Direct quantum mechanical/molecular mechanical simulations of two-dimensional vibrational responses: $N$-methylacetamide in water

Jonggu Jeon$^1$ and Minhaeng Cho$^{1,2,3}$

$^1$ Department of Chemistry, Korea University, Seoul 136-701, Korea
$^2$ Multidimensional Spectroscopy Laboratory, Korea Basic Science Institute, Seoul 136-713, Korea
E-mail: mcho@korea.ac.kr

*New Journal of Physics* 12 (2010) 065001 (27pp)
Received 16 February 2010
Published 4 June 2010
Online at http://www.njp.org/
doi:10.1088/1367-2630/12/6/065001

**Abstract.** Multidimensional infrared (IR) spectroscopy has emerged as a viable tool to study molecular structure and dynamics in condensed phases, and the third-order vibrational response function is the central quantity underlying various nonlinear IR spectroscopic techniques, such as pump–probe, photon echo and two-dimensional (2D) IR spectroscopy. In this paper, a new computational method is presented that calculates this nonlinear response function in the classical limit from a series of classical molecular dynamics (MD) simulations, employing a quantum mechanical/molecular mechanical (QM/MM) force field. The method relies on the stability matrix formalism where the dipole–dipole quantum mechanical commutators appearing in the exact quantum response function are replaced by the corresponding Poisson brackets. We present the formulation and computational algorithm of the method for both the classical and the QM/MM force fields and apply it to the 2D IR spectroscopy of carbon monoxide (CO) and $N$-methylacetamide (NMA), each solvated in a water cluster. The conventional classical force field with harmonic bond potentials is shown to be incapable of producing a reliable 2D IR signal because intramolecular vibrational anharmonicity, essential to the production of the nonlinear signal, is absent in such a model. The QM/MM force field, on the other hand, produces distinct 2D spectra for the NMA and CO systems with clear vertical splitting and cross peaks, reflecting the vibrational anharmonicities and the vibrational couplings between the underlying vibrational modes, respectively.

$^3$ Author to whom any correspondence should be addressed.
In the NMA spectrum, the coupling between the amide I and II modes is also well reproduced. While attaining the converged spectrum is found to be challenging with this method, with an adequate amount of computing it can be straightforwardly applied to new systems containing multiple chromophores with little modeling effort, and therefore it would be useful in understanding the multimode 2D IR spectrum of complex molecular systems.

Contents

1. Introduction 2
2. Theory and computation 4
   2.1. Theory 4
   2.2. Computational algorithm 7
   2.3. Systems and computational details 8
   2.4. Phase-matching consideration: photon echo spectroscopy 10
3. Linear and 2D IR spectra of CO in water 12
4. Simulated spectra of N-methylacetamide in water 16
   4.1. Dipolar linear response 16
   4.2. Third-order vibrational response function 19
   4.3. Apodization functions 21
   4.4. Absorptive 2D IR spectrum 23
5. Summary and a few concluding remarks 25
Acknowledgments 26
References 26

1. Introduction

Two-dimensional infrared (2D IR) spectroscopy is an experimental technique that measures the response of molecular systems to three incident IR pulses [1]. Because this method can reveal detailed molecular responses taking place during multiple time spans between incident and/or signal pulses, it provides information about the system dynamics, including vibrational mode anharmonicity, couplings between normal modes, and homogeneous and inhomogeneous line broadenings. The 2D IR spectrum can be analyzed in terms of peak shapes as well as peak positions and amplitudes in the 2D frequency space, and therefore a significant frequency resolution enhancement is naturally expected. Thus, 2D IR spectroscopy has been regarded as an optical analog of 2D NMR spectroscopy [2, 3], even though there are a few fundamental differences in the Hamiltonians determining the inter- and intra-molecular (or inter-spin) interaction-induced dynamics and the lineshapes of the resultant 2D spectra. Over the last decade, this spectroscopic tool has been widely used to elucidate the structure and dynamics of complex molecular systems, such as solvation and hydrogen bonding dynamics, chemical exchange reactions, and peptide and protein structure and dynamics [4–16].

Because the 2D spectral features are related to the underlying molecular properties in a complicated way, theoretical modeling and computation have been essential in their understanding and interpretation in terms of microscopic variables. A central molecular quantity determining the 2D IR spectrum is the third-order vibrational response function that connects...
the nonlinear polarization, an experimental observable, to the three input IR electric fields [12, 17]. Time-dependent perturbation theory yields the response function as nested quantum commutators of electric dipole operators at different times, and its evaluation requires a proper account of quantum states as well as system dynamics. Theoretical evaluation of the 2D IR spectrum therefore generally involves both quantum chemical electronic structure calculations (ESC) for the vibrational chromophores and molecular dynamics (MD) simulations accounting for dynamics of the interacting solvent molecules [15, 18]. This approach can reproduce the experimental spectrum quite accurately for individual vibrational modes, such as the amide I mode of peptides, and it has provided insights into the microscopic origins of the 2D spectral features. Despite the success of this hybrid ESC–MD method, it still relies on a number of approximations regarding the treatment of multiple coupled anharmonic oscillators and the quantum dynamical aspects of the system evolution. Also, this method requires a significant modeling effort on a system-by-system basis and is often limited to the description of a few specific vibrational modes. Thus, there have been many theoretical studies that explore the classical limit of the quantum response function that can be evaluated in a conceptually simpler and more general way with the classical MD simulation method [19–24].

The classical nonlinear response function can be obtained from its quantum mechanical counterpart by replacing the quantum commutators with classical Poisson brackets [20]. The classical nonlinear response function obtained with this approximation contains the so-called stability matrix, which is a measure of the deviation of phase space variables in time upon the perturbation of initial conditions. Due to the chaotic nature of classical trajectories of many-body systems, the stability matrix is an exponentially diverging quantity in time variables and attaining convergence in the nonlinear response function calculation becomes a challenging task [19, 25–28]. In fact, it is only recently that the classical nonlinear response function of a chaotic system has been demonstrated to be a well-behaved function [29–30].

Due to its classical nature, computational studies based on this approach had initially focused on the low-frequency intermolecular vibrational modes of bulk solvent systems. Examples include the fifth-order Raman spectroscopy of atomic liquids [22, 31], liquid carbon disulfide [23, 24, 32] and water [24]. Later, this method was also applied to the 2D IR surface spectroscopy of CO (carbon monoxide) vibration on a copper surface [33] and the 2D IR spectroscopy of intermolecular modes of bulk water [34, 35]. Except for the last application, the computational studies reported so far have been limited to second-order response spectroscopy [1].

In this paper, a new computational method to evaluate the classical limit of the third-order vibrational response function is developed using the stability matrix formalism. To the authors’ best knowledge, this is the first attempt at such a goal based on the stability matrix formulation combined with equilibrium MD simulation and a quantum mechanical/molecular mechanical force field. This study is also unique in its focus on high-frequency intramolecular vibrational modes. Because the nonlinear optical response vanishes in a harmonic system, accurate description of intramolecular 2D IR spectroscopy needs to be based on realistic anharmonic force fields for intramolecular vibrational motions. We employ both the conventional classical force field with harmonic bond potentials and fixed atomic charges and the combined quantum mechanical/molecular mechanical (QM/MM) force field [36] at the semiempirical PM3 level [37], and compare their differences. It will be shown that a reliable 2D IR spectrum can be obtained with the QM/MM force field. In this initial study, we apply the developed method on two model systems. The first one, a carbon monoxide (CO) molecule solvated in a water cluster,
is chosen as a simple model for an anharmonic oscillator. The second, an \(N\)-methylacetamide (NMA) molecule solvated in a water cluster, is a prototype model for peptides. The NMA has two prominent vibrational modes, amide I and amide II, whose appearances and coupling in the 2D IR spectrum have been studied both theoretically [38–41] and experimentally [42–46].

We first derive the classical expression of the third-order vibrational response function and describe the computational algorithm to evaluate it using equilibrium MD simulation. Complications arising from the QM/MM force field are highlighted and the procedure to obtain the absorptive 2D correlation spectrum from the time domain response function is also described in some detail. Then, the computed time domain response functions and 2D IR spectra are presented for the CO and NMA systems and they are compared with available experimental spectra. Finally, we assess the strengths and weaknesses of the method and discuss possible refinements for future studies.

2. Theory and computation

2.1. Theory

The third-order response function is defined as the expansion coefficient of third-order nonlinear polarization in terms of the incident electric fields [17], i.e.

\[
P^{(3)}(\mathbf{r}, t) = \int_0^\infty \mathrm{d}t_3 \int_0^\infty \mathrm{d}t_2 \int_0^\infty \mathrm{d}t_1 R^{(3)}(t_3, t_2, t_1) E(\mathbf{r}, t-t_3) E(\mathbf{r}, t-t_2-t_3) E(\mathbf{r}, t-t_3-t_2-t_1).
\]

(1)

In the case of time-resolved IR four-wave-mixing spectroscopy, the electric field \(E(\mathbf{r}, t)\) is given as the sum of three IR pulses that are temporally separated. Within the electric dipole approximation to radiation–matter interaction, the third-order response function is expressed as

\[
R^{(3)}(t_3, t_2, t_1) = \left( \frac{i}{\hbar} \right)^3 \theta(t_3)\theta(t_2)\theta(t_1) \langle \mu(t_1+t_2+t_3) [\mu(t_1+t_2), [\mu(t_1), [\mu(0), \rho(\infty)]]] \rangle,
\]

(2)

where \(\mu(t)\) is the dipole operator at time \(t\) in the interaction picture, which evolves in the absence of field–matter interaction, \(\theta(t)\) is the Heaviside step function and \(\rho(\infty)\) is the equilibrium density operator. The angular brackets represent an average over initial conditions.

In the classical limit, the quantum mechanical commutators are replaced by the Poisson brackets [20]

\[
[A, B] \rightarrow i\hbar \{A, B\}_{PB}
\]

(3)

and the response function in equation (2) becomes

\[
R^{(3)}_{\alpha\beta\gamma\delta}(t_3, t_2, t_1) = \beta \text{Tr} \left[ \left\{ \mu_\alpha(t_1+t_2+t_3), \mu_\beta(t_1+t_2) \right\}_{PB} \left( \left\{ \mu_\gamma(t_1), \mu_\delta(0) \right\}_{PB} - \beta \dot{\mu}_\gamma(t_1)\dot{\mu}_\delta(0) \right) \rho_{eq} \right],
\]

(4)

in the canonical ensemble with density matrix \(\rho_{eq}\). Here, \(\beta = 1/(k_B T)\) with the Boltzmann constant \(k_B\) and absolute temperature \(T\), the Greek subscripts denote Cartesian components, \(\mu\) now represents the dipole moment of the system and \(\dot{\mu}\) is its time derivative. Utilizing the invariance of Poisson brackets under canonical transformation [31, 47] and assuming that the
convenient to remove the time containing two and one stability matrix factor, respectively. For numerical computations, it is momenta at time $t$ arbitrary initial time, and introducing the reverse trajectory obtained by changing the sign of $t$ are related to the stability matrix of the trajectory. Thus, $\partial \mu$ numerator and denominator, where $\partial \mu_\alpha$ and $p_i$ are the $\alpha$ components of position and momentum, respectively, of atom $i$ and $N$ is the total number of atoms in the system. An expression equivalent to equation (5) has been previously derived in [48]. The partial derivatives with different time arguments in the numerator and denominator, $\partial \mu_\alpha(t_1 + t_2 + t_3)/\partial p_{i\lambda}(t_1 + t_2)$ and $\partial \tilde{\mu}_\delta(0)/\partial p_{j\eta}(t_1)$ in equation (5), are related to the stability matrix of the trajectory. Thus, $R^{(3)}_\mu(t_1, t_2, t_3)$ is the sum of two terms, containing two and one stability matrix factor, respectively. For numerical computations, it is convenient to remove the $t_1$ dependence in $p_{i\lambda}(t_1 + t_2)$ and to express the partial derivatives in a causal form. This can be achieved by shifting the time arguments by $t_0 - t_1$, where $t_0$ is an arbitrary initial time, and introducing the reverse trajectory obtained by changing the sign of momenta at time $t_0$. This leads to the following expression:

$$R^{(3)}_\mu(t_1, t_2, t_3) = \beta \mathrm{Tr} \left[ \rho_{eq} \sum_{i=1}^{N} \sum_{\lambda=x,y,z} \frac{\partial \mu_\alpha(t_1 + t_2 + t_3)}{\partial p_{i\lambda}(t_1 + t_2)} \frac{\partial \mu_\beta(t_1 + t_2)}{\partial q_{i\lambda}(t_1 + t_2)} \right] \times \left[ \sum_{j=1}^{N} \sum_{\eta=x,y,z} \frac{\partial \tilde{\mu}_\gamma (t_1)}{\partial q_{j\eta}(t_1)} \frac{\partial \tilde{\mu}_\delta (t_1)}{\partial p_{j\eta}(t_1)} + \beta \tilde{\mu}_\gamma (t_1) \tilde{\mu}_\delta (t_1) \right],$$

where $q_{i\alpha}$ and $p_{i\alpha}$ are the $\alpha$ components of position and momentum, respectively, and $N$ is the total number of atoms in the system. An expression equivalent to equation (5) has been previously derived in [48]. The partial derivatives with different time arguments in the numerator and denominator, $\partial \mu_\alpha(t_1 + t_2 + t_3)/\partial p_{i\lambda}(t_1 + t_2)$ and $\partial \tilde{\mu}_\delta(0)/\partial p_{j\eta}(t_1)$ in equation (5), are related to the stability matrix of the trajectory. Thus, $R^{(3)}_\mu(t_1, t_2, t_3)$ is the sum of two terms, containing two and one stability matrix factor, respectively. For numerical computations, it is convenient to remove the $t_1$ dependence in $p_{i\lambda}(t_1 + t_2)$ and to express the partial derivatives in a causal form. This can be achieved by shifting the time arguments by $t_0 - t_1$, where $t_0$ is an arbitrary initial time, and introducing the reverse trajectory obtained by changing the sign of momenta at time $t_0$. This leads to the following expression:

$$R^{(3)}_\mu(t_1, t_2, t_3) = \beta \mathrm{Tr} \left[ \rho_{eq} \sum_{i=1}^{N} \sum_{\lambda=x,y,z} \frac{\partial \mu_\alpha(t_1 + t_2 + t_3)}{\partial p_{i\lambda}(0; t_1 + t_2 + t_0)} \frac{\partial \mu_\beta(0; t_1 + t_2 + t_0)}{\partial q_{i\lambda}(0; t_1 + t_2 + t_0)} \right] \times \left[ \sum_{j=1}^{N} \sum_{\eta=x,y,z} \frac{\partial \tilde{\mu}_\gamma (0; t_1)}{\partial q_{j\eta}(0; t_1)} \frac{\partial \tilde{\mu}_\delta (0; t_1)}{\partial p_{j\eta}(0; t_1)} + \beta \tilde{\mu}_\gamma (0; t_1) \tilde{\mu}_\delta (0; t_1) \right],$$

where $x(t_2; t_1)$ denotes the quantity $x$ when the system has evolved for time $t_2$ from initial time $t_1$, and the overbar indicates the quantity is obtained from the reverse trajectory starting at time $t_0$. See figure 1 for the schematic view of trajectories involved. For later use, we rewrite equation (6) as follows:

$$R^{(3)}_\mu(t_1, t_2, t_3) = \beta \mathrm{Tr} \left[ \rho_{eq} \mathbf{M}_P(t_1, 0; t_1 + t_2) \cdot \mathbf{M}_Q^T(0; 0; t_1 + t_2) \right] \otimes \left[ -\mathbf{M}_P(0; 0; t_0) \cdot \mathbf{V}_P^T(1; 0; t_0) + \beta \mathbf{P}(0; t_0) \otimes \mathbf{P}(1; t_0) \right],$$

where $\mathbf{M}_X(t_2, t_1; t_0)$ and $\mathbf{V}_X(t_2, t_1; t_0)$ ($X = P, Q$) are $3 \times 3N$ dipole derivative and dipole velocity derivative matrices, respectively, defined as

$$\mathbf{M}_X(t_2, t_1; t_0) = \frac{\partial \mu(t_2; t_0)}{\partial x(t_1; t_0)} \quad (x = \mathbf{p}, \mathbf{q}),$$

$$\mathbf{V}_X(t_2, t_1; t_0) = \frac{\partial \tilde{\mu}(t_2; t_0)}{\partial x(t_1; t_0)} \quad (x = \mathbf{p}, \mathbf{q}).$$

Here, $\otimes$ denotes tensor direct product, superscript $T$ denotes matrix transpose and, once again, the overbar indicates reverse trajectory properties. Throughout this paper, we use the convention $(\partial x/\partial y)_{ab} = \partial x_a/\partial y_b$ for arbitrary vectors $\mathbf{x}$ and $\mathbf{y}$. 

New Journal of Physics 12 (2010) 065001 (http://www.njp.org/)
nonpolarizable force fields, since the solute dipole is given as \( d \) instantaneously affected by the position-perturbation of the solvent. However, for classical QM/MM ones, this does not lead to any computational simplification because the solute dipole lead to any computational simplification because the solute dipole

In equation (7), \( M_Q(0, 0; t_0 + t_2) \) and \( \tilde{M}_Q(0, 0; t_0) \) are just the vibrational transition dipoles evaluated at \( t_0 + t_2 \) and \( t_0 \), respectively. However, the other two terms \( M_P(t_3, 0; t_0 + t_2) \) and \( \tilde{V}_P(t_1, 0; t_0) \) are related to the stability matrix. If we introduce the stability matrix \( S(t_2, t_1; t_0) \) as follows,

\[
S(t_2, t_1; t_0) = \begin{pmatrix} S^{QQ}(t_2, t_1; t_0) & S^{QP}(t_2, t_1; t_0) \\ S^{PQ}(t_2, t_1; t_0) & S^{PP}(t_2, t_1; t_0) \end{pmatrix},
\]

(9)

\( M_P(t_2, t_1; t_0) \) and \( V_P(t_2, t_1; t_0) \) can be decomposed using the chain rule as

\[
M_P(t_2, t_1; t_0) = M_Q(0, 0; t_0) + \tilde{M}_Q(0, 0; t_0) + M_P(t_3, 0; t_0 + t_2) \cdot S^{QP},
\]

\[
V_P(t_2, t_1; t_0) = V_Q(0, 0; t_0) + \tilde{V}_P(t_1, 0; t_0) + V_P(t_3, 0; t_0 + t_2) \cdot S^{PP},
\]

(10)

for an arbitrary \( \tau \).

In actual computations, the full response function in equation (6) or (7) is found to converge very slowly due to the solvent contribution. Thus, we have used the solute response function instead, which is obtained by replacing the whole system dipole moment \( \mu \) with the solute dipole moment \( \delta d \) in equation (6). For general polarizable force fields including the QM/MM ones, this does not lead to any computational simplification because the solute dipole is instantaneously affected by the position-perturbation of the solvent. However, for classical nonpolarizable force fields, since the solute dipole is given as \( d(t) = \sum c_i q_i(t) \), where \( c_i \) and \( q_i \) are the atomic charge and the coordinate of the \( i \)th atom, the following relations hold:

\[
\frac{\partial d_s(t; t_0)}{\partial q_{i\beta}(t; t_0)} = \begin{cases} c_i \delta_{i\alpha} & (i \in \text{solute}) \\ 0 & (i \notin \text{solute}) \end{cases}
\]

(11)
The solute response function can then be simplified as

$$R^{(3)}_{\alpha \beta \gamma \delta}(t_3, t_2, t_1) = \beta \text{Tr} \left[ \rho_{eq} \sum_{i=1}^{N_S} c_i \frac{\partial d_i(t_3; t_0 + t_2)}{\partial p_{\beta}(0; t_0 + t_2)} \right. \times \left. \left( -\sum_{j=1}^{N_S} c_j \frac{\partial \tilde{d}_j(t_1 + t_0)}{\partial \tilde{p}_{\alpha}(0; t_0)} + \beta \tilde{d}_j(0; t_0) \tilde{d}_i(t_1; t_0) \right) \right]$$  \hspace{1cm} (12)

where the summations now run over $N_S$ solute atoms. For the rest of this paper, we confine our considerations to the solute response function and use the same $R^{(3)}(t_3, t_2, t_1)$ notation for it. For general polarizable force fields, it is given by equation (6)–(10) with $\mu$ redefined as the solute dipole moment. For nonpolarizable force fields, equation (12) provides its expression.

According to equation (7), the $t_1$ dependence of $R^{(3)}(t_3, t_2, t_1)$ is determined at the single trajectory level by the interference of the stability matrix term, $-\tilde{M}_Q(0, 0; t_0) \cdot \tilde{V}_p(t_1, 0; t_0)$, with the last term involving dipole velocities. It is then modulated in the statistical averaging by the first factor in equation (7), $\tilde{M}_P(t_1, 0; t_0 + t_2) \cdot \tilde{M}_Q(0, 0; t_0 + t_2)$. Likewise, the $t_3$ dependence of $R^{(3)}(t_3, t_2, t_1)$ is determined by the first factor in equation (7) at the single trajectory level and is modulated by the $t_1$-dependent factor during the averaging process. On the other hand, the $t_2$ dependence is determined by the equilibrium dynamics only. We can therefore deduce a rough correspondence between the quantum mechanical and classical nonlinear processes: the coherence evolutions during $t_1$ and $t_3$ in the quantum picture are related to the stability matrix evolutions governed by the instantaneous Hessian matrix [21] in the classical case and the population evolution during $t_2$ is to the ordinary Newtonian dynamics.

### 2.2. Computational algorithm

Numerical evaluation of the third-order response function is based on the working formula in equation (7). $R^{(3)}(t_3, t_2, t_1)$ has the form of the four-time correlation function (TCF) and, for its accurate evaluation, it is important to match the relative errors of various numerical derivatives at the same order in the step sizes of differential variables $t$, $p$, and $q$. In particular, the leap-frog algorithm employed in the MD simulation has errors of order $(\Delta t)^2$ [49] and the time derivatives should be evaluated at the same level of accuracy. Thus, time derivatives in equation (7) have been calculated from a single MD trajectory with the central difference scheme, i.e.

$$\tilde{\mu}(t_1; t_0) \approx \frac{\tilde{\mu}(t_1 + \Delta t; t_0) - \tilde{\mu}(t_1 - \Delta t; t_0)}{2\Delta t},$$  \hspace{1cm} (13)

and for the initial values,

$$\tilde{\mu}(0; t_0) \approx \frac{4\tilde{\mu}(\Delta t; t_0) - \tilde{\mu}(2\Delta t; t_0) - 3\tilde{\mu}(0; t_0)}{2\Delta t}.$$  \hspace{1cm} (14)

Both of these derivatives are subject to errors of order $(\Delta t)^2$. We employed a similar scheme for $M_{p,q}$, $V_{p,q}$ and $S$ that involve derivatives with respect to $p$ and $q$. However, these quantities cannot be evaluated from a single ordinary MD trajectory and it is necessary to calculate explicitly two additional surrogate trajectories with displaced initial $p$ or $q$ for each derivative. For example, $M_{p}(t_3, 0; t_0 + t_2)$ in equation (7) can be evaluated as

$$(M_{p})_{\alpha i}(t_3, 0; t_0 + t_2) \approx \frac{\mu_{\alpha i}^+(t_3; t_0 + t_2) - \mu_{\alpha i}^-(t_3; t_0 + t_2)}{2\Delta p_{\alpha i}} \quad (\alpha = x, y, z; \ i = 1, \ldots, 3N),$$  \hspace{1cm} (15)
where $\mu^+_a(t_3; t_0 + t_2)$ is the dipole moment at $t_0 + t_2 + t_3$ of the surrogate trajectory with initial momentum $p_i(t_0 + t_2)$ displaced by $\pm \Delta p_i$. Thus, we need $6N$ surrogate trajectories for $\mathbf{M}_p(t_3; 0; t_0 + t_2)$. This clearly demonstrates the advantage of eliminating the $t_1$ dependence in equation (5) and using equation (7) instead: the number of surrogate trajectories grows linearly with the number of initial times and the computational complexity becomes prohibitive if it includes $t_1$, which has a very short interval for the Fourier transform to obtain the 2D spectrum. $\mathbf{V}_p(t_1; 0; t_0)$ can be evaluated similarly by the combination of equations (13)–(15).

The algorithm described above can be used for the response function evaluation with fixed $t_2$ and variable $t_1$ and $t_3$. However, the waiting time dependence of the system response is of interest, which requires the response function at a series of different $t_2$ values. For this, we can exploit the following chain rule relation for the stability matrix [50],

$$\mathbf{S}(t_2, t_1; t_0) = \mathbf{S}(t_2, \tau; t_0) \cdot \mathbf{S}(\tau, t_1; t_0),$$  \hfill (16)

along with equation (10) to achieve significantly improved computational efficiency. Specifically, we divide the time span of $t_3$ into $N_b$ blocks and perform $N_b$ short MD simulations instead of a single long simulation over the entire $t_3$ time span. For $t_2 = 0$, this corresponds to calculating $6NN_b$ surrogate trajectories of length $t_3^{\text{max}}/N_b$ instead of $6N$ trajectories of length $t_3^{\text{max}}$. Then, $\mathbf{M}_p(t_3, 0; t_0 + t_2)$ at all values of $t_3$ can be obtained from equations (10) and (16). This does not provide any advantage for a single $t_2$ other than increased numerical accuracy [50]. However, this information can be reused in calculations at different $t_2$ values using the identity

$$\mathbf{M}_p(t_3, 0; t_0 + t_2) = \mathbf{M}_p(t_2 + t_3, t_2; t_0).$$  \hfill (17)

In the limit of small $t_2$ interval and long $t_3^{\text{max}}$, this block algorithm can be shown to be $N_b/2$ times more efficient than the simple one. Here, the factor 1/2 comes from the need to calculate position-perturbed surrogate trajectories in addition to momentum-perturbed ones (see equation (10)). Due to the added complexity associated with a larger number of trajectories, actual improvement is less dramatic but still significant: for the present computational conditions (for details, see below), the efficiency was improved by roughly an order of magnitude.

2.3. Systems and computational details

We apply the method described above to two model systems. The first one is a CO molecule solvated in a cluster of six water molecules and the second is a singly deuterated N-methylacetamide (d-NMA) molecule solvated in a 16-heavy-water cluster. In the AMBER 9 package [51] used in this study, QM/MM MD simulations of a bulk solution with a periodic boundary condition show much worse energy conservation than those of a cluster system and the stability matrix can be calculated accurately only in the cluster configuration. We have chosen a rather small number of water molecules in each system to reduce computational cost, but the employed systems still exhibit a significant solvation effect, as shown below.

In the all-classical description, the CO molecule was modeled with a classical force field including a harmonic bond potential. The bond force constant of $1.538 \times 10^3$ kcal (mol Å$^2$)$^{-1}$ and the equilibrium distance of 1.13 Å were used so that the resulting vibrational frequency (2310 cm$^{-1}$) and bond length (1.13 Å) match those of subsequent QM/MM MD at the semiempirical PM3 level (see below). The partial charges on atoms C and O were chosen as $\pm 0.1656|e|$ to reproduce the CO dipole moment of 0.9 D at the PM3/MM level. The Lennard-Jones parameters for CO were those of atom types C and O as defined in the AMBER ff99
The classical d-NMA molecule was modeled by the AMBER ff03 force field with the mass of amide hydrogen changed to 2 amu. For water molecules, a flexible three-site model derived from the SPC model was used for the CO cluster system and a flexible TIP3P model with a hydrogen mass of 2 amu for the d-NMA system. In QM/MM MD, the solute (CO or d-NMA) was described with the semiempirical PM3 theory and the solvent was described classically. To prevent molecules from diffusing out of the cluster, spherical capping potential was inevitably applied to all atoms in the system as implemented in the AMBER 9 package, with a force constant of 10 kcal (mol Å²)⁻¹ and radii of 4.5 Å (CO) and 6.6 Å (d-NMA).

Partial charges of QM atoms were calculated with the Mulliken population analysis and, from them, the solute dipole moment was determined. The AMBER 9 code was properly modified to save this information at three consecutive time steps (cf equations (13) and (14)) every 4 fs. When Fourier transformed, this 4 fs interval provides spectra up to 4100 cm⁻¹.

Each system was first equilibrated at 298 K using the Langevin thermostat and 2000 (1000 for CO) configurations were saved from a single constant NVT equilibrium trajectory with intervals of 1 ps (CO) or 10 ps (d-NMA). All MD simulations employed a time step size Δt of 0.5 fs. These 2000 configurations serve as starting points for the subsequent response function calculations. They were used in constant NVE simulations of 2.5 ps duration and the last 1.5 ps portions were taken as the reference trajectories along t₂ and t₃ (cf equation (7)). The starting point of this 1.5 ps portion corresponds to t = t₀ in the algorithm presented above and its length was chosen so that t₂ and t₃ of up to 1 and 0.5 ps can be covered, respectively. The initial condition for the reverse trajectory was prepared from this reference trajectory at t = t₀ by reversing signs of velocities. Here, care should be taken to account for the Δt/2 mismatch between position and velocity that is inherent in the leap-frog integrator. Specifically, the initial positions for the forward and reverse trajectories are identical at t = t₀ but the initial velocity for the reverse trajectory is the negative of the forward trajectory velocity at t = t₀ + Δt/2, which is simply reported in the MD program as the velocity at t = t₀ + Δt.

With this initial condition, the reverse trajectory was run for 0.5 ps, the maximum value of t₁. Position- and momentum-perturbed surrogate trajectories were then run for 0.5 ps each, starting with the coordinates and velocities saved in the forward and reverse reference trajectories. The displacement in momentum Δpᵢ was chosen as 0.020455 Å (ps amu)⁻¹ for all atoms, which results in different velocity displacement Δvᵢ for different atoms due to the mass scaling. The positional displacement Δrᵢ was chosen to be mass dependent as 0.0005 Å amu⁻¹. Note that the initial time of the forward surrogate trajectory is t₀ + t₂ in equation (7) and it needs to be started multiple times at every t₂ value, chosen to span 1 ps with a 4 fs increment. On the other hand, the reverse surrogate trajectories need to be started only once at t = t₀ (cf equation (7)). In the block algorithm introduced above, the 1.5 ps time span along t₂ + t₃ was divided into 75 blocks of length 20 fs and the forward trajectories were run in each block with a maximum length of 20 fs. The full response function was then reconstructed using equations (10), (16) and (17) in the range 0 ≤ t₁, t₃ ≤ 0.5 ps and 0 ≤ t₂ ≤ 1 ps in the 4 fs interval. During computation for the d-NMA system, it was found that a small number of samples with large amplitude spoil the average, and therefore 2% of samples with the largest mean square amplitude of the response function were deliberately discarded.
2.4. Phase-matching consideration: photon echo spectroscopy

To make contact with experiments, we consider the relation between the third-order response function and the absorptive 2D IR correlation spectrum. When the chromophore interacts with the incident light three times, the third-order nonlinear polarization \( P^{(3)}(r, t) \) can be induced in the system according to equation (1). In turn, this time-dependent polarization produces a signal electric field \( E^{(3)}(t) \) detected in experiment. The latter is given as an approximate solution of the Maxwell equation with \( P^{(3)}(r, t) \) as the source [1, 17],

\[
E^{(3)}_{ks}(t) \propto \frac{i \omega_k}{n(\omega_k)} P^{(3)}_{ks}(t),
\]

where \( E^{(3)}_{ks} \) and \( P^{(3)}_{ks} \) are the \((k_s, \omega_k)\) components in the plane wave expansion of the signal field and the polarization, respectively, and a perfect phase matching between them is assumed.

Incident light can be expressed as the sum of three light pulses with Gaussian envelope

\[
E(t) = \hat{e}_1 E_1 g_1(t + \tau + T) e^{ik_1 \cdot r - i \omega_1 t} + \hat{e}_2 E_2 g_2(t + T) e^{ik_2 \cdot r - i \omega_2 t} + \hat{e}_3 E_3 g_3(t) e^{ik_3 \cdot r - i \omega_3 t} + (\text{c.c.}),
\]

where \( g_i(t) = e^{-t^2/(2 \sigma_i^2)}/(\sigma_i \sqrt{2\pi}) \) with width \( \sigma_i \), c.c. denotes complex conjugate, \( \tau \) and \( T \) are the delay times between pulses and \( \omega_i \) is the carrier frequency of the \( i \)th pulse. When this is inserted in equation (1), \( P^{(3)}(r, t) \) can be expanded into \( 6^3 \) terms with distinct phases and frequencies:

\[
P^{(3)}(t) = \iiint d^3r_1 d^3r_2 d^3r_3 R^{(3)}(t_3, t_2, t_1): \left\{ \hat{e}_1 E_1 g_1(t + \tau + T - t_3) e^{ik_1 \cdot r - i \omega_1 (t-t_3)} + \hat{e}_2 E_2 g_2(t + T - t_3) e^{ik_2 \cdot r - i \omega_2 (t-t_3)} + \hat{e}_3 E_3 g_3(t - t_3) e^{ik_3 \cdot r - i \omega_3 (t-t_3)} + (\text{c.c.}) \right\}
\]

In an experiment, the signal field is detected along a specific direction given by one of \( k_s = \pm k_1 \pm k_2 \pm k_3 \). To each of these, only six terms of \( P^{(3)}(r, t) \) with matching phase factor will contribute. For example, the rephasing signal is detected along the direction \( k_R = -k_1 + k_2 + k_3 \) and the polarization responsible for it is given by

\[
P^{(3)}_{\text{reph}}(t) = \iiint dt_3 dt_2 dt_1 R^{(3)}(t_3, t_2, t_1): \left\{ \hat{e}_1 E_1^+ g_1(t + \tau + T - t_3) e^{-ik_1 \cdot r + i \omega_1 (t-t_3)} \right.
\]

\[
\times \hat{e}_2 E_2 g_2(t + T - t_3 - t_2) e^{ik_2 \cdot r - i \omega_2 (t-t_3-t_2)} \hat{e}_3 E_3 g_3(t - t_3 - t_2 - t_1) e^{ik_3 \cdot r - i \omega_3 (t-t_3-t_2-t_1)}
\]

\[
+ (5 \text{ other combinations}) \right\}.
\]

When the incident pulses are narrow, \( g_i(t) \) can be replaced with the Dirac delta function \( \delta(t) \) and the integral in equation (21) can be trivially carried out to obtain six terms. Among them, four terms contain the response function with negative time arguments and therefore vanish. In
this impulsive limit, therefore, the rephrasing polarization can be expressed as the sum of two terms:

\[ P_R^{(3)}(t) = [R^{(3)}(t, T, \tau) \hat{e}_3 \hat{e}_1 e^{i\omega_3t} e^{-i\omega_3 \tau} e^{i(\omega R - \omega_3)T} + R^{(3)}(t + T, -T, \tau + T) \hat{e}_3 \hat{e}_1 E_2 E_1^* e^{i\omega_3t} e^{-i\omega_3 \tau} e^{i(\omega R - \omega_3)T}] e^{i k_R \mathbf{r} - i\omega R t}, \]  

where \( \omega_R = -\omega_1 + \omega_2 + \omega_3 \). The second term in this expression has the response function with the second time argument of \(-T\) and therefore survives only at \(T = 0\). Considering the uncertainty associated with the impulsive limit, we disregard this single point contribution. However, the corresponding term in equation (21) may have a significant contribution if a finite pulse width is taken into account. Finally, combining equation (18) and the first term of equation (22), we obtain the rephrasing signal in terms of the response function as follows:

\[ E_R^{(3)}(t, T, \tau) \propto \frac{i \omega R}{n(\omega R)} R^{(3)}(t, T, \tau) \hat{e}_3 \hat{e}_1 E_3 E_2 e^{i\omega_3t} e^{-i\omega_3 \tau} e^{i(\omega R - \omega_3)T}. \]  

The 2D spectrum is obtained by the Fourier transform of this signal with respect to \(t\) and \(\tau\):

\[ S_R^{(3)}(\omega, \omega_3) = \int \int \mathrm{d}t \mathrm{d}\tau E_R^{(3)}(t, T, \tau) e^{i\omega t + i\omega_3 \tau}. \]  

By assuming that the three carrier frequencies \(\omega_i\) are the same, the rephrasing 2D spectrum can be expressed as

\[ S_R^{(3)}(\omega, \omega_3) \propto i \int \int \mathrm{d}t \mathrm{d}\tau R^{(3)}(t, T, \tau) \hat{e}_3 \hat{e}_1 E_3 E_2 e^{i(\omega_1 + \omega_2)T} e^{i(\omega_1 - \omega_3)\tau}, \]  

where \(\omega = \omega_R \approx \omega_0\). Similarly, the non-rephasing spectrum with \(k_{NR} = k_1 - k_2 + k_3\) can be derived from equation (20) as follows:

\[ S_{NR}^{(3)}(\omega, \omega_3) \propto i \int \int \mathrm{d}t \mathrm{d}\tau R^{(3)}(t, T, \tau) \hat{e}_3 \hat{e}_1 E_3 E_2 e^{i(\omega_1 + \omega_2)T} e^{i(\omega_1 + \omega_3)\tau}. \]  

Equations (25) and (26) show that the rephasing and non-rephasing signals can be obtained as the 2D Fourier transforms of the response function with different frequency factors. These expressions cannot be used directly, however, since the result depends on the pulse carrier frequency \(\omega\). Noting that the carrier frequency \(\omega\) is experimentally chosen to be close to the resonance frequency of the mode of interest, we proceed by defining the simple 2D Fourier transform of the signal as

\[ F(\omega, T, \omega_3) \equiv \int \int \mathrm{d}t \mathrm{d}\tau R^{(3)}(t, T, \tau) \hat{e}_3 \hat{e}_1 e^{i\omega_3t} e^{i\omega_3 \tau}. \]  

Then, equations (25) and (26) suggest that a large signal will result when \(\omega_3 \approx -\omega\) and \(\omega_3 \approx \pm \omega\) in the rephasing and non-rephasing spectra. Thus, we choose the rephasing signal as \(F(\omega_3, T, \omega_3)\) in the second quadrant of the \((\omega_3, \omega_3)\) plane and the non-rephasing signal in the third quadrant:

\[ S_R(\omega, T, \omega_3) \approx \text{Im} [F(-\omega, T, \omega_3)], \]  

\[ S_{NR}(\omega, T, \omega_3) \approx \text{Im} [F(-\omega, T, -\omega_3)]. \]  

Here, the imaginary parts are used since they correspond to the real part of the experimentally determined signal (note the extra factor of \(i\) in equation (18)). Finally, the absorptive 2D IR spectrum is obtained as their sum

\[ S_A(\omega, T, \omega_3) = S_R(\omega, T, \omega_3) + S_{NR}(\omega, T, \omega_3). \]
Figure 2. (a) Dipole–dipole TCF and (b) IR spectrum of the solvated CO molecule. Results from the classical force field (dashed line) and the QM/MM force field (solid) are shown. The displayed dipole–dipole TCF is the high-frequency components of the raw TCF (inset), obtained by removing frequency components lower than 1800 cm\(^{-1}\). The system temperature is about 300 K in both cases. Note that the amplitude of the IR spectrum of the classical CO system is multiplied by 50 for easier comparison with the QM/MM result.

In the following, the 2D spectra calculated by equations (28) and (29) are presented. We will only consider the all-parallel (zzzz) configuration of the input and signal electric fields. Although 1000–2000 initial conditions are expected to be sufficient in sampling different molecular orientations in the laboratory frame of reference, we have further taken the rotational average of \( R^{(3)}_{zzzz} \) to eliminate any potential bias. Unless otherwise noted, the last 0.05 ps part in each time variable of the response function was scaled down smoothly to zero with a Gaussian function prior to the Fourier transformation. We have also used a few different apodization methods and they will be discussed below.

3. Linear and 2D IR spectra of CO in water

We have calculated the linear and 2D IR spectra of the CO/water cluster system, employing both the full classical and the PM3/MM force fields to investigate the principal effects on the calculated spectra from the differences in the force fields used. The linear IR spectrum was obtained by taking the Fourier transform of the dipole–dipole TCF [54, 55] of the solute. Figure 2(a) shows the dipole–dipole TCF obtained with the two force fields. To examine
the high-frequency behavior of the TCF, we have removed the frequency components below 1800 cm\(^{-1}\) of the raw TCF with the Fourier transform method. The figure shows a decaying oscillation of the TCF, which appears to be similar between the classical and the QM/MM results. Turning to the raw TCF in the inset of figure 2, since the atomic partial charges of the classical force field were adjusted to reproduce the dipole moment of the QM/MM system, the two TCFs are very close to each other at time zero. However, the TCF with the classical force field decays slightly faster than that with the QM/MM force field and it even shows underdamped oscillation. The QM/MM IR spectrum in figure 2(b) reflects the high-frequency oscillation of the TCF as the main peak at 2316 cm\(^{-1}\), with a shoulder at about 2270 cm\(^{-1}\). In the classical spectrum, there appear two peaks at 2276 and 2328 cm\(^{-1}\). The fact that this splitting pattern is notably visible in the classical system, together with the underdamped oscillation of the classical TCF, indicates that the classical CO interacts very weakly with water compared to the quantum mechanical CO, even though their average dipole moments are very close to each other. We also note that the amplitude of the classical spectrum is more than 50 times weaker than the QM/MM one, which can be attributed to the lack of charge fluctuation or polarization effect in the former.

The 2D IR spectra of the CO/water system described with the classical and QM/MM force fields are displayed side-by-side in figures 3 and 4. Figure 3 shows the rephasing and non-rephasing 2D spectra obtained from each of the two terms of the response function (cf equation (6)). As expected, the rephasing signals overall exhibit a diagonal elongation of the main peak and the non-rephasing signals exhibit an anti-diagonal elongation [1, 11]. In the QM/MM spectra on the left column of figure 3, each signal consists of a single major peak and two very weak side peaks of opposite sign. Terms 1 and 2 of the response function in equation (7) produce almost equal and opposite peaks, with amplitudes of about \( \pm 0.025 \) and \( \pm 0.02 \) for the rephasing (figures 3(a) and (b)) and non-rephasing signals (figures 3(c) and (d)), respectively. Therefore, when they are combined, the total rephasing or non-rephasing signal will have a much smaller amplitude of \( \pm 0.003 \) to \( 0.004 \), as can be seen in the left column of figures 4(a) and (b). In an isolated harmonic oscillator, this cancelation would be perfect and no 2D signal should be observed. To obtain a reliable 2D signal, therefore, one needs to accurately calculate the small difference between the two large contributions. In figure 3, the spectra obtained with the classical force field appear to be more complicated (noisier) than their QM/MM counterparts. In particular, the non-rephasing signals on the right column of figures 3(c) and (d) show multiple peaks with both positive and negative amplitudes, unlike the QM/MM signals exhibiting predominantly a single peak. In addition, the maximum amplitudes of the classical signals are about three orders of magnitude smaller than the QM/MM ones. This roughly corresponds to the square of the amplitude difference observed in the linear spectra in figure 2(b) and is consistent with the \( \mu^2 \)- and \( \mu^4 \)-dependence of the linear and third-order responses, respectively. These seemingly weak and inconsistent signals show the inadequacy of the classical force field in describing the third-order intramolecular vibrational response. It should be noted that the C–O bond described by the present classical force field does contain a non-negligible anharmonic character due to the intermolecular interactions. However, figure 3

4 This frequency splitting of the CO stretch band is an artifact caused by the spherical capping potential that was used to keep the cluster in a spherical region. While this might be partly responsible for the complicated features observed in the 2D IR spectrum of the classical CO molecule in figures 3 and 4, it cannot be the cause of their apparent inconsistency. Moreover, this splitting is not pronounced in the QM/MM case and therefore does not affect the 2D spectra of the QM/MM systems appreciably. Thus, it does not compromise our major conclusions.
Figure 3. Individual terms of rephasing and non-rephasing spectra of CO. The results obtained with the QM/MM force field (left column) and the classical force field (right) are shown. (a) Rephasing signal obtained from term 1 of the response function. (b) Rephasing signal from term 2. (c) Non-rephasing signal from term 1. (d) Non-rephasing signal from term 2.
Figure 4. Rephasing, non-rephasing and absorptive spectra of CO. The results obtained with the QM/MM force field (left column) and the classical force field (right) are shown. (a) Rephasing signal. (b) Non-rephasing signal. (c) Absorptive spectrum.

shows that this solvent-induced anharmonicity fails to quantitatively produce a coherent 2D signal of intramolecular mode.

Figure 4 displays the total 2D spectra obtained as the sum of the corresponding two components in figure 3. Hereafter, we shall focus on the QM/MM spectra shown on the left column of figure 4. First, when the main diagonal peaks of individual rephasing signals in figures 3(a) and (b) are added together, they interfere with each other and produce a doublet in the total rephasing spectrum in figure 4(a). Similarly, the sum of the two individual
non-rephasing signals in figures 3(c) and (d) gives the total non-rephasing spectrum plotted in figure 4(b). This (positive–negative peak) doublet indicates that the signal peaks of the individual components (two terms in equation (7)) are slightly off-centered from each other. The peak separation between terms 1 and 2 is about 10 cm\(^{-1}\) in figure 3 and it produces a doublet splitting of about 50 cm\(^{-1}\) in figure 4, mainly in the vertical (\(\omega_t\)) direction. The absorptive 2D spectrum in figure 4(c), obtained as the sum of the two spectra in figures 3(a) and (b), shows this vertical splitting more clearly. A similar splitting has previously been observed in the classically calculated 2D IR spectrum of intermolecular modes of liquid water [34]. This vertical splitting of the diagonal peak is a well-known feature of 2D IR spectroscopy. Quantum mechanically, it originates from the frequency difference between the fundamental (0 → 1) transition and the overtone (1 → 2) transition and therefore is directly related to the anharmonicity of the mode [1]. It is remarkable that the same feature is observed in the present spectrum that is based on the classical limit of the response function. It is beyond the scope of this study to analyze the classical third-order response function and identify processes responsible for these phenomena. However, the comparative study of the QM/MM and all-classical CO/water systems presented here reveals that a realistic description of vibrational mode anharmonicity is essential to the simulation of 2D IR spectroscopy and the QM/MM force field can meet such a requirement.

4. Simulated spectra of \(N\)-methylacetamide in water

4.1. Dipolar linear response

We first study the solvation effect in the d-NMA/water system by comparing the QM/MM MD simulation results for an isolated d-NMA and a solvated d-NMA system. At 298 K, the isolated d-NMA described by the semiempirical PM3 force field was found to have an average dipole moment of 2.77D. When this solute was placed in a cluster of 16 classical water molecules, it exhibited significant polarization with an average dipole moment of 3.67D. Table 1 shows the time-averaged partial charges of the d-NMA atomic sites obtained from this QM/MM MD simulation and their changes from the values for the isolated d-NMA. This table shows that carbonyl oxygen has the largest polarization of \(-0.089e\), where \(e\) is the elementary charge, followed by N and C atoms. The magnitude of atomic charge fluctuation can be judged by the standard deviation of the fluctuating partial charges and it is in the order O, C and N. It is interesting to note that the nitrogen (N) atomic charge decreases by 0.032\(e\) (from \(-0.054\) to \(-0.022e\)) upon solvation. This indicates that the nitrogen site does not interact directly with water and its polarization is largely induced by the polarization changes of the other sites, especially oxygen. The deuterium (D) site shows a small polarization of 0.010\(e\), but its fluctuation is comparatively large with a standard deviation of 0.015\(e\). In Table 2, the charge correlations between different sites as measured by the quantity \(\langle \Delta c_i \Delta c_j \rangle\) are presented. According to this table, the most polarizable oxygen site appears to be negatively correlated with all other sites, with the O–N cross-correlation being the largest at \(-8.5\), which indicates a resonance structure formation. Unexpectedly, the O–C correlation of \(-1.6\) is smaller than the O–Me1 (\(-2.5\)) or O–Me2 (\(-1.8\)) correlations. Thus, N is by far the largest source of electrons for the oxygen polarization, although the other sites also contribute significantly, indicating a large degree of charge delocalization. The carbonyl carbon (C) charge correlates more strongly with that of the N atom than with O. This is presumably related to the well-known partial double bond character of the peptide C–N bond [56].
Table 1. Average partial charges of d-NMA atomic sites and their standard deviations from QM/MM MD. Polarization is the difference in partial charges from the values for an isolated d-NMA. Me1 and Me2 are methyl groups connected to C and N, respectively, and their charges are the sum of the partial charges of constituent atoms.

| Site | Average Charge (e) | Polarization | Standard deviation |
|------|--------------------|--------------|--------------------|
| Me1  | 0.065              | 0.013        | 0.017              |
| C    | 0.265              | 0.027        | 0.023              |
| O    | -0.467             | -0.089       | 0.039              |
| N    | -0.022             | 0.032        | 0.036              |
| D    | 0.084              | 0.010        | 0.015              |
| Me2  | 0.074              | 0.006        | 0.018              |

Table 2. Correlation between site charges of d-NMA from QM/MM MD. Tabulated are the averaged product of charge fluctuations $\langle \Delta c_i \Delta c_j \rangle$ (in $10^{-4}e^2$) for all site pairs. The square root of the diagonal entries corresponds to the standard deviation of the site charges shown in table 1.

| Site | Me1 | C  | O   | N   | D     | Me2 |
|------|-----|----|-----|-----|-------|-----|
| Me1  | 3.0 |    |     |     |       |     |
| C    | -0.76 | 5.1 |     |     |       |     |
| O    | -2.5 | -1.6 | 15.0 |     |       |     |
| N    | 0.46 | -2.9 | -8.5 | 13.0 |       |     |
| D    | -0.42 | 0.28 | -0.97 | -0.68 | 2.2 |     |
| Me2  | 0.26 | -0.21 | -1.8 | -1.1 | -0.38 | 3.3 |

We now turn to the dipolar linear response of the d-NMA system. Figure 5(a) compares the dipole–dipole TCF of d-NMA in the isolated and solvated states. The figure presents the high-frequency components of the TCF (main plot) obtained by high-pass filtering of the raw TCF data (inset). Near time zero, the two raw TCFs show a large difference, reflecting the different dipole moments mentioned above. Also, the fast initial decay until about 0.3 ps found in the solvated system TCF is absent in the isolated case and is therefore attributable to the collective motions of the solute on a sub-picosecond time scale, such as hindered rotation. However, the decay patterns of the two TCFs at a longer time are similar to each other, indicating little effect of the solvent on the dipole relaxation on picosecond and longer time scales. This can be related to the apparent absence of the inhomogeneous broadening effect in the 2D IR spectrum discussed below. The high-frequency behavior of the TCF can be seen in the main plot of figure 5(a), which displays the two TCFs after high-pass filtering at 1000 cm$^{-1}$ of the original data. It clearly shows the oscillations of the two TCFs due to intramolecular vibrations. After a Fourier transformation of the original TCF, we obtained the IR spectra in figure 5(b). The calculated IR spectra in this figure exhibit four peaks for both the isolated and the solvated d-NMAs.
Figure 5. (a) Dipole–dipole TCF and (b) IR spectrum of d-NMA described by the QM force field. Results for isolated d-NMA (solid line) and d-NMA solvated in a 16-water cluster (dashed line) are shown. The displayed dipole–dipole TCF is the high-frequency components of the raw TCFs (inset), obtained by removing frequency components lower than 1000 cm\(^{-1}\). The system temperature is about 300 K in both cases. The amplitudes of the two IR spectra are normalized at the amide I peak (1904 and 1871 cm\(^{-1}\) for isolated and solvated d-NMA, respectively).

Following our previous study [55], we have also calculated the power spectra of several NMA internal coordinates (data not shown) and compared them with the IR spectrum in figure 5(b) to identify internal modes responsible for each IR peak. This allowed us to assign the two high-frequency peaks as follows: (i) amide I’ (1904/1871 cm\(^{-1}\) in the isolated/solvated system): CO stretch, CNC bending; (ii) amide II’ (1440/1463 cm\(^{-1}\)): symmetric CN and NC(Me2) stretch, all bendings except CNC. The other two lower frequency peaks at 1375/1374 and 1333/1342 cm\(^{-1}\) originate from many backbone modes and therefore can be considered fairly delocalized. Note that existing studies on the amide II’ mode of d-NMA [46, 57] report the contribution of methyl deformation as well as CN stretch and ND bend to this mode. This difference presumably comes from the limitations of the semiempirical PM3 force field used in this study and can be improved in the future by employing a higher-level QM theory. Nevertheless, the present PM3 theory correctly predicts the frequency red shift of the amide I’ mode (1904 \(\rightarrow\) 1871 cm\(^{-1}\)) and the blue shift of the amide II’ mode (1440 \(\rightarrow\) 1463 cm\(^{-1}\)) on solvation of the d-NMA by water.
4.2. Third-order vibrational response function

The third-order vibrational response function was calculated for the d-NMA/water system from the PM3/MM MD simulations, and the result at zero waiting time ($t_2 = 0$) is shown in figure 6. First, figure 6(a) shows the response function $R^{(3)}_{zzzz}(t_3, 0, t_1)$ along the diagonal direction ($t_1 = t_3$). The two terms of the nonlinear response function in equation (7) have almost equal magnitude and opposite sign, and the total response function thus becomes much smaller as a result of large cancelation, destructive interference between the two echo fields. This has
already been observed in the 2D IR spectra of the CO/water system in figures 3 and 4, where the total rephasing and non-rephasing signals have much smaller amplitudes than their individual components. In the present case, the two individual terms of the response function tend to decay over the 0.5 ps time span. However, their cancelation at a long time is not perfect and the resultant total response function exhibits a diverging pattern, as shown in figure 6(b). This indicates that the response function is greatly affected by the diverging stability matrix after about 0.3 ps and the statistical averaging becomes rapidly inefficient at such a longer time. The behaviors of the response function along the $t_1$ and $t_3$ axes are quite different. According to figures 6(c) and (d), the response function tends to decay along the $t_1$ axis, while it shows a rapid divergence near 0.5 ps along the $t_3$-axis. This is reminiscent of the divergence predicted in the classical nonlinear response functions of a single anharmonic oscillator [28] that takes place along a specific direction in time arguments. However, the dynamics of the present d-NMA/water system is expected to be chaotic and the divergence observed here is likely due to imperfect sampling. With regard to this sampling issue, we note that the present computation requires CPU time in the order of $10^4$–$10^5$ h and, as figure 6 shows, it can produce a reasonably well-behaved response function up to 0.3–0.5 ps in each time argument. Improving upon this level of convergence would require significantly larger computational resources and may not be practical.

Despite the apparent divergence of the response function at long times, the 2D IR spectra obtained from it exhibit notable signals corresponding to the intramolecular vibrational modes of d-NMA. In figures 7(a) and (b), the rephasing and non-rephasing spectra from the two terms of the response function in equation (7) are displayed. Quite similar to the case of the CO/water system, term 1 of the response function produces predominantly positive peaks in the rephasing and non-rephasing signals with diagonal and anti-diagonal elongations, respectively, while term 2 produces very similar peaks, but with opposite sign. Each rephasing/non-rephasing spectrum consists of two diagonal peaks corresponding to the amide I’ and II’ modes, as well as two cross peaks at the off-diagonal positions representing finite coupling between the two modes. The total rephasing/non-rephasing spectrum in figure 7(c) shows a vertical splitting of the peaks. This indicates once again a certain mismatch in the peak positions between terms 1 and 2 spectra. The vertical splittings of amide I’ and II’ signals in figure 7(c) are similar at about 40 cm$^{-1}$. The two low-frequency peaks in the 1300 cm$^{-1}$ region observed in the linear spectrum of figure 5(b) are, however, not clearly resolved in the 2D spectrum due to their small amplitudes.

The non-rephasing spectrum presented in figure 7 was obtained by 2D Fourier transformation according to equation (28), which, in turn, is inferred from equation (26). However, by simply taking equation (28) as the non-rephasing signal, an unintended contribution from the phase-matched signal with $k_{III} = k_1 + k_2 - k_3$ is included since it results in the same frequency factor as the non-rephasing signal with $k_{NR} = k_1 - k_2 + k_3$ does in equation (26). The $k_{III}$ signal can be shown to contain an extra factor $\exp[i(\omega_2 + \omega_1)T]$ and therefore is expected to be a highly oscillating function of $T$. This has previously been recognized by Saito and co-workers [34], and their solution was to remove the high-frequency components in $T$ from the frequency domain signal $S_{NR}(\omega_1, T, \omega_\tau)$ in equation (28) by low-pass frequency filtering. We have taken a similar approach by applying the low-pass filtering in $T$ to the time domain response function $R^{(3)}(t, T, \tau)$ prior to the 2D Fourier transformation in $t$ and $\tau$. The resulting 2D spectra did not show any notable difference from the original and therefore is not presented here. However, this procedure may become important when studying...
Figure 7. Rephasing and non-rephasing spectra of d-NMA. The rephasing signal (left column) and the non-rephasing signal (right) are shown for the two terms of the response function and the total. (a) Contribution from term 1. (b) Contribution from term 2. (c) Sum of terms 1 and 2.

the $T$ dependence of the 2D IR spectrum, which is not the primary focus here due to the limited inhomogeneous broadening effect observed in the present cluster system (see below).

4.3. Apodization functions

The computation of the 2D IR spectrum from the time-domain vibrational response function requires a 2D discrete Fourier transformation. Since the response function obtained from the present method is greatly affected by the diverging stability matrix at long times, it is necessary to discard the data after a certain cutoff time to obtain a reliable spectrum. This would cause the artificial side peaks in the calculated spectrum that makes the interpretation difficult, as often observed in NMR spectroscopy when processing the free induction decay signal [2].
Figure 8. Effect of apodization on the rephasing and non-rephasing spectra of d-NMA. The rephasing signal (left column) and the non-rephasing signal (right) were obtained with different window functions. (a) Cosine window. (b) Kaiser window with $\theta = \pi/2$. (c) Kaiser window with $\theta = \pi$.

In the spectra presented above, the data up to 0.45 ps in each time variable were used unmodified and the remaining 0.05 ps portion was smoothly scaled down with a Gaussian function to minimize the truncation error. However, there are other apodization methods that provide a better resolution of the spectrum [2]. We have applied two kinds of well-known window functions to the 2D IR spectrum calculation. The first is the cosine window of the form $h(t) = \cos[\pi t/(2t_{\text{max}})]$, and the second is the Kaiser window involving the modified Bessel function of the first kind $I_0(t)$, $h(t) = I_0(\theta \sqrt{1 - (t/t_{\text{max}})^2})/I_0(\theta)$. Figure 8 shows the rephasing and non-rephasing spectra obtained with apodization by the cosine window and the Kaiser window with $\theta = \pi/2$ and $\pi$. In general, apodization suppresses the rapidly oscillating side peaks at the cost of broadening the main peak. Comparing the apodized spectra in figure 8 with the original ones in figure 7(c), we can see that the cosine window produces broader main peaks with even worse side peak oscillation and the Kaiser window with $\theta = \pi/2$ produces spectra of
similar quality to the original one. Significant improvement could be found only with the Kaiser window with \( \theta = \pi \), even though it makes the main peaks slightly broader. A similar tendency can be observed in the apodized absorptive 2D spectra in figure 9, where the Kaiser window with \( \theta = \pi \) provides the smallest side peak features without too much peak broadening.

4.4. Absorptive 2D IR spectrum

The absorptive 2D IR spectrum is obtained as the sum of the rephasing and non-rephasing signals and is often preferred in the literature, since the elongated dispersive tails in the latter two spectra are eliminated after summation and the signal becomes narrow and sharp [58]. The absorptive 2D spectrum of the d-NMA system was already introduced in figure 9 for the apodized response function and figure 10 provides the comparison between the original and apodized spectra. Also in figure 10, the linear spectrum is inserted on top of the 2D spectra for comparison.

As expected, we observe that the elongation of the four major peaks found in the rephasing/non-rephasing signals largely disappears in the absorptive 2D spectrum and the individual peak appears to be narrow and symmetric. The degree of vertical splitting is found to be about the same (~40 cm\(^{-1}\)) between the absorptive and the rephasing/non-rephasing spectra. While the absorptive spectrum exhibits distinctive amide I’ and II’ peaks consistent with the peak positions in the linear spectrum, the lower-frequency peaks in the linear spectrum are not well resolved. This can be better judged in figure 11, which presents the magnified view of the diagonal and lower-right cross-peak regions from figure 10(a). In figure 11(a), we can identify a cross peak between the amide II’ mode and the two lower-frequency modes on the upper-left corner, but the other features in the 1300–1400 cm\(^{-1}\) range are somewhat obscure. This is also the case in the cross-peak region between the amide I’ mode and the three low-frequency modes shown in figure 11(b).

The present results can be directly compared with the experimental 2D IR spectrum of the amide I’ and II’ modes of d-NMA reported by Tokmakoff and co-workers (figure 6 of [46]). We first note that the experimental d-NMA linear spectrum is different from the present result in
Figure 10. Absorptive 2D IR spectrum of d-NMA enhanced by apodization. (a) Original spectrum (Gaussian scaling on the tail part). (b) Spectrum enhanced with a Kaiser window with $\theta = \pi$. For comparison, the linear spectrum is also displayed above the 2D spectra.

Figure 11. Magnified views of (a) the amide II/III diagonal peak and (b) the amide I–amide II cross-peak regions in the absorptive 2D IR spectrum of d-NMA. The original spectrum is from figure 10(a).

the peak positions and relative frequency separation of amide I’ and II’ modes and the presence of a doublet in the amide II’ mode. These differences are expected to be dependent on the force field and therefore can be overcome by using QM theory of higher levels in the future. Apart from these, the present 2D IR spectrum agrees fairly well with the experimental spectra in terms of the vertical splitting patterns and the presence of cross peaks. The most notable difference is found in the elongation and tilt angle of the doublet peaks. The amide I’ peak in the
experimental spectrum shows a diagonal elongation and the nodal line of the doublet makes a finite angle with the horizontal line. In contrast, the present absorptive spectrum does not exhibit diagonal elongation and the nodal line of the doublet is almost horizontal. This indicates that the present d-NMA/water cluster system has a weaker inhomogeneous broadening effect than the bulk solution [38, 43]. This limitation should be overcome in future studies by properly taking into account the periodic boundary condition—note that the current state-of-the-art simulation methods are not quite reliable in this respect.

5. Summary and a few concluding remarks

We have presented a new computational scheme to calculate the third-order vibrational response function in the classical limit directly from the classical MD simulations employing the QM/MM force fields. A theoretical connection between response function and 2D IR spectroscopy has also been provided before. The method was then applied to the 2D vibrational spectroscopy of CO and d-NMA molecules, each solvated in a water cluster. The present calculation studies showed that the classical force field cannot provide a reliable 2D IR spectrum of intramolecular vibrations due to the harmonic nature of bond vibrations. The QM/MM force field, on the other hand, captures the anharmonicity of intramolecular vibrational modes and produces the 2D IR spectrum in good agreement with experiment. The method is particularly advantageous in that it provides the entire spectrum of the system vibrational modes, including the diagonal peaks due to individual modes and off-diagonal cross peaks reflecting inter-mode interactions, which is because the solute vibrations are determined by intrinsically anharmonic multidimensional QM potential function as well as by solute–solvent interaction potentials.

Limiting factors of the accuracy of the present method include the small size of the solvent system, the semiempirical QM theory employed and the fully classical nature of the response function. The 2D IR spectra presented in this study do not exhibit diagonal elongation of peaks, indicating that the inhomogeneous broadening effect may not be correctly reflected in the calculation. This can be attributed mainly to the fast solvent fluctuation expected in the cluster system. To reproduce a realistic solvent environment including slow dielectric responses, it would be necessary to use a larger solvent system. In this regard, an accurate implementation of the QM/MM MD simulation method with a periodic boundary condition would be instrumental in the future. The present method requires a significant amount of computation for the convergence of the response function. The number of trajectories needed for the stability matrix calculation scales linearly with the system degrees of freedom, and we expect the method to be practical for systems with up to a few hundred atoms. Therefore, algorithmic improvements would be required to deal with large biomolecules. It is also desirable to employ higher-level ab initio/DFT theories in the QM description. For example, the Hartree–Fock/MM MD method has been employed in our group for the computation of the vibrational circular dichroism (VCD) spectra of small peptides [59]. We are currently exploring the density functional-based tight binding (DFTB) method [60] for the 2D IR spectrum computation. Finally, on the theoretical side, a better understanding of the nature of classical nonlinear response functions [29, 61, 62] and possible semiclassical extensions [63, 64] would be an important subject of future study.

In conclusion, the present computational method combined with high-level QM force fields can be expected to provide a realistic 2D IR spectrum of vibrational chromophores in solution, including small organic molecules, peptides and, with improvement in computational efficiency, even polypeptides and proteins.
Acknowledgments

We thank Dr S Saito for helpful discussions. This work was supported by an NRF grant (no. 20090078897) and a KBSI grant (T30401) to MC.

References

[1] Cho M 2009 Two-Dimensional Optical Spectroscopy (Boca Raton: CRC Press)
[2] Ernst R R, Bodenhausen G and Wokaun A 1987 Nuclear Magnetic Resonance in One and Two Dimensions (Oxford: Oxford University Press)
[3] Wüthrich K 1986 NMR of Proteins and Nucleic Acids (New York: Wiley)
[4] Hamm P, Lim M and Hochstrasser R M 1998 J. Phys. Chem. B 102 6123–38
[5] Tanimura Y and Mukamel S 1993 J. Chem. Phys. 99 9496–511
[6] Park K and Cho M 1998 J. Chem. Phys. 109 10559–69
[7] Cho M 1999 Advances in Multi-Photon Processes and Spectroscopy ed S H Lin, A A Villaeys and Y Fujimura (Singapore: World Scientific Publishing) pp 229–300
[8] Mukamel S 2000 Annu. Rev. Phys. Chem. 51 691–729
[9] Cho M 2002 PhysChemComm 5 40–58
[10] Jonas D M 2003 Annu. Rev. Phys. Chem. 54 425–63
[11] Khalil M, Demirdöven N and Tokmakoff A 2003 J. Phys. Chem. A 107 5258–79
[12] Cho M 2008 Chem. Rev. 108 1331–418
[13] Zheng J, Kwak K and Fayer M D 2007 Acc. Chem. Res. 40 75–83
[14] Ganim Z, Chung H S, Smith A W, DeFlores L P, Jones K C and Tokmakoff A 2008 Acc. Chem. Res. 41 432–41
[15] Jansen T I C, Steinel T, Stromberg C, Corcelli S A, Lawrence C P, Skinner J L and Fayer M D 2004 J. Phys. Chem. A 108 1107–19
[16] Leegwater J A and Mukamel S 1995 J. Chem. Phys. 102 2365–71
[17] Mukamel S, Khidekel V and Chernyak V 1996 Phys. Rev. E 53 R1–4
[18] Saito S and Ohmine I 1998 J. Chem. Phys. 108 240–51
[19] Ma A and Stratt R M 2000 Phys. Rev. Lett. 85 1004–7
[20] Jansen T I C, Snijders J G and Duppen K 2000 J. Chem. Phys. 113 307–11
[21] Saito S and Ohmine I 2002 Phys. Rev. Lett. 88 207401
[22] Dellago C and Mukamel S 2003 Phys. Rev. E 67 035205
[23] Dellago C and Mukamel S 2003 J. Chem. Phys. 119 9344–54
[24] Noid W G, Ezra G S and Loring R F 2004 J. Phys. Chem. B 108 6536–43
[25] Kryvohuz M and Cao J 2006 Phys. Rev. Lett. 96 030403
[26] Malinin S V and Chernyak V Y 2008 Phys. Rev. E 77 025201
[27] Malinin S V and Chernyak V Y 2008 Phys. Rev. E 77 056201
[28] Ma A and Stratt R M 2002 J. Chem. Phys. 116 4962–71
[29] Saito S and Ohmine I 2003 J. Chem. Phys. 119 9073–87
[30] Nagata Y, Tanimura Y and Mukamel S 2007 J. Chem. Phys. 126 204703
[31] Yagasaki T and Saito S 2008 J. Chem. Phys. 128 154521
[32] Yagasaki T and Saito S 2009 Acc. Chem. Res. 42 1250–8
[33] Gao J 1996 Reviews in Computational Chemistry vol 7, ed K B Lipkowitz and D B Boyd (New York: VCH) pp 119–85
[34] Stewart J J P 1989 J. Comput. Chem. 10 209–20

New Journal of Physics 12 (2010) 065001 (http://www.njp.org/)
[38] Kwac K and Cho M 2003 J. Chem. Phys. 119 2256–63
[39] Kwac K, Lee H and Cho M 2004 J. Chem. Phys. 120 1477–90
[40] Dijkstra A G, Jansen T I C, Bloem R and Knoester J 2007 J. Chem. Phys. 127 194505
[41] Bloem R, Dijkstra A G, Jansen T I C and Knoester J 2008 J. Chem. Phys. 129 055101
[42] Zanni M T, Asplund M C and Hochstrasser R M 2001 J. Chem. Phys. 114 4579–90
[43] Woutersen S, Pfister R, Hamm P, Mu Y, Kosov D S and Stock G 2002 J. Chem. Phys. 117 6833–40
[44] Rubtsov I V, Wang J and Hochstrasser R M 2003 Proc. Nat. Acad. Sci. USA 100 5601–6
[45] DeCamp M F, DeFlores L, McCracken J M, Tokmakoff A, Kwac K and Cho M 2005 J. Phys. Chem. B 109 11016–26
[46] DeFlores L P, Ganim Z, Ackley S F, Chung H S and Tokmakoff A 2006 J. Phys. Chem. B 110 18973–80
[47] Goldstein H 1980 Classical Mechanics (Reading, MA: Addison-Wesley)
[48] Williams R B and Loring R F 2000 J. Chem. Phys. 113 1932–41
[49] Allen M P and Tildesley D J 1987 Comput. Simul. Liquids (Oxford: Clarendon)
[50] Garashchuk S and Light J C 2000 J. Chem. Phys. 113 9390–2
[51] Case D A et al 2006 AMBER 9 (San Francisco: University of California)
[52] Wu Y, Tepper H L and Voth G A 2006 J. Chem. Phys. 124 024503
[53] Mulliken R S 1955 J. Chem. Phys. 23 1833–40
[54] McQuarrie D A 1976 Statistical Mechanics (New York: Harper and Row)
[55] Yang S and Cho M 2005 J. Chem. Phys. 123 134503
[56] Besley N A 2004 J. Phys. Chem. A 108 10794–800
[57] Myshakina N S, Ahmed Z and Asher S A 2008 J. Phys. Chem. B 112 11873–7
[58] Khalil M, Demirdöven N and Tokmakoff A 2003 Phys. Rev. Lett. 90 047401
[59] Yang S and Cho M 2009 J. Chem. Phys. 131 135102
[60] Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S and Seifert G 1998 Phys. Rev. B 58 7260–8
[61] Noid W G and Loring R F 2004 J. Chem. Phys. 121 7057–69
[62] Kryvohuz M and Cao J 2005 Phys. Rev. Lett. 95 180405
[63] Noid W G, Ezra G S and Loring R F 2004 J. Chem. Phys. 120 1491–9
[64] Gruenbaum S M and Loring R F 2008 J. Chem. Phys. 129 124510

New Journal of Physics 12 (2010) 065001 (http://www.njp.org/)