Canonical Sampling Method for Initial Conditions for Reactive Flux Calculations Using Nosé-Hoover Chains

Song Hi Lee[*] and Youngshang Pak[†,*]

Department of Chemistry, Kyungsung University, Busan 608-736, Korea
[†]Department of Chemistry, Pusan National University, Busan 609-735, Korea
Received December 1, 2003

Canonical sampling method has been presented to generate the initial conditions for reactive flux studies of organic reactions in water. Velocity Verlet version of Nosé-Hoover chain dynamics algorithm has been employed to sample the initial conditions according to canonical distribution. The unstable normal mode of a transition state has been introduced to define a dividing plane separating reactant and product regions in reaction processes. This method has been implemented and tested for the case Diels-Alder reaction of methyl vinyl ketone (MVK) and cyclopentadiene (CPD) in water, providing a reliable tool for further reactive flux molecular dynamics studies in condensed media.

Key Words : Molecular dynamics simulation, Canonical sampling, Reactive flux, Nosé-Hoover chain dynamics, Diels-Alder reaction

Introduction

In order to carry out reactive flux (RF) calculations of chemical reactions in condensed phase, initial conditions should be generated by canonical distribution. Using the velocity scaling method of Berendsen *et al.*,[1] Otter and Briels[2] generated the initial conditions and carried out their RF studies of some isomerization reactions. Jang and Voth[3] proposed simple reversible molecular dynamics (MD) algorithms based on Nosé-Hoover chain (NHC) dynamics and successfully tested them for a model system of coupled harmonic oscillators. These algorithms can easily be extended to systems with a holonomic constraint, which is required in the sampling procedure with a proper choice of a dividing plane separating reactant and product regions. In the recent work of Otter and Briels,[2] the dividing plane was chosen by introducing a normal mode with a negative eigenvalue (*i.e.*, unstable normal mode) of a transition state. One of the advantages of this choice is that it is very closely related to intrinsic reaction coordinate (IRC) and gives a clear mathematical definition of the dividing plane, as was already discussed in their paper. Thus following their approach,[2] we have chosen to use the unstable normal mode of the transition state to define the RF plane as a holonomic constraint in our MD algorithm. Subsequently the transition state geometry and its Hessian matrix of a reactive system of interest need to be calculated and can be obtained by state-of art *ab initio* method with reasonable accuracy. Main purpose of this paper is to provide a canonical sampling method for the initial conditions for condensed phase RF studies, and we have implemented the algorithm based on velocity Verlet scheme (VV-1)[3] to a Diels-Alder reaction of methyl vinyl ketone (MVK) and cyclopentadiene (CPD) in water, which will be a subject of future *ab initio* RF MD studies.

[*]Co-Corresponding authors. Song Hi Lee (shlee@star.ks.ac.kr); Youngshang Pak(ypak@pusan.ac.kr)

Theory

In reactive flux calculations, a dynamic correction factor κ to a transition state (TST) rate constant κ^{TST} can be calculated by

$$\kappa \equiv \frac{k}{k^{\mathrm{TST}}} = \langle \theta_p[\xi(t_{pl}]\rangle_+ - \langle \theta_p[\xi(t_{pl}]\rangle_- , \tag{1}$$

where κ is the true rate constant, $\xi(t_{pl})$ is the reaction coordinate at plateau times t_{pl}, and θ_p is a step function, such that θ_p is 1 only if $\xi(t_{pl})$ is in the product region and 0 otherwise. The notations $<\cdots>_+$ and $<\cdots>_-$ in Eq. (1) denote averaging over a normalized flux weighted canonical distribution function with positive and negative initial reaction coordinate velocity, respectively. The normalized flux weighted distribution function at $\Gamma \equiv (\mathbf{x}, \mathbf{p}^{\mathbf{x}})$ in Cartesian coordinates is given by[4]

$$P^{(\pm)}(\Gamma)\mathbf{d}\Gamma = \frac{\mathbf{d}\Gamma\delta[\xi(0)-\xi^{\ddagger}]\theta_p[\pm\dot{\xi}(0)]|\dot{\xi}(0)|\exp(-\beta H)}{\int \mathbf{d}\Gamma\delta[\xi(0)-\xi^{\ddagger}]\theta_p[\pm\dot{\xi}(0)]|\dot{\xi}(0)|\exp(-\beta H)}, \tag{2}$$

where $\xi^{\ddagger}$ is a plane dividing the reactant and product regions of a reactive system consisting of N atoms, and $\xi(0)$ and $\dot{\xi}(0)$ are the reaction coordinate and velocity, respectively, at the initial time $t = 0$. In this work, bond length constraints have not been employed. Then the flux weighted distribution in Eq. (2) can also be written

$$P^{\pm}(\Gamma)\mathbf{d}\Gamma = \frac{\mathbf{d}\Gamma\delta(\xi-\xi^{\ddagger})\delta(\dot{\xi})\exp(-\beta H)d\dot{\xi}'|\dot{\xi}'|\theta_p(\pm\dot{\xi}')\exp\left(-\frac{1}{2}\beta\dot{\xi}'Z_{\xi}^{-1}\dot{\xi}'\right)}{\int \mathbf{d}\Gamma\delta(\xi-\xi^{\ddagger})\delta(\dot{\xi})\exp(-\beta H)\int_0^{\infty} d\dot{\xi}'|\dot{\xi}'|\theta_p(\pm\dot{\xi}')\exp\left(-\frac{1}{2}\beta\dot{\xi}'Z_{\xi}^{-1}\dot{\xi}'\right)}. \tag{3}$$

In Eq. (3), Z_ξ is defined as

$$Z_\xi = \sum_{i=1}^{N} \frac{1}{m_i} \frac{\partial \xi}{\partial \mathbf{x}_i} \cdot \frac{\partial \xi}{\partial \mathbf{x}_i}, \tag{4}$$

where m_i is the mass of the i-th atom. The average value of the step function $\theta_p[\xi(t)]$ in Eq. (2) sampled by the distribution of Eq. (3) exactly reproduces that of Eq. (2), and a proof goes as follows:

Denoting the step function $\theta_p[\xi(t)]$ to be $F(\xi, \dot{\xi}, t)$, the average value of F with the distribution of Eq. (3) is

$$\langle F \rangle_\pm = \frac{\int \mathbf{d}\Gamma \delta(\xi - \xi^\ddagger)\delta(\dot{\xi}) \exp(-\beta H) \int_0^\infty d\dot{\xi}' |\dot{\xi}'| \theta_p(\pm\dot{\xi}') \exp\left(-\frac{1}{2}\beta \dot{\xi}' Z_\xi^{-1} \dot{\xi}'\right) F(\xi, \dot{\xi}', t)}{\int \mathbf{d}\Gamma \delta(\xi - \xi^\ddagger)\delta(\dot{\xi}) \exp(-\beta H) \int_0^\infty d\dot{\xi}' |\dot{\xi}'| \theta_p(\pm\dot{\xi}') \exp\left(-\frac{1}{2}\beta \dot{\xi}' Z_\xi^{-1} \dot{\xi}'\right)}. \tag{5}$$

Let's introduce generalized coordinates and their conjugate momentums ($q_1, \cdots, q_{3N-1}, \xi, p_1^q, \cdots, p_{3N-1}^q, p^\xi$). Then using $\mathbf{dp^q} d\xi dp^\xi = Z_\xi^{-1} \mathbf{dp^q} d\xi d\dot{\xi}$, the change of the integration variables $\mathbf{dp^q} d\xi dp^\xi$ to $\mathbf{dp^q} d\xi d\dot{\xi}$ leads a partition of the kinetic energy term, such that

$$T = T_q + T_\xi \tag{6}$$

where T_q and T_ξ are the individual kinetic terms depending on $p_i^q (i = 1, \cdots, 3N-1)$ and $\dot{\xi}$, respectively. In particular note that $T_\xi = \frac{1}{2} \dot{\xi} Z_\xi^{-1} \dot{\xi}$. (See Ref. 2 for more details) As a consequence, Eq. (5) becomes

$$\langle F \rangle_\pm \propto \int \mathbf{dq} \cdot \mathbf{dp}^q d\xi d\dot{\xi} Z_\xi^{-1} \delta(\xi - \xi^\ddagger)\delta(\dot{\xi}) \exp[-\beta(T_q + T_\xi + V)] \times \int d\dot{\xi}' |\dot{\xi}'| \theta_p(\pm\dot{\xi}') \exp\left(-\frac{1}{2}\beta \dot{\xi}' Z_\xi^{-1} \dot{\xi}'\right) F(\xi, \dot{\xi}', t) \tag{7}$$

Integrating Eq. (7) with respect to $\dot{\xi}$ and changing the order of integration variable $\dot{\xi}'$ gives

$$\begin{aligned} \langle F \rangle_\pm &\propto \int \mathbf{dq} \cdot \mathbf{dp}^q d\xi d\dot{\xi}' Z_\xi^{-1} \delta(\xi - \xi^\ddagger) |\dot{\xi}'| \theta_p(\pm\dot{\xi}') \\ &\quad \times \exp\left[-\beta\left(T_q + V + \frac{1}{2}\beta \dot{\xi}' Z_\xi^{-1} \dot{\xi}'\right)\right] F \\ &= \int \mathbf{dq} \cdot \mathbf{dp}^q d\xi d\dot{\xi} Z_\xi^{-1} \delta(\xi - \xi^\ddagger) |\dot{\xi}'| \theta_p(\pm\dot{\xi}) \exp(-\beta H) F \\ &= \int \mathbf{dq} \cdot \mathbf{dp}^q d\xi dp^\xi \delta(\xi - \xi^\ddagger) |\dot{\xi}| \theta_p(\pm\dot{\xi}) \exp(-\beta H) F \\ &= \int \mathbf{d}\Gamma \delta(\xi - \xi^\ddagger) |\dot{\xi}| \theta_p(\pm\dot{\xi}) \exp(-\beta H) F . \end{aligned} \tag{8}$$

Furthermore the flux averages in Eq. (5) can be expressed in a more practical term:

$$\langle F \rangle_\pm = \frac{\int \mathbf{d}\Gamma P_{\xi,\xi}^c(\Gamma) \int_0^\infty d\dot{\xi}' P^{(\pm)}(\dot{\xi}'|\mathbf{x}) Z_\xi F}{\int \mathbf{d}\Gamma P_{\xi,\xi}^c(\Gamma) Z_\xi}, \tag{9}$$

with

$$P_{\xi,\xi}^c(\Gamma)\mathbf{d}\Gamma \equiv \frac{\mathbf{d}\Gamma \delta(\xi - \xi^\ddagger)\delta(\dot{\xi}) \exp(-\beta H)}{\int \mathbf{d}\Gamma \delta(\xi - \xi^\ddagger)\delta(\dot{\xi}) \exp(-\beta H)}, \tag{10}$$

$$P^{(\pm)}(\dot{\xi}'|\mathbf{x}) d\dot{\xi}' \equiv \frac{d\dot{\xi}' |\dot{\xi}'| \theta_p(\pm\dot{\xi}') \exp\left(-\frac{1}{2}\beta \dot{\xi}' Z_\xi^{-1} \dot{\xi}'\right)}{\int_0^\infty d\dot{\xi}' |\dot{\xi}'| \theta_p(\pm\dot{\xi}') \exp\left(-\frac{1}{2}\beta \dot{\xi}' Z_\xi^{-1} \dot{\xi}'\right)}, \tag{11}$$

where $P_{\xi,\xi}^c(\Gamma)$ is the probability distribution in Γ generated by MD simulations with the constraints of both $\xi = 0$ and $\dot{\xi} = 0$, and $P^{(\pm)}(\dot{\xi}'|\mathbf{x})$ is the conditional probability of the reaction velocity $\dot{\xi}'$ for a given value of $\mathbf{x}$. [The superscripts (+) and (–) represent the forward and backward reaction velocities, respectively.] For the sampling of a negative $\dot{\xi}'$, it can easily be generated by using $P^{(+)}(\dot{\xi}'|\mathbf{x})$ without resampling $P^{(-)}(\dot{\xi}'|\mathbf{x})$ and just reversing it. In a similar manner to Otter and Briels's approach,[2] the dividing plane for the reactive flux calculations is chosen to satisfy the following holonomic constraint:

$$h(x_1, \cdots, x_{3N}) = \xi = Q_1 = 0, \tag{12}$$

where Q_1 is a normal mode having a negative eigenvalue at the transition state. In order to generate the initial conditions according to $P_{\xi,\xi}^c(\mathbf{x},\mathbf{p}^\mathbf{x})$ in Eq. (3), one can use canonical MD algorithm with a holonomic constraint. Jang and Voth[3] proposed several reversible MD algorithm based on Nosé-Hoover chains to generate trajectories according to canonical distributions. In this work the velocity Verlet algorithm (VV-1) in Ref. 3 has been employed and briefly summarized in the next section. For more details of the algorithm, readers should refer to Ref. 3.

A. Velocity Verlet Algorithm with a Holonomic Constraint (VV-1). Let $x_i(\delta t)$, $v_i(\delta t)$, and $a_i(\delta t)$ be the position, velocity, and acceleration of a system consisting of N atoms at time step δt. The variables η_i, $v_{\eta i}$, and a_{η_i} are the ones corresponding to the i-th Nosé-Hoover chain associated with the system. The VV-1 algorithm consists of three steps:

Step 1:

$$v_i(\delta t/2) = \left\{ v_i(0) g_1^{(-)}(0) + \frac{\delta t}{2} a_i(0) \right\} / g_1^{(+)}(0) \tag{13}$$

$$\eta_{2k-1}(\delta t/2) = \eta_{2k-1}(0) + \frac{\delta t}{2} v_{\eta_{2k-1}}(0) \tag{14}$$

$$v_{\eta_{2k}}(\delta t/2) = \left\{ v_{\eta_{2k}}(0) g_{2k+1}^{(-)}(0) + \frac{\delta t}{2} a_{\eta_{2k}}(0) \right\} / g_{2k+1}^{(+)}(0), \tag{15}$$

Step 2:

$$x_i(\delta t) = x_i(0) + \delta t v_i(\delta t/2) \tag{16}$$

$$\eta_{2k}(\delta t) = \eta_{2k}(0) + \delta t v_{\eta_{2k}}(\delta t/2) \tag{17}$$

$$v_{\eta_{2k-1}}(\delta t) = \{ v_{\eta_{2k-1}}(0) G_{2k}^{(-)}(\delta t/2) + \delta t a_{\eta_{2k-1}}(\delta t/2) \} / G_{2k}^{(+)}(\delta t/2), \tag{18}$$

Step 3:

$$v_i(\delta t) = \left\{ v_i(\delta t/2) g_1^{(-)}(\delta t) + \frac{\delta t}{2} a_i(\delta t) \right\} / g_1^{(+)}(\delta t) \tag{19}$$

$$\eta_{2k-1}(\delta t) = \eta_{2k-1}(\delta t/2) + \frac{\delta t}{2} v_{\eta_{2k-1}}(\delta t) \tag{20}$$

$$v_{\eta_{2k}}(\delta t) = \left\{ v_{\eta_{2k}}(\delta t/2) g_{2k+1}^{(-)}(\delta t) + \frac{\delta t}{2} a_{\eta_{2k}}(\delta t) \right\} / g_{2k+1}^{(+)}(\delta t) , \tag{21}$$

where $G_j^{(\pm)}$ and $g_j^{(\pm)}$ at time t are defined as

$$G_j^{(\pm)}(\delta t) = 1 \pm \frac{\delta t}{2} v_{\eta_j}(\delta t) , \tag{22}$$

$$g_j^{(\pm)}(\delta t) = 1 \pm \frac{\delta t}{4} v_{\eta_j}(\delta t) . \tag{23}$$

In presence of the holonomic constraint [Eq. (12)], $v_i(\delta t/2)$, and $x_i(\delta t)$ are modified to give

$$v_i(\delta t/2) = v_i'(\delta t/2) + \lambda \frac{\delta t}{2m_i} \frac{\partial h}{\partial x_i} / g_1^{(+)}(0) \equiv v_i'(\delta t/2) + \lambda \alpha_{x_i}(0) , \tag{24}$$

$$x_i(\delta t) = x_i'(\delta t) + \lambda \frac{\delta t^2}{2m_i} \frac{\partial h}{\partial x_i} / g_1^{(+)}(0) \equiv x_i'(\delta t) + \lambda \delta t \alpha_{x_i}(0) , \tag{25}$$

where λ is a Lagrange multiplier, $v_i'(\delta t/2)$, and $x_i'(\delta t/2)$ are the velocity and position variables, respectively, calculated without the constraint at time $\delta t/2$. In general it is convenient to express normal modes of the system $\{Q_i\}$ in temrs of $3N-6$ curvilinear internal coordinates $\{S_j\}$:

$$Q_i = \sum_{j=1}^{3N-6} L_{i,j}^{-1} (S_j - S_{e,j}) , \tag{26}$$

where $L_{i,j}^{-1}$ are the elements of transformation matrix from the internal coordinates $\{S_j\}$ to the normal coordinates $\{Q_i\}$, and $\{S_{e,j}\}$ are the internal coordinates at the transition state. Then the reaction coordinate ξ can be written

$$\xi \equiv Q_1 = \sum_{j=1}^{3N-6} L_{i,j}^{-1} (S_j - S_{e,j}) . \tag{27}$$

In fact in Ref. 2, a reaction coordinate was expressed in terms of Cartesian displacement coordinates, which in turn requires calculations of gradients of a rotation matrix. The gradient calculations, however, can be avoided by use of Eq. (27) at the expense of computing **B**-matrix, which can be easily calculated by the **s**-vector method of Wilson *et al.*[5] With Eq. (27), $\alpha_{x_i}(0)$ in Eqs. (24) and (25) are calculated by

$$\begin{aligned} \alpha_{x_j}(0) &\equiv \frac{\delta t}{2} \frac{\partial h}{\partial x_j} / m_j g_1^{(+)}(0) = \frac{\delta t}{2} \sum_{i=1}^{3N-6} L_{1,i}^{-1} \frac{\partial S_i}{\partial x_j} / m_j g_1^{(+)}(0) \\ &= \frac{\delta t}{2} \sum_{i=1}^{3N-6} L_{1,i}^{-1} B_{i,j} / m_j g_1^{(+)}(0) , \end{aligned} \tag{28}$$

where $B_{i,j}$ are the elements of the **B**-matrix. Since Q_1 has been expressed in the internal coordinates [Eq. (27)], obtaining λ by a direct substitution of Eq. (25) into Eq. (12) is not straightforward in this case. Thus we expand S_i with respect to x_i' up to first order, which

$$S_i \approx S_i' + \sum_{j=1}^{3N} \frac{\partial S_i}{\partial x_j'} (x_j - x_j') = S_i' + \sum_{j=1}^{3N} B_{i,j}' (x_j - x_j') . \tag{29}$$

Then substituting Eq. (29) into Eq. (12) and solving for λ gives

$$\lambda \approx \frac{\Sigma_{i=1}^{3N-6} L_{1,i}^{-1} (S_i' - S_{e,i})}{\Sigma_{i=1}^{3N-6} \Sigma_{j=1}^{3N} L_{1,i}^{-1} B_{i,j}' \alpha_{x_j}(0) \delta t} . \tag{30}$$

An iteration procedure consisting of Eqs. (25), (29), and (30) is carried out until $\{x_i'\}^n$ become close to $\{x_i'\}^{n+1}$ within a convergence limit. Using the converged values of $\{x_i'\}^n$, λ is easily calculated and then used to determine the constraint corrected values of $v_i(\delta t/2)$ in Eq. (24). In addition the acceleration in the first Nosé-Hoover chain coupled to the i-th degree of freedom of the system is also affected by the constraint and obtained by using the modified equation:[3]

$$a_{\eta_{1(i)}}(\delta t/2) = [m_i v_i^2 - (1 - g_i^2 k_B T)] / \omega_{1(i)} , \tag{31}$$

where g_i is the fractional degree of freedom[3] substracted in the i-th direction and $\omega_{1(i)}$ is the first Nosé mass associated with the i-th degree of freedom (See the Ref. 3 for more details). With the normal mode constraint in Eq. (12), g_i^2 is given by

$$\begin{aligned} g_i^2 &= \left(\frac{1}{m_i} \frac{\partial h}{\partial x_i} \right)^2 / \sum_{k=1}^{3N} \left(\frac{1}{m_k} \frac{\partial h}{\partial x_k} \right)^2 \\ &= \left(\sum_{j=1}^{3N-6} L_{i,j}^{-1} B_{j,i} m_i^{-1} \right)^2 / \sum_{k=1}^{3N} \left(\sum_{j=1}^{3N-6} L_{i,j}^{-1} B_{j,k} m_k^{-1} \right)^2 \end{aligned} \tag{32}$$

Instead of associating NHCs to each of the i-th degree of freedom as in Eq. (31), an alternative expression to Eq. (31) can be obtained by coupling NHC to the whole system with $3N-1$ degree of freedom (*i.e.*, a reactive species with a constraint), yielding a simpler expression for the acceleration at time $\delta t/2$:

$$a_{\eta_1}(\delta t/2) = \left[\sum_{j=1}^{3N} m_i v_i^2 - (3N-1) k_B T \right] / \omega_1 . \tag{33}$$

Practically it is more efficient to have several NHCs coupled to the $3N-1$ degree of freedom and is straightforward to extend Eq. (33) to the case of multiple NHCs. In the VV-l algorithm,[3] the velocities at time δt also satisfy

$$\begin{aligned} \dot{h}(x_1, \ldots, x_{3N}) = \dot{Q}_1 &= \sum_{i=1}^{3N-6} L_{1,i}^{-1} \dot{S}_i \\ &= \sum_{i=1}^{3N-6} \sum_{j=1}^{3N} L_{1,i}^{-1} B_{i,j} v_j = 0 \end{aligned} \tag{34}$$

In a similar manner to Eq. (24), the velocities at time δt are calculated by introducing a new Lagrange multiplier λ', such that

$$v_i(\delta t) = v_i'(\delta t) + \lambda' \alpha_{x_i}(\delta t) . \tag{35}$$

Again substituting Eq. (35) into Eq. (34) results in a expression for λ'

$$\lambda' = \frac{\Sigma_{i=1}^{3N-6}\Sigma_{j=1}^{3N} L_{1,i}^{-1} B_{i,j} v_j'}{\Sigma_{i=1}^{3N-6}\Sigma_{j=1}^{3N} L_{1,i}^{-1} B_{i,j} \alpha_{x_j}(\delta t)}. \quad (36)$$

With the λ' value in Eq. (36), Eq. (35) is completely determined. Note that λ' is directly obtained without any iterative scheme.

B. Sampling of $\dot{\xi}$. The above MD algorithm generates the initial conditions based on $P^c_{\xi,\dot{\xi}}(\Gamma)$ in Eq. (10). Therefore after the canonical equilibrium of $P^c_{\xi,\dot{\xi}}(\Gamma)$ is reached, one has to reassign only the value of $\dot{\xi}$, simply using $P^{\pm}(\dot{\xi}|\mathbf{x})$ in Eq. (11). The value of Z_ξ in Eq. (4) is calculated by substituting $\xi = Q_1$ into Eq. (4) and using Eq. (27):

$$\begin{aligned} Z_\xi &= \sum_{j=1}^{3N-6}\sum_{k=1}^{3N-6} L_{1,i}^{-1} L_{1,k}^{-1} \sum_{i=1}^{3N} B_{j,i} m_i^{-1} B_{i,k}^T \\ &= \sum_{j=1}^{3N-6}\sum_{k=1}^{3N-6} L_{1,j}^{-1} G_{j,k} (L^{-1})_{k,1}^T \\ &= [\mathbf{L}^{-1}\cdot\mathbf{G}\cdot(\mathbf{L}^{-1})^T]_{1,1} \end{aligned} \quad (37)$$

where $G_{j,k}$ are the elements of the **G**-matrix,[5] which is defined tobe $\mathbf{G} \equiv \mathbf{B}\cdot\mathbf{M}^{-1}\cdot\mathbf{B}^T$. In general the **G**-matrix at an instant position can be written

$$\mathbf{G} = \mathbf{G}_0 + \mathbf{G}, \quad (38)$$

where $\mathbf{G}_0$ is the matrix evaluated at the transition state position and $\Delta\mathbf{G}$ is the remaining term depending on deviation from the transition state geometry. Thus Z_ξ can be rewritten

$$\begin{aligned} Z_\xi &= [\mathbf{L}^{-1}\cdot\mathbf{G}_0\cdot(\mathbf{L}^{-1})^T]_{1,1} + \mathbf{L}^{-1}\cdot\Delta\mathbf{G}_0\cdot(\mathbf{L}^{-1})^T]_{1,1} \\ &= 1 + Z_x \end{aligned} \quad (39)$$

The first term in Eq. (39) is 1, since $\mathbf{L}^T\cdot\mathbf{G}_0^{-1}\cdot\mathbf{L} = \mathbf{I}$ by the definition of **L** matrix.[6] The value of Z_ξ begins to deviate from 1, depending on the magnitude of ΔZ_ξ, as the molecular complex moves away from the transition state. This has also been pointed out by Otter and Briels,[2] using the Cartesian coordinate expression for Z_ξ.

Often one can choose a simple bond distance as the reaction coordinate (*i.e.*, $\xi = r$). In this case, Z_ξ is a simply diagonal component of the **G**-matrix corresponding to $\xi = r$, $\mathbf{G}_{\xi\xi}$. Therefore, noting that $\mathbf{G}_{\xi\xi} = m_i^{-1} + m_j^{-1}$, the expression of the average reactive flux in Eq. (9) is simplified to

$$\langle F\rangle_{\pm} = \int \mathbf{d}\Gamma P^c_{\xi,\dot{\xi}}(\Gamma)\int_0^\infty d\dot{\xi}' P^{(\pm)}(\dot{\xi}'|\mathbf{x})F. \quad (40)$$

After $\dot{\xi}(0)$ is sampled, it can be transformed into the corresponding Cartesian velocities using the following equation:[2]

$$\mathbf{v}_i = \mathbf{M}_i^{-1/2}\cdot\mathbf{f}\cdot\mathbf{l}_{i,1}\dot{\xi}, \quad (41)$$

where $\mathbf{l}_{i,1}$ is the 3-dimensional eigenvector transforming the reaction coordinate into the mass weighted Cartesian components of the *i*-th atom in a body fixed frame and $\mathbf{M}_i^{-1/2}$ is the 3×3 diagonal matrix with the same elements of $m_i^{-1/2}$. The matrix $\mathbf{l}_{i,1}$ in Eq. (41) is directly obtained by diagonalizing Hessian matrix in the mass weighted Cartesian coordinates at the transition state, and **f** denotes a 3×3 rotation matrix representing an instant body fixed frame without overall rotation and translation. In order to find the rotation matrix, the authors of Ref. 2 employed a numerical scheme based on the Newton-Raphson procedure. In this work, however, we have attempted to find such a matrix directly by introducing an Eckart frame[7,8] $\mathbf{f} = (\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3)$, such that

$$\mathbf{f} = \mathbf{F}\cdot(\mathbf{F}^T\cdot\mathbf{F})^{-1/2} \quad (42)$$

$$\mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3) \quad (43)$$

$$\mathbf{F}_\alpha = \sum_{i=1}^{N} c_{i\alpha} m_i^{1/2}\mathbf{x}_i \ (\alpha = 1, 2, 3), \quad (44)$$

where $c_{i\alpha}$ is one of the Cartesian components of the *i*-th atom at the transition state specified in a body fixed frame and $\mathbf{x}_i$ is an instantaneous position vector of the *i*-th atom. Note that $\mathbf{F}_i$ are the 3-dimensional column vectors. In particular, for a planar molecule, only two vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ are uniquely determined by the orthonomalization of $\mathbf{F}_1$ and $\mathbf{F}_2$ and $\mathbf{f}_3$ is obtained by $\mathbf{f}_3 = \mathbf{f}_1 \times \mathbf{f}_2$. For a linear molecule, only $\mathbf{f}_1$ is determined and the others are obtained by imposing the orthogonality of the frame. This procedure is quite efficient and easy to be implemented, since it requires only matrix inversion and diagonalization of the 3×3 Gram matrix $(\mathbf{F}^T\cdot\mathbf{F})$.[8] Using the current configuration of the system, the Eckart frame defined in Eqs. (42)-(44) is calculated once every time the reaction velocity $\dot{\xi}$ is sampled. After the pure Cartesian components of $\dot{\xi}$ are evaluated using Eq. (41), they are added to the existing velocities obtained by the MD runs described in the previous section to complete the sampling of Eq. (3). We can also implement the above algorithm forrm $P^c_{\xi,\dot{\xi}}(\Gamma)$ to a different sampling scheme developed by Carter *et al.*[9] In their method, the dynamic correction factor κ in Eq. (1) is written as

$$\kappa = \frac{\langle Z_\xi^{-1/2}\dot{\xi}(0)\theta_p[\xi(t_{pl}]\rangle_{\xi,M}}{\langle Z_\xi^{-1/2}\dot{\xi}(0)\theta_p[\dot{\xi}(0)]\rangle_{\xi,M}}, \quad (45)$$

where $<\cdots>_{\xi,M}$ denotes averaging over $P_\xi(\mathbf{x})\cdot P(\mathbf{p}^{\mathbf{x}}|\mathbf{x})$, each of which is defined in Eqs. (4) and (6), respectively, in Ref. 9. In this case, one can use only the configurational part sampled by our $P^c_{\xi,\dot{\xi}}(\Gamma)$ for $P_\xi(\mathbf{x})$ and then generate momentums using $P(\mathbf{p}^{\mathbf{x}}|\mathbf{x})$, which are just Maxwellian distributions in absence of the molecular constraints.

Implementation to Organic Reaction in Water

We have applied the above algorithm to generate the initial conditions for reactive flux calculations of methyl vinyl ketone (MVK) and cyclopentadiene (CPD) in water. The intramolecular potential energy surface $V^{intra}_{complex}$ for MVK and CPD complex (or solute) is obtained by a local harmonic approximation using the B3LYP/6-31+g* level of theory:

$$V^{intra}_{complex} = V_0 + \frac{1}{2}\sum_{i,j=1}^{3N-6} F_{i,j} dS_i dS_j, \quad (46)$$

where V_0 is the energy at the transition state, $F_{i,j}$ are the harmonic force constants in the internal coordinates, and dS_j are the internal displacement coordinates with respect to the transition state. Non-bonded interactions between the complex and water molecules consist of Lennard-Jones (LJ) and Coulomb interactions:

$$V_{i,j}^{inter} = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} + 4\varepsilon_{i,j}\left[\left(\frac{\sigma_{i,j}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{i,j}}{r_{ij}}\right)^{6}\right], \qquad (47)$$

where $\varepsilon_{i,j}$ and $\sigma_{i,j}$ are the LJ parameters and q_i are the partial charges on i-th atom. The values of Cornell *et al.*[10] are used for the LJ type interactions. The partial charges of the complex are obtained by using the restrained electrostatic potential (RESP)[11] fit at the same level of *ab initio* theory and the SPC/F2 parameters[12] are employed for the water molecules. In this work the GAUSSIAN 94 quantum chemistry program[13] has been used for the entire B3LYP calculations. At the B3LYP/6-31+g* level, the potential energy barriers from the reactant and the product to the transition state are predicted to be 16 kcal/mol and 32 kcal/mol, respectively. Using the same level of theory, Hessian matrix elements (66×66) at the transition state geometry are computed in terms of the mass weighted Cartesian coordinates and transformed into ones in the internal coordinates. The B3LYP/6-31+g* harmonic frequency of the unstable normal mode Q_1 is $431i$ cm^{-1}. A time step of 0.25 fs is chosen for the simulation. The optimized transition state complex obtained by B3LYP/6-31+g* is shown in Figure 1. For the MD simulation, the B3LYP/6-31+g* complex of MVK and CPD is surrounded by a total of 108 SPC/F2 water molecules in a cubic box with a size of 15.5 Å. In addition Nosé-Hoover chains of length 4 have been applied to both complex and water molecules at a target temperature of 300 K. The constraint MD simulation described in the previous section has been carried out for 20 ps for equilibration, and after then actual data are collected every 40 fs during a period of 40 ps. The total energy of the extended system (system + NHCs) has been plotted in Figure 2. The standard deviation of the total energy is 10^{-4} (a.u.), indicating that the energy is conserved reasonably well for the simulation. The average temperature for the simulation is 299.7 K and the average temperature fluctuation is observed to be within 4% of the target value. The ξ value that is constrained for the RF dividing plane has been maintained to be less than 10^{-10} (a.u.), which is usually achieved by 3 iteration steps. As was seen in Figure 3, sudden large fluctuations have been observed, but the magnitudes of these values are still small enough (within 10^{-10}) to be comfortably neglected. Figure 4 shows that the Z_ξ value is fluctuating around 1 as the complex structure of MVK + CPD deviates from the ransition state geometry. The $\langle Z_\xi \rangle$ value is calculated to be 1.004 ± 0.016. Note that the magnitude of Z_ξ plays an important role in determining a

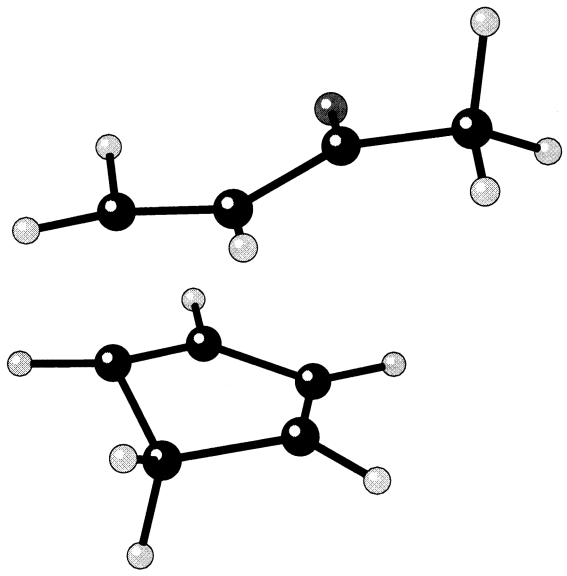

Figure 1. The optimized transition state complex for the reaction of methyl vinyl ketone (MVK) and cyclopentadiene (CPD) calculated by the B3LYP/6-31+g* method.

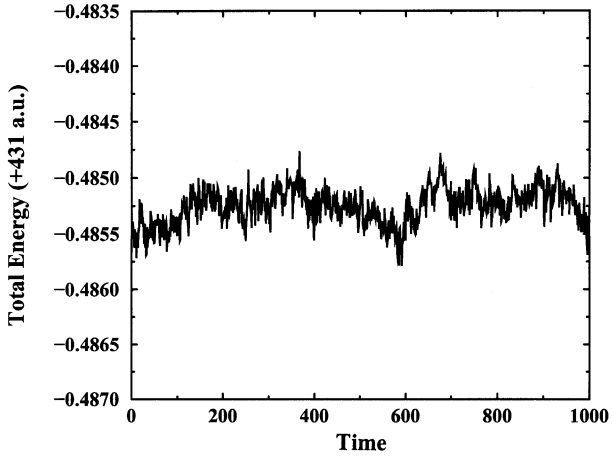

Figure 2. Total energy fluctuation of the extended system (system + NHCs) in a.u. during the 40 ps MD run. Each value of data is collected every 40 fs.

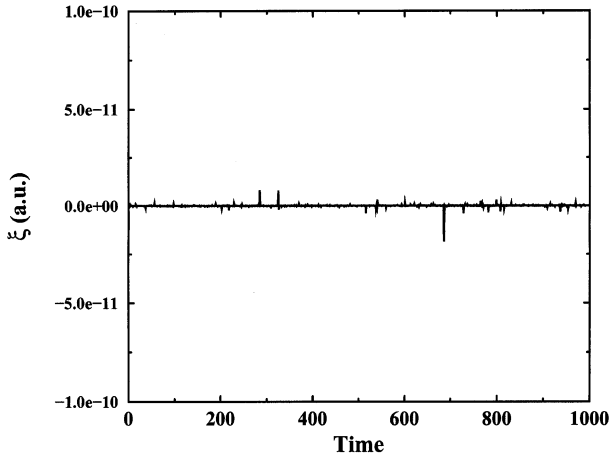

Figure 3. Plot of the unstable normal mode $\xi = Q_1$ during the 40 ps run. The Q_1 mode has an imaginary frequency of $401i$ cm^{-1} at the B3LYP/6-31+g* level of theory.

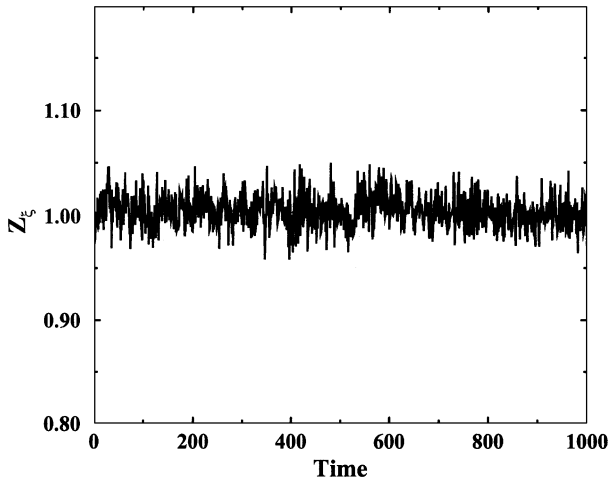

Figure 4. The fluctuation of the Z_ξ value during the 40 ps run. It provides a sampling width of reaction velocity $\dot{\xi}$ of for the $\dot{\xi}$ sampling.

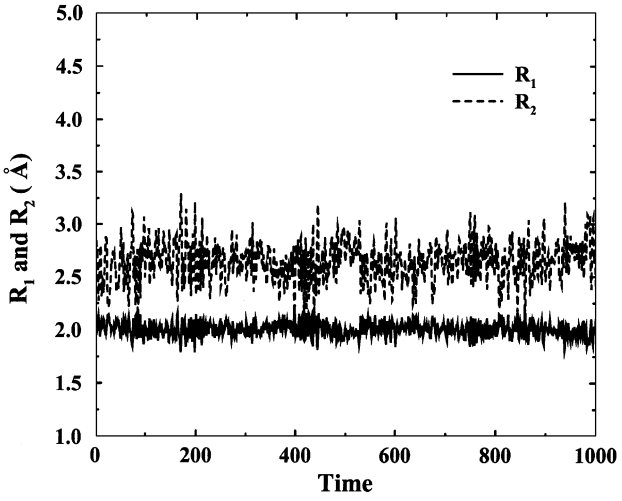

Figure 5. The fluctuations of the two C-C bonds that are involved in bond formation during 40 ps run. R_1 denotes the C-C bond not directly connected to the carbonyl group in the complex and R_2 is the other one connected to it.

width of sampling distribution of $P^{(\pm)}(\dot{\xi}|\mathbf{x})$ in Eq. (11), and that $Z_\xi^{-1/2}$ can be directly used as a biasing factor in Eq. (45). Since the Z_ξ value is nearly 1 with a relatively small fluctuation for this simulation, Eq. (9) can be approximated to be $\langle F\rangle_\pm \approx \int \mathbf{d}\Gamma P^c_{\xi,\dot{\xi}}(\Gamma) \int_0^\infty d\dot{\xi}' P^{(\pm)}(\dot{\xi}|\mathbf{x}) F$. The fluctuations of two C-C bond distances that are involved in bond formation or breaking are shown in Figure 5. The symbol R_1 denotes the C-C bond not directly connected to the carbonyl carbon and R_2 indicates the other C-C bond directly connected to it. The average fluctuations of R_1 and R_2 from the bond distances of the transition state (1.9948 Å and 2.6883 Å) are 0.05 Å and 0.15 Å, respectively, demonstrating somewhat larger fluctuation of R_2.

Summary

We have developed the NHC canonical sampling algorithm to generate the initial conditions for RF studies of organic reactions in condensed phase. This algorithm can be directly applied to classical RF MD studies of organic reactions, by constructing reactive potential energy surface, which has been an undergoing project. Also it is expected that this method can provide reliable initial conditions for future *ab initio* RF MD simulations to study solvent effects in organic reactions.

References

1. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
2. den Otter, W. K.; Briels, W. J. *J. Chem. Phys.* **1996**, *106*, 5494.
3. Jang, S.; Voth, G. A. *J. Chem. Phys.* **1998**, *107*, 9514.
4. Berne, B. J. In *Multiple Time Scales*; Brackbill, J. U.; Cohen, B. I., Eds.; Academic: New York, 1985.
5. Wilson, Jr., E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; Dover Publications: New York, 1955.
6. Papousek, D.; Aliev, M. R. *Molecular Vibrational-Rotational Spectra*; Elsevier: New York, 1982.
7. Eckart, C. *Phys. Rev.* **1935**, *47*, 552.
8. Louck, J. D.; Galbraith, H. W. *Rev. Mod. Phys*. **1976**, *48*, 69.
9. Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett*. **1989**, *156*, 472.
10. Cornell, W. D. *et al. J. Am. Chem. Soc.* **1995**, *117*, 5179.
11. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem*. **1993**, *97*, 10269.
12. Lobaugh, J.; Voth, G. A. *J. Chem. Phys*. **1997**, *106*, 2400.
13. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *GAUSSIAN 94*, Revision E.2; Gaussian, Inc.: Pittsburgh, PA, 1995.