PCCP

View Online / Journal Homepage
Dynamic Article Links

Cite this: DOI: 10.1039/c2cp42593a

www.rsc.org/pccp

PAPER

Geometry determination of complexes in a molecular liquid mixture using electron–vibration–vibration two-dimensional infrared spectroscopy with a vibrational transition density cube method

Rui Guo,[a] Shaul Mukamel[b] and David R. Klug*[a]

Received 27th July 2012, Accepted 5th September 2012
DOI: 10.1039/c2cp42593a

We demonstrate the use of a new vibrational transition density cube (VTDC) method for determining the geometry of complexes in a molecular liquid mixture from electron–vibration–vibration two-dimensional infrared (EVV 2DIR) spectra. The VTDC method was used to calculate the electrically-mediated intermolecular vibrational coupling and thereby the EVV 2DIR spectra. Using the 1 : 1 benzonitrile–phenylacetylene (BN–PA) liquid mixture as a test case, the new method leads to a distance of 3.60 Å between the interacting BN–PA pair, a much more accurate value than the distance previously obtained using a dipolar approximation for the electrical coupling. We also show that molecular dynamics simulations of the liquid mixture predict a modal geometry of complexation which agrees well with the geometry determined from the 2DIR data *via* VTDC analysis. We therefore conclude the combination of VTDC and EVV 2DIR data is a useful approach for the determination of the geometry of molecular complexes in the condensed phase.

I. Introduction

Two-Dimensional Infrared (2DIR) Spectroscopy is loosely analogous to Two-Dimensional Nuclear Magnetic Resonance Spectroscopy (2DNMR). 2DIR measures vibration–vibration coupling whereas 2DNMR measures spin–spin coupling. This analogy has raised the hope that it may be possible to solve molecular structures from 2DIR data as can be done using 2DNMR and that 2DIR may have some complementary advantages due to different sensitivities and coupling physics.[1]

A particular variant of 2DIR spectroscopy, known as EVV 2DIR has been in development for a number of years.[2–8] It has been applied to the study of biomolecular systems[4–6] particularly through its flexibility in terms of accessing coupled vibrations across the infrared and near-infrared spectrum and its resilience in the face of sample scatter. More relevant to this research is that we have previously shown that EVV 2DIR can be used to detect pure electrical coupling between molecular vibrations.[3,6,7] Through-space electrically-mediated coupling allows vibrations to interact even in the absence of a chemical bond.[7] This makes EVV 2DIR particularly suitable for the detection of the formation of molecular complexes in biological systems.[7] Non-covalent molecular complexation is the mainstay of biological function and regulation. Through numerous weak interactions, proteins bind to other proteins, nucleic acids, lipids, drugs and metabolites,[8] thus the capability to detect such molecular complexes is highly desirable.

Even more valuable is the capability to determine binding geometries of complexes. In a previous paper[7] we demonstrated that in principle it is possible to determine geometries of weakly coupled complexes with EVV 2DIR spectroscopy. Using a simple dipole–dipole model, geometrical parameters of the benzonitrile–phenylacetylene (BN–PA) dimer in a liquid mixture were determined. However, although the angle determined with this dipolar model was fairly good, the distance obtained (2.6 Å) was obviously too short for this system. These approximate geometries, although useful, are still not good enough for many applications, such as the rational design of drugs or the correct interpretation of binding motifs. It is therefore necessary to develop more accurate theoretical tools for the recovery of molecular interaction geometry from EVV 2DIR data. In this paper we went beyond the dipolar approximation in order to determine molecular interaction geometries more accurately and developed a robust method that will be generally applicable.

The distance between the BN–PA pair in the liquid mixture used here is only around 4 Å, roughly the same size as the two interacting molecular groups themselves. This is also typical of the distances found in interacting biomolecular chemical groups. For this class of problem, electronic distributions beyond dipoles will show non-negligible contributions to the electrostatic interaction between the two molecules and dipole–dipole interaction model will be inadequate.[9] It is therefore

[a] *Institute of Chemical Biology, Department of Chemistry, Imperial College London, London SW7 2AZ, UK. E-mail: d.klug@imperial.ac.uk*
[b] *Department of Chemistry, University of California, Irvine, California 92697-2025, USA*

necessary to add higher-order multipole interactions into our previous model for improved prediction and analysis of EVV 2DIR spectra. As higher and higher orders of multipole moments were included, theoretical results can be expected to approach the full and exact electrostatic interaction. However it is known the convergence of this process will be slow. Moreover, expressions for EVV 2DIR signals with higher-order multipoles will also become more complicated and cumbersome.

In this paper we adopted a straightforward approach to calculate from first principle the complete electrostatic coupling between molecules, without using any multipole expansion. Our scheme is based on the construction of so-called vibrational transitional density cubes (VTDCs) for each molecule and calculating the complete electrostatic coupling between them by integration. This approach is similar in spirit to that used by Krueger *et al*. for studying electronic coupling in a photosynthetic light harvesting complex.[10] VTDCs of the first-order have already been used to calculate vibrational mode coupling constants in polypeptides by Moran and Mukamel,[11] but further development in interaction-induced EVV 2DIR spectra presents new challenges. Since molecular vibrations involve nuclear motions, these VTDCs have some unique features and also their specific problems, which will be addressed in the present work. Following proper theoretical development, this new approach was applied to the BN–PA system as a test case and showed it indeed leads to much improved geometry determination from EVV 2DIR data. We also carried out molecular dynamics simulations of a 1 : 1 BN–PA liquid mixture, from which a better understanding of the liquid structure was obtained through various distribution functions. These MD simulations were then combined with our new VTDC approach to provide us with a realistic description of EVV 2DIR spectroscopy.

II. Computation details

A. Vibrational transition density cubes

As in our previous paper,[7] we consider the case when two closed-shell molecules A and B are interacting with each other electrostatically without any overlapping between their electronic distributions. The spinless electron density distribution of each molecule was denoted as $P(\boldsymbol{r})$ and Z_i is the nuclear charge of component atom i, with $i = 1\ldots N$. As we have shown previously, the electrostatic interaction between A and B contributes an extra term to the total Hamiltonian of the A–B complex and leads to induced EVV 2DIR signals. Instead of using multipole expansion for this interaction as in our previous work, the complete classical electrostatic interaction between these two charge distributions can be written as:

$$H' = -\sum_{i=1}^{N_A} {Z_i}^A \int_B \frac{P_B}{|{\boldsymbol{R}_i}^A - \boldsymbol{r}_B|} d\boldsymbol{r}_B - \sum_{j=1}^{N_B} {Z_j}^B \int_A \frac{P_A}{|{\boldsymbol{R}_j}^B - \boldsymbol{r}_A|} d\boldsymbol{r}_A + \iint_{AB} \frac{P_A P_B}{|\boldsymbol{r}_A - \boldsymbol{r}_B|} d\boldsymbol{r}_A d\boldsymbol{r}_B + \sum_{ij} \frac{{Z_i}^A {Z_j}^B}{|{\boldsymbol{R}_i}^A - {\boldsymbol{R}_j}^B|}, \quad (1)$$

which includes from left to right, attractive nuclei–electron (n–e) interactions, and repulsive electron–electron (e–e), nuclei–nuclei (n–n) interactions. H' couples the two molecules and hence any two localized vibrations on them.

Assuming the influence of intermolecular interaction on monomeric Hamiltonians ${H_A}^0$ and ${H_B}^0$ of molecules A and B is negligible, total Hamiltonian of the system can be written as: $H = {H_A}^0 + {H_B}^0 + H'$. Then one can expand the potential energy surface in terms of vibrational coordinates to obtain mechanical anharmonicities.[12] However for the study of intermolecular vibrational couplings, through-space-induced mechanical anharmonicity is negligible compared to through-space-induced electrical anharmonicity, as has been pointed out in our previous paper.[7] Expanding the dipole moment of the complex in vibrational coordinates, we have:

$$\mu = \mu_0 + \sum_s \frac{\partial \mu}{\partial Q_s} Q_s + \sum_t \frac{\partial \mu}{\partial Q_t} Q_t + \frac{1}{2}\left(\sum_s \frac{\partial^2 \mu}{\partial {Q_s}^2} {Q_s}^2 + \sum_t \frac{\partial^2 \mu}{\partial {Q_t}^2} {Q_t}^2 + \sum_s \sum_t \frac{\partial^2 \mu}{\partial Q_s \partial Q_t} Q_s Q_t \right) + \ldots .$$

Here Q_s is the vibrational coordinate of mode s of molecule A, and Q_t that of mode t of molecule B. The first two of the three electrical anharmonicities, $\frac{\partial^2 \mu}{\partial {Q_s}^2}$ and $\frac{\partial^2 \mu}{\partial {Q_t}^2}$, only contribute to intra-molecular EVV 2DIR signals of molecules A and B, while the last one $\frac{\partial^2 \mu}{\partial Q_s \partial Q_t}$ is fully responsible for the generation of inter-molecular vibrational crosspeaks, coupling vibrational mode s on molecule A with vibrational mode t on molecule B. Electrical dipole moment can be seen as the response of the molecular Hamiltonian toward an infinitesimal external electrical field.[13] Thus these induced electrical anharmonicities (IEAs) can be calculated by taking derivatives from both sides of eqn (1) with respect to an electric field E and two normal mode coordinates. In this case, the n–n term in eqn (1) makes no contribution to IEA and the contribution to IEA is simply the sum of contributions from the e–e and n–e terms, thus can be written as:

$$\left(\frac{\partial \mu_k}{\partial {Q_s}^A \partial {Q_t}^B}\right) = \left(\frac{\partial \mu_k}{\partial {Q_s}^A \partial {Q_t}^B}\right)_{ee} + \left(\frac{\partial \mu_k}{\partial {Q_s}^A \partial {Q_t}^B}\right)_{ne}.$$

For the sake of clarity, here the location of the vibrations is explicitly specified using superscripts. The e–e term contribution is:

$$\left(\frac{\partial \mu_k}{\partial {Q_s}^A \partial {Q_t}^B}\right)_{ee} = -\int d\boldsymbol{r}_B \frac{\partial^2 P_B}{\partial E_k \partial {Q_t}^B} \int d\boldsymbol{r}_A \frac{\partial P_A}{\partial {Q_s}^A} \frac{1}{|\boldsymbol{r}_A - \boldsymbol{r}_B|} - \int d\boldsymbol{r}_A \frac{\partial^2 P_A}{\partial E_k \partial {Q_s}^A} \int d\boldsymbol{r}_B \frac{\partial P_B}{\partial {Q_t}^B} \frac{1}{|\boldsymbol{r}_A - \boldsymbol{r}_B|}, \quad (2)$$

with $k = x, y, z$, the three Cartesian components. The contributions from the n–e terms are:

$$\left(\frac{\partial \mu_k}{\partial {Q_s}^A \partial {Q_t}^B}\right)_{ne} = \int_B d\boldsymbol{r}_B \frac{\partial^2 P_B}{\partial E_k \partial {Q_t}^B} \frac{\partial}{\partial {Q_s}^A} \left(\sum_{i=1}^{N_A} \frac{{Z_i}^A}{|{\boldsymbol{R}_i}^A - \boldsymbol{r}_B|} \right) + \int_A d\boldsymbol{r}_A \frac{\partial^2 P_A}{\partial E_k \partial {Q_s}^A} \frac{\partial}{\partial {Q_t}^B} \left(\sum_{j=1}^{N_B} \frac{{Z_j}^B}{|{\boldsymbol{R}_j}^B - \boldsymbol{r}_A|} \right).$$

The derivatives can be calculated easily with transformation matrices between normal modes and Cartesian coordinates, and then the total expression becomes:

$$\left(\frac{\partial \mu_k}{\partial Q_s{}^{\mathrm{A}} \partial Q_t{}^{\mathrm{B}}}\right)_{\mathrm{ne}} = -\sum_{i=1}^{N_{\mathrm{A}}} \int_{\mathrm{B}} \mathrm{d}\boldsymbol{r}_{\mathrm{B}} \frac{\partial^2 P_{\mathrm{B}}}{\partial E_k \partial Q_t{}^{\mathrm{B}}} \frac{Z_i{}^{\mathrm{A}}}{|\boldsymbol{R}_i{}^{\mathrm{A}} - \boldsymbol{r}_{\mathrm{B}}|^2} \frac{\partial \boldsymbol{R}_i{}^{\mathrm{A}}}{\partial Q_s{}^{\mathrm{A}}} \cdot \overline{(\boldsymbol{R}_i{}^{\mathrm{A}} - \boldsymbol{r}_{\mathrm{B}})} \quad (3)$$

$$-\sum_{j=1}^{N_{\mathrm{B}}} \int_{\mathrm{A}} \mathrm{d}\boldsymbol{r}_{\mathrm{A}} \frac{\partial^2 P_{\mathrm{A}}}{\partial E_k \partial Q_s{}^{\mathrm{A}}} \frac{Z_j{}^{\mathrm{B}}}{|\boldsymbol{R}_j{}^{\mathrm{B}} - \boldsymbol{r}_{\mathrm{A}}|^2} \frac{\partial \boldsymbol{R}_j{}^{\mathrm{B}}}{\partial Q_t{}^{\mathrm{B}}} \cdot \overline{(\boldsymbol{R}_j{}^{\mathrm{B}} - \boldsymbol{r}_{\mathrm{A}})}.$$

The derivative term: $\frac{\partial}{\partial Q_s{}^{\mathrm{A}}}\left(\sum_{i=1}^{N_{\mathrm{A}}} \frac{Z_i{}^{\mathrm{A}}}{|\boldsymbol{R}_i{}^{\mathrm{A}} - \boldsymbol{r}_{\mathrm{B}}|}\right)$, and similarly the term $\int \mathrm{d}\boldsymbol{r}_{\mathrm{B}} \frac{\partial P_{\mathrm{B}}}{\partial Q_t{}^{\mathrm{B}}} \frac{1}{|\boldsymbol{r}_{\mathrm{A}} - \boldsymbol{r}_{\mathrm{B}}|}$ in eqn (1), describes how electrostatic potential at position $\boldsymbol{r}_{\mathrm{B}}$ in molecule B, created by all the nuclei (or electrons) in molecule A, changes following the mode s vibration of molecule A. Using the IEA, EVV 2DIR signals can be calculated in the same fashion as shown in detail in our previous papers.[2,3,7]

There are two types of VTDCs in eqn (2) and (3) which were needed for the calculation of the induced electrical anharmonicity: the first-order VTDCs $\frac{\partial P}{\partial Q}$ and the second-order VTDCs $\frac{\partial^2 P}{\partial E \partial Q}$. The first-order VTDC describes how the electron density of a molecule changes with respect to its vibrational motion, while the second-order VTDC describes how it changes with respect to both the vibration and an infinitesimal electrostatic field. Detailed discussions of their properties can be found in Appendices A and B.

As shown in eqn (2) and (3), only second-order VTDCs are needed for the n–e term, while both first- and second-order VTDCs are required for the e–e term. All these were calculated through numerical finite differencing. For example the first-order VTDC for mode s on molecule A was calculated as:

$$\frac{\partial P_{\mathrm{A}}}{\partial Q_s{}^{\mathrm{A}}} = \frac{P_{\mathrm{A}}(\delta Q_s{}^{\mathrm{A}}) - P_{\mathrm{A}}(-\delta Q_s{}^{\mathrm{A}})}{2\delta Q_s{}^{\mathrm{A}}}; \quad (4)$$

To calculate this particular VTDC, first the ground state equilibrium structure of the molecule A was optimized and normal mode coordinates were obtained. Then a small step $\delta Q_s{}^{\mathrm{A}}$ of vibration along mode $Q_s{}^{\mathrm{A}}$ was added to the equilibrium geometry and an electron density cube $P_{\mathrm{A}}(\delta Q_s{}^{\mathrm{A}})$ at this geometry was generated. This was followed by subtracting the same small vibrational step from the equilibrium geometry and calculating another electron density cube $P_{\mathrm{A}}(-\delta Q_s{}^{\mathrm{A}})$ at the new geometry. Finally the VTDC was calculated using eqn (4) with the two density cubes.

The second-order VTDCs were calculated in a similar way, for example:

$$\frac{\partial^2 P_B}{\partial E_k \partial Q_t^B} = \frac{1}{4\delta Q_t^B \delta E_k} \cdot [P_B(\delta Q_t^B, \delta E_k) + P_B(-\delta Q_t^B, -\delta E_k) - P_B(\delta Q_t^B, -\delta E_k) - P_B(-\delta Q_t^B, \delta E_k)] \quad (5)$$

In this case a small electrostatic field steps δE_k along Cartesian axis k was also needed. All in all four separate calculations of electron densities incorporating both an electrostatic field and a vibrational step were needed to obtain a second-order VTDC.

To test the theory outlined above, again we chose the BN–PA dimer as our model system. In this case, the electrostatic interaction leads to the formation of a molecular complex, but the method is also applicable to cases when it is simply two chemical groups interacting with each other through-space. All VTDCs in this research were calculated at B3LYP/6-311++G(d,p) level using Gaussian03.[14] The vibrational step and the electric field step used in numerical differencing were chosen to be 0.01 Å and 0.005 a.u. respectively.

For vibrational couplings a coarse cube grid leads to very poor representation of VTDCs. The reason is that electronic density in a molecule form cusps around the component nuclei. When a molecule vibrates, its component atoms move along the normal mode coordinates with these cusps. When this happens, electron density will increase on one side of an atom, and decreases on the other side. This behaviour means that in an exact VTDC of a molecule there are always two cusps of opposite signs for each component atom, as shown in Fig. 1 for the simple case of hydrogen ion $H_2{}^+$. In another word, at the site of each atom, the value of a VTDC will switch from a large positive value to a large negative value. When a VTDC is represented in a discrete 3D grid, this fast-switching behaviour could cause numerical instability if not treated properly. If the grid of the density cube is not sampled finely enough, the integrations in eqn (2) and (3) become numerically unstable, jumping wildly between neighbouring sampling grids around each atom. On the other hand, in the regions between the component atoms, a VTDC changes much slowly, and using a fine grid in these regions is unnecessary and will make the integration prohibitively expensive. In this research, this dilemma was solved by using a two-tiered cube integration scheme: one cubic grid called the valence grid was first created

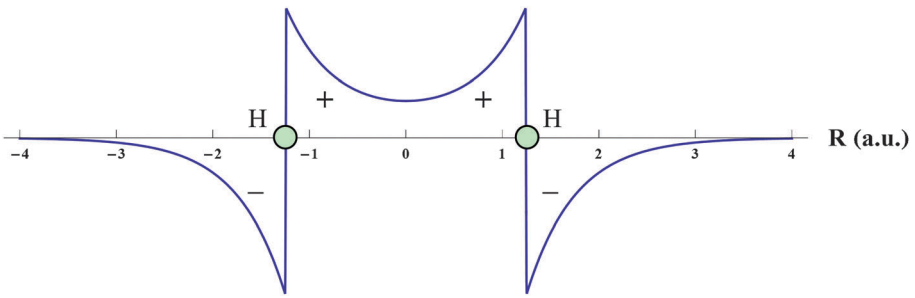

Fig. 1 Cusps in a first-order VTDC around atoms, shown in a section of the exact first-order VTDC for bond stretching vibration of hydrogen ion ($H_2{}^+$),[22] with plus and minus signs indicating increase or decrease of electron density.

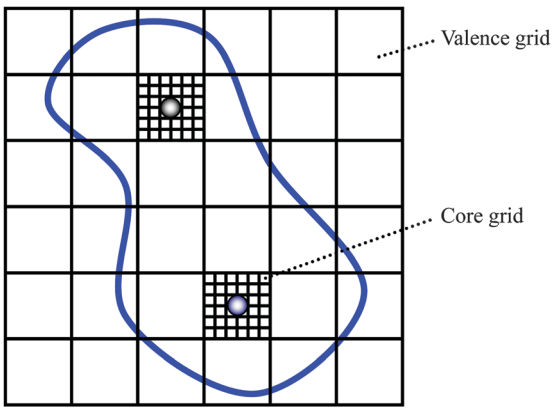

Fig. 2 Scheme of the two-tiered grids used for first-order VTDCs. For second-order VTDCs only valence grids were needed. See the text for details.

across the entire molecule, then around each component atom in the valence grid, a small region was singled out and a much finer grid, called a core grid, was generated. This is schematically shown in Fig. 2.

The double-cusp problem is mainly a problem for the first-order VTDCs. For the second-order VTDCs, actually it is largely alleviated by cancellation. When using eqn (5) to calculate a second order VTDC, the applied electrostatic field mostly induces polarizations to outlying electron distributions and has little effect to the core regions. Thus subtraction of the first two terms in the numerators on the right hand side largely cancelled out the cusps around the nuclei, removing the source of numerical instability persistent in first-order VTDCs. Thus only valence grid was used for second-order VTDCs.

For the results reported here, all the VTDCs are created on valence grids with 200 points on each side. This leads to 8 million grid points for each molecule, much more than those normally used in TDC method. The largest grid step used in this research is only about 0.062 Å, along the C≡C bond direction in PA. For first-order VTDCs, a 5 × 5 × 5 region around each atom in the valence grid was replaced by a core grid which is created with 100 points on each side, with the largest step only 0.0031 Å. A straightforward integration algorithm was used in our calculation, except for the integration in eqn (2) which was discussed in more detail in Appendix C. When core grids were used, the integration was carried out with both the valence grid and the set of core grids.

B. MD simulations of BN–PA mixture

Although gas-phase *ab initio* calculations of a molecular complex, such as the BN–PA dimer reported in our previous paper,[7] can tell us a lot about the nature of the interaction underlying the complexation, the structure of a liquid is obviously much more complicated than that encompassed in one single static structure. It makes much more sense to compare the 2DIR-determined structures with average structures obtained through molecular dynamics (MD) simulations. For this purpose, we also carried out molecular dynamics simulations of a 1 : 1 BN–PA liquid mixture.

A modified version of software package TINKER (version 5.1.09) was used for all MD simulations.[15] Modified OPLS-AA force field parameters were used for both BN and PA molecules.[16] Most of the force field parameters were used directly without any modification. However the torsional parameters for the –C–C≡N part of BN and –C–C≡CH part of PA were missing in OPLS-AA. These parameters were set to zero, as these groups are entirely linear. More seriously, when directly used for PA, the unpublished alkyne force field parameters in OPLS-AA leads to a molecule with a spurious total charge. This was corrected by setting up a new atomic type for the *ispo* aromatic carbon atom in PA, and setting its partial charge to zero. The other force parameters involving this new atomic type were the same as other aromatic carbon atoms without modification. With this modified force field, energy minimization of a BN–PA dimer reached an equilibrium structure very similar to our MP2 geometry. This can be seen partly as an indication that our modified force field is reasonable.

To achieve the correct density of the liquid mixture, 77 BN and 77 PA molecules were packed into a cubic box, each side of which is 30 Å and periodic boundary condition was applied. This led to an average density of 0.97 g cm^{-3}. The initial configuration of the system was created with packmol.[17] NPT simulation at 298.15 K and 1 atm was carried out with 1 fs step size. Particle Mesh Ewald (PME) summation was used with a cutoff of 15 Å. The system was first equilibrated for 2 ns, during which, parameters of the system were monitored to ensure proper equilibration. Then the production trajectory was run for another 0.9 ns to generate data reported here. A site–site distribution function was defined for the bond centres of C≡N and C≡C as $g_{CC-CN}(R)$ and obtained from the simulation. Since we were also interested in the correlation in the liquid mixture between the CN–CC distance R and the vectoral angle ϕ of C≡N relative to C≡C (for the definition of R and ϕ see Fig. 1(a) in ref. 7), a two-dimensional radial–angular distribution function was also calculated from the trajectory.

III. Results and discussion

A. VTDCs

First-order VTDCs for the C≡N stretching mode in benzonitrile and the C≡C stretching mode in phenylacetylene were calculated as described in Section II, including one valence VTDC and 13(14) core VTDCs for each molecule. Fig. 3 shows the valence VTDCs for both molecules. Here their transition densities are shown mostly concentrated around those atoms involved in the specific modes, *i.e.* the –C–C≡N group of BN, and the –C–C≡CH group of PA. In both cases, there are hardly any transition densities across the benzene rings, except for the *ipso* carbon atoms. The actual shape of transition density is qualitatively similar to the H_2^+ case shown in Fig. 1 and can be explained similarly. For example, when the C≡N bond in benzonitrile stretches, electron density between C and N atoms decreases, while increasing at the same time on both sides of C and N atoms.

As pointed out in Section II and shown in Fig. 1, when a molecule vibrates, electron difference density around nuclei shows distinctive double-cusp shapes, changing its sign quickly when sampled from one side of an atom to the other side. This behaviour is difficult to describe with a uniform grid enveloping

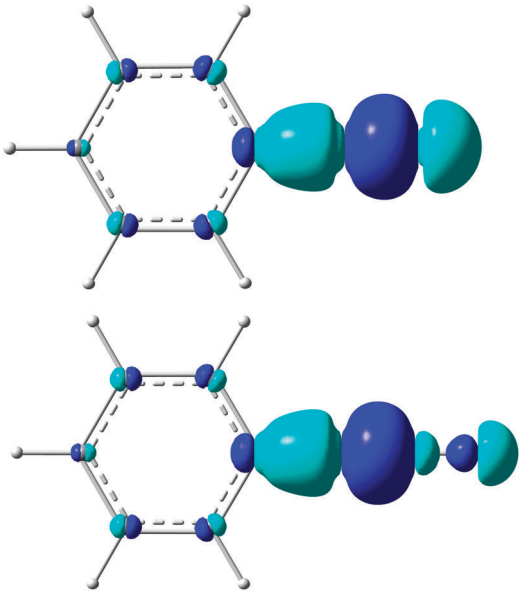

Fig. 3 First-order VTDCs for C≡N stretching of BN and C≡CH stretching of PA, drawn on an isosurface of 0.013 a.u.

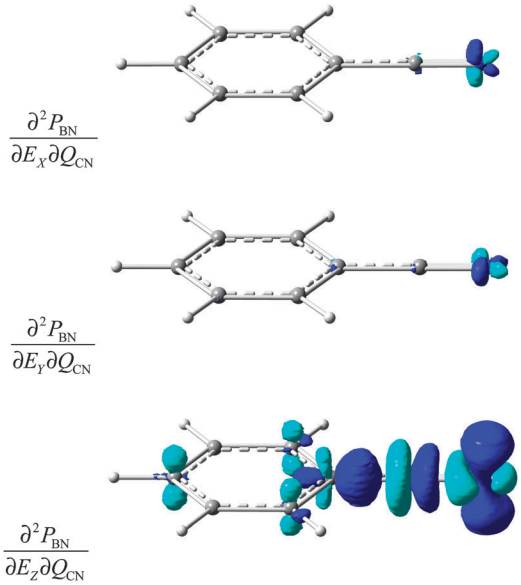

Fig. 4 Calculated second-order VTDCs for C≡N stretching of BN, drawn on an isosurface of 0.053 a.u.

the entire molecule. We solved this difficulty with a two-tiered grid scheme. Shown in Fig. 3 are just the valence grids we calculated, as the finer core grids around nuclei make them difficult to show in one figure.

First-order VTDCs are closely linked to vibrational transition moments. When the valence VTDCs shown in Fig. 3 were integrated along with its set of core VTDCs, according to eqn (A1) in Appendix A, first-order dipole moment derivatives of BN and PA can be easily calculated. These were given in Table 1 along with the published DFT values[7] and those measured from FTIR using pure liquid BN and pure liquid PA.

As shown in Table 1, our cube-integrated values for BN and PA are 4.71 and 9.76 times higher than the experimentally measured values. In the same fashion as usually applied in TDC method,[10] the calculated VTDCs were scaled with these factors before they were used to calculate the induced electrical anharmonicity in the BN–PA complex. Symmetric consideration also dictates that transition dipole of either BN or PA must have only non-zero z component which lies along the C≡N or C≡C bond direction. Although there are some small x and y components due to the nature of numerical integration, the values were too small to cause much concern once they were scaled.

Second-order VTDCs were also calculated for BN and PA, shown for BN in Fig. 4 and PA in Fig. 5. As explained in Section II, there is no need of core grids for second-order VTDCs, so there are only 3 second-order VTDCs for each molecule, with each cube corresponds to one of the three Cartesian components of the electric-field. The coordinate system was again chosen as the z axis lying along the C≡N or C≡C bond direction while the x axis is normal to the aromatic ring. In both the BN and PA cases, although there are some small transition densities in the E_X and E_Y cubes, the most significant of the transition densities obviously exist in the E_Z cube. Compared with the first-order VTDCs, the second-order cubes spread further out of the –C–C≡N part involved in the vibrations: the C atoms on the benzene ring opposite to and besides the *ipso* carbon atom also showed some transition density around them.

As pointed out in Appendix B, these second-order density cubes can also be called transition polarizability density cubes

Table 1 Comparison of experimentally measured first-order dipole moment derivatives and their Cartesian components calculated with VTDC method or DFT (in a.u.)

	DFT[a]	VTDC[b]	Exp.[c]
BN	(0.0, 0.0, 0.698)	(−0.007, −0.024, 1.098)	0.2333
PA	(0.0, 0.0, 0.256)	(0.007, −0.006, 0.540)	0.0553

[a] Calculated at B3LYP/6-311++G(d,p) level. [b] This work. [c] Magnitude only.

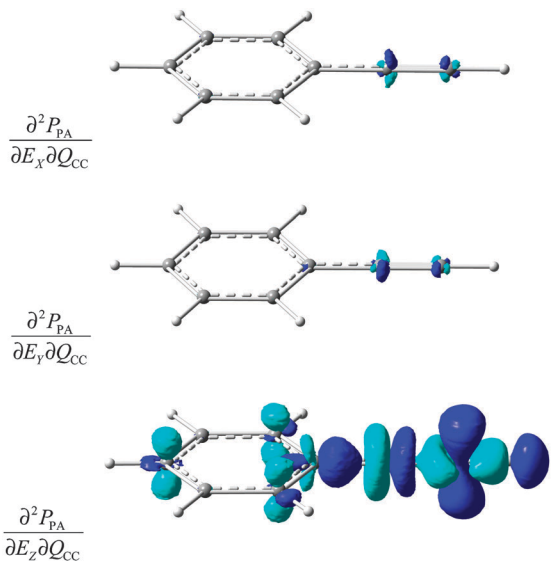

Fig. 5 Calculated second-order VTDCs for C≡CH stretching of PA, drawn on an isosurface of 0.053 a.u.

as all orders of transition polarizability can be obtained from them by integration. In particular the integrated transition dipole polarizabilities from these cubes are: $\begin{pmatrix} 0.94 & -0.66 & -0.45 \\ -0.76 & 1.96 & -0.33 \\ 0.02 & 0.21 & 79.68 \end{pmatrix}$ for BN and $\begin{pmatrix} -4.83 & -0.01 & 0.00 \\ 0.07 & 1.22 & 0.27 \\ 0.62 & -0.01 & 77.30 \end{pmatrix}$ for PA. These agreed quite well with the published DFT values[7] as: $\begin{pmatrix} 3.007 & 0 & 0 \\ 0 & 3.612 & 0 \\ 0 & 0 & 75.773 \end{pmatrix}$ for BN and $\begin{pmatrix} 4.524 & 0 & 0 \\ 0 & 4.548 & 0 \\ 0 & 0 & 73.755 \end{pmatrix}$ for PA, at least for the dominant $\frac{\partial \alpha_{ZZ}}{\partial Q}$ components. The smaller XX and YY terms were not reproduced very well, there are also small off-diagonal terms, most likely due to the nature of numerical integration. In the BN–PA case we considered in this paper, this poses no real problem. Since there are no available experimental $\frac{\partial \alpha}{\partial Q}$ values to compare to, our second-order VTDCs were used directly without scaling.

B. Geometry determination using induced electrical anharmonicity

The integrations in eqn (2) and (3) were carried out as described in more detail in Appendix C, and the results were summarised in Table 2. Compared to our previous dipole–dipole value for IEA, the magnitude of IEA has slightly more than doubled. This predominantly comes from the n–e term contributions, while the e–e terms only contributed a minor part.

Taking the newly calculated IEA into consideration, and applying the same procedure of geometry determination as described in more details in our previous paper[7] to the BN–PA system, a new value for the distance R between the C≡CH and C≡N bond centres can be determined. This is calculated to be 3.60 Å, which agrees extremely well with our MP2(full)/BSSE value at 3.68 Å, and a much improved distance over our previous results of 2.6 Å.

Defining an angular relationship between two molecules using interacting cubes is not obvious. However we can calculate a new PPS/PPP signal ratio with the new IEA, and then continue to use eqn (4) in our previous paper[7] to determine the angular relationship. The CC–CN vectoral angle ϕ was thus determined to be 31 degrees, which is exactly the same as was obtained from the previous dipole–dipole model.[7]

The new VTDC approach also gives us a chance to re-evaluate our dipole–dipole model, which, although deemed inadequate, managed to account for about 50% of the total induced electrical anharmonicity with much less computing burden. It also provides us with an easy and straightforward way to define the angular relationships between two interacting molecules. The natural approach for determining the geometry of molecular complexes is therefore to make a first pass approximation using the dipolar approximation and once the approximate geometry is determined to subsequently refine the structure using the more accurate VTDC approach. Actually this type of processes of approximation followed by refinement is common to most current biomolecular structural analysis methods.[18]

C. MD simulations and distribution functions

It is well known that even the simplest liquids have very complicated dynamic structures.[19] Therefore to further demonstrate that our scheme is realistic enough for real liquid mixtures, we also carried out MD simulations for the BN–PA system. The details of our simulations were given in Section II. Shown in Fig. 6 is a snapshot of the BN–PA system used in our MD simulation after its equilibration. To highlight interacting molecular groups we are interested in, all –C≡CH groups of phenylacetylene and –C≡N groups of benzonitrile were shown as CPK models, while one specific pair of BN and PA was highlighted in VDW model.

A site–site distribution function $g_{CC-CN}(R)$ was defined between each pair of –C≡CH and –C≡N bond centres over the entire MD trajectory, and was calculated and shown in Fig. 7. The two peaks on $g_{CC-CN}(R)$ indicates short-range order in the BN–PA mixture. The first peak of $g_{CC-CN}(R)$ appears at about 4.0 Å, where it shows that at this distance it is 1.5 times more likely to find a –C≡CH group around a –C≡N group and *vice versa*, compared to the non-interacting limit if the interaction between BN and PA were switched off. This distance is very close to the distance in our gas-phase structure of BN–PA dimer calculated at MP2 level[7] and that determined with our cube-integral scheme. Fig. 7 also showed a minimum at around 5.7 Å, which may be defined as the boundary of the first solvation shell of BN/PA.

Although site–site distribution functions, such as $g_{CC-CN}(R)$ in Fig. 7, are useful, they carry no information about the distribution of relative orientation between C≡C and C≡N bonds. In particular, the correlation between the CN–CC distance R and the orientation ϕ between the two bonds is of great interest to us. Thus a two-dimensional normalized R–ϕ distribution function $f_{CC-CN}(R,\phi)$ was calculated along the MD trajectory and shown in Fig. 8. Some interesting features can be observed in $f_{CC-CN}(R,\phi)$. First of all, it is obvious that distributions of R and ϕ are indeed correlated.

Table 2 Induced electrical anharmonicity (IEA) calculated with VTDC method and dipole–dipole model (in a.u.)

		VTDC	Dipole–dipole[7]
n–e Term		(0.0074 −0.0708 0.3025)	—
e–e Term	Valence	(0.0001 −0.0002 −0.0020)	—
	Core	(−0.0008 0.0016 −0.0144)	—
	Total	(−0.0007 0.0014 −0.0164)	—
Total IEA		(0.0067 −0.0694 0.2861)	(0.0021 −0.0055 0.1480)

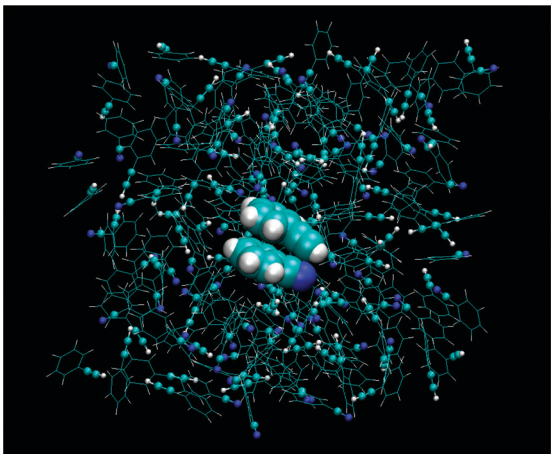

Fig. 6 A snapshot of the 1 : 1 BN–PA liquid mixture used in MD simulations, plotted with VMD.[23]

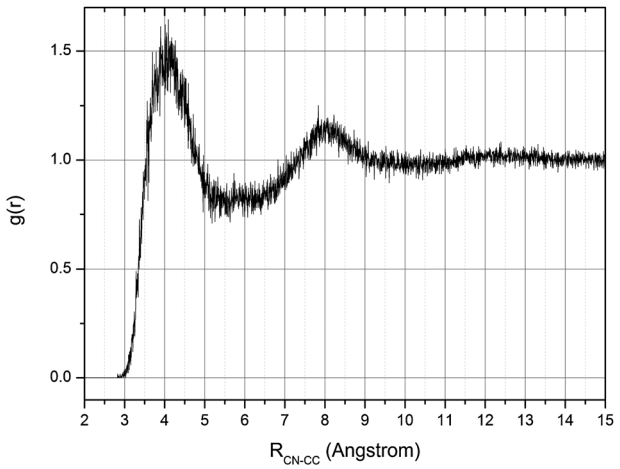

Fig. 7 Normalized CN–CC site–site distribution function $g_{CC-CN}(R)$ of the 1 : 1 BN–PA liquid mixture, calculated from MD simulations.

For example, there are reduced probabilities of finding –C≡CH and –C≡N groups in such geometries with R from 5.0 to 7.1 Å and ϕ from 35 to 120 degrees. On the other hand there is a region of enhanced probabilities of finding –C≡CH and –C≡N groups in geometries centring around R at 3.8 Å and ϕ from 25 to 50 degrees. This region agrees very well with our experimentally determined values of both distance and angle (3.60 Å and 31 degree), indicating that the experimentally determined geometry by EVV 2DIR spectroscopy correctly picked up the most possible or modal geometry in our MD simulation. The reduced probability valley exists most likely due to spatial hindering: since it is more possible for a neighbouring molecule taking up the "peak" geometry, it becomes much difficult for another molecule to assume the "valley" geometry. The seemingly enhanced probabilities close to the top and bottom edges (ϕ around 0 and 180 degrees) of $f_{CC-CN}(R,\phi)$ are largely artefacts due to a normalization factor of sin ϕ.

It should be noted that in this study we were only characterising the dynamical structural aspect of the BN–PA mixture. Solvation effects, which is largely a polarization effect of a solute molecule by electrostatic fields generated by surrounding solvent and other solute molecules, were not considered in this research. Further research along this direction is currently in progress.

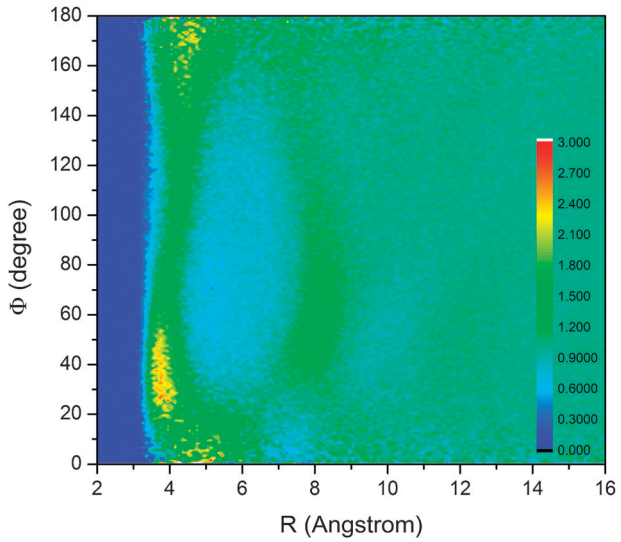

Fig. 8 Radial–angular distribution function $f_{CC-CN}(R,\phi)$ of the 1 : 1 BN–PA liquid mixture, calculated from MD simulations.

D. VTDC method with MD trajectory

The results presented above agree very well with our experimental results. However as we have shown in our MD simulation, neighbouring BN and PA molecules in a liquid mixture can actually assume many possible configurations, due to incessant thermal motion and weak interactions between them. For a more realistic simulation of EVV 2DIR experiments, the distribution of configurations must be properly addressed. One way to do this is to combine our VTDC technique with MD simulations.

From our MD trajectory of BN–PA mixture, a set of 100 snapshots of the periodic MD cell were taken out. Then in each snapshot our VTDC method was applied to calculate interaction-induced electrical anharmonicities for every possible combination of BN and PA molecules, within a cutoff limit of intermolecular distance of 10 Å. All in all about 90 000 BN–PA pairs were included in this calculation. Then for each of them, an EVV 2DIR signal was calculated. These give us the total 2DIR signal when coherently summed over the entire periodic cell.

To obtain any statistically meaningful results, a large number of interacting pairs of BN–PA molecules over long MD trajectories of the liquid mixture must be included in the calculation. This can be prohibitively expensive if the slower converging e–e terms in our VTDC scheme were included. To reduce the computational cost, a few simplifying measures had to be taken: first, only n–e terms in our cube integral scheme were taken into consideration. This is justified as the contributions from e–e terms were not significant, as shown in Table 2. Secondly, the second-order VTDCs needed for the n–e term calculation were truncated to a smaller grid size of 150 × 150 × 180, rather than the original size of 200 × 200 × 200. By doing so the total number of cells in a cube was halved, further reducing the computing burden. We found this practice had very little influence on the quality of the cubes. Actually if the IEA reported in Section III C was re-calculated using these truncated cubes, there was only a 3% change of its magnitude. The use of

truncated cubes also has the added benefit of avoiding spuriously huge values of induced electrical anharmonicities, as it was noticed that, although rare in our MD trajectory, one molecule can get so close to its partner that an overlap between the original cubes occurs, leading to unphysical values of induced electrical anharmonicities. Using truncated cubes effectively solved this problem.

On the other hand, although our dipolar model is less accurate than the VTDC technique, it is still of interest to us because of its faster computational speed and the physical picture it provides. Therefore the dipolar model is also used to calculate 2DIR signals using the same set of MD snapshots. For the sake of comparison with VTDC scheme, the same cutoff limit was used as well.

Shown in Fig. 9 is the distribution of magnitudes of the calculated induced electrical anharmonicities. Using our previously calculated values for our dipolar model and VTDC method (0.1481 and 0.2945 a.u.) as references, it can be seen that only a small proportion of BN–PA interactions lead to IEAs with a magnitude equal or higher, while a majority of the interactions only lead to very small induced electrical anharmonicities. This is not surprising as in our analysis every combination of BN and PA in MD trajectory was taken into consideration up to the cutoff limit, but presumably only those pairs lying close enough to each other can possibly interact strongly to generate contributing IEAs to EVV 2DIR signals. In Fig. 10 are the distributions of three Cartesian components of calculated IEAs in a body-fixed frame located on BN moiety of each BN–PA pair, using both dipole–dipole model and VTDC models. The *z*-axis of the body-fixed frame was chosen to be along C≡N bond direction, pointing from C to N. The most obvious feature in this figure is that VTDC-calculated induced electrical anharmonicities showed different distributions than those calculated with dipolar model. For the *x*- and *y*-components VTDC calculated distributions are much narrower than those calculated *via* dipolar model. While for the *z*-component, the distribution of the VTDC results is much wider. Since the same set of snapshots were used in dipolar and VTDC calculations,

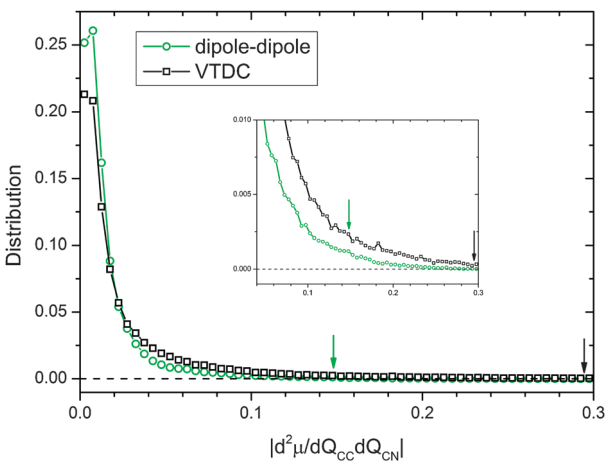

Fig. 9 Distributions of magnitudes of induced electrical anharmonicities calculated with dipole–dipole model and the VTDC method, also shown as an enlarged inset. The two arrows point out the corresponding reference values obtained with the gas-phase dimer model.

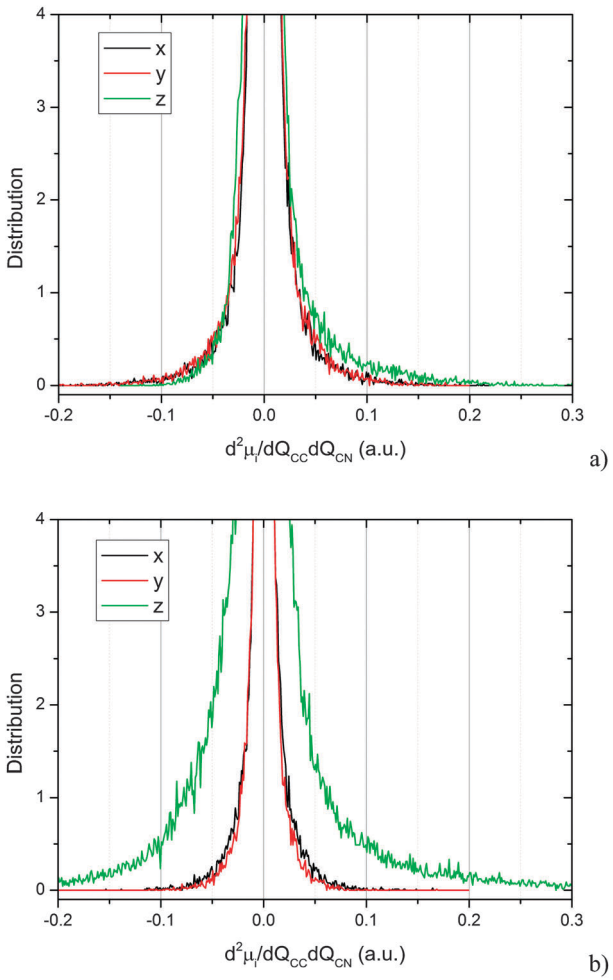

Fig. 10 Distributions of the *i* component (*i* = *x*, *y*, *z*) of induced electrical anharmonicities in a BN-fixed frame calculated with (a) dipole–dipole model and (b) VTDC method.

the reason of the differences in distributions of IEA components between dipolar and VTDC model must be the inclusion of all the higher order electrostatic interactions within VTDC model. The distribution of the *x*- and *y*-components in both dipolar and VTDC cases can be fitted nicely with Laplace distributions centred close to 0 and is largely symmetric. However the distributions of the *z*-component of IEAs in both cases are evidently asymmetric, with noticeably more distributions on the positive side (pointing from C to N) than those on the negative side (pointing from N to C). This is obviously a reflection of intrinsic point symmetry of BN and PA molecules, and is also in accord with the signs of values given in Table 2.

When calculated EVV 2DIR signals between all BN–PA pairs were summed over the entire periodic cell, the total signal detected in an experiment was obtained. This, divided by the number of BN–PA pairs in the simulation cell, was used to obtain an "effective" nonlinear susceptibility with all BN–PA pairs assuming a representative modal structure. For the current study we are more interested in signal ratios which experimentally can be measured reliably. PPS/PPP ratio is an important experimental observable which can be used to determine the angular relationship between the interacting pair. For the BN–PA system, these were measured experimentally to

Table 3 Calculated signal intensity ratios using cube integral method and dipole–dipole model with MD trajectory

	$\lvert\chi_{PPS}^{(3)}\rvert^2/\lvert\chi_{PPP}^{(3)}\rvert^2$		Ratio between CN–CC crosspeak and CN–CN crosspeak	
	CN–CC crosspeak (2239/4353 cm^{-1})	CC–CN crosspeak (2113/4353 cm^{-1})	PPP	PPS
VTDC + MD	0.826	0.154	0.133	0.759
Dipole–dipole + MD	0.258	0.635	0.037	0.066
Exp.[7]	0.417	0.368	0.73	—

be 0.417 ± 0.037 for the CN–CC crosspeak (at 2239/4353 cm^{-1}), and 0.368 ± 0.017 for the CC–CN crosspeak (at 2113/4353 cm^{-1}).[7] Using VTDC scheme, the squared magnitude of effective nonlinear susceptibility $|\chi^{(3)}|^2$ of the CN–CC crosspeak were calculated to be 4.48×10^{-30} (this and all the following values in units of $cm^6\ erg^{-2}$) for the case of PPP configuration and 3.69×10^{-30} in the PPS case. This gave us a PPS/PPP ratio of 0.826, which is much higher than the experimentally measured PPS/PPP ratio. For the CC–CN crosspeak, the values are 5.57×10^{-31} for the PPP case and 8.57×10^{-32} for the PPS case. This led to a PPS/PPP ratio of 0.154 which is too low compared to the experimental ratio.

The signal ratio between the CN–CC crosspeak and one of the two intra-molecular crosspeaks, the CN–CN crosspeak (2239/4478 cm^{-1}), was previously used with the dipolar model to determine the distance between the interacting BN–PA pair. Using quantum chemically calculated molecular properties, $|\chi^{(3)}|^2$ of the reference CN–CN crosspeak (2239/4478 cm^{-1}) was calculated to be 3.36×10^{-29} for the PPP polarization case and 4.87×10^{-30} for the PPS case. Compared to the values given above for the CN–CC crosspeak, we have the signal ratios between the two at 0.133 in the PPP case and 0.759 in the PPS case. However, the experimentally measured signal ratio in the PPP case is 0.73 ± 0.05. All these results were summarised in Table 3. Also shown in this table were the corresponding values calculated with the dipole–dipole model. Compared to the VTDC results in the table it is noteworthy that for both PPS/PPS ratios, the VTDC results seems to over-compensate the discrepancies between the dipole–dipole model and the experimental values by including full electrostatic interactions between BN and PA monomers. On the other hand, both models underestimated the signal ratio between the CN–CC crosspeak and the reference CN–CN crosspeak.

To check whether the discrepancies between our calculated signal ratios and experimental values were artefacts due to the relatively small MD system used in our simulation, we also carried out NPT MD simulations of a much larger BN–PA system, which contains 384 BN and PA molecules in a cubic box, with each side of 51.30 Å. It was first equilibrated for 1 ns, then a 0.5 ns production MD simulation was carried out. All the simulation conditions were chosen as the same as those described in Section II. Analysis of the trajectory of this larger MD system indicated exactly the same dynamics of the liquid mixture as our original MD model. Thus no improvement of theoretical signal ratios can be expected by using larger MD simulations. This strongly suggested that the discrepancies between our calculated theoretical signal ratios and reported experimental values are most likely inherent in the force field we used, which should be addressed in further studies.

V. Conclusion

In this paper, we have shown that an integration method based on vibrational transition density cubes (VTDCs) can be used for the calculation of induced electrical anharmonicities between two weakly interacting molecules. Rather than using any expansions, this method takes into consideration of the complete classical electrostatic interaction. An efficient two-tiered integration scheme was developed to speed up the integration. Using the benzonitrile–phenylacetylene liquid mixture as a test system, we showed that this newly developed method led to a much more accurate induced electrical anharmonicity between the –C≡N and –C≡CH stretching vibrations of the two molecules, and consequently improved accuracy in geometry determination from EVV 2DIR data. By taking into consideration of the complete electrostatic interaction between BN and PA, instead of counting on any multipole expansion, the distance between two interacting molecular groups was determined to be 3.60 Å. The angle between C≡N and C≡C bond directions was determined to be 31°. For a better understanding of the dynamic structure of the BN–PA liquid mixture, MD simulations were carried out and distribution functions characterising liquid structure were obtained. It was found the 2DIR-determined geometrical parameters agree quite well with the most probable geometry taken by one BN(PA) molecule and one of its neighbouring PA(BN) molecules in its first solvation shell. The new VTDC method and dipole–dipole model were also combined with MD simulations to provide us a more realistic description of EVV 2DIR signals. Limited by the currently available force fields, we nevertheless consider this combination of methods will provide the most powerful tools for our future studies of EVV 2DIR spectroscopy in condensed phase.

With the theoretical toolkit established in this and our previous paper,[7] a viable recipe for accurate geometry determination of non-bonded complexes from EVV 2DIR data can now be proposed. The procedure starts with first making a dipole–dipole approximation between the interacting complexes, and obtaining a set of approximate angles and distances of the complex. This will be followed by structural refinement using the VTDC approach outlined here. Future work will demonstrate the utility of this approach in molecular systems relevant to biology where interactions in the absence of covalent bonds are key contributions to biological function.

Appendix

A. First order VTDCs

Just like transition densities traditionally defined[20] for electronic transitions, the first-order VTDC defined here are closely linked

to vibrational transition moments. But unlike the former, the first-order VTDC alone cannot determine vibrational transition moments. Consider a molecule with an electron density distribution $P(\boldsymbol{r})$ and N nuclei with charges Z_i, $i = 1\ldots N$ located at $\boldsymbol{R}_i$. The molecular dipole moment operator is:

$$\boldsymbol{\mu} = -\int \boldsymbol{r}P(\boldsymbol{r})\mathrm{d}v + \sum_i Z_i\boldsymbol{R}_i.$$

Thus the first-order dipole derivative with respect to a normal mode Q_s will be:

$$\frac{\partial\boldsymbol{\mu}}{\partial Q_s} = -\int \boldsymbol{r}\frac{\partial P(\boldsymbol{r})}{\partial Q_s}\mathrm{d}v + \sum_i Z_i\frac{\partial\boldsymbol{R}_i}{\partial Q_s}, \qquad \text{(A1)}$$

which decides its corresponding vibrational transition dipole. Obviously here the first-order VTDC $\frac{\partial P(\boldsymbol{r})}{\partial Q_s}$ alone cannot determine the transition dipole. An extra term involving the nuclear contributions has to be included too, which is not generally zero.

Higher order transition moments can be obtained in similar ways. More generally consider a spherical l-pole moment defined as:

$$q_{\mathrm{lm}} = \int r^l C_{\mathrm{lm}}(\theta,\phi)P(\mathbf{r})\mathrm{d}v,$$

with C_{lm} being the modified spherical harmonics as defined by Brink and Satchler.[21] Then the l-pole moment of the molecule will be:

$$q_{\mathrm{lm}} = -\int r^l C_{\mathrm{lm}}(\theta,\phi)P(\boldsymbol{r})\mathrm{d}v + \sum_i Z_i R_i^l C_{\mathrm{lm}}(\theta_i,\phi_i)$$

Therefore the first-order derivative of vibrational transition l-pole moment will be:

$$\frac{\partial q_{\mathrm{lm}}}{\partial Q_s} = -\int r^l C_{\mathrm{lm}}(\theta,\phi)\frac{\partial P(\mathbf{r})}{\partial Q_s}\mathrm{d}v + \sum_i Z_i\frac{\partial}{\partial Q_s}[R_i^l C_{\mathrm{lm}}(\theta_i,\phi_i)]. \qquad \text{(A2)}$$

The nuclear derivatives in eqn (A1) and (A2) can be easily calculated from transformation matrix between normal mode coordinates and Cartesian coordinates. In theory, first-order VTDC combined with these nuclear derivatives completely determines the full set of vibrational transition moments of a molecule. Also note that for any molecule with a total charge C, *i.e.* $C = -\int P(\boldsymbol{r})\mathrm{d}v + \sum_i Z_i$. It follows that for a first order VTDC:

$$\int \frac{\partial P(\boldsymbol{r})}{\partial Q_s}\mathrm{d}v = 0.$$

This can be used as a test of the quality of VTDCs created.

B. Second-order VTDCs

The second-order VTDCs have a clearer physical meaning than their first-order counterparts, as transition polarizabilities can be solely determined by them. For the same molecule discussed in the above section, it can be shown that its electronic dipole polarizability is:

$$\alpha_k \equiv \frac{\partial\boldsymbol{\mu}}{\partial E_k} = -\int \boldsymbol{r}\frac{\partial P(\boldsymbol{r})}{\partial E_k}\mathrm{d}v,$$

as long as the electric field is not strong enough to significantly modify the molecular geometry. The transition dipole polarizability is determined by the first order derivative of α_k, which is:

$$\frac{\partial\alpha_k}{\partial Q_s} = \frac{\partial^2\boldsymbol{\mu}}{\partial E_k\partial Q_s} = -\int \boldsymbol{r}\frac{\partial^2 P(\boldsymbol{r})}{\partial E_k\partial Q_s}\mathrm{d}v,$$

thus can be completely decided by $\frac{\partial^2 P(\boldsymbol{r})}{\partial E_k\partial Q_s}$, the second-order VTDCs. In this sense, the second-order VTDC can also be called transition polarizability density cube. For the l-order moment as defined in Appendix A, the corresponding transition l-mode polarizability is:

$$\frac{\partial^2 q_{\mathrm{lm}}}{\partial E_k\partial Q_s} = -\int r^l C_{\mathrm{lm}}(\theta,\phi)\frac{\partial^2 P(\boldsymbol{r})}{\partial E_k\partial Q_s}\mathrm{d}v$$

Similar to the first-order VTDCs, the second-order VTDCs must also satisfy:

$$\int \frac{\partial^2 P(\boldsymbol{r})}{\partial E_k\partial Q_s}\mathrm{d}v = 0.$$

C. Integrating the e–e term contributions

With all the density cubes calculated, eqn (2) and (3) can be used to calculate the IEA. The n–e term integration in eqn (3) involves only one VTDC and can be easily carried out. However, the integration for e–e terms in eqn (2) involves two VTDCs, each having 8 million cells in our cases, and thus is extremely expensive computationally. The partition of the 3D space into valence and core regions means that the integration in eqn (2) now will become the sum of two terms: one only involves first-order core VTDCs:

$$\left(\frac{\partial\mu_k}{\partial Q_s{}^{\mathrm{A}}\partial Q_t{}^{\mathrm{B}}}\right)_{\mathrm{ee}}^{\mathrm{core}} = -\int \mathrm{d}\boldsymbol{r}_{\mathrm{B}}\frac{\partial^2 P_{\mathrm{B}}}{\partial E_k\partial Q_t{}^{\mathrm{B}}}\sum_i\int_{\mathrm{atom}\ i}\mathrm{d}\boldsymbol{r}_{\mathrm{A}}\left(\frac{\partial P_{\mathrm{A}}}{\partial Q_s{}^{\mathrm{A}}}\right)_i\frac{1}{|\boldsymbol{r}_{\mathrm{A}}-\boldsymbol{r}_{\mathrm{B}}|} - \int \mathrm{d}\boldsymbol{r}_{\mathrm{A}}\frac{\partial^2 P_{\mathrm{A}}}{\partial E_k\partial Q_s{}^{\mathrm{A}}}\sum_j\int_{\mathrm{atom}\ j}\mathrm{d}\boldsymbol{r}_{\mathrm{B}}\left(\frac{\partial P_{\mathrm{B}}}{\partial Q_t{}^{\mathrm{B}}}\right)_j\frac{1}{|\boldsymbol{r}_{\mathrm{A}}-\boldsymbol{r}_{\mathrm{B}}|} \qquad \text{(C1)}$$

and another involves only first-order valence VTDCs:

$$\left(\frac{\partial\mu_k}{\partial Q_s{}^{\mathrm{A}}\partial Q_t{}^{\mathrm{B}}}\right)_{\mathrm{ee}}^{\mathrm{valence}} = -\int \mathrm{d}\boldsymbol{r}_{\mathrm{B}}\frac{\partial^2 P_{\mathrm{B}}}{\partial E_k\partial Q_t{}^{\mathrm{B}}}\int \mathrm{d}\boldsymbol{r}_{\mathrm{A}}\left(\frac{\partial P_{\mathrm{A}}}{\partial Q_s{}^{\mathrm{A}}}\right)_v\frac{1}{|\boldsymbol{r}_{\mathrm{A}}-\boldsymbol{r}_{\mathrm{B}}|} - \int \mathrm{d}\boldsymbol{r}_{\mathrm{A}}\frac{\partial^2 P_{\mathrm{A}}}{\partial E_k\partial Q_s{}^{\mathrm{A}}}\int \mathrm{d}\boldsymbol{r}_{\mathrm{B}}\left(\frac{\partial P_{\mathrm{B}}}{\partial Q_t{}^{\mathrm{B}}}\right)_v\frac{1}{|\boldsymbol{r}_{\mathrm{A}}-\boldsymbol{r}_{\mathrm{B}}|}. \qquad \text{(C2)}$$

On first look, this partitioning makes the e–e term integration even more expensive: in the BN–PA case, although each core cube has about 1 million cells, there are now about 13 or 14 core cubes to integrate. However each core cube only occupies a tiny space around its corresponding atom (the longest variation is ±0.155 Å). If we denote the location of the atom as $\boldsymbol{R}_{\mathrm{A}}$ and the coordinate variation in the cube by $\delta\boldsymbol{r}_{\mathrm{A}}$, then we have:

$$|\boldsymbol{r}_{\mathrm{A}}-\boldsymbol{r}_{\mathrm{B}}| = |\boldsymbol{R}_i^{\mathrm{A}} + \delta\boldsymbol{r}_{\mathrm{A}} - \boldsymbol{r}_{\mathrm{B}}| = |\boldsymbol{R}_i^{\mathrm{A}} - \boldsymbol{r}_{\mathrm{B}} + \delta\boldsymbol{r}_{\mathrm{A}}| \sim |\mathbf{R}_i^{\mathrm{A}} - \boldsymbol{r}_{\mathrm{B}}|$$

for each core cube. The last approximation is valid if the distance $|\boldsymbol{R}_{\mathrm{A}} - \boldsymbol{r}_{\mathrm{B}}|$ is much larger than $\delta\boldsymbol{r}_{\mathrm{A}}$, which should be

true for most non-bonded interactions, *e.g.* the BN–PA interaction. A similar equation also exists for core cubes of molecule B. Then eqn (C1) becomes:

$$\left(\frac{\partial \mu_k}{\partial Q_s{}^{\mathrm{A}} \partial Q_t{}^{\mathrm{B}}}\right)_{\mathrm{ee}}^{\mathrm{core}} \simeq$$

$$-\int \mathrm{d}\boldsymbol{r}_{\mathrm{B}} \frac{\partial^2 P_{\mathrm{B}}}{\partial E_k \partial Q_t{}^{\mathrm{B}}} \sum_i \int_{\text{atom } i} \mathrm{d}\boldsymbol{r}_{\mathrm{A}} \left(\frac{\partial P_{\mathrm{A}}}{\partial Q_s{}^{\mathrm{A}}}\right)_i \frac{1}{|\boldsymbol{R}_{\mathrm{A}} - \boldsymbol{r}_{\mathrm{B}}|}$$

$$-\int \mathrm{d}\boldsymbol{r}_{\mathrm{A}} \frac{\partial^2 P_{\mathrm{A}}}{\partial E_k \partial Q_s{}^{\mathrm{A}}} \sum_j \int_{\text{atom } j} \mathrm{d}\boldsymbol{r}_{\mathrm{B}} \left(\frac{\partial P_{\mathrm{B}}}{\partial Q_t{}^{\mathrm{B}}}\right)_j \frac{1}{|\boldsymbol{r}_{\mathrm{A}} - \boldsymbol{R}_{\mathrm{B}}|}$$

$$= -\int \mathrm{d}\boldsymbol{r}_{\mathrm{B}} \frac{\partial^2 P_{\mathrm{B}}}{\partial E_k \partial Q_t{}^{\mathrm{B}}} \sum_i \frac{1}{|\boldsymbol{R}_i{}^{\mathrm{A}} - \boldsymbol{r}_{\mathrm{B}}|} \int_{\text{atom } i} \mathrm{d}\boldsymbol{r}_{\mathrm{A}} \left(\frac{\partial P_{\mathrm{A}}}{\partial Q_s{}^{\mathrm{A}}}\right)_i$$

$$-\int \mathrm{d}\boldsymbol{r}_{\mathrm{A}} \frac{\partial^2 P_{\mathrm{A}}}{\partial E_k \partial Q_s{}^{\mathrm{A}}} \sum_j \frac{1}{|\boldsymbol{r}_{\mathrm{A}} - \boldsymbol{R}_j{}^{\mathrm{B}}|} \int_{\text{atom } j} \mathrm{d}\boldsymbol{r}_{\mathrm{B}} \left(\frac{\partial P_{\mathrm{B}}}{\partial Q_t{}^{\mathrm{B}}}\right)_j$$

Since $\frac{1}{|\boldsymbol{R}_{\mathrm{A}} - \boldsymbol{r}_{\mathrm{B}}|}$ does not depend on $\boldsymbol{r}_{\mathrm{A}}$, it can be moved out of the inner integral. Now the inner integration is simply the sum of all cells in a core cube and only need to be carried out once, making the integral of eqn (C1) as fast as the n–e term integration.

That still leaves us with the integral in eqn (C2). Fortunately a large percentage of all the cells in a second-order VTDC were extremely small and did not contribute much to the integration. So when carrying out integration of eqn (C2), a cutoff limit was set to preselect those grid points which have values too small to be of any significance. For the results reported in this paper, a cutoff limit of 0.001 a.u. was used which still means 45–47% of total grid points being included in the e–e term integration. A series of test calculations with decreasing cutoff limits showed that although convergence was slow, enough accuracy has been reached for the purpose of geometry determination with the cutoff limit used. For the n–e term integration in eqn (3), complete second-order VTDCs were always used.

Acknowledgements

This research was supported by the Engineering and Physical Sciences Research Council (EPSRC) UK through the Single Cell Proteomics Project (EP/C54269X/1) and Proxomics Project (EP/I017887/1). SM gratefully acknowledges the support of the National Institute of Health through grant number GM-59230 and the National Science Foundation through Grant No. CHE-1058791.

References

1 W. Zhuang, T. Hayashi and S. Mukamel, *Angew. Chem., Int. Ed.*, 2009, **48**, 3750–3781.
2 P. M. Donaldson, R. Guo, F. Fournier, E. M. Gardner, L. M. C. Barter, C. J. Barnett, I. R. Gould, D. R. Klug, D. J. Palmer and K. R. Willison, *J. Chem. Phys.*, 2007, **127**, 114513.
3 P. M. Donaldson, R. Guo, F. Fournier, E. M. Gardner, I. R. Gould and D. R. Klug, *Chem. Phys.*, 2008, **350**, 201–211.
4 F. Fournier, E. M. Gardner, R. Guo, P. M. Donaldson, L. M. C. Barter, D. J. Palmer, C. J. Barnett, K. R. Willison, I. R. Gould and D. R. Klug, *Anal. Biochem.*, 2008, **374**, 358–365.
5 F. Fournier, E. M. Gardner, D. A. Kedra, P. M. Donaldson, R. Guo, S. A. Butcher, I. R. Gould, K. R. Willison and D. R. Klug, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 15352–15357.
6 F. Fournier, R. Guo, E. M. Gardner, P. M. Donaldson, C. Loeffeld, I. R. Gould, K. R. Willison and D. R. Klug, *Acc. Chem. Res.*, 2009, **42**, 1322–1331.
7 R. Guo, F. Fournier, P. M. Donaldson, E. M. Gardner, I. R. Gould and D. R. Klug, *Phys. Chem. Chem. Phys.*, 2009, **11**, 8417–8421.
8 R. Guo, M. Miele, E. M. Gardner, F. Fournier, K. M. Kornau, I. R. Gould and D. R. Klug, *Faraday Discuss.*, 2011, **150**, 161–174.
9 J. D. Jackson, *Classical electrodynamics*, Wiley, New York, 1999.
10 B. P. Krueger, G. D. Scholes and G. R. Fleming, *J. Phys. Chem. B*, 1998, **102**, 5378–5386.
11 A. Moran and S. Mukamel, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 506–510.
12 M. Cho, *Phys. Rev. A*, 2000, **61**, 023406.
13 A. D. Buckingham, *Adv. Chem. Phys.*, 1967, **12**, 107.
14 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03*, Revision E.01, Gaussian, Inc., Wallingford CT, 2004.
15 J. W. Ponder, *TINKER – Software Tools for Molecular Design*, Washington University, St. Louis, MO, 2011.
16 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.
17 L. Martinez, R. Andrade, E. G. Birgin and J. M. Martinez, *J. Comput. Chem.*, 2009, **30**, 2157–2164.
18 H. Friebolin and J. K. Becconsall, *Basic one- and two-dimensional NMR spectroscopy*, Wiley-VCH, Weinheim, 2005.
19 J.-P. Hansen and I. R. McDonald, *Theory of simple liquids*, Elsevier/Academic Press, Amsterdam/Boston, 2007.
20 R. McWeeny, *Methods of molecular quantum mechanics*, Academic Press, London; San Diego, 1989.
21 D. M. Brink and G. R. Satchler, *Angular momentum*, Clarendon Press, Oxford, 1962.
22 F. L. Pilar, *Elementary quantum chemistry*, McGraw-Hill, New York, 1990.
23 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.