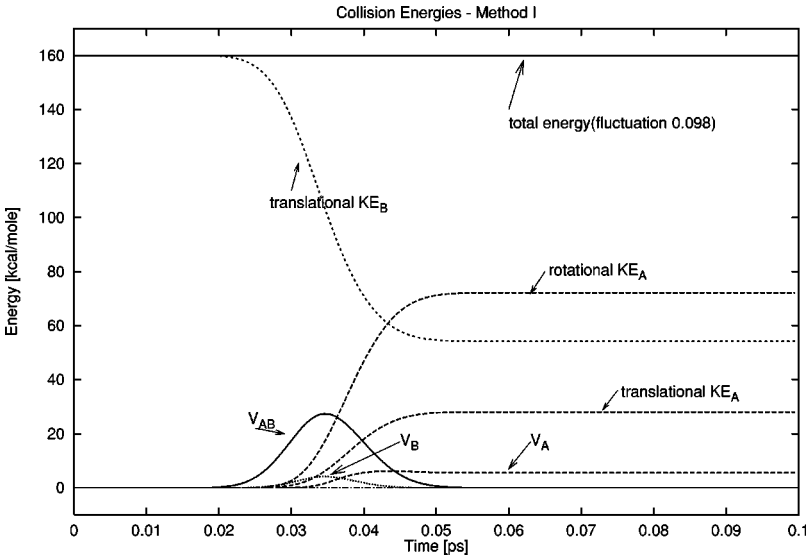

FIG. 5. Method I energy profile: Total energy and potential energy V_{AB} are in solid line; energies associated with molecule A are in dashed line; and energies associated with molecule B are in dotted line.

This method is also time-reversible and preserves linear and angular momentum exactly. This follows from the same arguments as given for Method I. However, it requires the computation of

$$\tilde{G}(q)=G(q)-K^{-1}\frac{\partial}{\partial q}\lambda(q,p=0). \tag{61}$$

This is expensive and leads to complex update directions in Eq. (57). It also cannot be implemented within the standard SHAKE iteration. However, the situation can be remedied by using the related modified force-field method.[1,16]

Finally, one might consider discretization of Eqs. (37)–(39). This time we obtain

$$q_{n+1}=q_n+\Delta t M^{-1}p_{n+1/2}, \tag{62}$$

$$p_{n+1/2}=p_n-\tfrac{1}{2}\Delta t\nabla_q V(q_n)-\tfrac{1}{2}\Delta t G(q_n)^t\lambda_n, \tag{63}$$

$$0=\tilde{g}(q_{n+1}), \tag{64}$$

$$p_{n+1}=p_{n+1/2}-\tfrac{1}{2}\Delta t\nabla_q V(q_{n+1})-\tfrac{1}{2}\Delta t G(q_{n+1})^t\mu_n, \tag{65}$$

$$0=G(q_{n+1})M^{-1}p_{n+1}, \tag{66}$$

which, similar to algorithms (40)–(44), can be solved by a proper modification of the standard SHAKE iteration. However, one has to expect a significant drift in energy (and perhaps other problems) due to nonconservation of energy in the analytical formulation.

V. NUMERICAL EXPERIMENTS

A. Collision of two diatomic molecules

Examining collisions is a useful first step in evaluating the quality and properties of an integration method. Let us first consider a collision of two hypothetical diatomic molecules. Figure 2 shows the collision model. The first molecule (A) is at rest initially and has two atoms with atomic

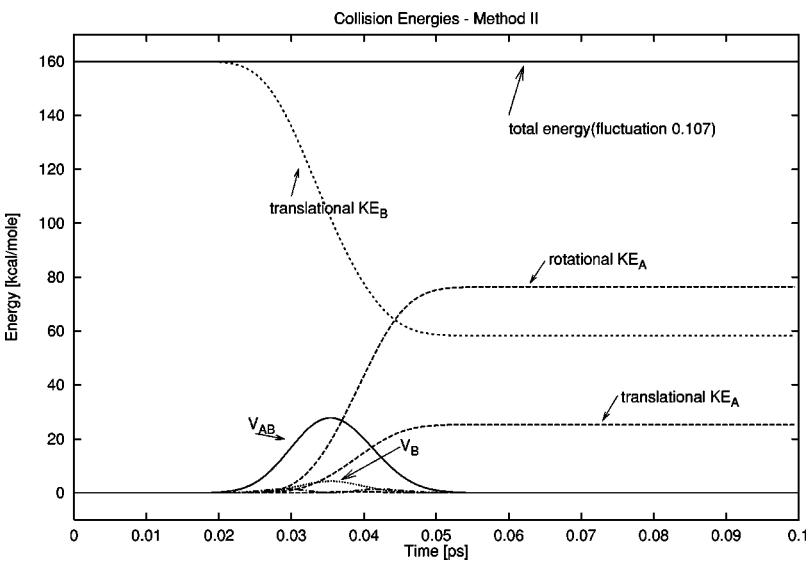

FIG. 6. Method II energy profile: Total energy and potential energy V_{AB} are in solid line; energies associated with molecule A are in dashed line; and energies associated with molecule B are in dotted line.

FIG. 7. The initial condition of a two water system.

mass 1 and 3. The second molecule (B) travels toward the first at a velocity of 81.82 Å/ps and has two identical atoms with atomic mass 2. Both molecules have the same harmonic force constant, 400 Kcal/Å^2, and bond length, 2 Å, for their bonds. The interaction potential between atoms 2 and 3 is anharmonic with

$$\mathcal{V}=\tfrac{1}{6}(20)(2-r)^6, \tag{67}$$

when their distance is less than 2 Å and zero otherwise.

This model was chosen in part for its simplicity, while including all major types of energies (translational, rotational, vibrational, and intramolecular energies). In this model we chose the integration time step as 0.1 femtosecond (fs) since the duration of the collision is roughly 20 fs.

The simulation was carried out with four methods: free dynamics, SHAKE, Method I, and Method II. Figures 3, 4, 5, and 6 show the energy profiles for simulated collision using free dynamics, SHAKE, Method I, and Method II, respectively.

In the figures, we plotted the total energy, translational kinetic energy for each molecule, rotational and vibrational kinetic energy for each molecule, and bond energy for each molecule. The interaction potential energy between two molecules and energy fluctuations for the total energy during the time course are 0.151, 0.134, 0.098, and 0.107 Kcal/mol for free dynamics, SHAKE, Method I, and Method II, respectively. So, all methods yield very good total energy conservation. The rms energy fluctuations are very small. The difference in bond energy and rotational energy profile between the free dynamics and SHAKE are obvious and expected. In comparing these four trajectories, we first note that all fully conserve energy and momentum. In comparing the free dynamics with the hard constraints (SHAKE), the free dynamics has additional internal vibrational energy which has primarily been removed from the rotational energies. The collision durations are comparable (where $V_{AB}>5$ Kcal/mol). With Method I, the collision duration has been increased but the final energy partitioning is very similar to the free dynamics. Note that molecule A's rotation at the end of the simulation causes the bond to stretch due to the centrifugal force. This arises since the geometry optimization in the solution of the nonlinear equations (40)–(44) includes the system's kinetic energy as well as the potential energy.

The trajectory for Method II has a final energy partition which is almost identical to that of the rigid system (SHAKE) although the collision time is significantly longer. Both soft constraint simulations have very similar collision durations. This analysis provides a view that Method I better reproduces the behavior of a free dynamics simulation while Method II, though also allowing flexibility, better retains some of the characteristics of the rigid (SHAKE) system.

B. Collision of two water molecules

The second collision system demonstrates the behavior of water molecules. The two water molecules were initially 15 Å apart, which was beyond the cutoff distance 12 Å, and hence there is no initial interaction between them. We prepared the system first with d_{2h} symmetry and then skewed one water 0.35 Å along the *Y*-axis in order to induce rotation during the collision. The dynamics simulation starts with one water at rest and with the other moving toward it with a velocity of 14.5 Å/ps along the *X*-axis (Fig. 7). After the collision, both water molecules spin away from each other.

Again, four sets of trajectories (two soft constraint methods, rigid water, and free dynamics) were generated and analyzed. The integration time step was 1 fs and total simulation length was 1.2 ps, which is long enough to observe the collision and separation. The total energy as a function of time is plotted in Fig. 8. The larger ''spikes'' in the collision with

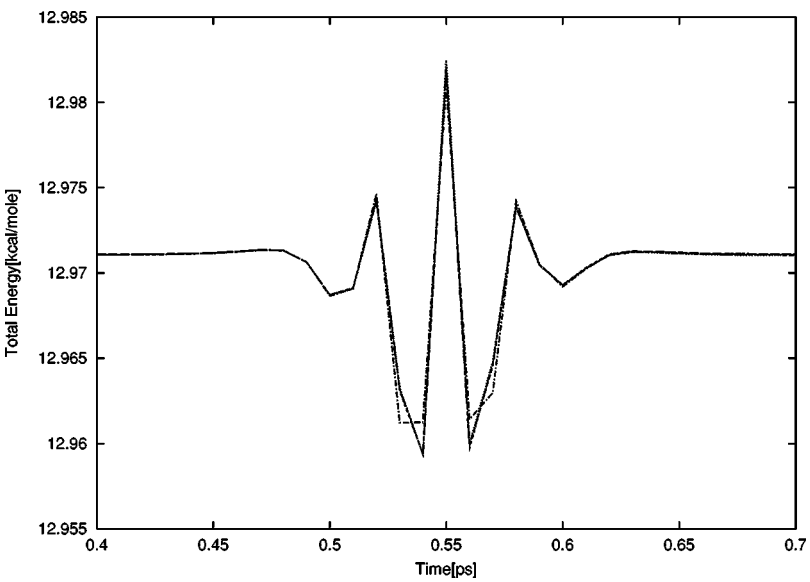

FIG. 8. Total energy as function of time during two-water collision.

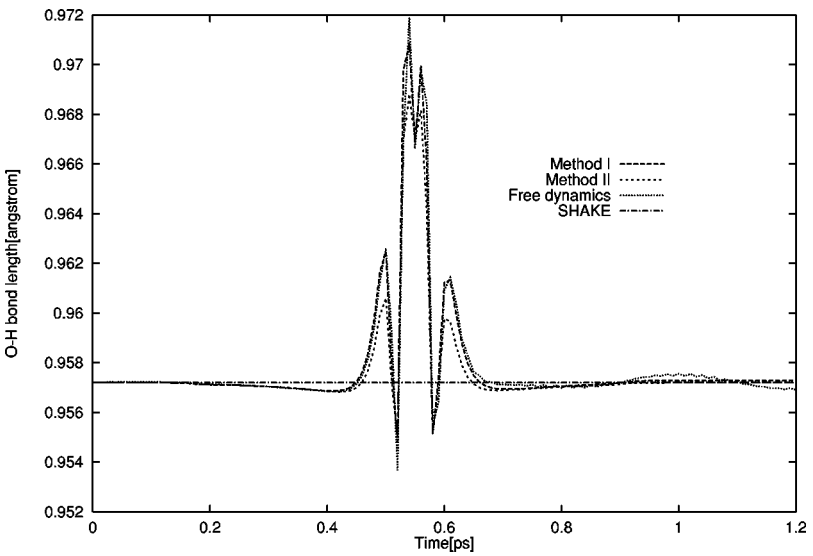

FIG. 9. One of O–H bond length as function of time during two-water collision.

SHAKE indicate that the collision is harder [larger $\nabla^2 V(q)$ along the collision coordinate] than those of the flexible constraint or unconstrained simulations.

On a detailed scale, the large spikes in total energy during the collision are artifacts of the integration procedure in that the simple sum of potential and kinetic energy does not well represent the simulation conserved quantity. A simple ''high frequency correction'' term added to the total energy removes these artifacts:

$$E(hfc)_n=\frac{1}{3}\left(\frac{p^t_{n-1/2}M^{-1}p_{n-1/2}}{2}-\frac{p^t_n M^{-1}p_n}{2}+V(q_{n-1})-V(q_n)\right)+\frac{1}{12}\left(\frac{p^t_{n-3/2}M^{-1}p_{n-3/2}}{2}-\frac{p^t_{n+1/2}M^{-1}p_{n+1/2}}{2}\right). \tag{68}$$

This correction term is generally applicable to any Verlet[20] based integration scheme to remove all of the total energy fluctuation due to high frequency harmonic motion. Figure 8 demonstrates that all four methods well conserve energy for collisions of triatomic systems.

It is also interesting to examine the change of a bond length during the water collision simulation. Figure 9 shows the bond length of an O–H bond in one of the water molecules as a function of time during the collision simulation.

The results are as expected. The bond length in the SHAKE method remains at its parameter value, throughout, while it undergoes remarkable change in the other three methods. Again, Method I mimics free dynamics method better than does Method II. More significantly, after the collision when two water molecules fall outside their interaction range the bond length in free dynamics method oscillates while it stretches in Method I, but returns to its minimum energy parameter value in Method II.

C. Simulation of a periodic water box

The MD simulation on a box of 216 SPC (Ref. 21) water molecules with periodic boundary was performed for all models. The initial coordinates were derived from an equilibrated configuration of TIP4P (four-site water model of Jorgenson)[22] at a density of 0.0333 molecules/Å^{-3}. The box edge length was 18.6476 Å. The flexible SPC model used here is identical to the rigid SPC model except for two bond energy terms $[V_b=450(r-1.0)^2]$ and one angle energy term $[V_\theta=55(\theta-109.47122)^2]$. There was no attempt to optimize the properties of the flexible SPC model employed here through further parameter refinement. This was done to facilitate a comparison between the four simulation methods, not to develop a viable water model. Long range electrostatic and van der Waals interaction were truncated at 9 Å using a force-shift approach.[4,23] The simulated temperature was 300 K. The velocities were scaled every 1000 steps during equilibration to meet the target 300 K should the temperature fall out of 300±10 K window during the first 10 ps, after which the simulation was run without intervention for another 200 ps.

Four methods, soft constraint Method I and II, free dynamics and SHAKE, were used to carry out the simulation using the same protocol with the integration time step of 1 fs and 2 fs. A larger time step, 4 fs, was also used to carry out the simulation with Method I and SHAKE. Unconstrained Verlet dynamics is unstable at 4 fs time step due to high frequency vibrations.[24]

The methods were implemented in CHARMM[25] and run on HP9000/735 workstations. The bond length convergence value of 10^{-6} Å was used for both Method I and Method II. Relaxing this convergence value will greatly reduce the computational time, but at the expense of energy conservation and reversibility. The detailed nature of this degradation has not yet been explored. The simulation time for Method I is

TABLE I. The rms fluctuations for the total energy.

Time step	Method I	Method II	Free dynamics	SHAKE
1 fs	0.035	0.041	0.667	0.032
2 fs	0.150	0.124	3.255	0.122
4 fs	0.823	not done	unstable	0.756

TABLE II. The average temperature and energy distribution per water (all energies are in Kcal/mol).

Time step	Method	Total	Kinetic	Potential	Bond & angle	Elec. & vdW	ΔH_{vap}	T(K)	ΔT(K)
1 fs	Method I	−8.503	1.795	−10.298	0.697	−10.995	−10.9	301.8	9.9
	Method II	−8.509	1.794	−10.303	0.648	−10.952	−10.9	301.6	9.8
	Free Dyn.	−6.648	2.675	−9.323	1.614	−10.937	−10.8	299.5	7.7
	SHAKE	−7.589	1.800	−9.389	0.0	−9.389	−10.0	302.6	9.7
2 fs	Method I	−8.375	1.831	−10.206	0.689	−8.150	−10.8	307.8	10.0
	Method II	−8.474	1.798	−10.272	0.645	−10.917	−10.9	302.2	10.0
	Free Dyn.	−6.170	2.702	−8.872	1.771	−10.643	−10.4	302.6	7.8
	SHAKE	−7.630	1.783	−9.413	0.0	−9.413	−10.0	299.8	9.9
4 fs	Method I	−8.303	1.827	−10.130	0.682	−8.052	−10.7	307.1	9.9
	SHAKE	−7.537	1.794	−9.331	0.0	−9.331	−9.9	301.6	9.6

roughly 13 times more expensive than for free dynamics. This increase reflects that, on average, 13 energy evaluations are performed at each time step. For Method II the simulation time was 65 times longer. This increase is due to additional large scale matrix manipulation at each iteration which increased each iteration by an additional factor of 5. There has not been an attempt to optimize the performance of these integrators and improvement of two-fold could be achieved with some recoding and improvements in the optimization algorithms and convergence criteria. The cost for Method I is almost entirely due to additional energy calls. Method II, in addition to the energy calls, involves the expensive calculation of complex update directions in Eq. (57).

1. Energy

The total energies as functions of time are nearly constant for all four simulations. The rms fluctuations for the total energy (without the high frequency correction) during the 200 ps simulations are listed in Table I. It is clear that both Method I and Method II give a very good energy conservation with minimal energy fluctuation and drift.

Also calculated, from the last 200 ps simulation, are the average values of total energy, kinetic energy, potential energy, bond and angle energy, electric and van der Waals energy, and ΔH_{vap} (Ref. 22) for a single water as well as the average temperature and its fluctuation, ΔT, over last 200 ps. The results are listed in Table II. All simulations have 1293 $(6N_{water}-3)$ degrees of freedom except for the free dynamics which has 1941 $(9N_{water}-3)$ degrees of freedom. This accounts for the 3 kT larger in kinetic energy per water, 6 kT larger in total energy per water, and 3 kT larger in the potential energy per water. The additional degrees of freedom also account for the small temperature fluctuation. All methods except SHAKE can lower total energy by relaxing the internal geometry. This accounts for the larger potential energy and weaker ΔH_{vap} of this system relative to the soft constraint methods. Note that both methods well reproduce the ΔH_{vap} observed with free dynamics.

TABLE III. The diffusion constant in units of 10^{-5} cm^2 s^{-1} calculated from the four 1 fs time step simulations.[a]

Group	Method I	Method II	Free dynamics	SHAKE
1	2.90	3.00	2.97	3.85
2	2.78	2.85	3.15	3.97
3	2.77	3.29	2.83	3.87
4	2.59	3.10	2.79	4.55
average	2.76	3.06	2.94	4.06
std dev	0.13	0.17	0.16	0.33

[a]The experimental value is 2.3×10^{-5} cm^2 s^{-1} at 298 K (Ref. 28).

2. Properties of bulk water

Some of the important water properties can be calculated from our simulation of water box with periodic boundaries. Comparing these properties from different methods with the experimental data, one should see how well a method simulates the water system. It should be noted that the calculated value of transport properties is a sensitive test to the quality of an integration method.[26,27] We calculated here the self-diffusion constant, oxygen–oxygen radial distribution, and dielectric constant of water.

The self-diffusion constant, D, is usually calculated by the Einstein relation[10]

$$6Dt=\langle|\mathbf{r}_i(t)-\mathbf{r}_i(0)|^2\rangle, \tag{69}$$

where $\mathbf{r}_i(t)$ is the position of the ith water molecule at time t and the brackets denote the ensemble average. In practice, the ensemble average can be computed by averaging over the N particles in the simulation, i.e.,

$$\langle|\mathbf{r}_i(t)-\mathbf{r}_i(0)|^2\rangle=\frac{\Sigma_{i=1}^{N}(\mathbf{r}_i(t)-\mathbf{r}_i(0))^2}{N}. \tag{70}$$

We calculated the self-diffusion constant from the last 200 ps simulations of a 216 water box. To examine the variability, we divided 216 water molecules into four groups, each with 54 water molecules. For each group, the $\langle|\mathbf{r}_i(t)-\mathbf{r}_i(0)|^2\rangle$ vs. t time series was calculated in a interval of 50 ps. To get a better result the time series was actually averaged over 16 50 ps intervals in our 200 ps simulation starting at $t_0=0,10,20,\ldots,150$. The time series was then linear fitted and the self-diffusion constant D was calculated from the slope which is $6D$. Ten values of D were calculated from the four groups of the water molecules as well as their average. The standard deviation was estimated by

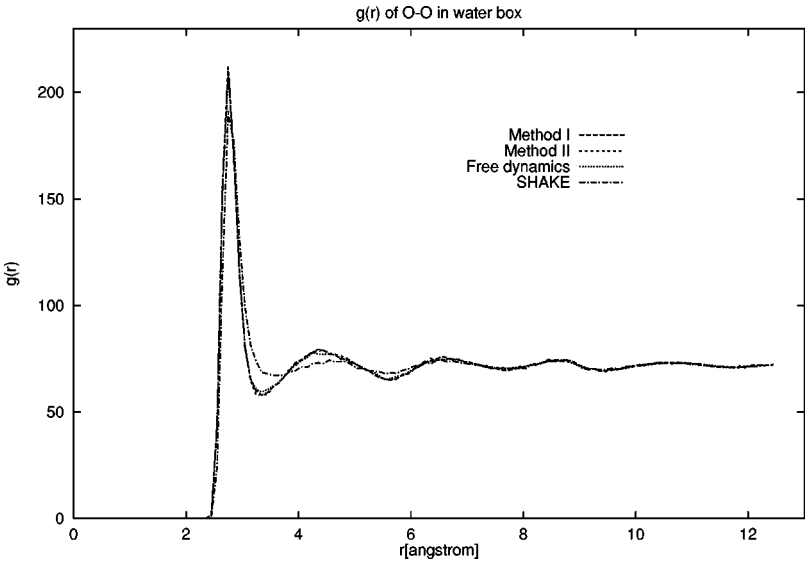

FIG. 10. Oxygen–oxygen radial distribution function evaluated from the four simulations.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N_s}(D_i - \bar{D})^2}{N_s}}, \tag{71}$$

where N_s, which is 4 in this case, is the number of samples. Table III lists the results.

With the experimental value at $2.3\times10^{-5}\,(\mathrm{cm^2\,s^{-1}})$ (Ref. 28) and earlier simulation reports of $3.3-4.5\times10^{-5}\,(\mathrm{cm^2\,s^{-1}})$ (Refs. 29–33) with a rigid SPC (Ref. 21) model, the effects of flexibility seem reasonable. The flexible methods and free dynamics yield similar diffusion constants.

Figure 10 shows oxygen–oxygen radial distribution function, $g(r)$, evaluated from our four simulations. Two soft constraint methods, Methods I and II, give a surprisingly good agreement with the free dynamics method. Their $g(r)$ are almost superimposed with that from the free dynamics method.

The dielectric constant of water, ϵ, was calculated using the relation[10]

$$\epsilon = 1 + \frac{4\pi}{3Vk_bT}\left(\left\langle\left|\sum_{i=1}^{N}\mu_i\right|^2\right\rangle - \left\langle\sum_{i=1}^{N}\mu_i\right\rangle^2\right), \tag{72}$$

TABLE IV. The dielectric constant at 300 K calculated from the four 1 fs time step simulations.[a]

Ensemble	Method I	Method II	Free dynamics	SHAKE
1	68.68	89.53	72.15	53.76
2	68.79	89.51	72.48	53.84
3	68.80	89.09	72.60	53.98
4	68.70	88.88	72.26	53.08
5	68.49	89.15	72.06	53.94
6	68.34	89.47	72.41	53.82
7	68.25	89.50	72.89	53.77
8	68.33	89.28	73.07	53.68
9	68.40	89.10	72.65	53.65
10	68.49	89.25	72.19	53.74
average	68.53	89.27	72.48	53.83
std dev	0.19	0.21	0.31	0.14
200 ps av	68.53	89.27	72.48	53.82

[a]The experimental value is 78.54 at 298 K (Ref. 34).

where μ_i is the dipole moment of a water molecule, N is the number of water, and the brackets denote the ensemble average. In each of the four simulations, 20 000 trajectory frames were obtained from the 200 ps production run by saving the coordinates every 0.01 ps. We first calculated the dielectric constant using these 20 000 trajectory frames as one ensemble. The results are listed in the last row of Table IV. To estimate the error bars, we divided the 20 000 frames into ten ensembles by grouping trajectory frames 0.1 ps interval apart into an ensemble. Ten values of the dielectric constant were calculated from the ensembles as well as their average. The standard deviation was estimated by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N_s}(\epsilon_i - \bar{\epsilon})^2}{N_s}}, \tag{73}$$

where $N_s = 10$ in this case. The results are listed in Table IV.

Although the dielectric constant is sensitive to integration techniques and simulation conditions (temperature and pressure), it is a fundamental property of water behavior which is important in any study involving solvation. The experimental value is 78.54 at 298 K (Ref. 34) which is reasonably well reproduced by free dynamics and the soft constraint methods. Other simulations with SPC water model report the dielectric constant 68 ± 7,[33] which is somewhat larger than the SHAKE value reported here. The difference may be due to density differences or long range electrostatic cutoff effects.

We also calculated the average magnitude of water molecular dipole moment and its fluctuation from the four simulation methods. The average is over all water molecules from ten randomly selected water trajectory snapshots saved from

TABLE V. The average water molecular dipole moment magnitude and its fluctuation in Debye.

	Method I	Method II	Free dynamics	SHAKE
Average dipole	2.4240	2.4193	2.4131	2.2740
fluctuation	0.0506	0.0499	0.1362	0.0000

the simulations. Since each snapshot contains 216 water molecules, the average is over 2160 water dipole moments. The results are listed in Table V.

As expected, the water molecules of the SHAKE simulation have a constant magnitude of dipole moment determined only by the rigid geometry. Both soft constraint methods have an increased dipole moment magnitude that almost exactly matches the dipole average from the free dynamics, however the fluctuation is significantly smaller due to the fact that molecular vibration (primarily angle bending) has been eliminated. The variation of dipole moments in the soft constraint methods is due to the diversity of the local environment about around water molecules.

VI. CONCLUSION

Two new integration methods are presented that allow the deformation of high frequency degrees of freedom without the necessity of including high frequency fluctuations and the associated (nonphysical) high frequency heat capacity. It has been demonstrated that both of these methods give essentially the same general behavior as that observed with free dynamics when examining transport properties and other properties that depend on the water structure. Both methods are energy conserving and reversible. Method II is also symplectic but involves a greater iteration cost. Both methods are easily extended to complex systems such as proteins.

Although these methods do allow a larger time step, in the same spirit as bond length SHAKE constraints, these methods are more expensive due to the optimization at each iteration. The results presented in this article demonstrate the high level of energy conservation that can be achieved when the optimization is fully converged. The nature of the degradation of these methods with an incomplete optimization in order to reduce computational costs is the subject of further work (Note: The cost is proportional to the number of iterations per time step.) The use of this flexible constraint method is likely to allow a larger time step than SHAKE, especially at higher temperature, due to a ''softening'' at collisions as demonstrated in the collision examples presented here.

The most promising aspect of this work is that these methods can be used in conjunction with other integration techniques that require an iterative optimization at each time step, such as the self-consistent electronic polarization methods[6–8] or iterative combined QM/MM hybrid simulations.[9] The optimization steps required by these methods can be combined with those required for flexible constraints in a mannner that minimizes the computational costs. In other words, the additional energy evaluations that are required can be used for both optimizations. The methods presented here can be extended to a Drude oscillator model for electronic polarization where the oscillations are replaced by an optimization to determine positions. We believe that ultimate utility of these soft constraint methods will be found with such an approach, where all high frequency fluctuations (both electronic and bond-stretching) can be eliminated, but yet still be allowed to respond to the local environment.

[1] T. Schlick, J. Comput. Phys. **151**, 1 (1999).
[2] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[3] H. C. Andersen, J. Comput. Phys. **52**, 24 (1983).
[4] W. F. van Gunsteren and M. Karplus, Macromolecules **15**, 1528 (1982).
[5] W. F. van Gunsteren, Mol. Phys. **40**, 1015 (1980).
[6] D. Vanbelle and S. J. Wodak, J. Am. Chem. Soc. **115**, 647 (1993).
[7] L. X. Dang, J. E. Rice, J. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **113**, 2481 (1991).
[8] M. Sprik, J. Phys. Chem. **95**, 2283 (1991).
[9] M. Ben-Nun and T. J. Martinez, Chem. Lett. **290**, 289 (1998).
[10] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University, New York, 1987).
[11] B. Leimkuhler and R. D. Skeel, J. Comput. Phys. **112**, 117 (1994).
[12] V. I. Arnold, *Mathematical Methods in Classical Mechanics*, 2nd ed., Graduate Texts in Mathematics (Springer-Verlag, Berlin, 1989), Vol. 60.
[13] M. Fixman, Proc. Natl. Acad. Sci. USA **71**, 3050 (1974).
[14] S. Reich, Physica D **89**, 29 (1995).
[15] S. Reich, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **33(2)**, 475 (1996).
[16] S. Reich, Numer. Algorithms **19**, 213 (1998).
[17] S. Reich, Physica D **76**, 375 (1994).
[18] M. Q. Zhang and R. D. Skeel, J. Comput. Chem. **16**, 365 (1995).
[19] E. Barth, K. Kuczera, B. Leimkuhler, and R. D. Skeel, J. Comput. Chem. **16**, 1192 (1995).
[20] L. Verlet, Phys. Rev. **159**, 98 (1967).
[21] H. J. C. Berendsen, J. P. M. Postma, W. F. von Gunsteren, and J. Hermans, *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, Holland, 1981), p. 331.
[22] W. L. Jorgensen, J. D. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[23] P. Steinbach and B. R. Brooks, J. Comput. Chem. **15**, 667 (1994).
[24] R. W. Pastor, B. R. Brooks, and A. Szabo, Mol. Phys. **65**, 1409 (1988).
[25] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[26] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, J. Chem. Phys. **103**, 4613 (1995).
[27] S. E. Feller, R. W. Pastor, A. Rojnuckarin, S. Bogusz, and B. R. Brooks, J. Phys. Chem. **100**, 17011 (1996).
[28] M. Krynicki, C. D. Green, and D. W. Sawyer, Discuss. Faraday Soc. **66**, 199 (1978).
[29] O. Teleman, B. Jönsson, and S. Engström, Mol. Phys. **69**, 193 (1987).
[30] J. L. Barrat and I. R. McDonald, Mol. Phys. **70**, 535 (1990).
[31] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[32] A. Wallqvist and O. Teleman, Mol. Phys. **74**, 515 (1991).
[33] H. E. Alper and R. M. Levy, J. Chem. Phys. **91**, 1242 (1989).
[34] R. C. Weast, M. J. Astle, and W. H. Beyer, *CRC Handbook of Chemistry and Physics*, 67th ed. (CRC Press, Boca Raton, FL, 1986).