University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Martin Centurion Publications | Research Papers in Physics and Astronomy

2014

Reconstruction of three-dimensional molecular structure from diffraction of laser-aligned molecules

Jie Yang
University of Nebraska—Lincoln, s-jyang6@unl.edu

Varun Makhija
Kansas State University, vmakhija@phys.ksu.edu

Vinod Kumarappan
Kansas State University, vinod@phys.ksu.edu

Martin Centurion
University of Nebraska-Lincoln, martin.centurion@unl.edu

Follow this and additional works at: https://digitalcommons.unl.edu/physicscenturion

 Part of the Atomic, Molecular and Optical Physics Commons

Yang, Jie; Makhija, Varun; Kumarappan, Vinod; and Centurion, Martin, "Reconstruction of three-dimensional molecular structure from diffraction of laser-aligned molecules" (2014). *Martin Centurion Publications*. 18.
https://digitalcommons.unl.edu/physicscenturion/18

This Article is brought to you for free and open access by the Research Papers in Physics and Astronomy at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Martin Centurion Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Reconstruction of three-dimensional molecular structure from diffraction of laser-aligned molecules

Jie Yang, Varun Makhija, Vinod Kumarappan, and Martin Centurion

Citation: Structural Dynamics **1**, 044101 (2014); doi: 10.1063/1.4889840
View online: http://dx.doi.org/10.1063/1.4889840
View Table of Contents: http://scitation.aip.org/content/aca/journal/sdy/1/4?ver=pdfcov
Published by the American Crystallographic Association, Inc.

Articles you may be interested in

Structural optimization of molecular clusters with density functional theory combined with basin hopping
J. Chem. Phys. **137**, 134106 (2012); 10.1063/1.4755994

Molecular structure determination from x-ray scattering patterns of laser-aligned symmetric-top molecules
J. Chem. Phys. **131**, 131101 (2009); 10.1063/1.3245404

Linear-scaling implementation of molecular response theory in self-consistent field electronic-structure theory
J. Chem. Phys. **126**, 154108 (2007); 10.1063/1.2715568

Reconstruction of Cracks from Eddy Current Signals Using Genetic Algorithm and Fuzzy Logic
AIP Conf. Proc. **820**, 775 (2006); 10.1063/1.2184605

IMRT verification by three-dimensional dose reconstruction from portal beam measurements
Med. Phys. **29**, 1847 (2002); 10.1118/1.1494988

Reconstruction of three-dimensional molecular structure from diffraction of laser-aligned molecules

Jie Yang,[1] Varun Makhija,[2] Vinod Kumarappan,[2] and Martin Centurion[1]
[1]*University of Nebraska—Lincoln, Lincoln, Nebraska 68588, USA*
[2]*James R. Macdonald Laboratory, Department of Physics, Kansas State University, Manhattan, Kansas 66506, USA*

(Received 28 March 2014; accepted 30 June 2014; published online 9 July 2014)

Diffraction from laser-aligned molecules has been proposed as a method for determining 3-D molecular structures in the gas phase. However, existing structural retrieval algorithms are limited by the imperfect alignment in experiments and the rotational averaging in 1-D alignment. Here, we demonstrate a two-step reconstruction comprising a genetic algorithm that corrects for the imperfect alignment followed by an iterative phase retrieval method in cylindrical coordinates. The algorithm was tested with simulated diffraction patterns. We show that the full 3-D structure of trifluorotoluene, an asymmetric-top molecule, can be reconstructed with atomic resolution. © *2014 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License*. [http://dx.doi.org/10.1063/1.4889840]

Most known molecular structures have been determined using X-ray diffraction from crystallized samples. For molecules that cannot be crystallized, structural determination in the gas phase becomes important. Gas electron diffraction (GED) has been a successful method for structure determination for molecules in the gas phase. However, it provides only 1-D information (interatomic distances), which limits the complexity of structures that can be retrieved. Laser alignment breaks the spherical symmetry of randomly oriented molecules in the gas phase, creating a possible solution for 3-D imaging without crystallization.[1,2] In 1-D alignment, a linearly polarized laser pulse is used to align the most polarizable axis of the molecule. For symmetric-top molecules, the molecule is free to spin with respect to this axis, while for asymmetric-top molecules, the laser affects the angular distribution about two molecular axes.[3] In 3-D alignment, either two laser pulses or an elliptically polarized laser pulse are used to align all three axes.[4,5] Most studies of structure determination from diffraction of aligned molecules have focused on 1-D alignment because it provides a higher degree of alignment. Existing reconstruction algorithms[6–12] are limited by several factors. First, they require an extremely high degree of alignment, which is not compatible with current diffraction experiments. In order to reach a high degree of alignment it is necessary to start with very cold molecules that can only be produced at densities that are several orders of magnitude below those needed for diffraction experiments. Second, each method is limited to a specific type of molecule. Third, only a 2-D projection of the 3-D structure can be reconstructed unless the symmetry of the molecule can be assumed to be known. Diffraction from a single molecule could in principle be used for 3-D imaging with atomic resolution; however, in general it is not possible to accumulate enough scattering events before the molecule is damaged (either with X-rays or electrons). X-ray free-electron lasers (X-FELs) could potentially be operated in single-molecule diffraction mode if sufficiently short and intense pulses become available.[13]

Structure determination from diffraction of aligned molecules is currently an area of great interest.[6–12,14–16] Even for the case of perfect 1-D alignment, a general method for structural reconstruction has not been demonstrated. The difficulty lies in the rotational averaging over all azimuthal angles with respect to the alignment axis.[3] Iterative phase retrieval in cylindrical coordinates can be used to reconstruct an azimuthal projection of molecules with a high degree of rotational symmetry.[6,7] A holographic method has been applied to symmetric-top molecules in

 © Author(s) 2014

which one of the atoms has a scattering cross section that is much higher than the others.[8] An angular-correlation-function method[9–12] can be used to reconstruct the axial projection of an object[11,12] and the 3-D structure of objects of cylindrical symmetry[9,10] with 1-D perfect alignment. It has been recently shown that angular correlation functions are able to reconstruct the 3-D structure from randomly oriented objects with helical symmetry[14] or icosahedral symmetry,[17] provided that single-shot diffraction patterns from a single molecule or a few molecules are available. Recently, the holographic method, combined with a genetic algorithm, has been used to reconstruct an azimuthal projection of a symmetric-top molecule from experimental data.[2]

In this Letter, we demonstrate a two-step method for retrieving the 3-D structure of trifluorotoluene ($C_6H_5CF_3$), an asymmetric-top molecule, using two diffraction patterns: one corresponding to partial 1-D alignment and the other to randomly oriented molecules. Perfect alignment corresponds to a delta-function-like angular distribution, while partially aligned molecules have a continuous angular distribution. The diffraction pattern of partially aligned molecules can be treated as a convolution of the diffraction pattern of perfectly aligned molecules with the angular distribution. The first step in our method is a deconvolution to retrieve the diffraction pattern corresponding to perfectly aligned molecules from the two input diffraction patterns, using a genetic algorithm.[2,18] The second step is to reconstruct the molecular structure from the deconvolved diffraction pattern. Briefly, an iterative phase retrieval algorithm in 3-D cylindrical space is used to reconstruct the full 3-D molecular structure. The genetic algorithm used in the first step is applicable for the case where a single axis of the molecule is aligned. For asymmetric-top molecules, 1-D alignment affects the angular distribution about two molecular axes. We were able to apply the genetic algorithm by selecting a time at which the angular distribution is narrow along the most polarizable axis and approximately random along the second-most polarizable axis.

The genetic algorithm converts the problem of achieving perfect alignment experimentally into a problem that is much easier to solve—achieving partial alignment with a known angular distribution. The genetic algorithm deconvolution requires the diffraction patterns and the angular distribution as inputs. The angular distribution can be measured experimentally or calculated if the moments of inertia and polarizability of the molecule are known.[16,19] If the angular distribution is not known, an estimated distribution can be used as a starting point, and then it is optimized by the genetic algorithm until the distribution that best matches the data is found. The result with lowest final error gives the best retrieval pattern.[19] Since most published reconstruction algorithms assume a pattern corresponding to perfect alignment to begin with, their functions are comparable to our second step, the iterative phase retrieval. Our iterative phase retrieval is *ab initio*—it does not require any structural information of the molecule, as opposed to standard GED and ultrafast electron diffraction (UED) methods were an initial model of the structure is needed.[20] This advantage will be particularly important for determining the structure of intermediate states in ultrafast reactions, for which it is difficult to obtain accurate theoretical models. In this algorithm, multiple cylindrical harmonics are used to resolve the symmetry of different groups in the molecule, and a 3-D atomicity constraint is applied to retrieve the full 3-D structure with atomic resolution. The atomicity constraint, first implemented in 1952,[21] takes advantage of the fact that the scattering arises from a limited number of point-like objects.

The temporal evolution of the angular distribution of the molecules was calculated in order to simulate the diffraction pattern corresponding to partial alignment. The calculation of the time-dependent rotational wave function for an asymmetric top molecule following an aligning laser pulse has been detailed in prior work.[16] Briefly, we solve the time-dependent Schrodinger equation (TDSE) for a rigid rotor interacting with a non-resonant laser pulse. The laser pulse excites only rotational transitions through the polarizability interactions from the initial ground vibronic state, and centrifugal distortion is not taken into account. A thermal distribution of initial asymmetric top states is used, and the lack of rotational symmetry of trifluorotoluene implies that each rotational state is equally weighted. Focal volume averaging is neglected.

The time-dependent Hamiltonian is

$$H = H_{rot} - \frac{1}{2}\mathbf{E}(t) \cdot (\alpha \mathbf{E}(t)), \tag{1}$$

where $H_{rot} = AJ_a^2 + BJ_b^2 + CJ_c^2$ is the field-free Hamiltonian for an asymmetric rigid rotor, α is the lab frame polarizability tensor of the molecule, and $\mathbf{E}(t)$ is the envelope of the laser pulse in the polarization direction, assumed to be Gaussian. The polarizabilities for trifluorotoluene are assumed to be same as that of toluene found in Ref. 22, and the rotational constants A, B, and C can be found in Ref. 23. Since the laboratory frame polarizability tensor contains both the Euler angles θ and χ, the wave function also depends on both angles and is calculated in the symmetric top basis. We follow the convention in Zare,[24] where θ, χ, and φ are the Euler angles, denoting the polar angle, the azimuthal angle in the molecular frame, and the azimuthal angle in the laboratory frame, respectively. The Boltzmann-weighted squares of the propagated wave functions for each initial state are summed to give the angular distribution ρ as a function of time during and after the pulse

$$\rho(\theta, \chi; t) = \sum_{J_0, \tau_0, M_0} e^{-E_{J_0,\tau_0,M_0}/kT} |\Psi_{J_0,\tau_0,M_0}(\theta, \chi; t)|^2, \tag{2}$$

where J_0, τ_0, and M_0 are the asymmetric top angular momentum indices for the initial states in the thermal distribution, E_{J_0,τ_0,M_0} is the field-free energy for a particular initial state, and $\Psi_{J_0,\tau_0,M_0}(\theta, \chi; t)$ is the time-dependent wave function for the same initial state. k is Boltzmann's constant and T is the rotational temperature. The distribution function ρ is calculated numerically on a 20×20 grid in θ and χ from the wave function.

The simulated laser pulse duration of 50 fs with a peak intensity of 15 TW/cm^2 and linear polarization lies in the regime of impulsive alignment. An initial rotational temperature of 0.1 K results in a peak alignment of $\langle\cos^2\theta\rangle = 0.58$, where "$\langle\rangle$" indicates averaging over all molecules. While the simulations were performed for impulsive alignment, it is expected that with adiabatic alignment a comparable degree of alignment could be reached with initial temperatures higher than 10 K, although in this case the distribution along χ might be less uniform. The temporal evolution of $\langle\cos^2\theta\rangle$ and $\langle\cos^2\chi\rangle$ is shown in Fig. 1(a). We simulate the diffraction patterns using the angular distribution at a time of 1.5 ps before the peak alignment, where $\langle\cos^2\theta\rangle = 0.56$ and $\langle\cos^2\chi\rangle = 0.45$, indicated by the vertical line in Fig. 1(a). $\langle\cos^2\theta\rangle = 1$ and $\langle\cos^2\chi\rangle = 1$ correspond to perfect alignment, while $\langle\cos^2\theta\rangle = 1/3$ and $\langle\cos^2\chi\rangle = 1/2$ correspond to a random distribution of the corresponding angles. This difference is because θ is a polar angle and χ is an azimuthal angle. The angular distribution $\rho(\theta, \chi)$ at this time is shown in Fig. 1(b). At this time, the distribution approximates single-axis alignment, to which the genetic algorithm can be applied. Even though the angular distribution along χ is not quite random, the reconstruction was successful. The diffraction patterns of $C_6H_5CF_3$ were simulated using a structure that was calculated using *ab initio* molecular orbital techniques.[25] The atomic scattering cross sections and scattering phases are listed in Ref. 26. Simulation of diffraction patterns from imperfectly aligned molecules is described in detail in Appendix A.

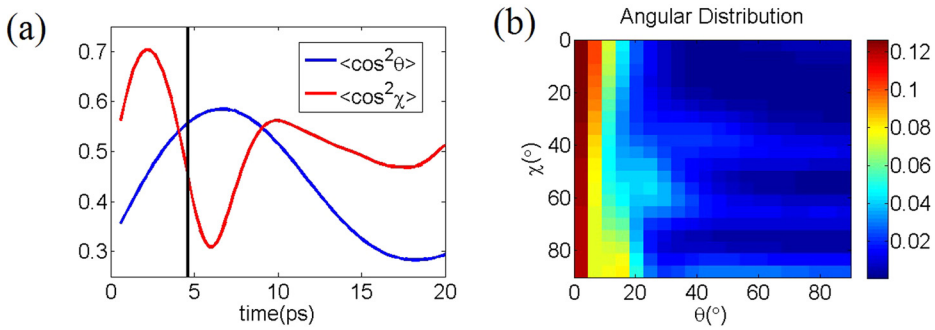


FIG. 1. (a) The time evolution of $\langle\cos^2\theta\rangle$ and $\langle\cos^2\chi\rangle$. The vertical line corresponds to the selected time of diffraction. (b) The 2-D angular distribution $\rho(\theta, \chi)$ at the selected time. Only 0°–90° in each angle is shown due to symmetry in the laser-alignment process.

In this manuscript, I_{random}, $I_{perfect}$, and $I_{partial}$ are used to represent the normalized diffraction intensities corresponding to randomly oriented, perfectly single-axis-aligned and partially single-axis-aligned molecules, respectively. All patterns are normalized to the sum of the scattering intensity over all the atoms in the molecule. For the simulation of $I_{perfect}$, $I_{partial}$, and I_{random}, a large set of single-molecule diffraction patterns are calculated for each possible orientation (θ, φ, χ) of the molecule, and the patterns are combined according to the angular distribution. The simulated $I_{perfect}$, $I_{partial} - I_{random}$, and I_{random} are shown in Fig. 2. We use a diffraction-difference pattern $I_{partial} - I_{random}$ that is commonly used in experiments to remove background noise.[20] The alignment axis is vertical in Fig. 2. The diffraction patterns in Figs. 2(a)–2(c) are displayed as a function of s, where $s = \frac{4\pi}{\lambda}\sin\left(\frac{\alpha}{2}\right)$ is the momentum change of diffracted electrons, λ is the electron wavelength, and α is the scattering angle. The diffraction pattern for partial alignment (Fig. 2(b)) shows significant blurring compared to the one from perfect alignment (Fig. 2(a)). We assume a 19 Å^{-1} s coverage in all patterns, which can be achieved experimentally with an electron beam with a kinetic energy of 100 keV and a detector with an angular aperture of 6.4°.

We have used a genetic algorithm that combines the patterns $I_{partial}$ and I_{random} to retrieve a pattern $I_{retrieved}$ that approximates $I_{perfect}$. For single-axis alignment, if the angular distribution $\rho(\theta)$ and $I_{perfect}$ are known, the corresponding $I_{partial}$ can be obtained by convolving $I_{perfect}$ with a point-spread function (PSF).[18] The PSF uses a known $\rho(\theta)$ to map each pixel in $I_{perfect}$ onto multiple pixels in $I_{partial}$, which causes the blurring seen in $I_{partial}$. It is impossible to find $I_{perfect}$ from $I_{partial}$ via a simple deconvolution with the PSF, since the PSF here is not shift-invariant. The PSF mapping depends only on $\rho(\theta)$, i.e., it does not depend on the structure.

In the genetic algorithm, the simulated patterns $I_{partial}$ and I_{random} with the corresponding PSFs are taken as the input. We start with a uniform pattern as the first guess of $I_{retrieved}$. In each iteration, a small random change is made in $I_{retrieved}$. The PSF mapping is then used to generate an $I'_{partial}$ and I'_{random} from the $I_{retrieved}$. The error is defined as the difference between the measured and the guessed patterns. Only changes that reduce the error are kept. The solution converges after about 10^5 iterations. The algorithm was run 20 times and the results were averaged to produce $I_{retrieved}$. The retrieved scattering amplitude $\sqrt{I_{retrieved}}$ is shown in Fig. 3(a), and the simulated $\sqrt{I_{perfect}}$ is shown in Fig. 3(b) for comparison. Figure 3 illustrates that the main features in $I_{perfect}$, such as the brightness and position of the local maxima, are retrieved accurately.

We have tested the genetic algorithm with different levels of shot noise. The same pattern is retrieved if there are at least 40 scattering events per pixel in a 70×70 pixels diffraction pattern shown in Figure 3. For example, using an electron source with an energy of 100 keV, the minimum number of scattered electrons in the diffraction pattern is 4×10^7. For a gas jet with a molecular density of $10^{15}\,\text{cm}^{-3}$ and a diameter of 200 μm, 10^{10} incident electrons are required. For an experiment operating at a repetition rate of 10 kHz and 2×10^3 electrons per pulse this means a total acquisition time of 500 s. The quality of the retrieved pattern starts to degrade for lower signal levels.

Once $I_{retrieved}$ was found, an iterative phase retrieval algorithm was used to determine the structure. For single-axis alignment, it is simpler to use cylindrical coordinates. For a single

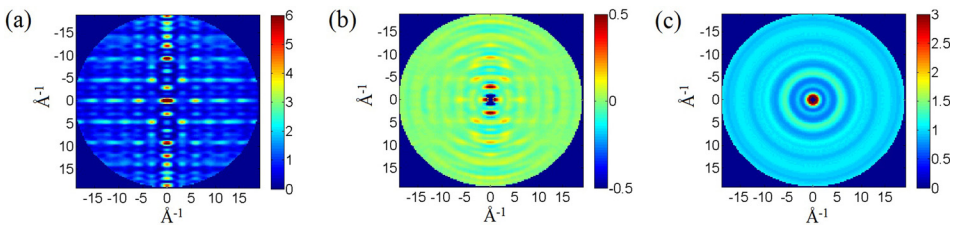


FIG. 2. The simulated (a) $I_{perfect}$; (b) $I_{partial} - I_{random}$; (c) I_{random}. The central region corresponding to $s < 0.6\,\text{Å}^{-1}$ in (b) and (c) are zeroed to simulate experimental conditions.

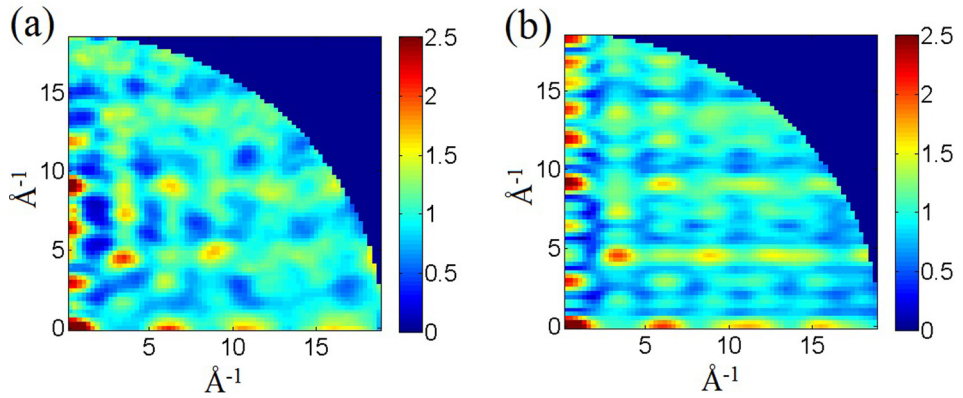


FIG. 3. (a) The genetically retrieved diffraction amplitude $\sqrt{I_{retrieved}}$. (b) The simulated diffraction amplitude $\sqrt{I_{perfect}}$.

molecule, the scattering amplitude can be expressed in terms of cylindrical reciprocal space coordinates (R, ψ, ζ)

$$F(R, \psi, \zeta) = \sum_m G_m(R, \zeta) \exp(im\psi), \tag{3}$$

where $G_m(R, \zeta)$ is the m_{th}-order expansion in cylindrical harmonics.[6,7] For single-axis-aligned molecules, the rotational averaging removes the dependence of the diffracted intensity $I(R, \zeta)$ on the azimuthal angle ψ

$$I(R, \zeta) = \langle |F(R, \psi, \zeta)|^2 \rangle_\psi = \sum_m |G_m(R, \zeta)|^2. \tag{4}$$

For a given structure, $G_m(R, \zeta)$ can be calculated by

$$G_m(R, \zeta) = \sum_k f_k \exp[i(\zeta z_k - m\varphi_k)] i^m J_m(Rr_k), \tag{5}$$

where f_k is the form factor of the k^{th} atom, (r_k, φ_k, z_k) gives the coordinates of the k^{th} atom in cylindrical coordinates, and J_m is the Bessel function of order m.[6,7]

The reciprocal and real space expansion functions are connected by Fourier-Hankel transforms

$$G_m(R, \zeta) = \frac{1}{\sqrt{2\pi}} \int\int r g_m(r, z) i^m J_m(Rr) \exp(i\zeta z) dr dz, \tag{6}$$

$$g_m(r, z) = \frac{1}{\sqrt{2\pi}} \int\int R G_m(R, \zeta) i^{-m} J_m(Rr) \exp(-i\zeta z) dR d\zeta, \tag{7}$$

where (r, φ, z) are the coordinates in the molecular frame and $g_m(r, z)$ is the m^{th}-order expansion of the object in cylindrical harmonics. We selected a square area in $I_{retrieved}$ to compute the fast Fourier-Hankel transform. The selected area corresponds to $13.5\,\text{Å}^{-1} \times 13.5\,\text{Å}^{-1}$ in reciprocal space, which results in a $0.23\,\text{Å} \times 0.23\,\text{Å}$ pixel size in real space. A central circle with a radius of $s = 0.6\,\text{Å}^{-1}$ is zeroed to simulate experimental conditions. The full 3-D object in real space $f(r, \varphi, z)$ can be reconstructed by summing over all the g_m

$$f(r, \varphi, z) = \sum_m g_m(r, z) \exp(im\varphi). \tag{8}$$

Each g_m can be obtained from the full 3-D object $f(r, \varphi, z)$

$$g_m(r,z) = \int f(r,\varphi,z)\exp(-im\varphi)d\varphi. \tag{9}$$

Each g_m contains different information about the structure. g_0 is the projection of the molecule on the z-r plane. For $m \neq 0$, each g_m contains the part of the object with m-fold rotational symmetry. Trifluorotoluene contains a benzene ring (2-fold rotational symmetry) and a trifluoromethyl group (3-fold rotational symmetry). The $m = \pm 2$, 4, 6,. orders contain the structure of the benzene ring, while the $m = \pm 3$, 6, 9,… orders contain the structure of CF_3. Other orders, such as $m = \pm 1$, 5, 7, are zero.

In a conventional phase retrieval algorithm, the amplitude is known and only the phase needs to be retrieved. Here, only the total diffracted intensity $I(R,\zeta)$ is known, and the amplitude and phase of each G_m must be retrieved. We include all orders up to $m = \pm 3$ in the algorithm. In order to construct the full 3-D structure, at least the $m = 0$, ± 2, ± 3 orders need to be included. We also included the $m = \pm 1$ orders, which are zero for trifluorotoluene, to ensure that the algorithm is run without any assumptions on the symmetry of the molecule.

The steps in the retrieval algorithm are similar to those found in Ref. 7. For the initial guess, we distributed the pattern uniformly over all m orders, $|G_m(R,\zeta)| = \sqrt{I(R,\zeta)/7}$, and associated each pattern with a random phase. In each iteration of the loop, an inverse Fourier-Hankel transform is applied on each G_m to calculate g_m using Eq. (7), and a 3-D object $f(r,\varphi,z)$ is generated by combining all g_m using Eq. (8). The real space constraints (described below) are applied on $f(r,\varphi,z)$, and then a new set of g_m are computed using Eq. (9). Finally, the G'_m are computed by Fourier-Hankel transform of g_m (Eq. (6)), and a Fourier space constraint is applied on G'_m to calculate G_m. The flowchart of the algorithm is shown in Appendix B.

The Fourier space constraint is

$$G_m(R,\zeta) = \sqrt{\frac{I(R,\zeta)}{\sum_m |G'_m(R,\zeta)|^2}} \times G'_m(R,\zeta). \tag{10}$$

The algorithm is run for 1600 iterations, separated in 20 cycles with 30 iterations of the Hybrid Input-Output algorithm (HIO)[27] followed by 50 iterations of Charge-Flipping (CF) algorithm[28,29] in each cycle. The feedback coefficient β is set to 0.9 for both algorithms.

The real space operations of the HIO and CF are given by

$$f_{n+1}(r,\varphi,z) = \begin{cases} f'_n(r,\varphi,z) & \text{if } f'_n(r,\varphi,z) \text{ satisfies the constraints} \\ f_n(r,\varphi,z) - \beta f'_n(r,\varphi,z) & \text{if } f'_n(r,\varphi,z) \text{ violates the constraints (HIO)} \\ -\beta f'_n(r,\varphi,z) & \text{if } f'_n(r,\varphi,z) \text{ violates the constraints (FC)} \end{cases}, \tag{11}$$

where f'_n is the 3-D object in the n^{th} iteration obtained using Eq. (8).

We have used three real space constraints: a loose support, positivity in the real part of f, and atomicity. It has been shown that using the support and positivity constraints, the azimuthal projection of the object g_0 can be retrieved.[7] We have introduced the additional constraint of atomicity, which has allowed us to reconstruct all 7 g_m. We used a fixed loose support of 30×9 ($z \times r$) pixels. The size of the object (excluding the hydrogen atoms) in g_0 is 25×6 ($z \times r$) pixels. The hydrogen atoms are not expected to be reconstructed because their scattering amplitude is much lower than that of carbon and fluorine. The third dimension φ is set to be 20 pixels deep. The second constraint, positivity of the real part of the object, was verified numerically using the known atomic scattering phases.[26]

The atomicity constraint is applied by converting the object from cylindrical coordinates $f(r,\varphi,z)$ to Cartesian coordinates $f(x,y,z)$. The local maxima are considered to be the atoms, i.e., a pixel is identified as an atom if its value is larger than the value of all 6 adjacent pixels and above a threshold that is set at 7% of the brightest pixel. A $3 \times 3 \times 3$ voxel area centered at the identified atom is considered to be satisfying the atomicity constraint. Finally, the selected voxels are converted back to cylindrical coordinates. The atomicity constraint is

updated every 20 iterations. To make the convergence faster, we apply a 2-D atomicity constraint in the last 10 out of every 100 iterations. The 2-D atomicity is a local-maxima search in g_0, the projection of molecule in the z-r plane. In 2-D atomicity, a pixel that is brighter than all 4 adjacent pixels is considered a local-maximum. The brightest 15 local maxima are identified as 2-D atoms. For each 2-D atom, an adjacent 3×3 pixels area is considered as satisfying the 2-D atomicity. The 2-D atomicity applies uniformly on the third dimension φ. More 2-D atoms are kept than the actual number of 2-D atoms in g_0, which is six. Note that an off-axis 2-D atom might represent one or more atoms in 3-D, depending on the rotational symmetry. Applying the atomicity constraint requires the atoms to have a separation of at least 2 pixels. In our case, the minimum bond length excluding hydrogen atoms is 1.32 Å, which corresponds to more than 5 pixels.

We define an error in real space E_{Real} that is used for monitoring the convergence of the algorithm

$$E_{Real} = 1 - \frac{\sum_{(r,\varphi,z)\ni S,P,A} |f(r,\varphi,z)|}{\sum_{(r,\varphi,z)\ni S} |f(r,\varphi,z)|}, \tag{12}$$

where S, P, and A represent the loose support, positivity, and atomicity constraints, respectively. $\sum_{(r,\varphi,z)\ni X} |f(r,\varphi,z)|$ is the sum of the absolute value of f over all pixels that satisfy the constraint X. The g_m corresponding to the minimum of E_{Real} is returned as the reconstruction result. The algorithm was run 20 times, and the 2-D projection of the structure g_0 was reconstructed successfully 15 out of 20 times. Out of the 15 successful 2-D reconstructions, 9 exhibit the same rotational symmetries, and those are considered successful 3-D reconstructions. The average of the 9 successful reconstructions is taken as the final reconstruction result.

Figures 4(a)–4(d) show the absolute value of the reconstructed g_m for $m = 0$, 1, 2, and 3, respectively. For $g_{m\neq 0}$, the discrete Hankel transform limits the reconstruction to $r > 0$.[30]

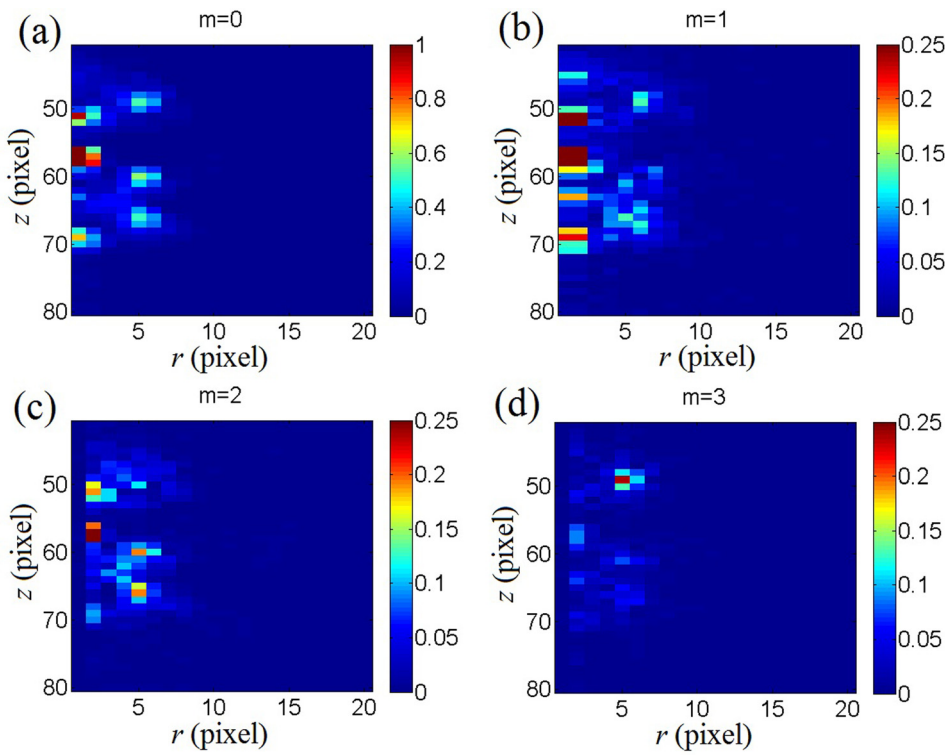


FIG. 4. (a)–(d) The reconstructed $g_0, g_1, g_2,$ and g_3 expansion of the structure. Each pixel is 0.23 Å.

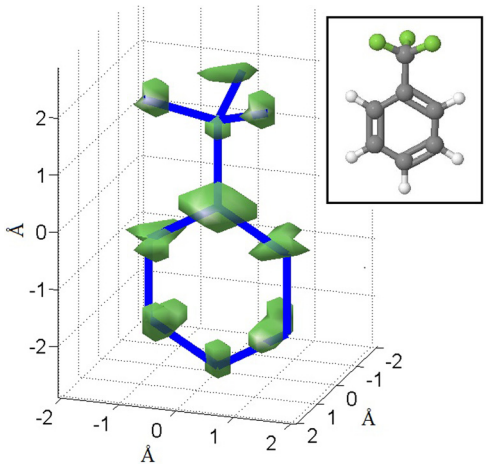


FIG. 5. The isosurface rendering of the reconstructed 3-D molecular structure. The overlapped blue sticks show the frame of the molecule with non-hydrogen atoms in both ends of each stick. The ball-and-stick model of the molecule[31] is shown in the inset.

Each off-axis 2-D atom in g_0 holds a certain rotational symmetry that can be retrieved from the $g_{m\neq 0}$. For example, at the location of the upper off-axis atom, g_3 is brighter than g_1 and g_2, indicating a 3-fold rotational symmetry. At the location of the two lower off-axis atoms g_2 is brighter than g_1 and g_3, indicating a two-fold rotational symmetry. In the 3-D reconstruction, for the on-axis atoms, only g_0 is used, and, for each off-axis atom, g_0 and the $g_{m\neq 0}$ that is brightest at that position are used. Only an area of 3×3 pixels is kept around the atom for the $g_{m\neq 0}$.

Figure 5 shows the retrieved 3-D object, along with the ball-and-stick model of the molecule shown in inset. The relative orientation between the benzene ring and the trifluoromethyl group depends on the relative phase between g_2 and g_3. It was not possible to accurately retrieve this relative orientation, most likely because the diffraction pattern is not very sensitive to this angle. Once the angle is fixed, the position of each atom deviates by less than 0.1 Å from the theoretical values. The structure is clearly atomized, and both the benzene ring and the trifluoromethyl group can be easily identified. For this reconstruction, including rotational symmetries up to $m = \pm 3$ was sufficient. More complex molecules might require more orders to be included and, thus, more information to be retrieved from a single diffraction pattern. Extending the method to more complex molecules might require providing additional information, for example, by increasing the coverage of the diffraction pattern or using inputs from theoretical models of the structure.

In summary, we have demonstrated a two-step method for retrieving the full 3-D molecular structure from diffraction patterns of partially aligned asymmetric top molecules. A genetic algorithm was used to retrieve a diffraction pattern corresponding to perfect alignment from patterns corresponding to partially aligned molecules. Then an iterative phase retrieval algorithm in cylindrical coordinates was used to reconstruct the full 3-D molecular structure from the retrieved diffraction pattern. We have successfully tested this algorithm on a small asymmetric-top molecule with simulated diffraction patterns. The data were simulated using a 50 fs alignment pulse with 15 TW/cm^2 peak intensity and molecules with a 0.1 K initial rotational temperatures. We would expect the algorithm work for shot noise level of no less than 40 scattering events per pixel, on a 70×70 pixel detector. Our method has the advantage that it does not require perfect alignment, which makes it compatible with diffraction experiments. In addition, our iterative phase retrieval is *ab initio*—it does not require any structural information of the molecule, this will be particularly important for determining the structure of intermediate states in ultrafast reactions. Moreover, the full 3-D structure can be reconstructed.

The work at UNL was supported by Chemical Sciences, Geosciences, and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy under

Grant No. DE-SC0003931. The work at KSU was supported by the same agency under Grant No. DE-FG02-86ER13491. This work was completed utilizing the Holland Computing Center of the University of Nebraska.

APPENDIX A: SIMULATION OF PARTIAL ALIGNMENT DIFFRACTION PATTERNS

Consider an experimental setup shown in Figure 6, where a single molecule is located at origin, aligned along the Y axis. An electron beam is propagating along Z axis, and the scattering polar angle and azimuthal angle are denoted as θ and φ. Assume the molecule consists of N atoms, the position of the *nth* atom in the laboratory frame is given by a vector $\vec{r_n}$, and the atomic scattering amplitude is $f_n(\theta)$. In this work, we use the amplitude and phase of atomic scattering in Ref. 26, and $\vec{r_n}$ is given by Ref. 25.

The diffraction of this single molecule is described a spherical wave, and the pattern record by a camera is given by the following equation:

$$I(\theta,\varphi) = \left|\sum_{n=1}^{N} f_n(\theta)\exp(ik|\vec{r_n} - \vec{r}(\theta,\varphi)|)\right|^2 \Big/ r^2(\theta,\varphi), \tag{A1}$$

where k is the wave number of the incident electron beam and $\vec{r}(\theta,\varphi)$ is a vector pointing from the origin to the camera pixel oriented at (θ,φ).

For a molecule that is not perfectly aligned, assume it is oriented along (α,β) with respect to the alignment axis. Here, α is the polar angle to Y axis, β is the azimuthal angle, or the angle to X axis in X-Z plane. The only parameter need to be changed in the previous formula is the position of atoms $\vec{r_n}$. This can be achieved by a rotation

$$I(\theta,\varphi) = \left|\sum_{n=1}^{N} f_n(\theta)\exp(ik|\hat{R}(\alpha,\beta)\cdot\vec{r_n} - \vec{r}(\theta,\varphi)|)\right|^2 \Big/ r^2(\theta,\varphi), \tag{A2}$$

where $\hat{R}(\alpha,\beta)$ is a rotation matrix that rotates the molecule from the initial orientation to (α,β).

For an ensemble of molecules, the total diffraction pattern is an incoherent sum of diffraction patterns of each molecule. For a given angular distribution $F(\alpha,\beta)$, the diffraction pattern can be written as

$$I(\theta,\varphi) = \int_0^{\pi}\int_0^{2\pi}\left|\sum_{n=1}^{N} f_n(\theta)\exp\left(ik|\hat{R}(\alpha,\beta)\cdot\vec{r_n} - \vec{r}(\theta,\varphi)|\right)\right|^2 \frac{F(\alpha,\beta)}{r^2(\theta,\varphi)}d\beta d\alpha. \tag{A3}$$

In our simulation, we discretely sample the angular distribution $F(\alpha,\beta)$ and use sums to replace integrals in Eq. (A3)

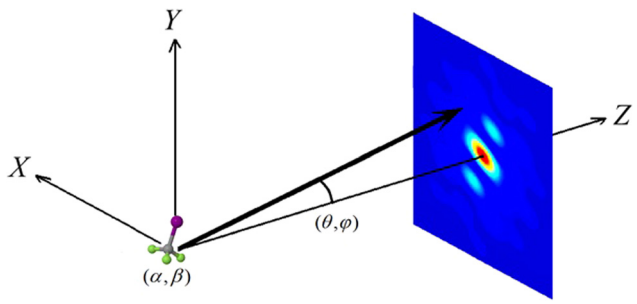


FIG. 6. The diagram of single molecule scattering.

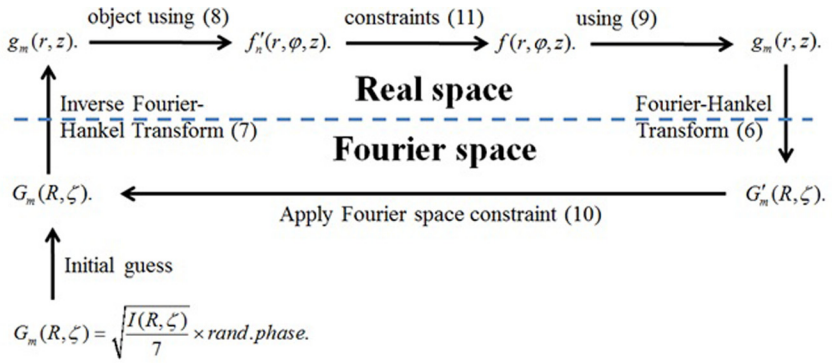


FIG. 7. The iterative phase retrieval algorithm flowchart.

$$I(\theta,\varphi) = \sum_{\alpha,\,\beta} \left| \sum_{n=1}^{N} f_n(\theta) \exp\left(ik|\hat{R}(\alpha,\beta)\cdot\vec{r_n} - \vec{r}(\theta,\varphi)|\right) \right|^2 \frac{F(\alpha,\beta)}{r^2(\theta,\varphi)}\,. \tag{A4}$$

APPENDIX B: FLOWCHART OF THE ITERATIVE PHASE RETRIEVAL ALGORITHM

Figure 7 shows the flowchart of the iterative phase retrieval algorithm.

[1]J. C. H. Spence, K. Schmidt, J. S. Wu, G. Hembree, U. Weierstall, B. Doaka, and P. Fromme, Acta Crystallogr., Sect. A: Found. Crystallogr. **61**, 237–245 (2005).
[2]C. J. Hensley, J. Yang, and M. Centurion, Phys. Rev. Lett. **109**, 133202 (2012).
[3]H. Stapelfeldt and T. Seideman, Rev. Mod. Phys. **75**, 543 (2003).
[4]J. J. Larsen, K. Hald, N. Bjerre, H. Stapelfeldt, and T. Seideman, Phys. Rev. Lett. **85**, 2470 (2000).
[5]K. K. Lee, D. M. Villeneuve, P. B. Corkum, A. Stolow, and J. G. Underwood, Phys. Rev. Lett. **97**, 173991 (2006).
[6]D. K. Saldin, V. L. Shneerson, D. Starodub, and J. C. H. Spence, Acta Crystallogr., Sect. A: Found. Crystallogr. **66**, 32–37 (2010).
[7]D. Starodub, J. C. H. Spence, and D. K. Saldin, Proc. SPIE **7800**, 78000O (2010).
[8]P. J. Ho, D. Starodub, D. K. Saldin, V. L. Shneerson, A. Ourmazd, and R. Santra, J. Chem. Phys. **131**, 131101 (2009).
[9]D. K. Saldin, V. L. Shneerson, M. R. Howells, S. Marchesini, H. N. Chapman, M. Bogan, D. Shapiro, R. A. Kirian, U. Weierstall, K. E. Schmidt, and J. C. H. Spence, New J. Phys. **12**, 035014 (2010).
[10]D. K. Saldin, H. C. Poon, M. J. Bogan, S. Marchesini, D. A. Shapiro, R. A. Kirian, U. Weierstall, and J. C. H. Spence, Phys. Rev. Lett. **106**, 115501 (2011).
[11]D. K. Saldin, H. C. Poon, V. L. Shneerson, M. Howells, H. N. Chapman, R. A. Kirian, K. E. Schemidt, and J. C. H. Spence, Phys. Rev. B. **81**, 174105 (2010).
[12]V. Elser, New J. Phys. **13**, 123014 (2011).
[13]R. Neutze, R. Wouts, D. Spoel, E. Weckert, and J. Hajdu, Nature **406**, 752 (2000).
[14]H.-C. Poon, P. Schwander, M. Uddin, and D. K. Saldin, Phys. Rev. Lett. **110**, 265505 (2013).
[15]P. R. Reckenthaeler, M. Centurion, W. Fuß, S. A. Trushin, F. Krausz, and E. E. Fill, Phys. Rev. Lett. **102**, 213001 (2009).
[16]S. Pabst, P. J. Ho, and R. Santra, Phys. Rev. A **81**, 043425 (2010).
[17]D. K. Saldin, H.-C. Poon, P. Schwander, M. Uddin, and M. Schmidt, Opt. Express **19**, 17318–17335 (2011).
[18]J. Yang, C. J. Hensley, and M. Centurion, Proc. SPIE **8845**, 884502 (2013).
[19]X. Ren, V. Makhija, and V. Kumarappan, Phys. Rev. A **85**, 033405 (2012).
[20]R. Srinivasan, V. Lobastov, C. Ruan, and A. Zewail, Helv. Chim. Acta. **86**, 1761–1799 (2003).
[21]D. Sayre, Acta Crystallogr. **5**, 60–65 (1952).
[22]V. V. Ilyushin, L. B. Favero, W. Caminati, and J. Grabow, ChemPhysChem **11**, 2589–2593 (2010).
[23]R. Chelli and P. Procacci, J. Chem. Phys. **117**, 9175 (2002).
[24]R. N. Zare, *Angular Momentum: Understanding Spatial Aspects in Chemistry and Physics* (Wiley-Interscience, New York, 1988).
[25]P. C. Chen and C. W. Wu, J. Phys. Chem. **99**, 15023–15027 (1995).
[26]H. L. Cox and R. A. Bonham, J. Chem. Phys. **47**, 2599 (1967).
[27]J. R. Fienup, Appl. Opt. **21**, 2758 (1982).
[28]J. S. Wu, U. Weierstall, J. C. H. Spence, and C. T. Koch, Opt. Lett. **29**, 2737 (2004).
[29]G. Oszlanyi and A. Suto, Acta Crystallogr., Sect. A: Found. Crystallogr. **64**, 123–134 (2008).
[30]M. Leutenegger, *MATLAB Toolbox,* see http://www.mathworks.com/matlabcentral/fileexchange/13371-hankel-transform (2006).
[31]Jmol: an open-source Java viewer for chemical structures in 3D, see http://www.jmol.org/.