







Applying Bayesian inference and deterministic anisotropy to retrieve the molecular structure $|\Psi(\mathbf{R})|^2$ distribution from gas-phase diffraction experiments

Kareem Hegazy^{1,2}, Varun Makhija³, Phil Bucksbaum^{1,2,4}, Jeff Corbett⁵, James Cryan², Nick Hartmann⁶, Markus Ilchen^{2,7,8}, Keith Jobe⁵, Renkai Li⁹, Igor Makasyuk⁵, Xiaozhe Shen⁵, Xijie Wang⁵, Stephen Weathersby⁵, Jie Yang¹⁰ & Ryan Coffee^{2,6}

Currently, our general approach to retrieving molecular structures from ultrafast gas-phase diffraction heavily relies on complex ab initio electronic or vibrational excited state simulations to make conclusive interpretations. Without such simulations, inverting this measurement for the structural probability distribution is typically intractable. This creates a so-called inverse problem. Here we address this inverse problem by developing a broadly applicable method that approximates the molecular frame structure $|\Psi(\mathbf{R}, t)|^2$ distribution independent of these complex simulations. We retrieve the vibronic ground state $|\Psi(\mathbf{R})|^2$ for both simulated stretched NO_2 and measured N_2O . From measured N_2O , we observe 40 mÅ coordinate-space resolution from 3.75 \AA^{-1} reciprocal space range and poor signal-to-noise, a 50X improvement over traditional Fourier transform methods. In simulated NO_2 diffraction experiments, typical to high signal-to-noise levels predict 100–1000X resolution improvements, down to 0.1 mÅ. By directly measuring the width of $|\Psi(\mathbf{R})|^2$, we open ultrafast gas-phase diffraction capabilities to measurements beyond current analysis approaches. This method has the potential to effectively turn gas-phase ultrafast diffraction into a discovery-oriented technique to probe systems that are prohibitively difficult to simulate.

¹ Department of Physics, Stanford University, Stanford, CA 94305, USA. ² Stanford PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. ³ Department of Chemistry and Physics, University of Mary Washington, Fredericksburg, VA 22401, USA. ⁴ Department of Applied Physics, Stanford University, Stanford, CA 94305, USA. ⁵ SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. ⁶ Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. ⁷ European XFEL, Holzkoppel 4, 22869 Schenefeld, Germany. ⁸ Universität Kassel, Heinrich-Plett-Str. 40, 34132 Kassel, Germany. ⁹ Department of Engineering Physics, Tsinghua University, Beijing 100084, China. ¹⁰ Department of Chemistry, Tsinghua University, Beijing 100084, China. ✉email: KHegazy@slac.stanford.edu; Coffee@slac.stanford.edu

Ultrafast molecular gas-phase diffraction, from either x-rays^{1,2} or electrons^{3–6}, is a vital tool for retrieving time-dependent molecular structures. In elastic molecular gas-phase diffraction experiments, x-rays or electrons scatter off of electrons and nuclei, with differing proportionality. Each pairwise atomic distance creates a pattern of scattered x-rays or electrons as a function of their transverse momentum q . The measured diffraction pattern is the sum of all such contributions, this is orientationally averaged over the lab frame ensemble distribution. We lose pairwise directional information and thus the ability to explicitly distinguish individual atomic distances. Consequently, directly inverting diffraction patterns for the molecular structure is generally intractable, this is a so-called inverse problem. Typically, we avoid this inverse problem and retrieve both the molecular structures and the molecular frame orientations by simulating the forward excited state process. These are generally time-dependent ab initio electronic and vibrational excited state simulations that explore a large parameter space (rovibration, structure, and electronic state) with trajectory bifurcations due to effects like conical intersections^{7–10}. We refer to such simulations as complex simulations, that are typically validated through comparisons with measured diffraction patterns or pair-distribution functions (PDFs – a weighted histogram of pairwise distances). Consequently, ultrafast gas-phase diffraction is generally limited by the ability to perform these complex simulations. We aim to expand diffraction measurements for high-resolution reconstructions of molecular structure probability distribution $|\Psi(\mathbf{R}, t)|^2$ without relying on complex molecular dynamics simulations by effectively solving this inverse problem with a statistical interpretation.

A variety of studies sought to reduce reliance on complex simulations, but are either limited in the systems they address or quickly run into the curse of dimensionality. Fourier transforming the time dependence exposes dissociative and vibronic signals^{11–13} but it is insensitive to classes of isomerizations. Methods employing ensemble anisotropy have garnered much interest^{14–23} yet they struggle to get sub-Angstrom resolution and the full 3d structure for generic molecular structures. Optimization methods, while capable of exposing large-scale motion, are susceptible to local minima²¹. Pattern matching measured data against sampled isomers^{24–26} becomes intractable for moderately large molecules due to the curse of dimensionality. For example, a molecule with N_{atoms} atoms has $3N_{\text{atoms}} - 6$ degrees of freedom. To independently sample each degree of freedom 10 times would require $10^{3N_{\text{atoms}} - 6}$ structures, becoming intractable for molecules with 7 or more atoms. Simulations reduce the structure-space of isomers to select, but this trade-off requires previous knowledge²⁴ that potentially imparts biases.

We employ insights from molecular ensemble anisotropy methods, applied statistics, and machine learning principles to address the inverse problem and the curse of dimensionality to approximate the molecular structure probability density $|\Psi(\mathbf{R}, t)|^2$. It is important to note that instead of sampling individual molecular structures and comparing single structures to the measured data, we are sampling entire $|\Psi(\mathbf{R}, t)|^2$ probability distributions. We access the molecular frame by decomposing measured data onto anisotropic components. Then, we iteratively approximate $|\Psi(\mathbf{R}, t)|^2$ with a statistical approach uniquely suited for high repetition-rate diffraction facilities. We observe that resolution strongly improves with signal-to-noise (SNR) much faster than increasing the q range beyond moderate values. Unlike the PDF approach, this method retrieves the molecular distances and angles required to define a unique molecular structure.

In our method, we recover the molecular frame through time-dependent ensemble anisotropy^{20,27–32}. One rotates into the molecular frame with the lab frame Euler angles $\theta_1^{(\text{lf})}$ (polar), $\phi_1^{(\text{lf})}$

(azimuthal), and $\chi_1^{(\text{lf})}$ (Fig. 1a). An induced rotational wavepacket creates ensemble anisotropy given by $|\Psi(\theta_1^{(\text{lf})}, \phi_1^{(\text{lf})}, \chi_1^{(\text{lf})}, t)|^2$. Axis distribution moments (ADMs)^{29,33,34} are the coefficients in the Wigner D matrix expansion of $|\Psi(\theta_1^{(\text{lf})}, \phi_1^{(\text{lf})}, \chi_1^{(\text{lf})}, t)|^2$

$$A_{mk}^l(t) = \frac{2l+1}{8\pi^2} \langle \Psi(t) | D_{mk}^l(\phi_1^{(\text{lf})}, \theta_1^{(\text{lf})}, \chi_1^{(\text{lf})}) | \Psi(t) \rangle. \quad (1)$$

These ADMs (Eq. (1)) describe the ensemble of molecular frame orientations with respect to the lab frame. When

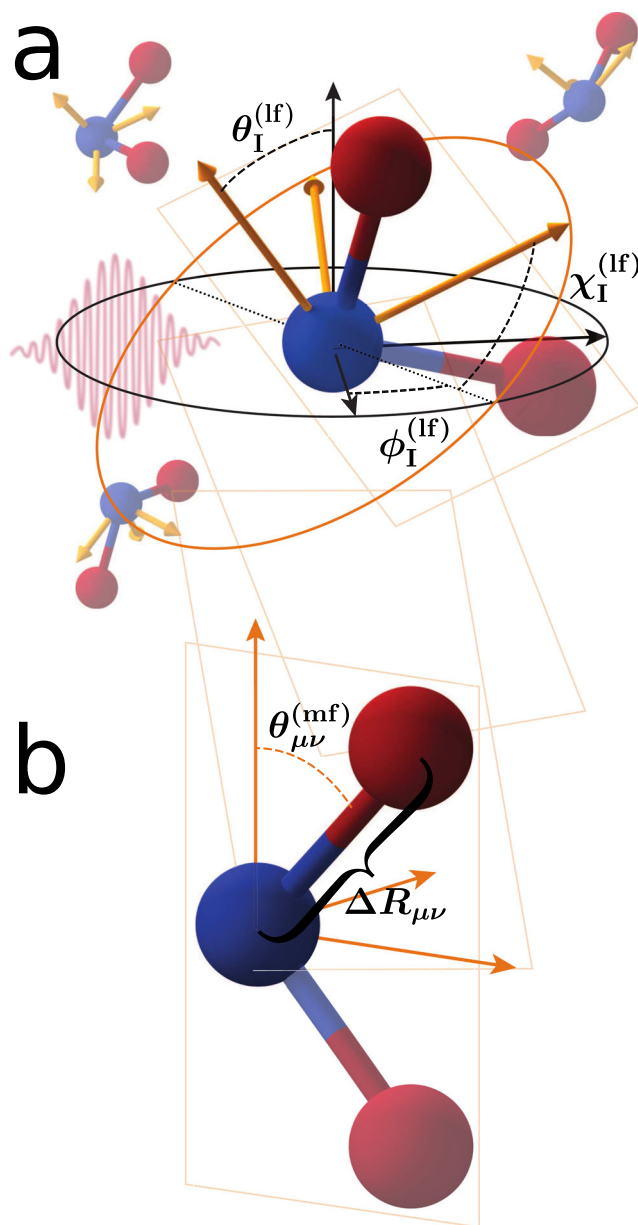


Fig. 1 Correspondence between the lab and molecular frame. Our analysis considers each pairwise distance independently and we define the origin of both the lab and molecular frames by one of the pairwise vectors. For the highlighted NO bond, the nitrogen atom (blue) defines the origin. **a** The lab frame is defined by the laser polarization ($\hat{\mathbf{z}}$) and propagation direction ($\hat{\mathbf{y}}$). **b** The molecular frame is defined by the molecule's rovibronic ground state principal moments of inertia, where the molecular A, B, and C axes define $\hat{\mathbf{z}}^{(\text{mf})}$, $\hat{\mathbf{y}}^{(\text{mf})}$, and $\hat{\mathbf{x}}^{(\text{mf})}$. Here the NO is described by $\Delta R_{\mu\nu}$, $\theta_{\mu\nu}^{(\text{mf})}$, and $\phi_{\mu\nu}^{(\text{mf})}$ which correspond to its distance, polar angle, and azimuthal angle respectively. One accesses the molecular frame by rotating the lab frame by the Euler angles $\theta_1^{(\text{lf})}$, $\phi_1^{(\text{lf})}$, and $\chi_1^{(\text{lf})}$.

calculating the ADMs, the l , m , and k are difference and sum of quantum numbers between rotational eigenstates, respectively for the total angular momentum, the projection onto the lab frame z -axis, and the projection onto the molecular frame z -axis. These ADMs transform the lab frame into the molecular frame by decomposing the measured lab frame anisotropy into $C_{lmk}(q)$ coefficients, which are dependent on molecular frame pairwise distances and angles $\Delta\mathbf{R}_{\mu\nu} = \mathbf{R}_\mu - \mathbf{R}_\nu = [\Delta R_{\mu\nu}, \theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})}]$ (Fig. 1b). The PDF is not directly sensitive to these angles. After impulsively aligning the molecular ensemble, Fig. 2 illustrates how transient anisotropy (panels b and c) provides constraints on these Euler angles and consequently the molecular frame (panels d–g). For example, at 39.25 ps the anisotropy provides simultaneous constraints on $\theta_1^{(\text{lf})}$ and $\chi_1^{(\text{lf})}$. At 39.68 ps, $\chi_1^{(\text{lf})}$ (the molecular frame azimuthal plane) is highly constrained. At 39.85 ps the ensemble is well localized in $\theta_1^{(\text{lf})}$, resolving measurements along the molecular frame z -axis. Here, $P(\phi_1^{(\text{lf})})$ is uniform due to cylindrical symmetry imparted by a linearly polarized pulse.

To effectively invert the molecular diffraction pattern and approximate $|\Psi(\mathbf{R}, t)|^2$, we use Bayesian inference. Bayesian inference describes a class of statistical inference techniques using Bayes's Theorem to update one's model based on observed data³⁵. We first approximate $|\Psi(\mathbf{R}, t)|^2$ as the probability distribution $P(\mathbf{R}, t|\Theta, C)$, which is parameterized by the molecular structure degrees of freedom Θ . Using Bayesian inference, we then relate $P(\Theta|C)$ to the measured molecular diffraction pattern. With this framework, we use Markov-chain Monte Carlo (MCMC) techniques to build $P(\Theta|C)$ and tackle the curse of dimensionality by efficiently sampling structures most consistent with the measured $C_{lmk}(q)$. This method is unbiased and naturally avoids regions in our sampling space that are inconsistent with the $C_{lmk}(q)$. We retrieve $P(\mathbf{R}, t|\Theta, C)$ with neither the PDF nor complex molecular dynamics simulations since we will analytically relate the molecular frame pairwise distances and angles to the $C_{lmk}(q)$. Further intuition is provided in Supplementary Note 4 and Ref. 36.

Instead of complex molecular dynamics simulations this method has fewer simulation requirements. In this method's simplest form, when probing structural dynamics it only requires the much more tractable simulation of the rovibronic ground state structure to define the molecular frame. When measuring the equilibrium vibronic ground state, one does not require a priori knowledge of the structure they wish to find. This is because each sampled structure will define a new molecular frame. When using anisotropy components, we require time-dependent rotational simulations for the ADMs. This requires rotational constants and the molecular polarizability, all of which can be measured or calculated from the rovibronic ground state structure. When applying this method to excited states, we require the transition dipole, which is also measured or calculated from the rovibronic ground state structure. As discussed later, depending on the desired accuracy, one must select a functional form for $P(\mathbf{R}, t|\Theta, C)$ based on a priori knowledge of the excitation or use normal distributions as a “first-order” approximation.

In this manuscript, we validate these principles by retrieving $|\Psi(\mathbf{R})|^2$ for the vibronic ground states of both simulated NO₂ and measured N₂O rotational wavepackets. Here NO₂, an asymmetric top, serves as a test case to show our method's broad capabilities and behavior under various experimental conditions. Furthermore, we validated these capabilities with measured N₂O data from the ultrafast MeV electron diffraction facility at SLAC (UED). We chose these molecules to specifically be amenable to conventional methods since triatomics do not suffer significantly from the curse of dimensionality. In this lower dimensional realm, we benchmark and validate our method against

conventional methods with intentions to later expand to larger molecules. In the following, all simulations and equations correspond to ultrafast electron diffraction experiments but are easily extended to x-ray diffraction.

In this work, we rigorously and qualitatively describe this method in addition to quantitatively benchmarking both its advantages and shortcomings. We provide intuition and mathematically describe how induced anisotropy accesses the molecular frame structural angles ($\theta_{\mu\nu}^{(\text{mf})}$ and $\phi_{\mu\nu}^{(\text{mf})}$) and how to retrieve this molecular frame structure using Bayesian inference. We evaluate this method on simulated and measured data, showing how $P(\mathbf{R}|\Theta, C)$ significantly improves upon the traditional Fourier limited PDF. Firstly, $P(\mathbf{R}|\Theta, C)$ unambiguously defines a unique molecular mean structure without complex molecular dynamics simulations. This is generally not possible from the PDF alone. Secondly, we report pairwise distance resolutions of order 10 mÅ and down to 0.1 mÅ from measured and simulated data, respectively. These resolutions are respectively a factor of 50 and 1000 times smaller than their corresponding PDF resolutions. Thirdly, we investigate this method's behaviors and systematic errors as a function of experimental factors and analysis choices. We find this procedure depends more strongly on signal-to-noise than it does by extending measured momentum transfer. Fourthly, we demonstrate how this method expands ultrafast gas-phase diffraction experiments to quantitatively measure additional parameters, such as the width of $|\Psi(\mathbf{R}, t)|^2$. Lastly, we describe how one can apply this method to excited state dynamics. With these advancements, this method has the potential to expand ultrafast gas-phase diffraction into a more discovery-oriented technique, one that is free of complex excited state simulation limitations and is applicable to currently inaccessible molecular systems.

Results

Both the simulated NO₂ and measured N₂O diffraction patterns are from the SLAC UED facility⁶. Elastic electron diffraction is primarily sensitive to the nuclei. Diffraction from electronic transients occurs within the removed low q regions, removed because of experimental conditions. We use the independent atom approximation because we are primarily concerned with the nuclear structure. Given an anisotropic distribution of molecules, the measured diffraction pattern $\langle I(\mathbf{q}, t) \rangle_{\text{rigid}}$ (derived in Supplementary Note 2) is given by

$$\langle I(\mathbf{q}, t) \rangle_{\text{rigid}} = \mathcal{I} \left(\sum_{\mu} |f_{\mu}(q)|^2 + \sum_{\mu, \nu: \mu \neq \nu} \text{Re} \left\{ f_{\mu}(q) f_{\nu}^*(q) \sum_l \frac{32\pi^3 l!}{2l+1} \times \sum_{m,k} (-1)^k \right. \right. \quad (2)$$

$$\left. \left. \underbrace{Y_l^m(\theta_q^{(\text{lf})}, \phi_q^{(\text{lf})})}_{\text{Lab Frame}} \underbrace{|\Psi(0)| j_l(q\Delta R_{\mu\nu}) Y_l^{-k}(\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})})}_{\text{Molecular Frame Structure}} \underbrace{|\Psi(0)| \mathcal{A}_{mk}^l(t)|_{\text{rigid}}}_{\text{Anisotropy}} \right\} \right)$$

where $f_{\mu}(q)$ is the scattering amplitude of the μ^{th} atom, $j_l(q\Delta R_{\mu\nu})$ are the spherical Bessel functions of the first kind, \mathcal{I} is the diffraction beam intensity, and the momentum transfer vector is given by $\mathbf{q} = [q, \theta_q^{(\text{lf})}, \phi_q^{(\text{lf})}]$. Here we applied the rigid rotor approximation. The simulated and measured diffraction patterns that Eq. (2) describe are shown in Fig. 3 (row a).

The $C_{lmk}(q)$ coefficients (Eq. (3)) distill the molecular frame information from Eq. (2)

$$C_{lmk}(q) = \mathcal{I} \sum_{\mu, \nu: \mu \neq \nu} \text{Re} \left\{ f_{\mu}(q) f_{\nu}^*(q) (-1)^k \frac{32\pi^3 l!}{2l+1} \times \underbrace{|\Psi(0)| j_l(q\Delta R_{\mu\nu}) Y_l^{-k}(\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})})}_{\text{Molecular Frame Structure}} |\Psi(0)| \right\} \quad (3)$$

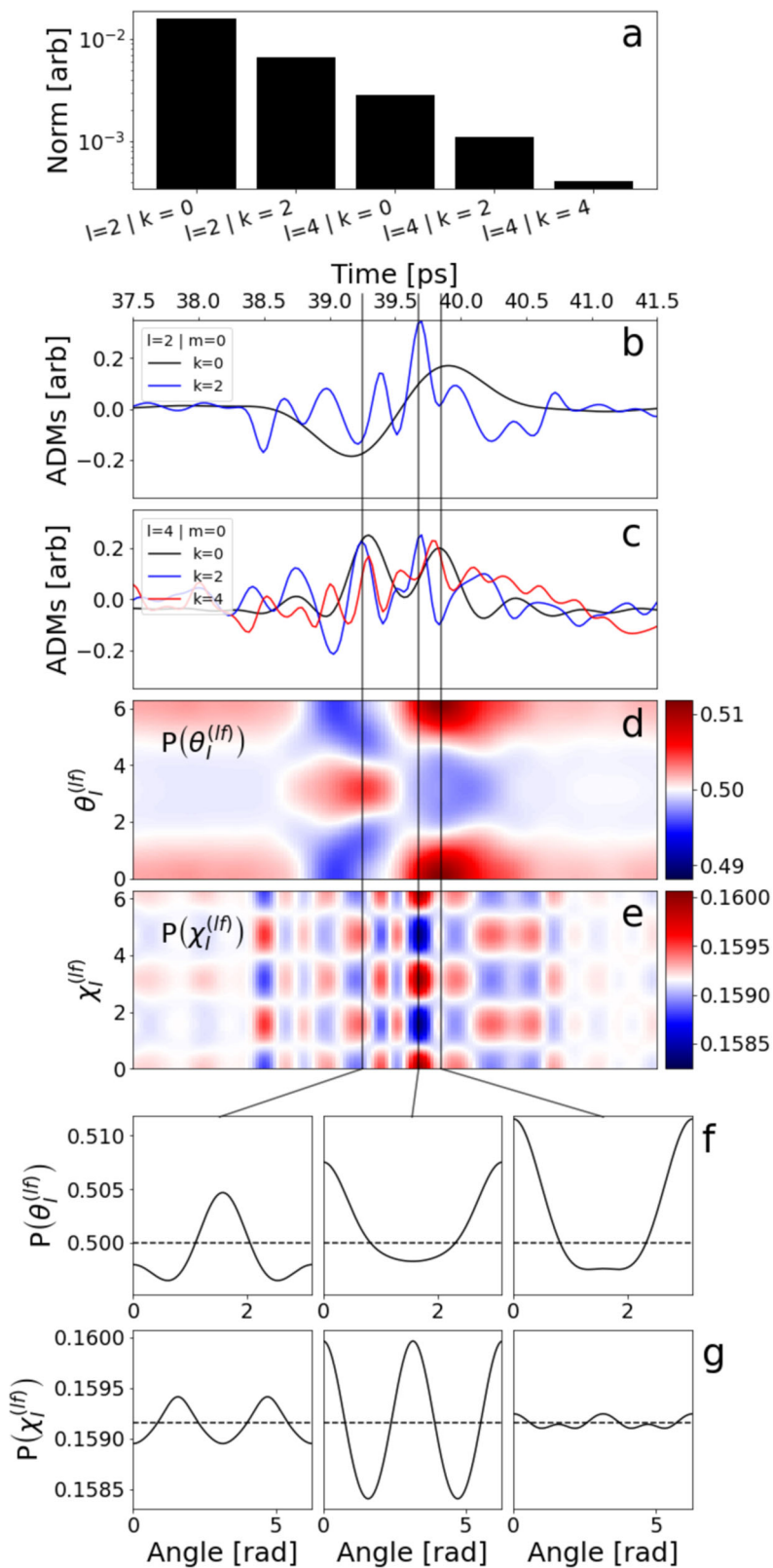


Fig. 2 Axis distribution moments and ensemble anisotropy. The Axis distribution moments (ADMs) encapsulate the ensemble anisotropy which provides various constraints on the molecular frame as a function of time. The ADMs are parameterized by the three angular momentum quantum numbers l , m , and k which correspond to the total angular momentum, the projection along the lab frame (\hat{z}) axis, and projection along the molecular frame (\hat{z}) axis respectively. **a** We show the square norms of the ADMs and **(b, c)** highlight the time dependence of these normalized ADMs. **d, e** We show the time-dependent ensemble anisotropy probability distributions for $\theta_l^{(lf)}$ and $\chi_l^{(lf)}$, respectively. **f, g** We show illustrative line-outs of these Euler angle distributions for $\theta_l^{(lf)}$ and $\chi_l^{(lf)}$, respectively, with isotropy indicated by the dashed lines.

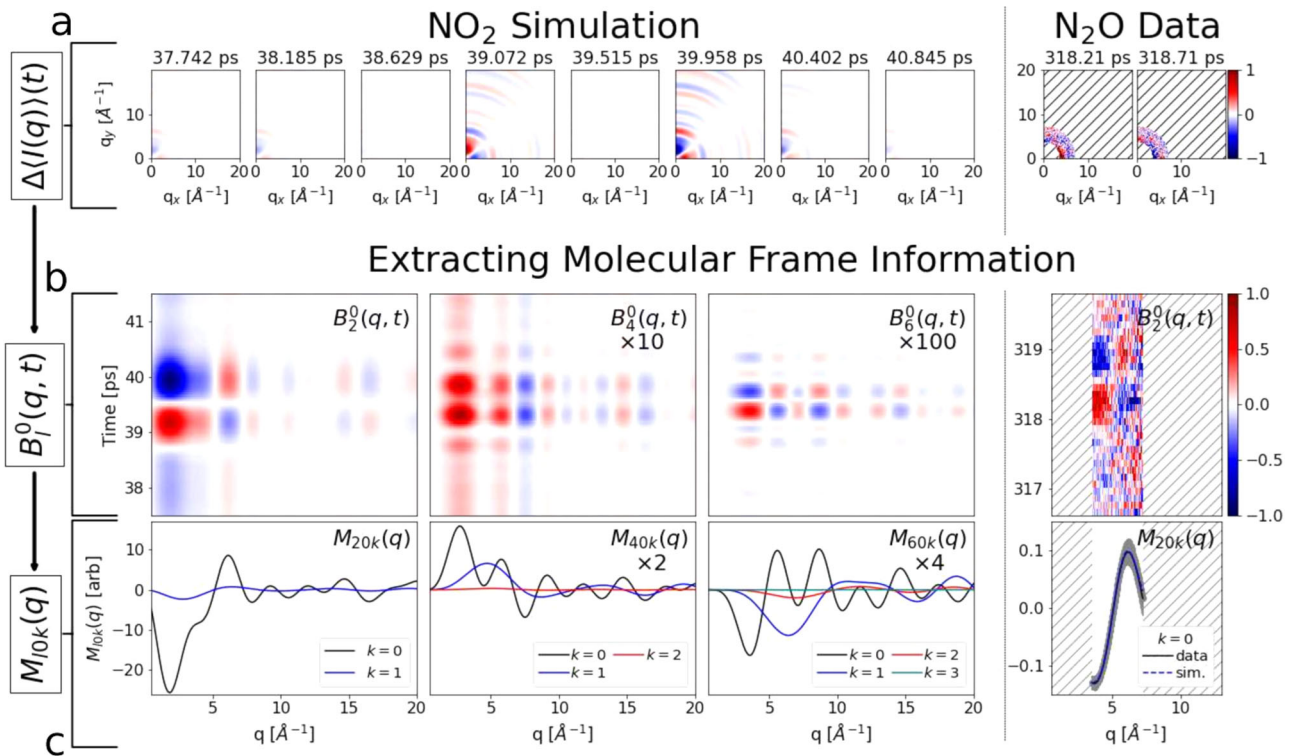


Fig. 3 Analysis to access the molecular frame signal. To access the molecular structure term, in the molecular frame, one must remove the lab frame anisotropy dependence and fit onto the ADMs. For the NO₂ simulation (left) and N₂O data (right), we illustrate the analysis steps. **a** One first measures the difference diffraction pattern $\langle \Delta I(\mathbf{q}) \rangle(t)$. **b** Removing the detector angular dependence, one retrieves time-dependent lab frame anisotropy components $B_l^m(q, t)$. **c** Removing the time-dependent ensemble anisotropy (ADMs) yields the molecular frame $M_{lmk}(q)$ coefficients. All as described in the text. We note that in the N₂O data (right) we have limited visibility of data due to experimental limitations illustrated by the hash lines area.

$$C_{lmk}^{(\text{calc})}(q) = \mathcal{I} \int \sum_{\mu, \nu: \mu \neq \nu} \text{Re} \left\{ f_{\mu}(q) f_{\nu}^*(q) (-1)^k \frac{32\pi^3 i^l}{2l+1} \right. \\ \left. \times \underbrace{j_l(q\Delta R_{\mu\nu}) Y_l^{-k}(\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})})}_{\text{Molecular Frame Structure}} \right\} P(\mathbf{R}|\Theta, C) d\mathbf{R} \quad (4)$$

$$M_{lmk}(q) = \frac{C_{lmk}(q)}{\sum_{\mu} |f_{\mu}(q)|^2} \quad (5)$$

where Eq. (3) is our measurement and Eq. (4) describes our measurement in terms of the desired Θ parameters. The modified $M_{lmk}(q)$ in Eq. (5) (which remove the q^{-4} dependence) are shown in Figs. 4b and 3 (row c), respectively for NO₂ and N₂O. To get these $C_{lmk}(q)$ we simulated the ADMs, accounting for centrifugal distortion in the N₂O case.

To determine the functional form of $P(\mathbf{R}|\Theta, C)$, we define our simulated stretched NO₂ molecule in the ground vibrational state and we observe that 99.99% of the N₂O molecules occupy the vibrational ground state (Supplementary Note 9). The normal distribution, $P^{(\mathcal{N})}(\mathbf{R}|\Theta, C)$ (Eq. (9)), is a good description of both our NO₂ and N₂O vibronic ground state $|\Psi(\mathbf{R})|^2$ distribution as the normal distribution is the ground state eigenfunction of the harmonic oscillator.

$$P(\mathbf{R}|\Theta, C) \approx |\Psi(\mathbf{R})|^2 \quad (6)$$

$$P^{(\delta)}(\mathbf{R}|\Theta, C) = \delta(\Theta^{(\delta)} - \mathbf{R}) \quad (7)$$

$$\Theta^{(\delta)} = [\langle \text{NO}^{(1)} \rangle, \langle \text{NO}^{(2)} \rangle, \langle \angle \text{ONO} \rangle] \quad (8)$$

$$P^{(\mathcal{N})}(\mathbf{R}|\Theta, C) = \frac{1}{\sqrt{2\pi}^{N_{\text{dof}}} \prod_{i=0}^{i < N_{\text{dof}}} \Theta_{2i+1}^{(\mathcal{N})}} \\ \times \exp \left\{ -\frac{1}{2} \sum_{i=0}^{i < N_{\text{dof}}} \left(\frac{\Theta_{2i}^{(\mathcal{N})} - \mathbf{R}_i}{\Theta_{2i+1}^{(\mathcal{N})}} \right)^2 \right\} \quad (9)$$

$$\Theta^{(\mathcal{N})} = [\langle \text{NO}^{(1)} \rangle, \sigma(\text{NO}^{(1)}), \langle \text{NO}^{(2)} \rangle, \sigma(\text{NO}^{(2)}), \langle \angle \text{ONO} \rangle, \sigma(\angle \text{ONO})] \quad (10)$$

The Θ parameters are the means and standard deviations of pairwise distances and angles that parameterize this multivariate normal distribution (Eq. (10)).

Using the statistical nature of our experiment, we effectively invert Eq. (4) for the probability distribution of Θ parameters given the measured $C_{lmk}(q)$ ($P(\Theta|C)$). Here, $P(\Theta|C)$ is the posterior distribution that is proportional to the likelihood function

$$P(C|\Theta) = \left[\prod_{lmk,q} \frac{1}{\sigma_{lmk}(q) \sqrt{2\pi}} \right] \\ \times \left\{ \exp -\frac{1}{2} \sum_{lmk,q} \left(\frac{C_{lmk}(q) - C_{lmk}^{(\text{calc})}(q, \Theta)}{\sigma_{lmk}(q)} \right)^2 \right\} \quad (11)$$

which is the probability of observing the measured $C_{lmk}(q)$ coefficients assuming the Θ parameters are true. In Eq. (11) the $\sigma_{lmk}(q)$ are the standard error of the mean from the distribution of measured $C_{lmk}(q)$ coefficients. Applying an MCMC technique to Eq. (11), we build $P(\Theta|C)$ and find the optimal Θ parameters (Θ^*) for $P(\mathbf{R}|\Theta, C)$ as described in the Methods section.

In our main result, we illustrate our method's efficacy by retrieving Θ^* and consequently $P^{(\mathcal{N})}(\mathbf{R}|\Theta^*, C)$ from both

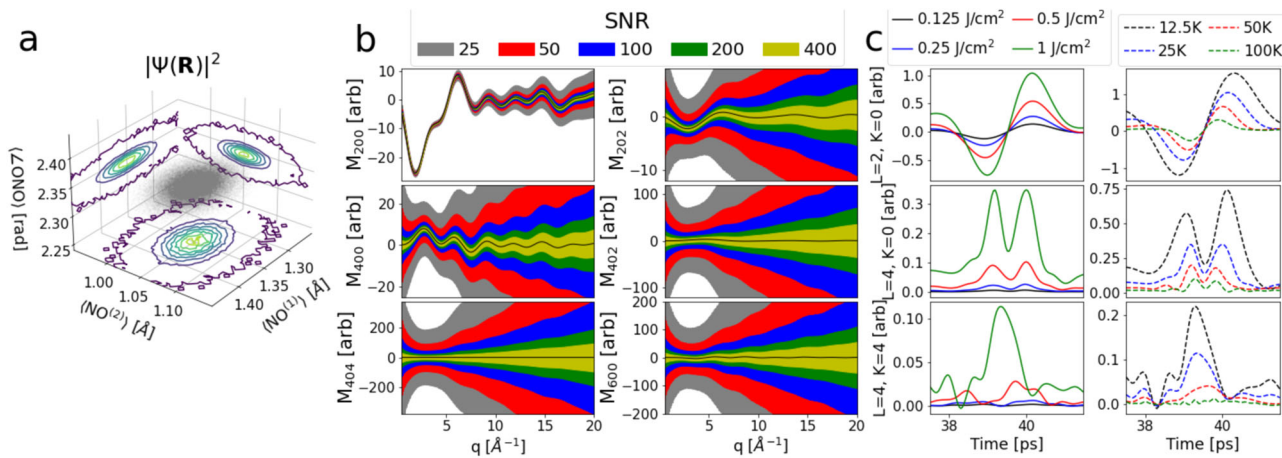


Fig. 4 Simulated NO₂ data at various experimental conditions. For simulated NO₂ we defined a $|\Psi(\mathbf{R})|^2$ distribution, from which we calculated the $C_{lmk}(q)$ under various experimental conditions. **a** The simulated NO₂ distribution is a multivariate normal distribution that we use to calculate the simulated NO₂ responses ($C_{lmk}(q)$ and $M_{lmk}(q)$). **b** The $M_{lmk}(q)$, shown for various signal-to-noise ratios (SNR) for the case of an ensemble temperature of 100 K and kick fluence of 1 J/cm², isolate the molecular frame structure. **c** The ADMs have two dependencies: pump strength (constant ensemble temperature of 25 K) on the left and temperature (constant pump fluence of 1 J/cm²) on the right.

Table 1 Retrieved molecular frame structure parameters for simulated NO₂.

Θ Parameters	Input	$P^{(\mathcal{A})}(\Theta C)$		$P^{(\delta)}(\Theta C)$	
		Θ^*	σ^Θ	Θ^*	σ^Θ
$\langle \text{NO}^{(1)} \rangle$ [Å]	1.35	1.3500	0.0005	1.3509	0.0004
$\sigma(\text{NO}^{(1)})$ [Å]	0.03	0.030	0.004	–	–
$\langle \text{NO}^{(2)} \rangle$ [Å]	1.05	1.0500	0.0006	1.0485	0.0005
$\sigma(\text{NO}^{(2)})$ [Å]	0.02	0.020	0.007	–	–
$\langle \angle \text{ONO} \rangle$ [rad]	2.34	2.340	0.001	2.3401	0.0007
$\sigma(\angle \text{ONO})$ [rad]	0.01	0.01	0.02	–	–

Our approximation of $|\Psi(\mathbf{R})|^2$ ($P(\mathbf{R}|\Theta, C)$) is parameterized by molecular frame distances, angles, and their corresponding widths (Θ parameters). The optimal parameters, denoted as Θ^* , correspond to the mode of $P(\Theta|C)$. We provide the retrieved Θ^* parameters along with their corresponding resolutions for the simulated NO₂. The input Θ parameters are those used to simulate the NO₂ $C_{lmk}(q)$ coefficients with a signal-to-noise ratio (SNR) of 400. The retrieved Θ^* parameters are those found when applying $P^{(\mathcal{A})}(\mathbf{R}|\Theta, C)$ and $P^{(\delta)}(\mathbf{R}|\Theta, C)$ to the NO₂ $C_{lmk}(q)$ simulated using $P^{(\mathcal{A})}(\mathbf{R}|\Theta, C)$. The σ^Θ values are the resolution of Θ^* and the uncorrelated widths of $P^{(\mathcal{A})}(\Theta|C)$ and $P^{(\delta)}(\Theta|C)$, respectively.

Table 2 Retrieved molecular frame structure parameters for measured N₂O.

	Θ^* Literature	Θ^*	σ^Θ
$\langle \text{N}^{\text{T}}\text{N}^{\text{C}} \rangle$ [Å]	1.128	1.14	0.04
$\sigma(\text{N}^{\text{T}}\text{N}^{\text{C}})$ [Å]	–	0.08	0.03
$\langle \text{N}^{\text{C}}\text{O} \rangle$ [Å]	1.184	1.18	0.04
$\sigma(\text{N}^{\text{C}}\text{O})$ [Å]	–	3×10^{-8}	0.03
$\langle \angle \text{NNO} \rangle$ [rad]	3.142	3.14	0.06
$\sigma(\angle \text{NNO})$ [rad]	–	6×10^{-12}	0.06

We provide the optimal molecular frame pairwise distance and angle parameters (Θ^*) for the measured N₂O dataset. The Θ^* parameters correspond to the mode of $P(\Theta|C)$. The resolution of Θ^* (σ^Θ) is the standard deviation of the 1d uncorrelated projection of $P^{(\mathcal{A})}(\Theta|C)$. We also provide the corresponding literature values for the vibronic ground state^{37, 38}, denoted as $\Theta^*_{\text{Literature}}$.

simulated NO₂ and measured N₂O $C_{lmk}(q)$ coefficients. We further investigate our method's behavior and sensitivity to varying experimental conditions for the simulated NO₂ system. Finally, we show how our Bayesian inference method significantly improves real-space resolution.

Molecular structure distribution retrieval. To retrieve $P^{(\mathcal{A})}(\mathbf{R}|\Theta^*, C)$ for simulated NO₂ and measured N₂O data we first built the posteriors $P^{(\mathcal{A})}(\Theta|C)$ shown in Fig. 5a and c, respectively. Figure 5b and d show $P^{(\mathcal{A})}(\mathbf{R}|\Theta^*, C)$ for NO₂ and the simulated PDF for N₂O, respectively. Tables 1 and 2 give the extracted Θ^* (the most probable Θ parameters) and σ^Θ , respectively, for N₂O and NO₂. For the NO₂ simulation, the SNR is 400. For NO₂, $P^{(\mathcal{A})}(\Theta|C)$'s resolution (σ^Θ) for the nuclear distances and angles is ~ 0.5 mÅ and fully encompasses the ground truth values. Despite the largely flat $\langle \angle \text{ONO} \rangle$ distribution, Θ^* still converges on the ground truth values. For N₂O data, the retrieved $P^{(\mathcal{A})}(\Theta|C)$ encompasses the previously measured results of the vibronic ground state^{37, 38}. The resolution of this distribution is of order 10 mÅ even with our limited q range of $[3.5, 7.25]$ Å^{−1} and the very poor SNR. Moreover, the retrieved $\langle \angle \text{NNO} \rangle$ is π and we

resolve the ~ 50 mÅ difference between the N^TN^C and N^CO bond distances (Table 2). The retrieved widths $\sigma(\text{N}^{\text{T}}\text{N}^{\text{C}})$ and $\sigma(\angle \text{NNO})$ are unphysical due to the limited q range, as discussed later. Compared to the PDF (Fig. 5d), with a ~ 2 Å Fourier resolution, this method improves resolution by a factor of 50. In the PDF, the missing low and high q components produce ringing artifacts in this inverse Fourier transform because of the incomplete Fourier space. This confuses the PDF results as they are not positive definite and falsely indicate population at large distances.

We observe (Fig. 5a and c) that Θ^* does not correspond to the mean or mode of most 1-dimensional projections of $P^{(\mathcal{A})}(\Theta|C)$. This is due to the nonlinearity and correlations of $P(\Theta|C)$ in Θ space. This illustrates the importance of finding Θ^* in this correlated space since the structure parameters are indeed correlated.

Exploring experimental effects and systematics. The measured q range is a critical component of gas-phase ultrafast diffraction, determining the information content and the PDF's resolution. When expanding this range, Figs. 6a and 7, we observe resolution (σ^Θ) improvements only until ~ 8 Å^{−1}, after which it plateaus. This indicates that after a modest q range our method is not very sensitive to further increases. The false correlations between Θ parameters (Fig. 7e), still, continue to decline as we increase this range. The plotted correlation in Fig. 7e is between all 6 Θ

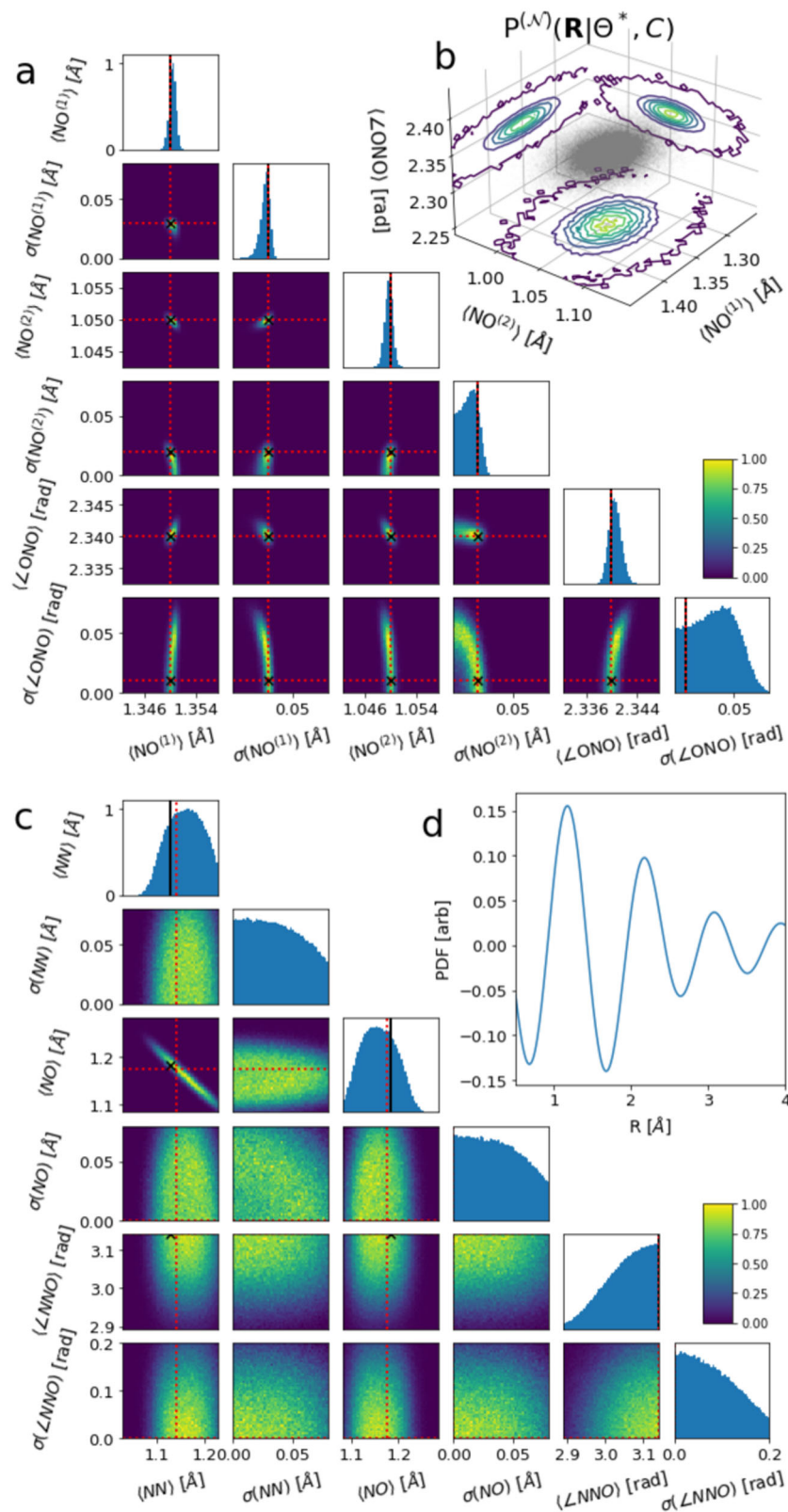


Fig. 5 Retrieving $P^{(N)}(\Theta|C)$, $P^{(N)}(R|\Theta^*, C)$, and the molecular structure parameters. We successfully retrieve the multivariate posterior $P^{(N)}(\Theta|C)$ for NO₂ and N₂O from which we find Θ^* . From $P^{(N)}(\Theta|C)$, we observe the distribution of Θ parameters: the mean and standard deviations of the pairwise distances and angles that define $P^{(N)}(R|\Theta, C)$. **a** We show the 1d and 2d projections of $P^{(N)}(\Theta|C)$ distributions for the simulated NO₂ response. **b** The recovered $P^{(N)}(R|\Theta^*, C)$ is what we compare to the simulated $|\Psi(R)|^2$ in Fig. 4a. The red dashed lines indicate the retrieved mode (Θ^*), while the black "X" and solid black lines indicate the ground truth, respectively. **c** We show the 1d and 2d projections $P^{(N)}(\Theta|C)$ distributions for N₂O data, though only using the $C_{200}(q)$ contribution. The black "X" and solid black lines indicate previously measured values for N₂O^{37,38}. **d** For comparison, we provide the simulated Pairwise Distribution Function (PDF) from the same q range.

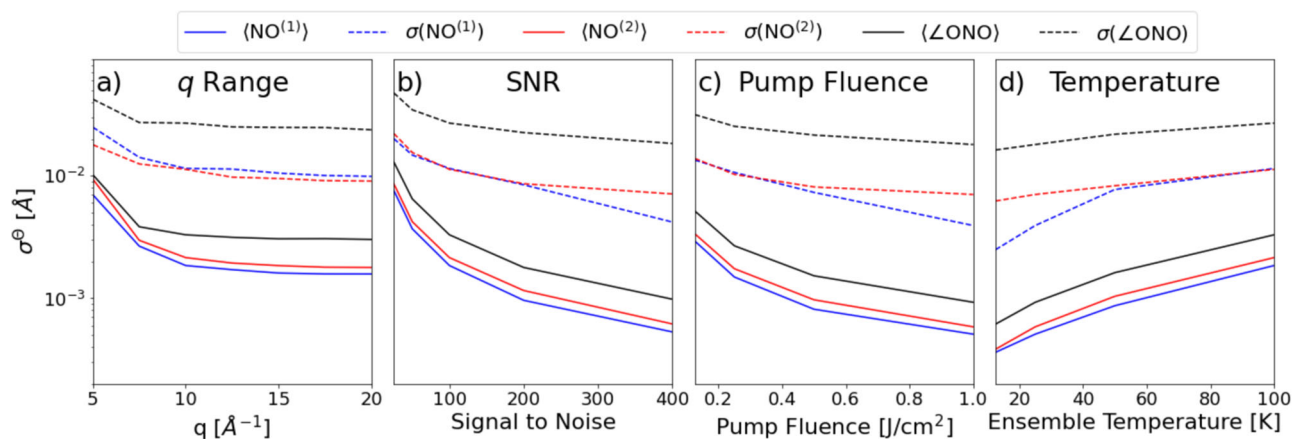


Fig. 6 The effects of various experimental parameters on $P^{(N)}(\Theta|C)$. Varying experimental parameters affect the resolution (width) of $P^{(N)}(\Theta|C)$, but our method is most sensitive to the measured signal-to-noise ratio (SNR). The SNR is the geometric mean of $C_{000}(q)/\sigma_{000}(q)$ between $0.5 < q < 4 \text{ \AA}^{-1}$. The uncorrelated widths of $P^{(N)}(\Theta|C)$, denoted by σ^Θ , change with increasing the (a) q range and (b) the SNR. The axis distribution moments, and consequently σ^Θ , are dependent on (c) pump fluence (ensemble temperature of 25 K) and (d) the molecular ensemble temperature (pump fluence of 1 J/cm^2).

parameters. The correlations seen in Fig. 7a and c are termed false correlations since the simulated $|\Psi(\mathbf{R})|^2$ is a multivariate normal distribution with a diagonal covariance matrix. Increasing the measured reciprocal range q provides more information about the system and reduces these correlations, seen in Fig. 7a, c, and e. Regardless of the sampled q ranges, we still retrieve the ground truth Θ parameters and consequently $P^{(N)}(\mathbf{R}|\Theta, C)$: Fig. 7b and d.

When varying the SNR, Fig. 6b, σ^Θ rapidly decreases with increasing SNR. Increasing SNR by an order of magnitude decreases σ^Θ by an order of magnitude for pairwise distances and angles. This strong and continuous dependence indicates that our method is sensitive to SNR due to our statistical interpretation. Although $P^{(N)}(\Theta|C)$ becomes more peaked, the general shape from the correlations does not change since higher SNR improves resolution but does not add more information, in terms of the q range.

Increasing the induced rotational coherence and lowering the ensemble temperature rapidly improves resolution (Fig. 6c and d) similar to increasing SNR. In Fig. 6c, the gas was at 25 K while varying the rotational coherence. In Fig. 6c, the pump fluence was 1 J/cm^2 while varying the ensemble temperature. Increasing the rotational coherence and decreasing the temperature increases the magnitude and complexity of the ADMs (Fig. 4c). This is because higher average pump fluences induce larger rotational coherence and lowering the ensemble temperature diminishes the spread of initial rotational states that incoherently interfere. The result is an increase in signal, a larger SNR, and consequently the similarly continuous behavior in Fig. 6b.

Generally, when varying the q range, SNR levels, pump fluence, and ensemble temperature we find the pairwise distances' σ^Θ to be of order 1 m\AA ; for the width parameters, σ^Θ is order 10 m\AA . Our retrieved Θ^* values are generally within a relative error of $\sim 10^{-7}$ and $\sim 10^{-3}$ from the ground truth values for structural and width parameters, respectively. This resolution is often ~ 100 times better than PDF-based methods because our statistical treatment is highly sensitive to SNR.

Aside from experimental parameters, we investigate systematics induced by incorrectly selecting the functional form of $P(\mathbf{R}|\Theta, C)$. We assert the simulated NO_2 vibronic ground state $|\Psi(\mathbf{R})|^2$ distribution is a multivariate normal distribution (Fig. 4a). We evaluate both $P^{(N)}(\Theta|C)$ (Eq. (9)) and $P^{(\delta)}(\Theta|C)$ (Eq. (7)) on this simulation, and in Fig. 8 we compare their 1d projections as a function of q range. The $P^{(N)}(\Theta|C)$ distribution consistently encompasses the correct values, but the $P^{(\delta)}(\Theta|C)$ distribution

fails to do so for q ranges of $[0.5, 7.5]$, $[0.5, 10]$, and $[0.5, 12.5] \text{ \AA}^{-1}$. These q ranges are shown in Fig. 8c. This is because $P^{(\delta)}(\mathbf{R}|\Theta, C)$ assumes a single molecule response can describe a signal averaged over an ensemble of structures. With increasing q ranges, $P^{(\delta)}(\Theta|C)$ converges in an unstable fashion on the ground truth (Fig. 8b), unlike the smooth convergence in $P^{(N)}(\Theta|C)$ (Fig. 8a). We note that for NO_2 , retrieving $P^{(\delta)}(\Theta|C)$ is ~ 100 times faster than $P^{(N)}(\Theta|C)$, which respectively take order 10 s to 1 min and 1 h to 1 day on 10 CPUs. This is because $P^{(\delta)}(\mathbf{R}|\Theta, C)$ doesn't have to sum over structures in Eq. (4). Supplementary Note 7 and Ref. 36 provides plots and further discussion of these results.

Effects of Bayesian inference. Our method retrieves the labeled pairwise distances with ~ 100 times better resolution than the PDF. This is due to our statistical treatment using Bayesian inference where each lmk and q contribution is itself an independent probability distribution; each is an experiment of its own. The Metropolis-Hastings algorithm's (MHA) discrimination power grows exponentially with more $C_{lmk}(q)$, which increases the magnitude of the negative exponent in the relative ratio of likelihood functions $P(C|\Theta)$ (Eq. (11)). Our method therefore heavily relies on $\sigma_{lmk}(q)$ and $C_{lmk}(q)$ (seen in Fig. 6b). Statistical noise increases $\sigma_{lmk}(q)$, making $P(\Theta|C)$ wider (Fig. 6b), while systematic errors in $C_{lmk}(q)$ shifts the centroid of $P(\Theta|C)$ (Fig. 5c). Supplementary Note 5 describes our method for consistently accounting for both statistical and systematic errors. The PDF error adds in quadrature in $\sigma_{lmk}(q)$; its scale is set by the largest error bar and disproportionately suffers from poorly measured data points. Conversely, MHA amplifies the contribution of high precision measurements while reducing contributions from poorly measured data points by weighting each term in the likelihood by $1/\sigma_{lmk}(q)$ (Eq. (11)).

Our Bayesian inference approach expands the utility of gas-phase ultrafast diffraction to measure previously inaccessible variables. Given $P(\mathbf{R}|\Theta, C)$ is a generic function parameterized by Θ , one can introduce variables through Θ by selecting a $P(\mathbf{R}|\Theta, C)$ that depends on them. Here, we expanded the measurable parameters of gas-phase ultrafast diffraction to include the width of $|\Psi(\mathbf{R})|^2$ in $P^{(N)}(\mathbf{R}|\Theta, C)$, shown in Fig. 5 and given in Table 1. Depending on one's system and desired accuracy, a priori knowledge is needed to select the form of $P(\mathbf{R}|\Theta, C)$, e.g. harmonic oscillator eigenstates for vibrational excited states. Outside of the vibronic ground state, $P^{(N)}(\mathbf{R}|\Theta, C)$

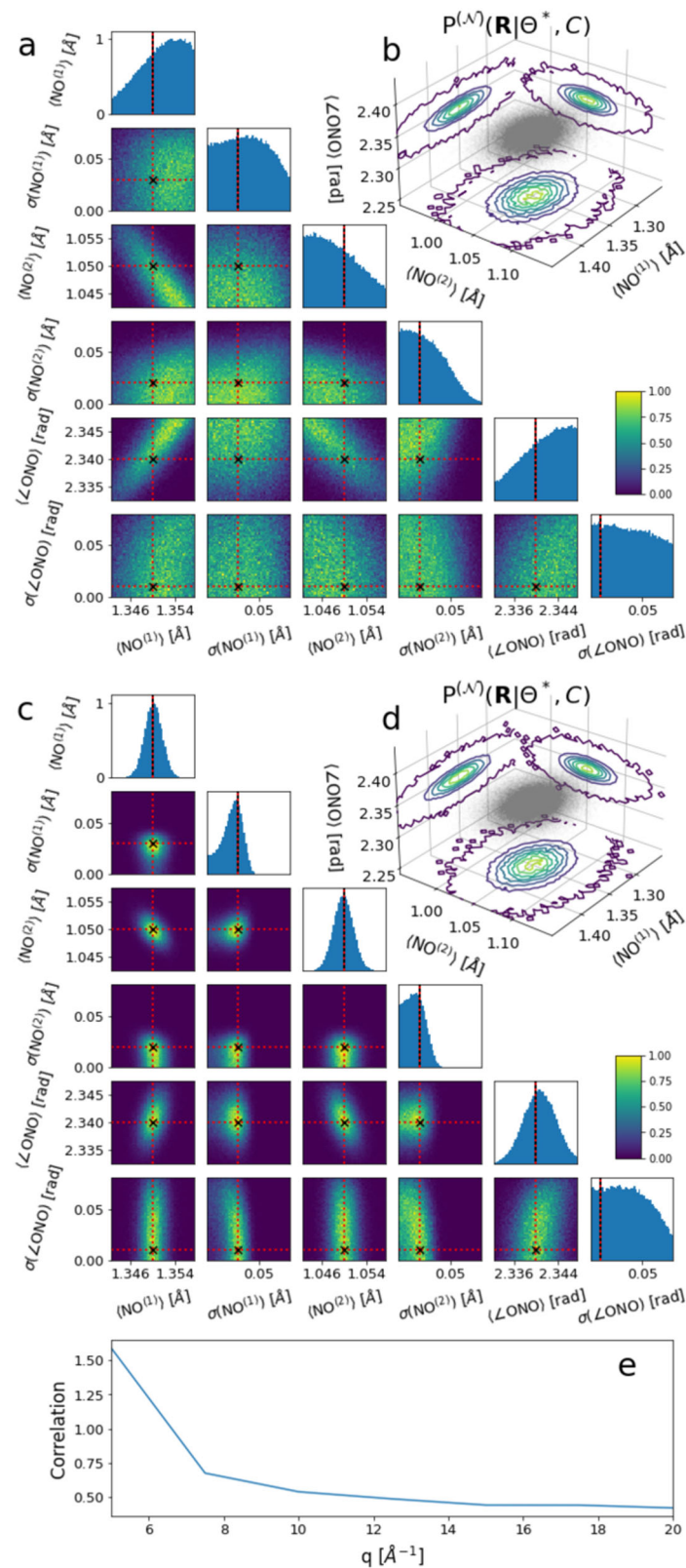


Fig. 7 Effects of varying the measured q range on $P^{(N)}(\Theta|C)$. Varying the measured q range affects false correlations in $P^{(N)}(\Theta|C)$ for NO_2 ; a larger reciprocal space provides more information and dampens false correlations. **a** We show the 1d and 2d projections of $P^{(N)}(\Theta|C)$ and **(b)** the corresponding $P^{(N)}(\mathbf{R}|\Theta^*, C)$ for a limited q range of $[0.5, 5] \text{ \AA}^{-1}$. The red dashed lines illustrate Θ^* , while the black "X" and solid lines indicate the ground truth values. Similarly, we show **(c)** the 1d and 2d projections of $P^{(N)}(\Theta|C)$ **(d)** the corresponding $P^{(N)}(\mathbf{R}|\Theta^*, C)$ for the broader q range of $[0.5, 20] \text{ \AA}^{-1}$. **e** We highlight the correlation between all Θ parameters as a function of q -range. We note the decrease in correlations with larger q , which is further illustrated by the decreasing widths and correlations in the 1d and 2d $P^{(N)}(\Theta|C)$ projections with higher q .

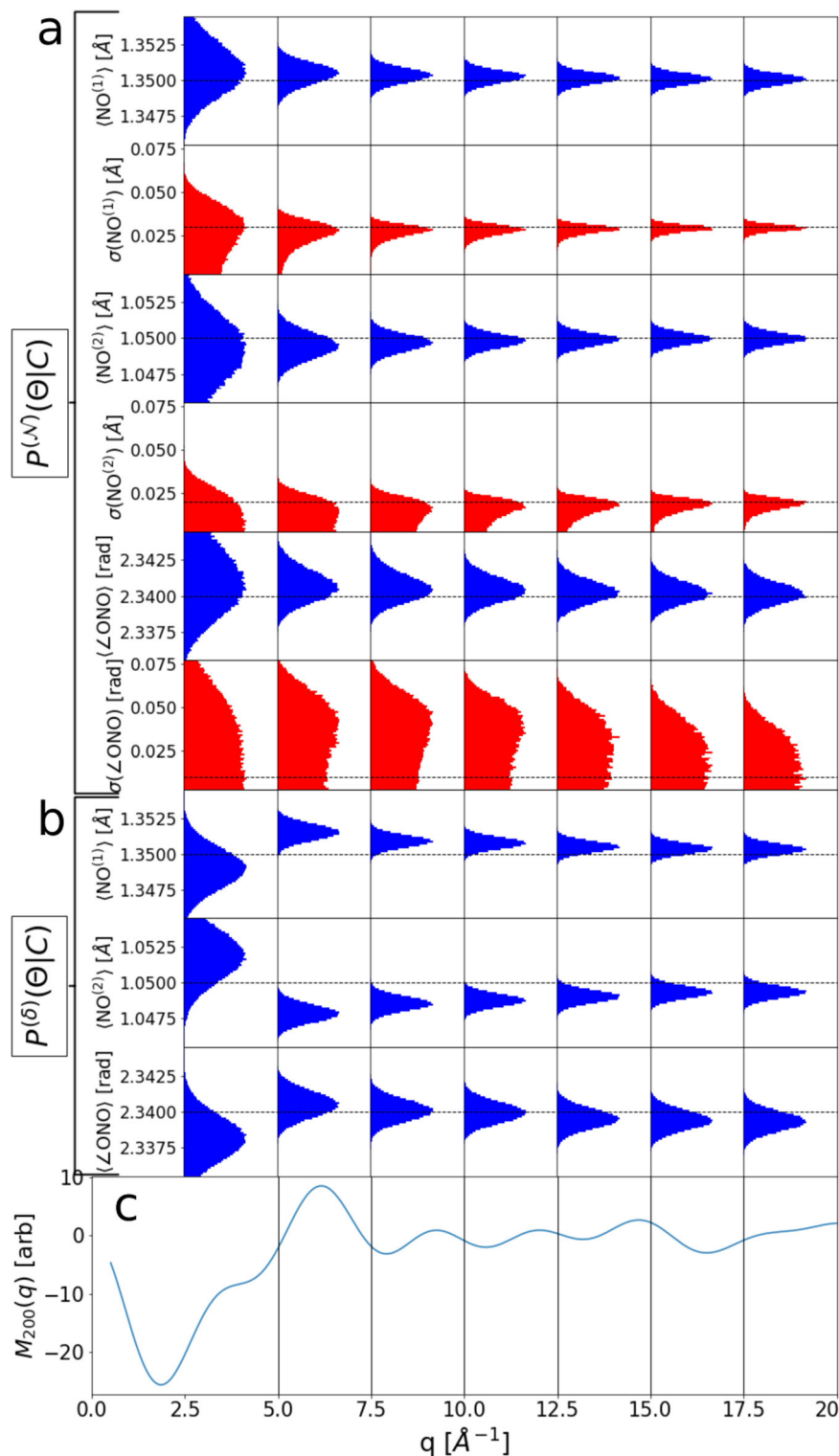


Fig. 8 Systematic errors from selecting incorrect $|\Psi(\mathbf{R})|^2$ distributions. The $P^{(\delta)}(\Theta|C)$ distribution suffers from a q dependent systematic error stemming from the false assumption that a single structure describes the results measured from an ensemble. Here we show the 1d projections of (a, b) $P^{(\delta)}(\Theta|C)$ as a function of (c) the measured q range and a signal-to-noise ratio (SNR) of 400. Each column indicates a different q range starting at 0.5 Å⁻¹ with the end of said q range indicated by the rightmost border of that column. The dashed lines are the ground truth values. c We used the same simulated $C_{200}(q)$ coefficient for both posteriors. Its intersected by the black lines indicate the upper q range of each column.

is a “first-order” measurement of the $|\Psi(\mathbf{R})|^2$ width. It also reduces the systematic effects of assuming a single structure ($P^{(\delta)}(\Theta|C)$) as illustrated in Fig. 8. This was the case for our measured N₂O data where our q range of [3.5, 7.25] Å⁻¹ is

insufficient to resolve the width of $|\Psi^{(N_2O)}(\mathbf{R})|^2$. Therefore, the widths become nuisance parameters used to avoid these systematic errors. Still, $P^{(\delta)}(\Theta|C)$ is accurate on the 10 mÅ scale and runs ~100 times faster than $P^{(N)}(\Theta|C)$. Therefore,

$P^{(\delta)}(\mathbf{R}|\Theta, C)$ serves as an intermediate test analysis before switching to the normal or any other distribution. For very large molecules with many degrees of freedom, $P^{(\delta)}(\mathbf{R}|\Theta, C)$ may be the only tractable method.

The MHA performs an unbiased search through Θ space guided by the $C_{lmk}(q)$ coefficients and correlates each Θ parameter. Our method is model independent and does not suffer from model bias as might be a concern for conventional methods. Limited q range artificially introduces correlations between Θ parameters. Since Θ is the minimal set of parameters to define $P(\mathbf{R}|\Theta, C)$, we expect the parameters to be uncorrelated. Figure 7 shows how adding information by extending the q range decreases false correlations. For the N_2O data, we observe these false correlations between $\langle \text{N}^{\text{T}}\text{N}^{\text{C}} \rangle$ and $\langle \text{N}^{\text{C}}\text{O} \rangle$ (Fig. 5c). Simultaneously evaluating all Θ parameters leverages well-resolved parameters to constrain poorly resolved parameters. For example, the long OO bond (or $\angle \text{ONO}$) in our asymmetric NO_2 is the best constrained parameter as it produces the most q oscillations. The MHA removes structures where the two NO distances are inconsistent with the well-resolved OO distance. These correlations similarly help find Θ^* , as observed with N_2O , where the $P^{(\delta)}(\Theta|C)$ uncorrelated widths do not distinguish the $\text{N}^{\text{T}}\text{N}^{\text{C}}$ and $\text{N}^{\text{C}}\text{O}$ bonds but Θ^* does.

The width of $P(\Theta|C)$ (σ^Θ) relies heavily on SNR rather than increasing q range (Fig. 6b), which is ideal since it is generally prohibitively difficult to change the q range at ultrafast diffraction facilities and easier to reduce the SNR by taking more measurements³⁹. This is because smaller $\sigma_{lmk}(q)$ makes it less probable for the MHA to visit Θ parameters with larger residuals. For the PDF, the resolution is $2\pi/\Delta q$, or 1.26, 0.63, and 0.31 Å for q ranges of 5, 10, and 20 Å⁻¹ respectively, which is roughly 100 to 1000 times larger than our observed resolution for simulated NO_2 at typical to high SNR, respectively. For the measured N_2O data with a very poor SNR and 0.04 Å resolution, we observe a 50X improvement over the 1.7 Å Fourier resolution. This agrees with our simulated results that have more than a factor of 2 better SNR and indicates we may observe these 100–1000X improvements in future measurements. Our method, therefore, lends itself well to high repetition-rate machines, such as the upcoming LCLS II. We note that increasing the q range above 8 Å⁻¹ has a larger effect on the width parameters (Fig. 6a).

Discussion

In the following, we provide intuition about and describe how this method is able to approximate $|\Psi(\mathbf{R})|^2$ while significantly improving upon real-space resolution. We first provide intuition for how induced anisotropy accesses the molecular frame structural angles $\theta_{\mu\nu}^{(\text{mf})}$ and $\phi_{\mu\nu}^{(\text{mf})}$. We then provide a brief intuitive discussion, that compliments the Methods section, of how our Bayesian inference approach inverts $\langle I(\mathbf{q}, t) \rangle$ for Θ while improving upon resolution. Finally, we introduce methods to evaluate excited electronic state dynamics.

To provide intuition for the distinct angular terms, we condense and label the reference frames from Eq. (2)

$$\langle I(q) \rangle_{\text{mol}} = \dots \sum_{m,k} (-1)^k Y_l^m \left(\underbrace{\theta_q^{(\text{lf})}, \phi_q^{(\text{lf})}}_{\text{Lab Frame}} \right) \underbrace{\mathcal{A}_{lmk}^l(t)}_{\text{Anisotropy}} \underbrace{|\Psi(t)\rangle j_l(q\Delta R_{\mu\nu}) Y_l^{-k} \left(\underbrace{\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})}}_{\text{Molecular Frame Structure}} \right) |\Psi(t)\rangle}_{\text{Molecular Frame Structure}} \quad (12)$$

Equation (12) highlights the anisotropic contributions at each level of this method. The molecular frame structure component separates into pairwise distance ($j_l(q\Delta R_{\mu\nu})$) and angular ($Y_l^{-k}(\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})})$) terms. The former governs the q dependence

and the latter is the angular decomposition of the molecular structure which acts as a scaling parameter. The ensemble anisotropy $\mathcal{A}_{lmk}^l(t)$ acts as a key from the measured lab frame anisotropy ($Y_l^m(\theta_q^{(\text{lf})}, \phi_q^{(\text{lf})})$) to the molecular frame structure by coupling these two reference frames. Similar derivations^{40–42} exist but do not stress the dependence on the 3d molecular frame coordinates; Ref. 40 is not treated fully quantum mechanically as done here in Supplementary Note 2. Anisotropy is required for our method to have an explicit dependence on the pairwise angles. Without anisotropy, $C_{000}(q)$ has no explicit angular dependence (Eq. (3)), just like the PDF.

Stronger impulsive alignment produces a broader coherent rotational wavepacket which exhibits higher amplitude signals with more variations (Fig. 4c). Larger amplitude ADMs improve $C_{lmk}(q)$ SNR by lifting higher order coefficients up out of the noise, resulting in similar resolution improvements to only increasing SNR, shown in Fig. 6c. Increasing the number of $C_{lmk}(q)$ coefficients improves the $\theta_{\mu\nu}^{(\text{mf})}$ and $\phi_{\mu\nu}^{(\text{mf})}$ resolution since each $C_{lmk}(q)$ provides a new angular constraint via $Y_l^{-k}(\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})})$ (Eq. (3)).

One can produce fast signal variations with an initially broad hot thermal ensemble. Writing coherence onto hotter molecular ensembles produces weak but fast varying ADMs, shown in Fig. 4c. Figure 6d shows how quickly the resolution worsens at higher temperatures. When fitting the ADMs to $B_l^m(q, t)$, one ideally measures particular points that include two separate regions where the ADMs have high variation and sufficiently before and after the prominent anisotropy signal where their magnitude dampens. One need not strictly measure the entire transient rotational signal.

To simulate the ADMs one will need to measure the rotational constants or calculate them from the vibronic ground state structure. Measured constants remove structural biases potentially induced by calculating these coefficients from a simulated or presumed structure and decouple the rotational signal from the MHA sampling. When simulating or inducing molecular tumbling is prohibitively difficult, one may use the induced anisotropy from the dipole alignment of the initial photo-excitation. This method can be made more general as our Bayesian inference approach does not require anisotropy and is applicable to the traditionally used isotropic component.

With Bayesian inference, we use data to effectively invert $\langle I(\mathbf{q}, t) \rangle$ for Θ . We use the $C_{lmk}(q)$ coefficients to independently constrain $P(\Theta|C)$, from which we find Θ^* to parameterize $P(\mathbf{R}|\Theta^*, C)$. The $P(\mathbf{R}|\Theta^*, C)$ distribution, which approximates $|\Psi(\mathbf{R})|^2$, provides the most probable (and unique) molecular structure. Traditionally, the PDF, being the inverse Fourier transform of $q M_{000}(q)$, is at best a weighted histogram of unlabeled pairwise distances from which one generally cannot obtain a unique structure. Since our measurements necessarily exclude q all the way to 0, and the strong signal drop-off limits high q measurements, our q range is always limited. These limitations obfuscate the PDF interpretations by introducing sinusoidal systematics that result in negative probabilities, e.g. in Fig. 5d where we do not expect any distance above 2.3 Å. Therefore, we typically simulate $|\Psi(\mathbf{R})|^2$ with a priori knowledge and validate simulation against the measured PDF. Our method instead uncovers the globally optimal parameters (Θ^*) from the data for a given $P(\mathbf{R}|\Theta, C)$. This requires only the initial vibronic ground state structure, simulations of the coherent rotational wavepacket when using $C_{lmk}(q)$ for $l > 0$, and for excited state dynamics one additionally needs relevant transition dipole moments. As made clear by comparing Fig. 5b and d, the $P(\mathbf{R}|\Theta^*, C)$ distribution is significantly more information-rich than the PDF, e.g. it provides the 3d molecular structure and width of the $|\Psi(\mathbf{R})|^2$. This method

thus has the potential to shift ultrafast diffraction to a discovery method applicable even to systems that extend beyond the scope of theory.

We find that building $P(\Theta|C)$ to later find its mode (Θ^*) and its resolution (σ^Θ) is more informative and robust than using a gradient-based optimization routine to find Θ^* and its precision. In either case, an optimization routine is used to find Θ^* , but given $P(\Theta|C)$ our method starts near the global minima and is more robust to local minima. If either routine finds a local minimum, one can avoid reporting misleading results by citing the resolution of $P(\Theta|C)$ (σ^Θ) as its error. Since σ^Θ is the standard deviation of all Θ parameters consistent with the data, it is a conservative estimate that very likely encompasses the global minimum. The precision, used by an optimization routine, is determined by the loss landscape around Θ^* and is unaware of the entire Θ distribution. The $P(\Theta|C)$ distribution can also inform the experimentalist which values are best measured, which ones are correlated, and potentially how to improve the experimental apparatus through the false correlations and widths in Figs. 5a, c and 7a, c. One does so by varying experimental parameters, in simulation, to determine how isolated and resolved Θ parameters become.

Our method is broadly applicable to diffraction experiments with laser excitation, including dynamics from excited electronic states. Laser excitation imparts one or more units of angular momentum providing at least $C_{20k}(q)$. From low SNR N_2O data we see the $C_{200}(q)$ alone recovers ~ 40 mÅ resolution. The primary difficulty with extending our method to excited states dynamics lies in isolating the ADMs in rovibronically coupled systems at sufficiently long timescales. Since the principle moments of inertia change with the structure, one must reorient the altered excited state structure by adding three molecular frame Euler angles to the Θ parameters Supplementary Note 2. The generally much wider excited state $|\Psi(\mathbf{R}, t)|^2$ dampens $C_{lmk}(q)$ coefficients and reduces the need for extended q . We discuss two variants to isolate the ADMs, a time-separable method and an isotropic method.

The time-separable method introduces a separation of time scales by assuming the ADMs are relatively stationary during the vibronic motion. This approximation is analogous to the Born-Oppenheimer approximation. For a single excitation pulse, the dipole selection rule introduces ensemble anisotropy independent of the difficulty to create a rotational wavepacket:

$$\langle I(\mathbf{q}) \rangle_{\text{sep}}^{(1)}(t) \approx \mathcal{I} \left(\sum_{\mu} |f_{\mu}(q)|^2 + \sum_{\mu, \nu: \mu \neq \nu} \text{Re} \left\{ f_{\mu}(q) f_{\nu}^*(q) \sum_l \frac{32\pi^3 l!}{2l+1} \sum_{m_1, m_2} (-1)^{m_1} Y_l^{m_1}(\theta_q^{(\text{lf})}, \phi_q^{(\text{lf})}) \right. \right. \\ \left. \left. \times \sum_{n, n'} \tilde{\mathcal{A}}_{m_2 m_1}^{(2)l}(n, n'; \tau) \langle \psi_{\text{el-vib}}^{n'}(t) | j_l(q \Delta R_{\mu\nu}) Y_l^{-m_1}(\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})}) | \psi_{\text{el-vib}}^n(t) \rangle \right\} \right). \quad (13)$$

Here, $\tilde{\mathcal{A}}_{m_2 m_1}^{(1)l}(n, n')$ are the ADMs calculated with the rovibronic ground state structure, the ground rovibronic transition dipole, and evaluated immediately after laser excitation. This requires knowledge of either the transition dipole moment or the Frank-Condon factor and the electronic transition dipole moment.

To further constrain $P(\Theta|C)$, one can couple to more $C_{lmk}(q)$ coefficients by introducing a precursor pulse that excites a rotational wavepacket. This precursor pulse, assumed to be a rotational Raman impulse, is chosen to have a negligible effect on the vibronic system thus maintaining consistency with our separation of timescale approximation. The Raman impulse first induces rotational coherence. Following the Raman impulse, the system evolves for a rotational time τ , at this point the vibronic excitation pulse arrives. One would measure the vibronic dynamics over a small window ($t \ll \tau$). This is repeated for different orientations by scanning the delay τ over an

appreciable portion of the rotational evolution. This window, measured by t , is typically of order picosecond or less such that the ADMs do not appreciably change. The measured diffraction images are given by

$$\langle I(\mathbf{q}) \rangle_{\text{sep}}^{(2)}(t, \tau) \approx \mathcal{I} \left(\sum_{\mu} |f_{\mu}(q)|^2 + \sum_{\mu, \nu: \mu \neq \nu} \text{Re} \left\{ f_{\mu}(q) f_{\nu}^*(q) \sum_l \frac{32\pi^3 l!}{2l+1} \sum_{m_1, m_2} (-1)^{m_1} Y_l^{m_1}(\theta_q^{(\text{lf})}, \phi_q^{(\text{lf})}) \right. \right. \\ \left. \left. \times \sum_{n, n'} \tilde{\mathcal{A}}_{m_2 m_1}^{(2)l}(n, n'; \tau) \langle \psi_{\text{el-vib}}^{n'}(t) | j_l(q \Delta R_{\mu\nu}) Y_l^{-m_1}(\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})}) | \psi_{\text{el-vib}}^n(t) \rangle \right\} \right). \quad (14)$$

where n labels the vibronic states, $|\psi_{\text{el-vib}}^n(t)\rangle$ is the vibronic state (assumed unknown), $\tilde{\mathcal{A}}_{mk}^{(2)l}(n, n'; \tau)$ are the modified ADMs, and t is the arrival time of the probe after the second excitation pulse. These modified ADMs consider the angular momentum transfer by the vibronic excitation photon and require the vibronic transition dipole moments of the corresponding vibronic state. One then follows the above analysis procedure for each time t . In such an experiment, one should measure the ensemble anisotropy without the vibronic excitation pulse to find the best-fit ADMs. Supplementary Note 2 further describes our separation of timescale approximation and provides the derivations for Eqs. (13) and (14).

The isotropic method uses only the $C_{000}(q, t)$ term, similar to conventional analyses. Since $\tilde{\mathcal{A}}_{00}^{(a)0}(n, n'; t, \tau)$ becomes a constant absorbed by \mathcal{I} , this method can be applied to single (Eq. (13)) and double pulse (Eq. (14)) experiments. The $C_{000}(q, t)$ term only implicitly depends on the pairwise angles through $\Delta R_{\mu\nu}$. This is in contrast to the explicit pairwise angle dependence in the higher order $C_{lmk}(q)$ terms. Our statistical treatment likely provides adequate pairwise angle resolution because we have more pairwise distances than are required to specify a unique structure.

For a Raman-inducing precursor pulse, one will likely use a combination of the isotropic and time-separable methods. For fast dynamics, one would use the time-separable method for small windows shortly following the rotation time τ . Longer-lived dynamics can be retrieved by the isotropic method. When retrieving $P(\Theta|C)$, in either case, one initiates the MHA with the vibronic ground state Θ^* parameters at the first temporal measurement. For each subsequent time step one initiates MHA with the Θ^* parameters from the previous time step.

Electronic and vibrational excited state wavepackets bifurcate into multiple states, e.g. at conical intersections, causing $P(\mathbf{R}, t|\Theta^*, C)$ to bifurcate as well. We account for these different states by

$$P(\mathbf{R}, t|\Theta^*, C) = \sum_i^{N_{\text{ex}}} c_i P(\mathbf{R}, t|\Theta_i^*, C) \quad (15)$$

where N_{ex} is the number of excited state distributions with appreciable population. Conical intersections will induce bifurcations that spawn a new distribution that adds to N_{ex} . In this way we consider this method to be fully data-driven since we can change our theoretical description (c_i) based on data alone.

Thus far we have only considered diffraction consistent with the independent atom approximation and all the equations above have been derived under this approximation. Recently, diffraction beyond the independent atom approximation has been observed in both electron⁴³ and x-ray diffraction⁴⁴. Under such conditions, this method must be modified by either re-deriving the above equations to consider these effects or by accounting for this signal in the $C_{lmk}(q)$ coefficients. For MeV electron diffraction, inelastic scattering is limited to the low $q < 1 \text{ \AA}^{-1}$ region and can be easily removed from the $C_{lmk}(q)$ coefficients. For x-ray diffraction beyond the independent atom approximation, contributions from excited Rydberg states create a constant offset after the initial

signal turn-on that spans the entire q -range^{24,44}. Due to the diffuse nature of the Rydberg state this signal does not vary appreciably in time and can be subtracted out.

Conclusion

We have shown that our method can approximate $|\Psi(\mathbf{R})|^2$ with $P(\mathbf{R}|\Theta^*, C)$ for the vibronic ground states of NO₂ and N₂O. In simulation, we retrieve ~0.5 mÅ resolution for NO₂. From measured N₂O UED data, we retrieve ~40 mÅ resolution despite a short q range of [3.5, 7.25] Å⁻¹ and very poor SNR. Compared to PDF-based methods, this returns the labeled pairwise distances and angles with 50 and 100–1000 times better resolution in measurement and simulation respectively. In spite of similar bond distances and atomic scattering amplitudes for NO₂ and N₂O, our method distinguishes these distances. We begin to resolve the $\langle \text{N}^{\text{T}}\text{N}^{\text{C}} \rangle$ and $\langle \text{N}^{\text{C}}\text{O} \rangle$ distances in our low SNR and narrow q range UED measurement. These results are highly encouraging and illustrate the viability of our Bayesian inference approach. They also inspire further expansion into excited state dynamics. The code repository⁴⁵ contains the algorithms used for this work and instructions on how to reproduce these results. It also contains instructions on how to run this analysis and templates for applying this method to new molecules.

This Bayesian inference approach is best suited for gas-phase ultrafast diffraction instruments that have high SNR such as high repetition-rate free electron facilities, e.g. LCLS-II-HE. Resolution quickly improves with SNR considerably faster than if one increases q beyond ~8 Å⁻¹. Nevertheless, larger q ranges improve resolution for widths of $|\Psi(\mathbf{R})|^2$ and diminish false correlations between Θ parameters.

Our general method has the potential to become commonplace for ultrafast gas-phase diffraction measurements due to its broad applicability and its independence from complex excited state simulations. In this work, we validated its use for standard pump-probe setups. One can extend this method to excited state dynamics either with or without anisotropy. Our isotropic method is well suited for current pump-probe setups that generally focus on the isotropic component. This method greatly benefits from deterministic anisotropy that can either be induced by impulsive Raman or by the dipole moment selection from the excitation pulse. Beyond ultrafast gas-phase diffraction, one can apply this general framework to other classes of experiments, e.g. the previously mentioned photo-electron experiments^{27–29,31,32}. This is done by deriving the molecular frame response (Eq. (2)) and applying this Bayesian inference approach.

Given its broad applicability, high resolution, amenability to various measurements, and independence from complex molecular dynamic simulations, our method has the potential to effectively turn ultrafast gas-phase molecular diffraction into a discovery-oriented technique. This method can retrieve a unique molecular structure distribution for general molecules with $\lesssim 10$ mÅ. Moreover, because our method is parameterized by Θ , we have the opportunity to expand the scope of ultrafast gas-phase diffraction into previously inaccessible measurements. For instance, we demonstrated the use of this parameterization to measure the width of $|\Psi(\mathbf{R}, t)|^2$; this width is important in the excited state where single structures lose their

meaning. This method unlocks our ability to study larger and more complex systems that are currently too difficult to simulate.

Methods

Our method can be subdivided into three principal concepts. Firstly, we use ensemble anisotropy, described by the ADMs, to access the molecular frame by projecting the data onto anisotropic components. Secondly, we select a model, $P(\mathbf{R}|\Theta, C)$, to approximate $|\Psi(\mathbf{R})|^2$ and develop our statistical approach to solve for Θ using Bayesian inference. That is, through the statistical nature of our measurement we use Bayesian inference to effectively invert the diffraction signal for Θ . Lastly, we take our statistical description and use MCMC techniques to solve for $P(\Theta|C)$ to retrieve the optimal Θ parameters (Θ^*). The code used for this analysis⁴⁵ can be run to reproduce the following results or adapted for other molecules.

Extracting molecular frame information. We describe our analysis procedure for a system given an induced deterministic ensemble anisotropy under experimental conditions at the SLAC MeV ultrafast electron diffraction facility (UED)⁶. Our generic pump-probe setup is similar to most ultrafast diffraction setups, consisting of an 800 nm Ti:Sapphire pump laser and a 120 fs FWHM electron bunch probe. For the simulated NO₂ results, we consider using a single 10 TW/cm² 800 nm pump pulse to impulsively induce a coherent rotational wave packet and probing it within a window of high anisotropy variation: [37.5, 41.5] ps. For the measured N₂O sample, a train of 8 identical 800 nm pulses (40 fs duration and 5×10^{12} W/cm² irradiance) separated by full quantum revivals induced such rotational wavepacket⁴⁶. We measured the first field free full quantum revival over a window of ~3 ps. We masked q regions [0, 3.5] Å⁻¹ and above 7.25 Å⁻¹ due to ellipticity in the imaging of the diffraction pattern and poor signal-to-noise, respectively. Linearly polarized pump pulses induce azimuthal symmetry, which sets $m = 0$ in Eq. (1) ($P(\phi_l^{(\text{lf})}, t) = 1/2\pi$), while the Raman excitation of the wavepacket requires l being even in Eq. (1).

We define anisotropy in two equivalent ways and quantify it through the ADMs. Firstly, anisotropy is defined by a non-zero projection of the measured diffraction pattern onto any Y_l^m with even $l > 0$ for a given Δq range. Secondly, anisotropy exists when there is a non-zero $\mathcal{A}_{mk}^l(t)$ for $l > 0$. To calculate the ADMs, one must know the rotational constants (A, B, C) and ideally the centrifugal distortion (D) constants, as well as the differential polarizability, which can be calculated from the known ground state structure or measured from Raman spectroscopy. For N₂O, we used the measured rotational constants^{47,48} to model the rotational wavepacket for the fitted ensemble temperature and laser intensity described in Supplementary Note 1. We note other methodologies to calculate the ADMs^{33,34,49}. Supplementary Note 1 describes both our calculation of the ADMs and our search for the best-fit ADMs.

We access the molecular pairwise distances and angles in the molecular frame. Using the ADMs and the Independent Atom Approximation, we relate measured lab frame anisotropy in diffraction patterns, $\langle I(\mathbf{q}, t) \rangle$, to the molecular structure

$$\langle I(\mathbf{q}, t) \rangle = \mathcal{I} \left(\sum_{\mu} |f_{\mu}(q)|^2 + \sum_{\mu, \nu: \mu \neq \nu} \text{Re} \left\{ f_{\mu}(q) f_{\nu}^*(q) \sum_l 4\pi i^l \right. \right. \\ \left. \left. \sum_{m,k} (-1)^k Y_l^m \left(\theta_q^{(\text{lf})}, \phi_q^{(\text{lf})} \right) \langle \Psi(t) | D_{mk}^l \left(\phi_l^{(\text{lf})}, \theta_l^{(\text{lf})}, \chi_l^{(\text{lf})} \right) j_l(q \Delta R_{\mu\nu}) Y_l^{-k} \left(\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})} \right) | \Psi(t) \rangle \right\} \right) \quad (16)$$

Lab Frame Ensemble Anisotropy Molecular Frame Structure

In Eq. (16), derived in Supplementary Note 2, $f_\mu(q)$ is the scattering amplitude of the μ th atom, $j_l(q\Delta R_{\mu\nu})$ are the spherical Bessel functions of the first kind, \mathcal{I} is the diffraction beam intensity, and the momentum transfer vector is given by $\mathbf{q} = [q, \theta_q^{(\text{lf})}, \phi_q^{(\text{lf})}]$. The difference vector $\Delta\mathbf{R}_{\mu\nu} = \mathbf{R}_\mu - \mathbf{R}_\nu = [\Delta R_{\mu\nu}, \theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})}]$ is the molecular frame pairwise distance and

$$B_l^m(q, t) = \int_0^\pi \left\langle I(\mathbf{q}(q, \theta^{(\text{d})}), t) \right\rangle_{\text{rigid}} Y_l^m(\theta_q^{(\text{lf})}(q, \theta^{(\text{d})}), \phi_q^{(\text{lf})}(q, \theta^{(\text{d})})) \sin(\theta_q^{(\text{lf})}(q, \theta^{(\text{d})})) d\theta^{(\text{d})} \\ = \mathcal{I} \sum_{\mu, \nu; \mu \neq \nu} \text{Re} \left\{ \underbrace{f_\mu(q) f_\nu^*(q)}_{\text{Molecular Frame Structure}} \underbrace{\frac{32\pi^3 i^l}{2l+1} (-1)^k \langle \Psi(0) | j_l(q\Delta R_{\mu\nu}) Y_l^{-k}(\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})})}_{\text{Anisotropy}} \right\} |\Psi(0)\rangle \mathcal{A}_{mk}^l(t) \Big|_{\text{rigid}} \Big\} \quad (17)$$

angles between the μ th and ν th atoms, illustrated in Fig. 1b. Equation (16) shows how the ensemble anisotropy connects the lab frame to the molecular frame structure. Directly accessing the molecular frame pairwise angles $(\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})})$ requires anisotropy and is otherwise inaccessible through the PDF and isotropic contributions alone. This is evident by isolating the isotropic component ($l = 0, m = 0, k = 0$) which sets $Y_0^0(\theta_{\mu\nu}^{(\text{mf})}, \phi_{\mu\nu}^{(\text{mf})}) = 1/(2\sqrt{\pi})$.

For our method, we describe optimal representations of the lab and molecular frames used in Eq. (16). The molecular frame is defined by the molecule's principal moments of inertia before laser excitation with the $\hat{\mathbf{z}}^{(\text{mf})}$, $\hat{\mathbf{x}}^{(\text{mf})}$, and $\hat{\mathbf{y}}^{(\text{mf})}$ corresponding to the principle moments of inertia in decreasing order: A, B, and C respectively. This necessitates knowledge of the rovibronic ground state structure when one is measuring an excited rovibronic structure. When looking at the $\Delta\mathbf{R}_{\mu\nu}$ contribution, we isolate the μ th and ν th atoms while ignoring other atoms and translate the atom pair such that \mathbf{R}_ν defines the origin. This is highlighted in Fig. 1b where the nitrogen is translated to the origin. This translation allows us to define the pairwise angles and derive Eq. (16). Since we are concerned with a difference in locations $\Delta\mathbf{R}_{\mu\nu}$, Eq. (16) is invariant under such molecular frame translations. In the lab frame, the laser polarization defines $\hat{\mathbf{z}}^{(\text{lf})}$ and the propagation direction of the probe pulse defines $\hat{\mathbf{y}}^{(\text{lf})}$. The measured signals in the lab frame, on a 2D detector, are defined by detector parameters $q = |\mathbf{q}|$ and the azimuthal angle $\theta^{(\text{d})}$ defined by $\hat{\mathbf{z}}^{(\text{lf})}$. Supplementary Note 2 describes how to rewrite \mathbf{q} in terms of the detector coordinates. For small angle scattering at UED $\theta_q^{(\text{lf})} \approx \theta^{(\text{d})}$ and $\phi_q^{(\text{lf})} \approx 0$.

The primary difficulty of working with Eq. (16) comes from the expectation value including both the ensemble anisotropy and molecular frame structure. We want to separate the ensemble anisotropy into the ADMs. This isolates the time-dependent molecular structure term that we would like to retrieve. By doing this, we only require more tractable molecular rotation simulations with respect to the known rovibronic ground state structure in order to retrieve the time-dependent molecular structure. Otherwise, as Eq. (16) is written, it requires a priori knowledge of exactly the unknown time-dependent structures for which we are solving. In this work, we describe various ways to do this under common experimental conditions.

Focusing on the vibronic ground state of NO_2 , we can separate the ADMs and molecular structure contribution in Eq. (16) by applying a rigid rotor approximation. Equation (2) is the general form, which we adapt to our specific case by setting $m = 0$ and replacing $\theta_q^{(\text{lf})} \approx \theta^{(\text{d})}$ and $\phi_q^{(\text{lf})} \approx 0$. The resulting lab frame measurements are shown in Fig. 3a.

To retrieve $P(\mathbf{R}|\Theta, C)$, we first isolate the molecular frame structure terms from Eq. (2) with a series of fits. The first fit removes the initial diffraction beam intensity (\mathcal{I}), described in Supplementary Note 8. The second fit projects out the measured lab frame anisotropy $(Y_l^m(\theta_q^{(\text{lf})}, \phi_q^{(\text{lf})}))$ from Eq. (2) by fitting the angular dependence of the measured diffraction.

This yields the time (t) and q dependent $B_l^m(q, t)$ coefficients shown in Fig. 3b. The third fit isolates the molecular frame information by fitting out the time dependence of $B_l^m(q, t)$ with the simulated ADMs, $\mathcal{A}_{mk}^l(t)$. The resulting coefficients, $C_{lmk}(q)$, relate measured data to the molecular frame pairwise structure. Here, $M_{lmk}(q)$ are the modified $C_{lmk}(q)$ coefficients that compensate for the rapid q^{-4} falloff in the electron scattering amplitudes. Figure 3c shows the retrieved $M_{lmk}(q)$ for both the simulated and measured data. For the N_2O data, the poor signal-to-noise precludes all contributions except $C_{200}(q)$. Depending on the data quality and degree of orthogonality in the ADMs, one may need to employ regularization to retrieve physical fit values. Regularization adds a fitting cost to extraneous coefficients, thus minimizing the impact of non-orthogonal ADMs. Supplementary Note 3 provides a further discussion on fitting the ADMs and regularization.

The standard error of the mean $\sigma_{lmk}(q)$ for each $C_{lmk}(q)$ is calculated from a distribution of measured $C_{lmk}(q)$ coefficients. For the N_2O data, Supplementary Note 5 describes the data processing and retrieval of $\sigma_{lmk}(q)$. For the NO_2 simulation, we add Poisson noise to the diffraction patterns and propagate that noise through the lab frame anisotropy and ADM fit (see Supplementary Section Supplementary Note 5).

Applying Bayesian Inference. We approximate $|\Psi(\mathbf{R})|^2$ with the probability distribution $P(\mathbf{R}|\Theta, C)$, which is parameterized by Θ and conditioned on the observed $C_{lmk}(q)$ coefficients. This requires one to choose a functional form of $P(\mathbf{R}|\Theta, C)$ dependent on the system's state and the desired degree of accuracy. Depending on the desired accuracy and precision of the desired results, this requires varying degrees of a priori knowledge. For example, one may choose a multivariate delta function for a single molecule response, a normal distribution to model the ground vibrational states, or harmonic oscillator eigenfunctions to describe arbitrary individual vibrational states.

The Θ parameters include the $3N_{\text{atom}} - 6$ structural degrees of freedom (N_{dof}) needed to define a unique molecular structure, and the width parameters in the case of $P^{(\wedge)}(\mathbf{R}|\Theta, C)$. Here, Θ has the minimal number of parameters needed to define $P(\mathbf{R}|\Theta, C)$, and adding redundant parameters can significantly alter one's results.

Having isolated the molecular frame structure terms ($C_{lmk}(q)$) and chosen $P(\mathbf{R}|\Theta, C)$, we apply Bayesian Inference to address the diffraction inverse problem^{35,36,50} by effectively inverting $C_{lmk}(q)$ to approximate $|\Psi(\mathbf{R})|^2$. With Bayes rule,

$$P(\Theta|C) = \frac{P(C|\Theta)P(\Theta)}{P(C)} \quad (18)$$

we use the statistical nature of our measurement to analytically relate the desired Θ parameters to the measured $C_{lmk}(q)$. In

Eq. (18), $P(\Theta|C)$ is the posterior distribution we wish to build. The likelihood $P(C|\Theta)$ relates the measured data to the Θ parameters and is the probability of observing $C_{lmk}(q)$ given the parameters Θ . Here, $C_{lmk}^{(calc)}(q, \Theta)$ are the calculated $C_{lmk}(q)$ coefficients, and $\sigma_{lmk}(q)$ are the standard errors of the means for $C_{lmk}(q)$. The prior, $P(\Theta)$ contains our a priori knowledge of the system, and in this work is used to constrain Θ to physicality (e.g., $\Theta > 0$ and $\angle \text{ONO} < \pi$). This is because we do not assume any prior knowledge or simulations of the system. Calculating the marginal likelihood $P(C)$ is generally, and in our case, intractable. Further intuition regarding how the statistical nature of our measurement allows us to invert for Θ is described in ref. 36.

Given the functional forms of $P(\Theta|C)$, $P(C|\Theta)$, and the presumed functional form of $P(\mathbf{R}|\Theta, C)$, we now find the globally optimal Θ parameters (Θ^*) by building $P(\Theta|C)$ and finding its mode. To converge on the mode of $P(\Theta|C)$, one must use the correlations between the Θ parameters by building $P(\Theta|C)$ in the full Θ -space rather than sampling each parameter individually. Consequently, we must next address the curse of dimensionality.

Solving for the high dimensional model parameters Θ . We retrieve $P(\Theta|C)$ with the Metropolis-Hastings algorithm (MHA) from the following system of equations:

$$C_{lmk}(q) = \int H_{lmk}(q, \mathbf{R}) |\Psi(\mathbf{R})|^2 d\mathbf{R} \quad (19)$$

$$C_{lmk}^{(calc)}(q, \Theta) = \int H_{lmk}(q, \mathbf{R}) P(\mathbf{R}|\Theta, C) d\mathbf{R} \quad (20)$$

$$H_{lmk}(q, \mathbf{R}) = \text{IRe} \left\{ (-1)^k \frac{32\pi^3 l}{2l+1} \times \sum_{\mu, \nu; \mu \neq \nu} |f_{\mu}(q)| |f_{\nu}(q)| j_l(q\Delta R_{\mu\nu}) Y_l^{-k}(\theta_{\mu\nu}^{(mf)}, \phi_{\mu\nu}^{(mf)}) \right\}. \quad (21)$$

We note the high dimensionality and complexity of Eq. (20), which is a system of order 10 equations, each with order 100 terms, embedded in an order 100-dimensional space of measurements in q . This must be evaluated on a Θ -dimensional space of all possible molecular structures and width parameters. The MHA is chosen for its ability to retrieve probability distributions from high dimensional integral equations^{50,51} like Eq. (20).

The MHA is designed to efficiently and preferentially sample regions of Θ -space proportional to the agreement with data, spending the vast majority of its time sampling regions of high probability (agreement). The MHA builds $P(\Theta|C)$ by accumulating Θ parameters based on their relative posteriors

$$\frac{P(\Theta'|C)}{P(\Theta|C)} = \frac{P(C|\Theta')}{P(C|\Theta)} \quad (22)$$

where Θ and Θ' are both physical, and the prior and the marginal likelihood cancel out. We note Eq. (22), and hence the MHA, is theory independent and is analogous to a random walk guided by the relative agreement of neighboring Θ parameters to the data. For instance, if the likelihood of Θ is 2 times larger than Θ' , the MHA will sample twice as many structures around Θ than Θ' . Similarly, if the likelihood for Θ is 1000 times larger than for Θ' , then the MHA will effectively remove structures around Θ' from the search space. Reference⁵⁰, the MHA python package⁵⁰ used in this work, and Ref. 36 give detailed descriptions of combining Bayesian inference and the MHA. Supplementary Note 4 describes our use of the MHA and Bayesian inference in greater detail and

how one can introduce physical intuition, or a priori knowledge, into the MHA.

This method ultimately yields the following three results; a distribution of Θ parameters (the posterior $P(\Theta|C)$), the optimal set of model parameters (Θ^*), and a parameterized probability of molecular structures $P(\mathbf{R}|\Theta^*, C)$. For each individual Θ parameter, where the i^{th} parameter is denoted as Θ_i , we calculate its resolution as the standard deviation of the projection of $P(\Theta|C)$ onto Θ_i . This resolution, σ^{Θ} , is the one-dimensional standard deviation after marginalizing over all other parameters, which removes the correlations between Θ parameters. That is, if one randomly draws some parameters Θ from $P(\Theta|C)$, the distribution of parameter Θ_i will have a width of σ^{Θ} . In this work, we focus on how Bayesian inference and Eq. (16) effectively invert data for $P(\mathbf{R}|\Theta^*, C)$ via an unambiguous and sharp $P(\Theta|C)$. It is this $P(\Theta|C)$ and its width (resolution) that are our figures of merit for the inversion. The accuracy of Θ^* depends on one's method for finding the mode, of which there are many methods. The precision of Θ^* is a function of its local region. The mean and mode of said marginalized distribution will likely not correspond to Θ^* , since Θ^* is the mode of the full Θ -space distribution. We find Θ^* via a simple mode search algorithm described in Supplementary Note 6.

The measured q range, the induced rotational wavepacket, and the $\sigma_{lmk}(q)$ are vital in determining the width, shape, and parameter correlations of $P(\Theta|C)$. To investigate such dependencies we first define a $|\Psi(\mathbf{R})|^2$ distribution for NO₂ to calculate $C_{lmk}(q)$. Figure 4a and Table 1 show and describe this distribution, respectively. Measuring more diffraction patterns increases the signal-to-noise ratio (SNR) by reducing $\sigma_{lmk}(q)$ which scales as $1/\sqrt{N}$. Here, the SNR is the geometric mean of $C_{000}(q)/\sigma_{000}(q)$ between $0.5 < q < 4 \text{ \AA}^{-1}$. Figure 4b illustrates the $C_{lmk}(q)$ coefficients used in this analysis with the following SNRs based on previous UED⁵² and x-ray¹² diffraction experiments. Unless otherwise stated, the standard configuration of experimental parameters for our NO₂ results is a q range of $[0.5, 10] \text{ \AA}^{-1}$, a SNR of 100, a pump fluence of 1 J/cm^2 and a 100 K ensemble temperature.

Data availability

The UED N₂O data used in this analysis will be provided by the corresponding authors upon reasonable request. The simulated NO₂ data, $C_{lmk}(q)$, can be calculated by the supplied analysis code in ref. 45.

Code availability

The code used in this analysis can be found in ref. 45. Here, one will find a detailed description of the code and how to run it in order to reproduce the NO₂ results. This repository also includes templates for one to apply this algorithm to new molecules.

Received: 26 October 2022; Accepted: 5 October 2023;

Published online: 13 November 2023

References

1. Stankus, B. et al. Advances in ultrafast gas-phase x-ray scattering. *J. Phys. B At. Mol. Opt. Phys.* **53**, 234004 (2020).
2. Minitti, M. P. et al. Imaging molecular motion: femtosecond x-ray scattering of an electrocyclic chemical reaction. *Phys. Rev. Lett.* **114**, 255501 (2015).
3. Williamson, J. C., Cao, J., Ihee, H., Frey, H. & Zewail, A. H. Clocking transient chemical changes by ultrafast electron diffraction. *Nature* **386**, 159–162 (1997).
4. Ewbank, J. et al. Instrumentation for gas electron diffraction employing a pulsed electron beam synchronous with photoexcitation. *Rev. Sci. Instrum.* **63**, 3352–3358 (1992).
5. Ewbank, J. et al. Time-resolved gas electron diffraction study of the 193-nm photolysis of 1,2-dichloroethenes. *J. Phys. Chem.* **97**, 8745–8751 (1993).

6. Shen, X. et al. Femtosecond gas-phase mega-electron-volt ultrafast electron diffraction. *Struct. Dyn.* **6**, 054305 (2019).
7. Ben-Nun, M., Quenneville, J. & Martínez, T. J. Ab initio multiple spawning: photochemistry from first principles quantum molecular dynamics. *J. Phys. Chem. A* **104**, 5161–5175 (2000).
8. Siegbahn, P. E. M., Almlöf, J., Heiberg, A. & Roos, B. O. The complete active space scf (casscf) method in a newton-raphson formulation with application to the hno molecule. *J. Chem. Phys.* **74**, 2384–2396 (1981).
9. Mai, S., Marquetand, P. & Gonzalez, L. Nonadiabatic dynamics: The sharc approach. *WIREs Comput. Mol. Sci.* **8**, e1370 (2018).
10. Meyer, H.-D., Manthe, U. & Cederbaum, L. The multi-configurational time-dependent Hartree approach. *Chem. Phys. Lett.* **165**, 73–78 (1990).
11. Ware, M. R., Glowina, J. M., Natan, A., Cryan, J. P. & Bucksbaum, P. H. On the limits of observing motion in time-resolved x-ray scattering. *Phil. Trans. R. Soc. A* <https://doi.org/10.1098/rsta.2017.0477> (2019).
12. Ware, M. R., Glowina, J. M., Al-Sayyad, N., O'Neal, J. T. & Bucksbaum, P. H. Characterizing dissociative motion in time-resolved x-ray scattering from gas-phase diatomic molecules. *Phys. Rev. A* **100**, 033413 (2019).
13. Bucksbaum, P. H., Ware, M. R., Natan, A., Cryan, J. P. & Glowina, J. M. Characterizing multiphoton excitation using time-resolved x-ray scattering. *Phys. Rev. X* **10**, 011065 (2020).
14. Saldin, D. K., Shneerson, V. L., Starodub, D. & Spence, J. C. H. Reconstruction from a single diffraction pattern of azimuthally projected electron density of molecules aligned parallel to a single axis. *Acta Crystallograph. Sec. A* **66**, 32–37 (2010).
15. Starodub, D., Spence, J. C. H. & Saldin, D. K. Reconstruction of the electron density of molecules with single-axis alignment. in *Image Reconstruction from Incomplete Data* (eds. Bones, P. J., Fiddy, M. A. & Millane, R. P.) (SPIE, 2011).
16. Ho, P. J. et al. Molecular structure determination from x-ray scattering patterns of laser-aligned symmetric-top molecules. *J. Chem. Phys.* **131**, 131101 (2009).
17. Elser, V. Three-dimensional structure from intensity correlations. *J. Phys.* **13**, 123014 (2011).
18. Poon, H.-C., Schwander, P., Uddin, M. & Saldin, D. K. Fiber diffraction without fibers. *Phys. Rev. Lett.* **110**, 265505 (2013).
19. Reckenthaeler, P. et al. Time-resolved electron diffraction from selectively aligned molecules. *Phys. Rev. Lett.* **102**, 213001 (2009).
20. Pabst, S., Ho, P. J. & Santra, R. Computational studies of x-ray scattering from three-dimensionally-aligned asymmetric-top molecules. *Phys. Rev. A* **81**, 043425 (2010).
21. Yang, J., Makhija, V., Kumarappan, V. & Centurion, M. Reconstruction of three-dimensional molecular structure from diffraction of laser-aligned molecules. *Struct. Dyn.* **1**, 044101 (2014).
22. Hensley, C. J., Yang, J. & Centurion, M. Imaging of isolated molecules with ultrafast electron pulses. *Phys. Rev. Lett.* **109**, 133202 (2012).
23. Wilkin, K. J. et al. Ultrafast electron diffraction from transiently aligned asymmetric top molecules: rotational dynamics and structure retrieval. *Struct. Dyn.* **9**, 054303 (2022).
24. Stankus, B. et al. Ultrafast x-ray scattering reveals vibrational coherence following rydberg excitation. *Nat. Chem.* **11**, 716–721 (2019).
25. Natan, A. Real-space inversion and super-resolution of ultrafast scattering. *Phys. Rev. A* **107**, 023105 (2023).
26. Natan, A. et al. Resolving multiphoton processes with high-order anisotropy ultrafast x-ray scattering. *Faraday Discuss.* **228**, 123–138 (2021).
27. Makhija, V., Ren, X., Gockel, D., Le, A.-T. & Kumarappan, V. Orientation resolution through rotational coherence spectroscopy. *Atom. Phys.* <https://doi.org/10.48550/arXiv.1611.06476> (2016).
28. Marceau, C. et al. Molecular frame reconstruction using time-domain photoionization interferometry. *Phys. Rev. Lett.* **119**, 083401 (2017).
29. Gregory, M., Hockett, P., Stolow, A. & Makhija, V. Towards molecular frame photoelectron angular distributions in polyatomic molecules from lab frame coherent rotational wavepacket evolution. *J. Phys. B Atom. Mol. Opt. Phys.* **54**, 145601 (2021).
30. Mikosch, J. et al. Channel- and angle-resolved above threshold ionization in the molecular frame. *Phys. Rev. Lett.* **110**, 023004 (2013).
31. Sándor, P. et al. Angle-dependent strong-field ionization of halomethanes. *J. Chem. Phys.* **151**, 194308 (2019).
32. Sándor, P. et al. Angle dependence of strong-field single and double ionization of carbonyl sulfide. *Phys. Rev. A* **98**, 043425 (2018).
33. Stolow, A. & Underwood, J. G. Time-resolved photoelectron spectroscopy of nonadiabatic dynamics in polyatomic molecules. *Adv. Chem. Phys.* <https://doi.org/10.1002/9780470259498.ch6> (2008).
34. Underwood, J. G. & Reid, K. L. Time-resolved photoelectron angular distributions as a probe of intramolecular dynamics: connecting the molecular frame and the laboratory frame. *J. Chem. Phys.* **113**, 1067–1074 (2000).
35. Box, G. & Tiao, G. *Bayesian. Inference in Statistical Analysis* (Wiley, 2011).
36. Hegazy, K. *Merging Ultrafast Gas-Phase Diffraction Experiment, Theory, and Machine Learning for a New Look at Molecular Dynamics* (Stanford University Press, 2023).
37. Herzberg, G. *Electronic Spectra and Electronic Structure of Polyatomic Molecules* 1st edn, Vol. 2 (Van Nostrand, 1966).
38. Teffo, J.-L. & Chédin, A. Internuclear potential and equilibrium structure of the nitrous oxide molecule from rovibrational data. *J. Mol. Spect.* **135**, 389–409 (1989).
39. Ma, L. et al. Ultrafast x-ray and electron scattering of free molecules: a comparative evaluation. *Struct. Dyn.* **7**, 034102 (2020).
40. Baskin, J. S. & Zewail, A. H. Oriented ensembles in ultrafast electron diffraction. *Chem. Phys. Chem.* **7**, 1562–1574 (2006).
41. Xiong, Y. et al. Retrieval of the molecular orientation distribution from atom-pair angular distributions. *Phys. Rev. A* **106**, 033109 (2022).
42. Parrish, R. & Martinez, T. Ab initio computation of rotationally-averaged pump-probe x-ray and electron diffraction signals. *J. Chem. Theory Comput.* **15**, 1523–1537 (2019).
43. Yang, J. et al. Simultaneous observation of nuclear and electronic dynamics by ultrafast electron diffraction. *Science* **368**, 885–889 (2020).
44. Yong, H. et al. Observation of the molecular response to light upon photoexcitation. *Nature Communications* **11**, 2157 (2020).
45. GitHub. *Bigr Github Repository*. <https://github.com/khegazy/BIGR> (2022).
46. Cryan, J. P., Bucksbaum, P. H. & Coffee, R. N. Field-free alignment in repetitively kicked nitrogen gas. *Phys. Rev. A* **80**, 063412 (2009).
47. Toth, R. A. Line-frequency measurements and analysis of n₂o between 900 and 4700 cm⁻¹. *Appl. Opt.* **30**, 5289–5315 (1991).
48. Bohlin, A., Kindeya, A., Nordstrom, E. & Bengtsson, P.-E. Validation of a rotational coherent anti-stokes raman scattering model for n₂o at temperatures from 295 k to 796 k. *J. Spectr.* **43**, 604–610 (2012).
49. Hockett, P. General phenomenology of ionization from aligned molecular ensembles. *J. Phys.* **17**, 023069 (2015).
50. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: The MCMC hammer. *Publ. Astron. Soc. 125*, 306 (2013).
51. Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (1970).
52. Wolf, T. J. A. et al. The photochemical ring-opening of 1,3-cyclohexadiene imaged by ultrafast electron diffraction. *Nat. Chem.* **11**, 504–509 (2019).

Acknowledgements

Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The UED work was performed at SLAC MeV-UED, which is supported in part by the DOE BES SUF Division Accelerator and Detector research and development program, the LCLS Facility, and SLAC under contract Nos. DE-AC02-05-CH11231 and DE-AC02-76SF00515. Markus Ilchen acknowledges funding by the Volkswagen Foundation for a Peter-Paul-Ewald Fellowship. We thank Markus Gühr for his help in setting up the SLAC gas phase UED experiment. We thank Theodore Vecchione for his help operating the SLAC UED facility.

Author contributions

K.H., V.M. and R.C. conceived of this method. K.H. derived Eq. (16) with help from V.M., and V.M. derived its extension to excited state dynamics in Eq. (14). K.H. derived and implemented the Bayesian Inference approach and wrote the manuscript. J.C. and P.B. provided critical feedback and advice about the analysis and this paper. R.C. managed the project. K.H., R.C., M.L., J.Y., X.S., R.L., S.W., X.W., J.C., N.H., K.J., I.M. and J.C. helped run the SLAC UED experiment and collect the data used in this work. V.M. simulated the NO₂ axis distribution moments.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42005-023-01420-9>.

Correspondence and requests for materials should be addressed to Kareem Hegazy or Ryan Coffee.

Peer review information *Communications Physics* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023