Bayesian inference and deterministic anisotropy for molecular geometry retrieval in gas-phase diffraction experiments

UED Opportunities for Dynamical Imaging of Materials
11/6/2023

Kareem Hegazy, Varun Makhija, Phil Bucksbaum, Jeff Corbett, James Cryan, Nick Hartmann, Markus Ilchen, Keith Jobe, Renkai Li, Igor Makasyuk, Xiaozhe Shen, Xijie Wang, Stephen Weathersby, Jie Yang, **Ryan Coffee**

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Presentation Objective and Contents

We introduce a data driven approach to rigorously retrieve molecular frame geometries while reducing simulation requirements.

1. Current approaches and the curse of dimensionality
2. Anisotropy reveals the molecular frame (MF)
3. Bayesian Inference
   a. Overcoming the curse of dimensionality
   b. Metropolis Hastings Algorithm
   c. Results
   d. Systematic errors from single geometry assumptions
4. Future Application to Excited State Geometric Dynamics
5. Summary

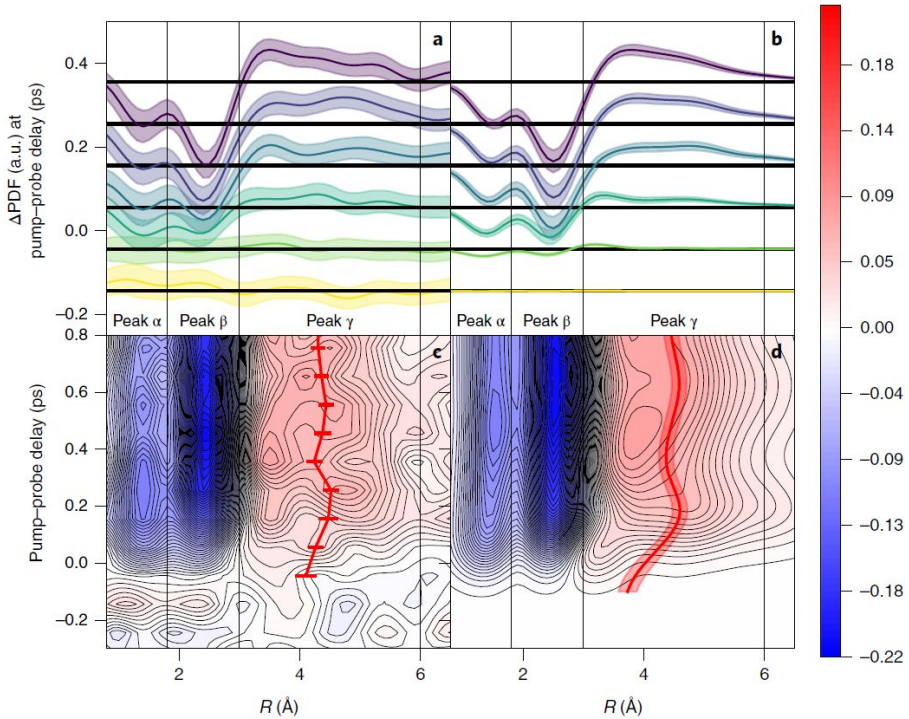
Current Molecular Geometry Retrieval Methods
Traditional and New Methods

PULSE

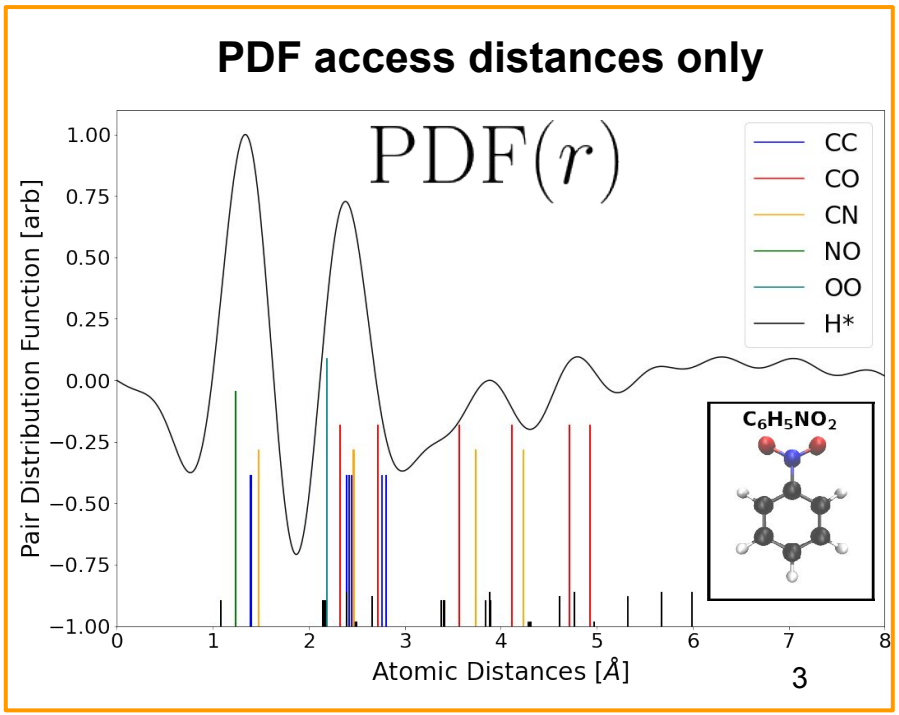
SLAC

Traditional Method

- Pair distribution function (PDF)
- Compare PDF, or diffraction, with theory to interpret the molecular geometry transience



Wolf, T J A et al. "The photochemical ring-opening of 1,3-cyclohexadiene imaged by ultrafast electron diffraction." *Nature chemistry* vol. 11,6 (2019): 504-509. doi:10.1038/s41557-019-0252-7



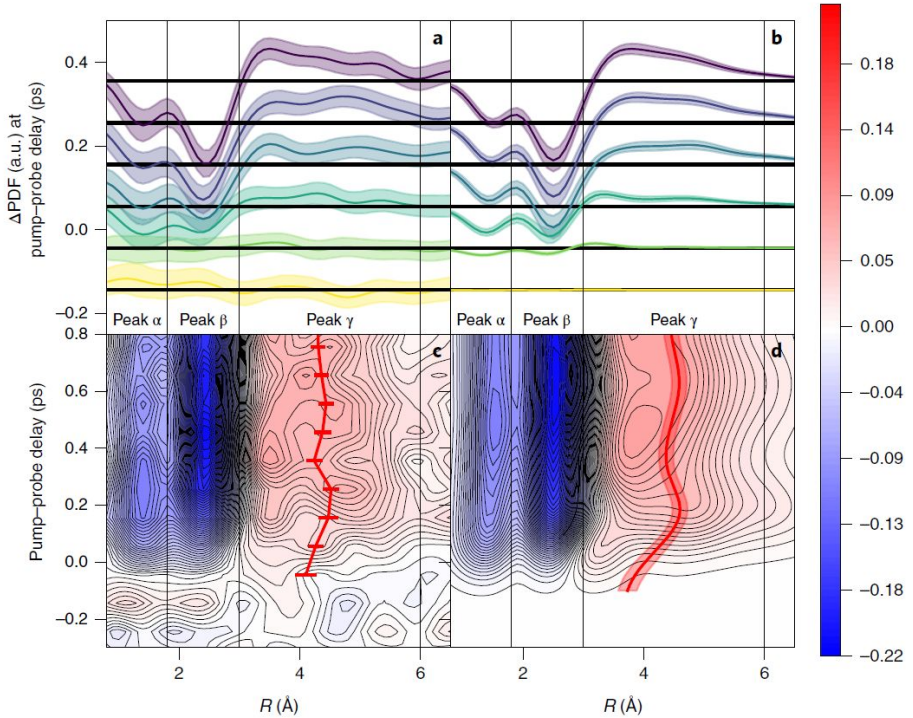
Current Molecular Geometry Retrieval Methods
Traditional and New Methods

PULSE

SLAC

Traditional Method

- Pair distribution function (PDF)
- Compare PDF, or diffraction, with theory to interpret the molecular geometry transience

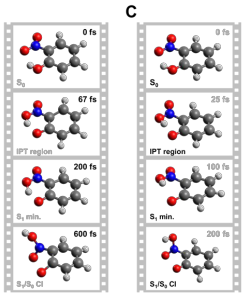


Wolf, T J A et al. "The photochemical ring-opening of 1,3-cyclohexadiene imaged by ultrafast electron diffraction." *Nature chemistry* vol. 11,6 (2019): 504-509. doi:10.1038/s41557-019-0252-7

Data Focused/Driven Method

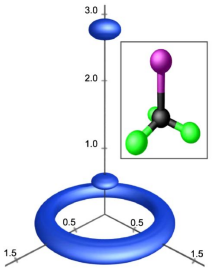
- Use ML to optimize primary features
- Use molecular alignment to access the molecular frame of symmetric tops
- Use many simulated geometries to statistically improve precision

Nune et al. retrieved the primary molecular geometry changes with a genetic algorithm. Each atom was allowed to move within intuitive constraints.



Nunes, J P F et al. "Imaging proton transfer dynamics in o-nitrophenol by Ultrafast Electron Diffraction" Submitted to Journal

Hensley et al. accessed the molecular frame of symmetric tops at the peak of alignment.



Hensley, Christopher J et al. "Imaging of isolated molecules with ultrafast electron pulses." Physical review letters vol. 109,13 (2012): 133202. doi:10.1103/PhysRevLett.109.133202

Stankus et al. compared excited state measurements to many simulations to form a distribution of likely atomic distances and angles.

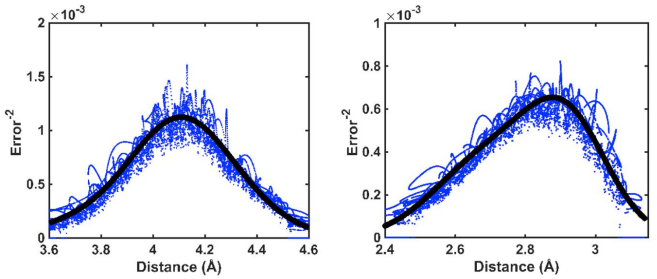


Fig. 4: The inverse of the least-squares fitting error as a function of the given structural parameter and the fits as described in the text. Left: O-C5 distance at 0.05 ps delay time. Right: O-N distance at 2.65 ps delay time.

Stankus, B., Yong, H., Zotev, N. et al. Ultrafast X-ray scattering reveals vibrational coherence following Rydberg excitation. Nat. Chem. 11, 716–721 (2019). https://doi.org/10.1038/s41557-019-0291-0

Current Molecular Geometry Retrieval Methods
Curse of Dimensionality

PULSE

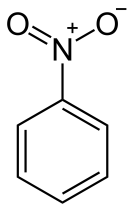
SLAC

The number of geometries to sample (S) grows exponentially, for a grid search

$$S = R^{3N-6}$$

R: Number of sample points per dimension
N: Number of atoms
V: Percent of volume

Nitrobenzene: 21 dimensions (ignoring H)



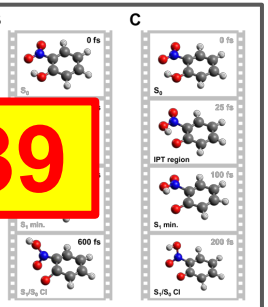
S	R	V|R=10	V|R=4
10^6	1.93	10^{-14}%	2.3×10^{-5}%
10^9	2.40	10^{-12}%	2.3×10^{-3}%

Data Focused/Driven Method

- Use ML to optimize primary features
- Use molecular alignment to access the molecular frame of symmetric tops
- Use many simulated geometries to statistically improve precision

Nunes et al. retrieved the primary molecular geometry cha with a genetic algorithm. Ea atom was allowed to move within intuitive constraints.

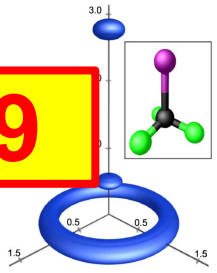
39



Nunes, J P F et al. "Imaging proton transfer dynamics in o-nitrophenol by Ultrafast Electron Diffraction" Submitted to Journal

Hensley et al. accessed th molecular frame of symmetric at the peak of alignment.

9



Hensley, Christopher J et al. "Imaging of isolated molecules with ultrafast electron pulses." Physical review letters vol. 109,13 (2012): 133202. doi:10.1103/PhysRevLett.109.133202

Stankus et al. compared excited state measurements to many simulations to form a distribution of likely atomic distances and angles.

15

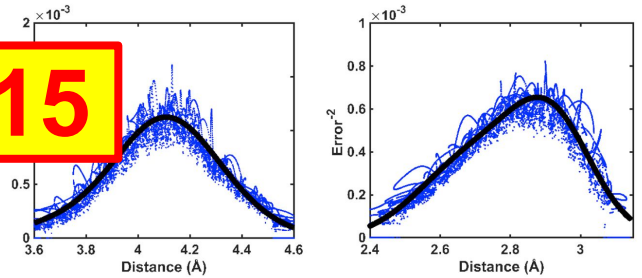
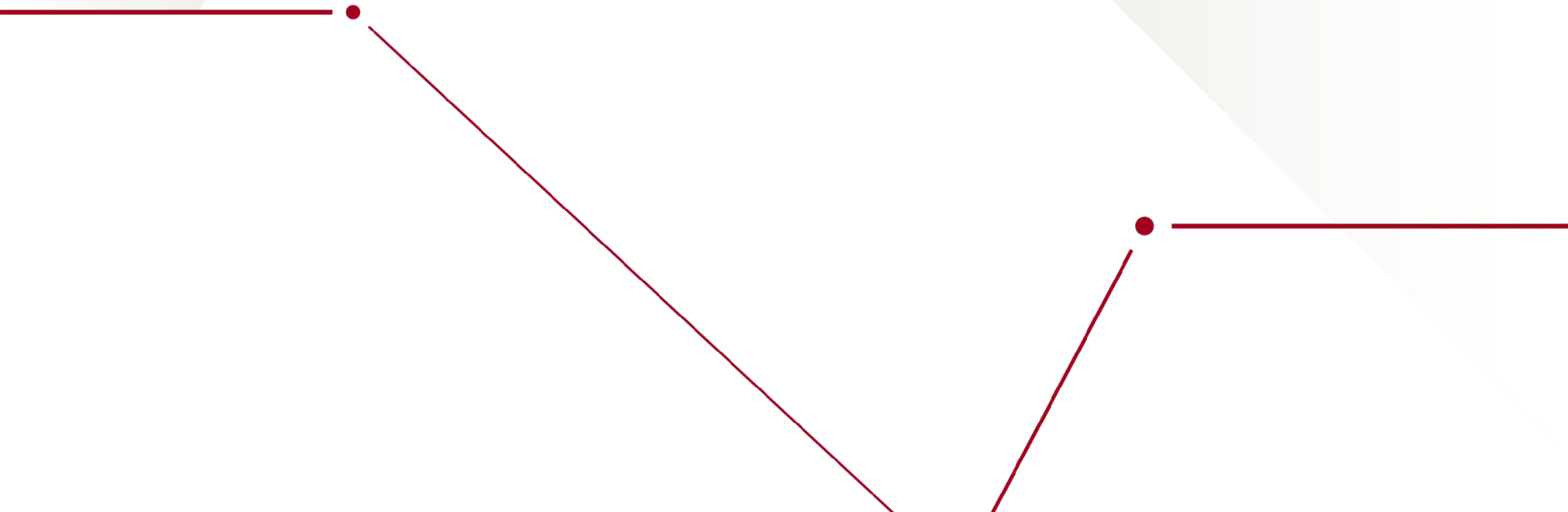


Fig. 4: The inverse of the least-squares fitting error as a function of the given structural parameter and the fits as described in the text. Left: O-C5 distance at 0.05 ps delay time. Right: O-N distance at 2.65 ps delay time.

Stankus, B., Yong, H., Zotev, N. et al. Ultrafast X-ray scattering reveals vibrational coherence following Rydberg excitation. Nat. Chem. 11, 716–721 (2019). https://doi.org/10.1038/s41557-019-0291-0

Anisotropy Reveals the Molecular Frame (MF)



Anisotropy Reveals the Molecular Frame (MF)

- Anisotropy provides constraints on molecular frame (MF) degrees of freedom.
- Combining many measurements (constraints) allows one to retrieve the MF.
- Application in photo-electron spectroscopy
  - V. Makhija, et. al., (2016), arXiv:1611.06476 [physics.atom-ph]
  - C. Marceau, et. al., Phys. Rev. Lett. 119, 083401 (2017)
  - M. Gregory, et. al., (2020), arXiv:2012.04561 [physics.chem-ph]
- Use a stretched NO_2: an asymmetric top

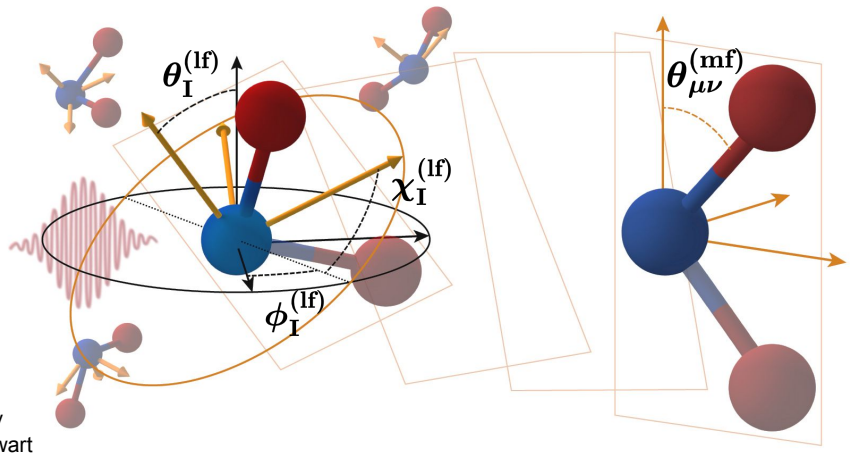
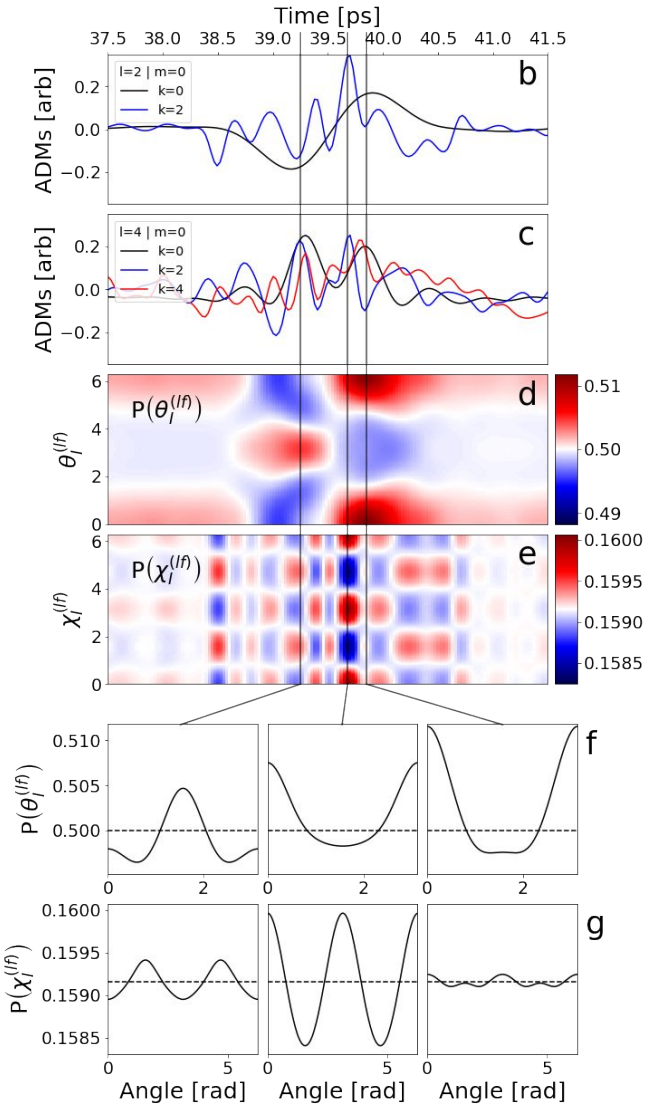


Illustration by Gregory Stewart



NO_2 rotational wavepacket anisotropy signature

Accessing the MF via Deterministic Anisotropy

$$\langle I(\mathbf{q},t)\rangle = \mathcal{I}\Bigg(\sum_{\mu} |f_\mu(q)|^2 + \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\Bigg\{ f_\mu(q) f_\nu^*(q) \sum_{l} 4\pi i^l \times \sum_{m,k} (-1)^k \underbrace{Y_l^m\left(\theta_q^{(\mathrm{lf})}, \phi_q^{(\mathrm{lf})}\right)}_{\text{Lab Frame}} \langle \Psi(t)| \underbrace{D_{mk}^l\left(\phi_{\mathrm{I}}^{(\mathrm{lf})}, \theta_{\mathrm{I}}^{(\mathrm{lf})}, \chi_{\mathrm{I}}^{(\mathrm{lf})}\right)}_{\text{Ensemble Anisotropy}} \underbrace{j_l(q\Delta R_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right)}_{\text{Molecular Frame Structure}} |\Psi(t)\rangle \Bigg\} \Bigg)$$

Independent atom approximation

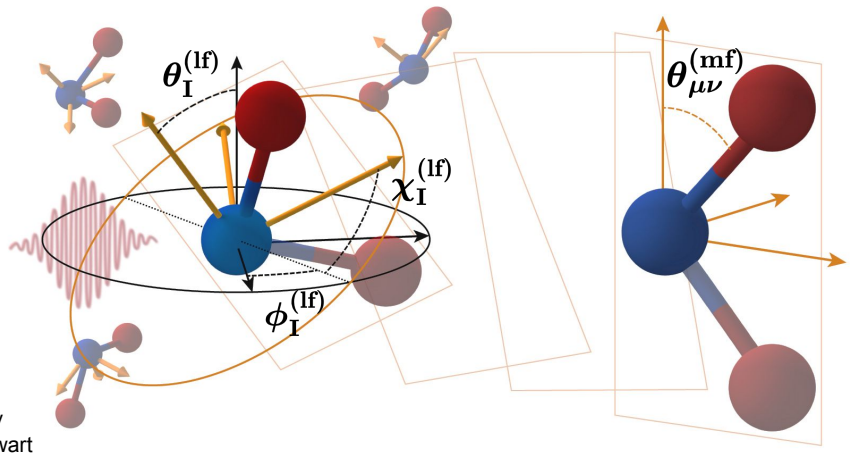


Illustration by
Gregory Stewart

Accessing the MF via Deterministic Anisotropy

$$\langle I(\mathbf{q},t)\rangle = \mathcal{I}\Bigg(\sum_{\mu} |f_\mu(q)|^2 + \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\Bigg\{ f_\mu(q) f_\nu^*(q) \sum_l 4\pi i^l \times \sum_{m,k} (-1)^k \underbrace{Y_l^m\left(\theta_q^{(\mathrm{lf})}, \phi_q^{(\mathrm{lf})}\right)}_{\text{Lab Frame}} \langle\Psi(t)| \underbrace{D_{mk}^l\left(\phi_{\mathrm{I}}^{(\mathrm{lf})}, \theta_{\mathrm{I}}^{(\mathrm{lf})}, \chi_{\mathrm{I}}^{(\mathrm{lf})}\right)}_{\text{Ensemble Anisotropy}} \underbrace{j_l(q\Delta R_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right)}_{\text{Molecular Frame Structure}} |\Psi(t)\rangle \Bigg\}\Bigg)$$

Independent atom approximation

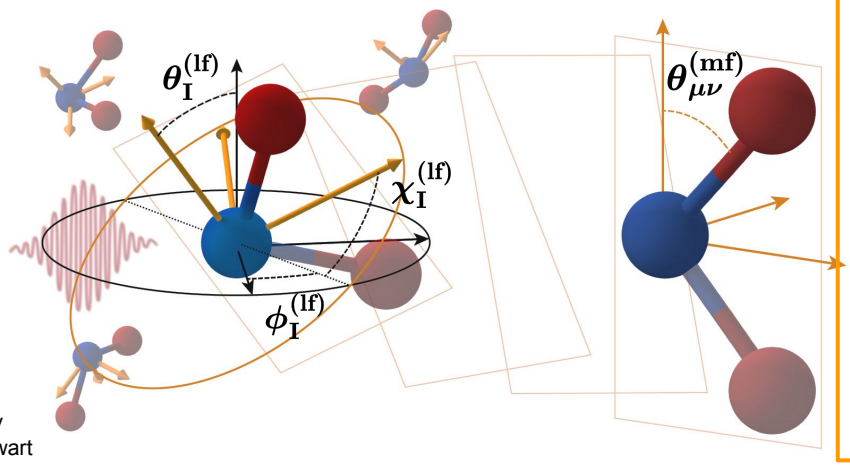
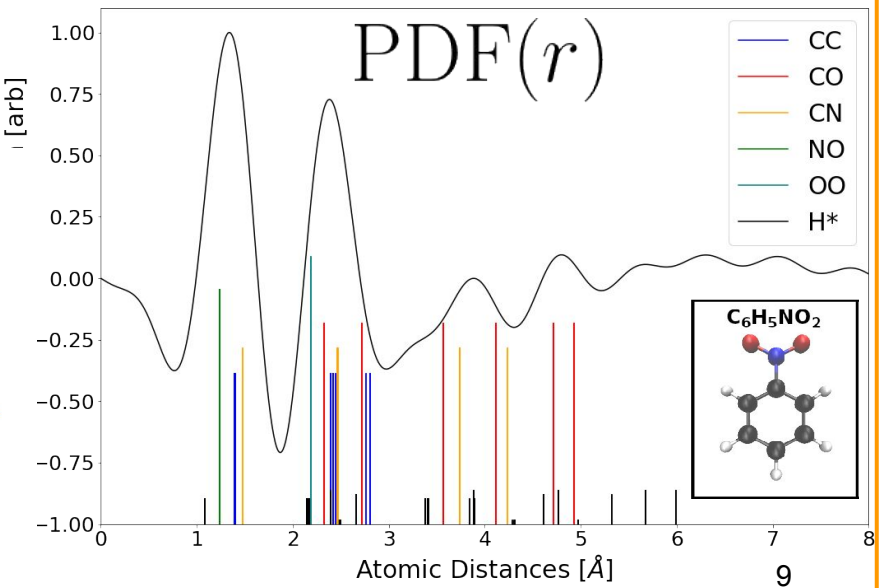


Illustration by Gregory Stewart

PDF access distances only



Accessing the MF via Deterministic Anisotropy

$$\langle I(\mathbf{q},t)\rangle = \mathcal{I}\Bigg(\sum_{\mu} |f_\mu(q)|^2 + \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\Bigg\{ f_\mu(q) f_\nu^*(q) \sum_l 4\pi i^l \times \sum_{m,k} (-1)^k \underbrace{Y_l^m\left(\theta_q^{(\mathrm{lf})}, \phi_q^{(\mathrm{lf})}\right)}_{\text{Lab Frame}} \langle \Psi(t)| \underbrace{D_{mk}^l\left(\phi_{\mathrm{I}}^{(\mathrm{lf})}, \theta_{\mathrm{I}}^{(\mathrm{lf})}, \chi_{\mathrm{I}}^{(\mathrm{lf})}\right)}_{\text{Ensemble Anisotropy}} \underbrace{j_l(q\Delta R_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right)}_{\text{Molecular Frame Structure}} |\Psi(t)\rangle \Bigg\}\Bigg)$$

Independent atom approximation

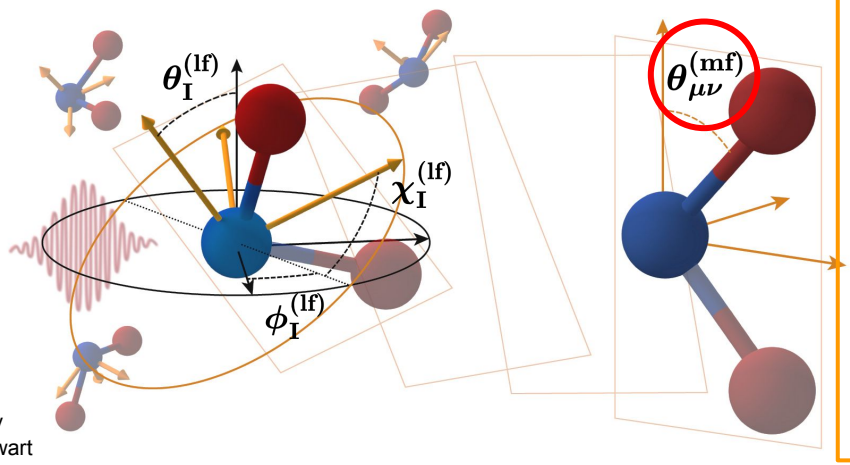
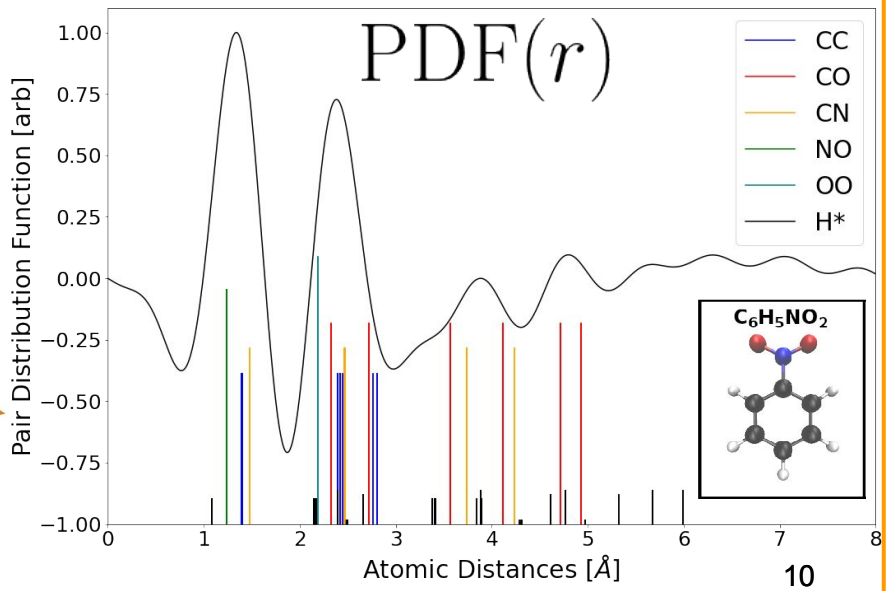


Illustration by Gregory Stewart

PDF access distances only



Accessing the MF via Deterministic Anisotropy

$$\langle I(\mathbf{q},t)\rangle = \mathcal{I}\Bigg(\sum_{\mu} |f_\mu(q)|^2 + \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\Bigg\{ f_\mu(q) f_\nu^*(q) \sum_l 4\pi i^l \times \sum_{m,k} (-1)^k \underbrace{Y_l^m\left(\theta_q^{(\mathrm{lf})}, \phi_q^{(\mathrm{lf})}\right)}_{\text{Lab Frame}} \langle \Psi(t)| \underbrace{D_{mk}^l\left(\phi_{\mathrm{I}}^{(\mathrm{lf})}, \theta_{\mathrm{I}}^{(\mathrm{lf})}, \chi_{\mathrm{I}}^{(\mathrm{lf})}\right)}_{\text{Ensemble Anisotropy}} \underbrace{j_l(q\Delta R_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right)}_{\text{Molecular Frame Structure}} |\Psi(t)\rangle \Bigg\}\Bigg)$$

Independent atom approximation

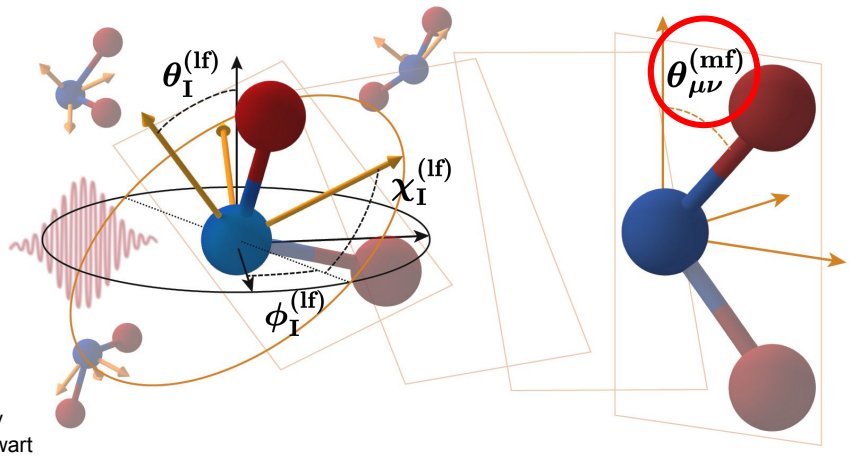


Illustration by Gregory Stewart

Explicit dependence on molecular frame angles provides an opportunity to develop methods that extract these angles, and hence a unique geometry, from data alone.

Accessing the MF via Deterministic Anisotropy
Rigid Rotor

PULSE

SLAC

Rigid Rotor Approximation

$$\langle I(\mathbf{q},t)\rangle_{\text{rigid}} = \mathcal{I}\Bigg(\sum_{\mu}|f_\mu(q)|^2 + \sum_{\mu,\nu:\mu\neq\nu}\text{Re}\Bigg\{f_\mu(q)f_\nu^*(q)\sum_l\frac{32\pi^3 i^l}{2l+1}$$

$$\times\sum_{m,k}(-1)^k\underbrace{Y_l^m\left(\theta_q^{(\text{lf})},\phi_q^{(\text{lf})}\right)}_{\text{Lab Frame}}\underbrace{\langle\Psi(0)|j_l(q\Delta R_{\mu\nu})Y_l^{-k}\left(\theta_{\mu\nu}^{(\text{mf})},\phi_{\mu\nu}^{(\text{mf})}\right)|\Psi(0)\rangle}_{\text{Molecular Frame Structure}}\underbrace{\mathcal{A}_{mk}^l(t)\big|_{\text{rigid}}}_{\text{Anisotropy}}\Bigg\}\Bigg)$$

Remove LF Y_l^m by fitting measured anisotropy

$$B_l^m(q,t) = \mathcal{I}\sum_{\mu,\nu:\mu\neq\nu}\text{Re}\Bigg\{f_\mu(q)f_\nu^*(q)\frac{32\pi^3 i^l}{2l+1}(-1)^k$$

$$\times\underbrace{\langle\Psi(0)|j_l(q\Delta R_{\mu\nu})Y_l^{-k}\left(\theta_{\mu\nu}^{(\text{mf})},\phi_{\mu\nu}^{(\text{mf})}\right)|\Psi(0)\rangle}_{\text{Molecular Frame Structure}}\underbrace{\mathcal{A}_{mk}^l(t)\big|_{\text{rigid}}}_{\text{Anisotropy}}\Bigg\}$$

Remove time dependence by fitting ADMs

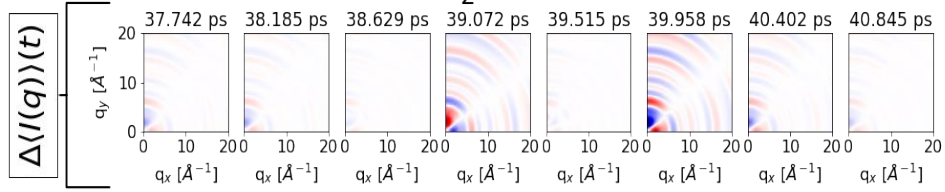
$$C_{lmk}(q) = \mathcal{I}\sum_{\mu,\nu:\mu\neq\nu}\text{Re}\Bigg\{f_\mu(q)f_\nu^*(q)(-1)^k\frac{32\pi^3 i^l}{2l+1}$$

$$\times\underbrace{\langle\Psi(0)|j_l(q\Delta R_{\mu\nu})Y_l^{-k}\left(\theta_{\mu\nu}^{(\text{mf})},\phi_{\mu\nu}^{(\text{mf})}\right)|\Psi(0)\rangle}_{\text{Molecular Frame Structure}}\Bigg\}$$

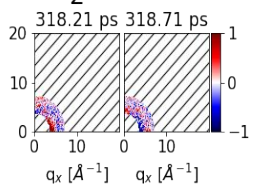
$$M_{lmk}(q) = \frac{C_{lmk}(q)}{\sum_\mu|f_\mu(q)|^2}.$$

$\Delta\langle I(q)\rangle(t)$

NO_2 Simulation

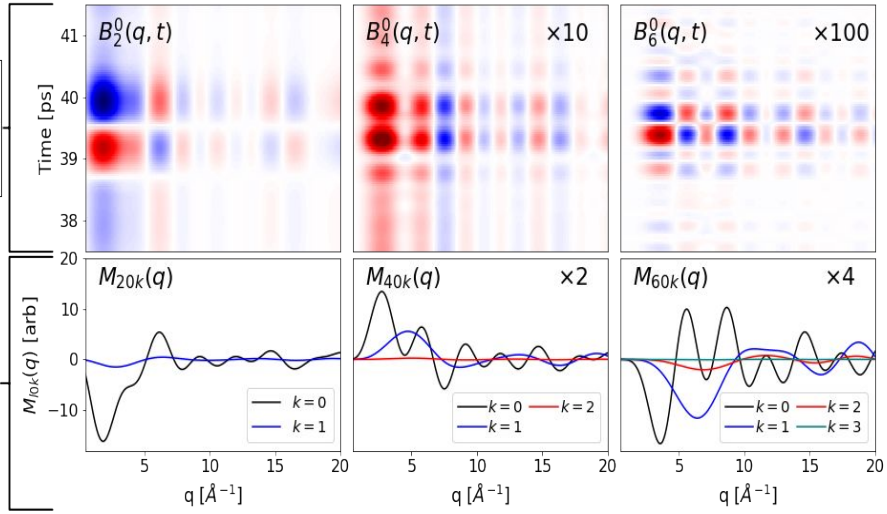


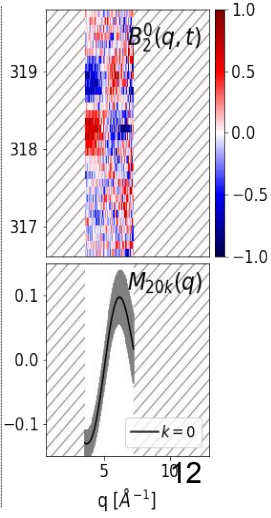
N_2O Data



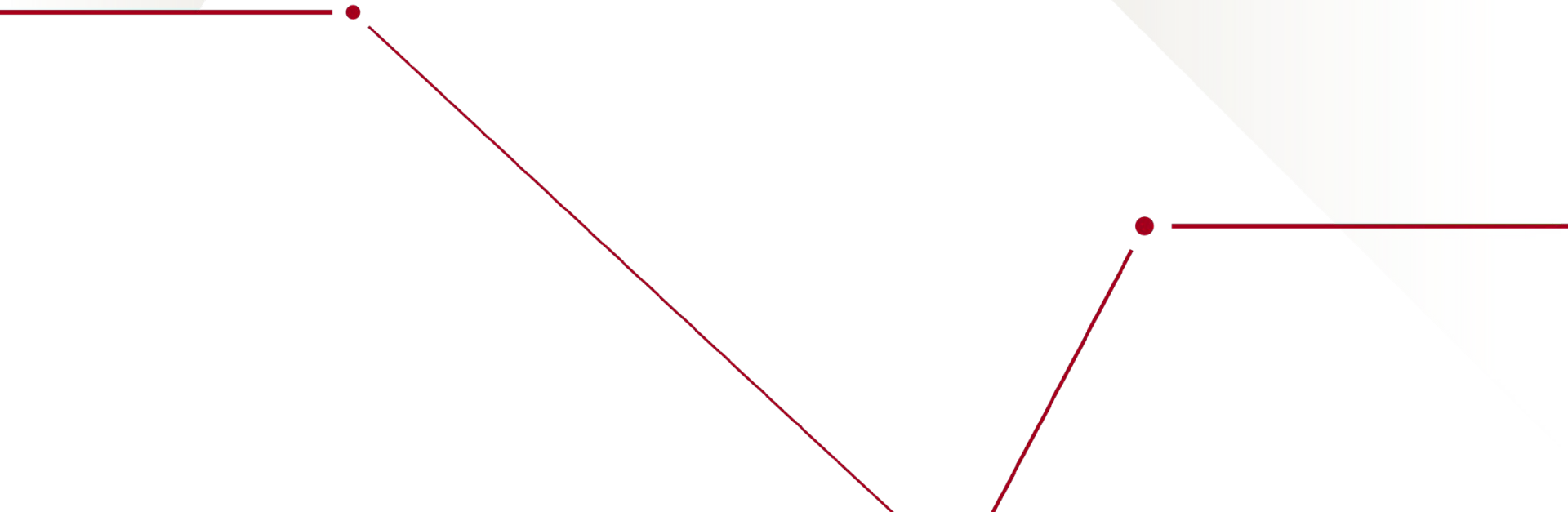
$B_l^0(q,t)$

Extracting Molecular Frame Information





Bayesian Inference



Bayesian Inferencing
Reframing the Problem

PULSE

SLAC

What we have: Molecular frame representations for the nuclear geometry with explicit dependence on pair-wise distances and angles

What we want: Invert the integral equation for $|\Psi(\mathbf{R}^{(mf)})|^2$

How we do it: Approximate $|\Psi(\mathbf{R}^{(mf)})|^2$ with a chosen distribution and solve for the model parameter ($\boldsymbol{\theta}$) distribution P($\boldsymbol{\theta}$|C)

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

Bayesian Inference

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

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

Delta Distribution:

$$P^{(\delta)}(\mathbf{R}|\,\boldsymbol{\Theta}, C) = \delta\left(\boldsymbol{\Theta}^{(\delta)} - \mathbf{R}\right)$$

$$\boldsymbol{\Theta}^{(\delta)} = \left[\langle \mathrm{NO}^{(1)}\rangle, \langle \mathrm{NO}^{(2)}\rangle, \langle \angle \mathrm{ONO}\rangle\right]$$

Normal Distribution:

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

$$\boldsymbol{\Theta}^{(\mathcal{N})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \sigma\left(\mathrm{NO}^{(1)}\right), \langle \mathrm{NO}^{(2)}\rangle, \sigma\left(\mathrm{NO}^{(2)}\right), \langle \angle \mathrm{ONO}\rangle, \sigma\left(\angle \mathrm{ONO}\right)\right]$$

Bayesian Inferencing
Reframing the Problem

What we have: Molecular frame representations for the nuclear geometry with explicit dependence on pair-wise distances and angles

What we want: Invert the integral equation for $|\Psi(\mathbf{R}^{(mf)})|^2$

How we do it: Approximate $|\Psi(\mathbf{R}^{(mf)})|^2$ with a chosen distribution and solve for the model parameter ($\boldsymbol{\theta}$) distribution P($\boldsymbol{\theta}$|C)

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

Bayesian Inference

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

When finding **θ** one can make novel measurements

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

Delta Distribution:

$$P^{(\delta)}(\mathbf{R}|\,\boldsymbol{\Theta}, C) = \delta\left(\boldsymbol{\Theta}^{(\delta)} - \mathbf{R}\right)$$

$$\boldsymbol{\Theta}^{(\delta)} = \left[\langle \mathrm{NO}^{(1)}\rangle, \langle \mathrm{NO}^{(2)}\rangle, \langle \angle \mathrm{ONO}\rangle\right]$$

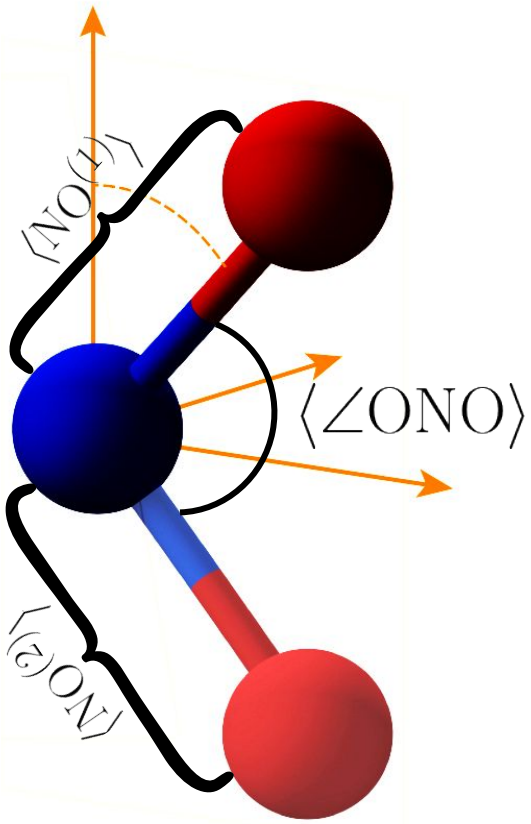
Normal Distribution:

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

$$\boldsymbol{\Theta}^{(\mathcal{N})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \sigma\left(\mathrm{NO}^{(1)}\right), \langle \mathrm{NO}^{(2)}\rangle, \sigma\left(\mathrm{NO}^{(2)}\right), \langle \angle \mathrm{ONO}\rangle, \sigma(\angle \mathrm{ONO})\right]$$

Modeling $|\Psi(\mathbf{R}^{(mf)})|^2$ and Search Parameters

- The **θ** parameterization allows for novel measurements
  - Degrees of freedom to specify a unique geometry
  - Bond distances and angles
  - Width of the wave packet σ(⟨…⟩)



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

Delta Distribution:

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

$$\mathbf{\Theta}^{(\delta)} = \left[\langle \mathrm{NO}^{(1)}\rangle, \langle \mathrm{NO}^{(2)}\rangle, \langle \angle \mathrm{ONO}\rangle\right]$$

Normal Distribution:

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

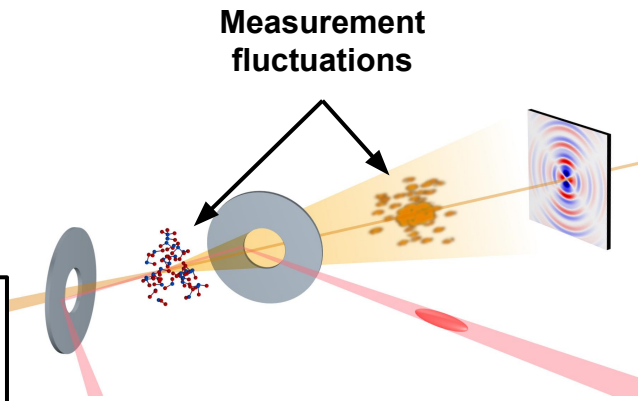
$$\mathbf{\Theta}^{(\mathcal{N})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \sigma\left(\mathrm{NO}^{(1)}\right), \langle \mathrm{NO}^{(2)}\rangle, \sigma\left(\mathrm{NO}^{(2)}\right), \langle \angle \mathrm{ONO}\rangle, \sigma(\angle \mathrm{ONO})\right]$$

PULSE

SLAC

Bayes Rule

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



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

Delta·Distribution:

$$P^{(\delta)}(\mathbf{R}|\,\boldsymbol{\Theta}, C) = \delta\left(\boldsymbol{\Theta}^{(\delta)} - \mathbf{R}\right)$$

$$\boldsymbol{\Theta}^{(\delta)} = \left[\langle \mathrm{NO}^{(1)}\rangle, \langle \mathrm{NO}^{(2)}\rangle, \langle \angle \mathrm{ONO}\rangle\right]$$

Normal Distribution:

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

$$\boldsymbol{\Theta}^{(\mathcal{N})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \sigma\left(\mathrm{NO}^{(1)}\right), \langle \mathrm{NO}^{(2)}\rangle, \sigma\left(\mathrm{NO}^{(2)}\right), \langle \angle \mathrm{ONO}\rangle, \sigma\left(\angle \mathrm{ONO}\right)\right]$$

Applying Bayesian Inferencing
Bayesian Inferencing

PULSE

SLAC

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

Prior Distribution

$$P(\Theta) = \text{constant}$$

- The probability of observing these **Θ** parameters
  - Relies on a priori chemical knowledge and chemical intuition
- A constant prior is chosen to have an unbiased **Θ** search
  - For unphysical values the prior is 0 e.g., negative bond distance

Applying Bayesian Inferencing
Bayesian Inferencing

PULSE

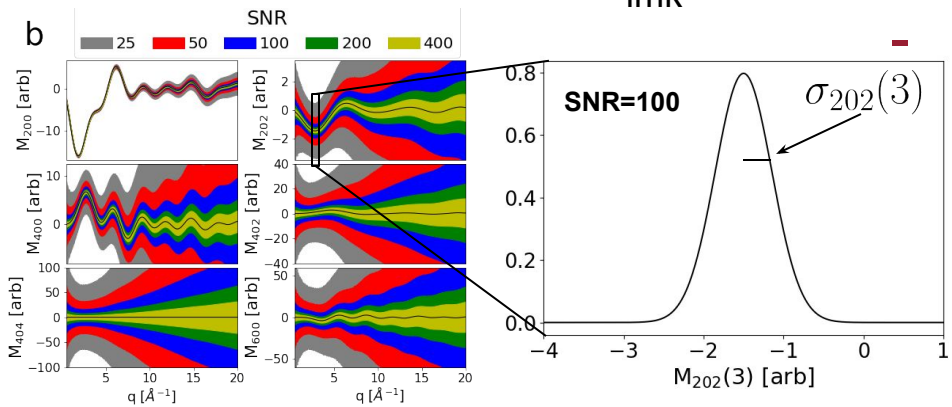
SLAC

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

Likelihood Distribution

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

- Each measured $C_{lmk}(q)$ is it's own Gaussian probability distribution



By the central limit theorem we expect $C_{lmk}(q)$ to be Gaussian

- Each $C_{lmk}(q)$ is a distribution of weighted sums of the measured pixels
- $C_{lmk}(q)$ distribution has > 100 entries

PULSE

SLAC

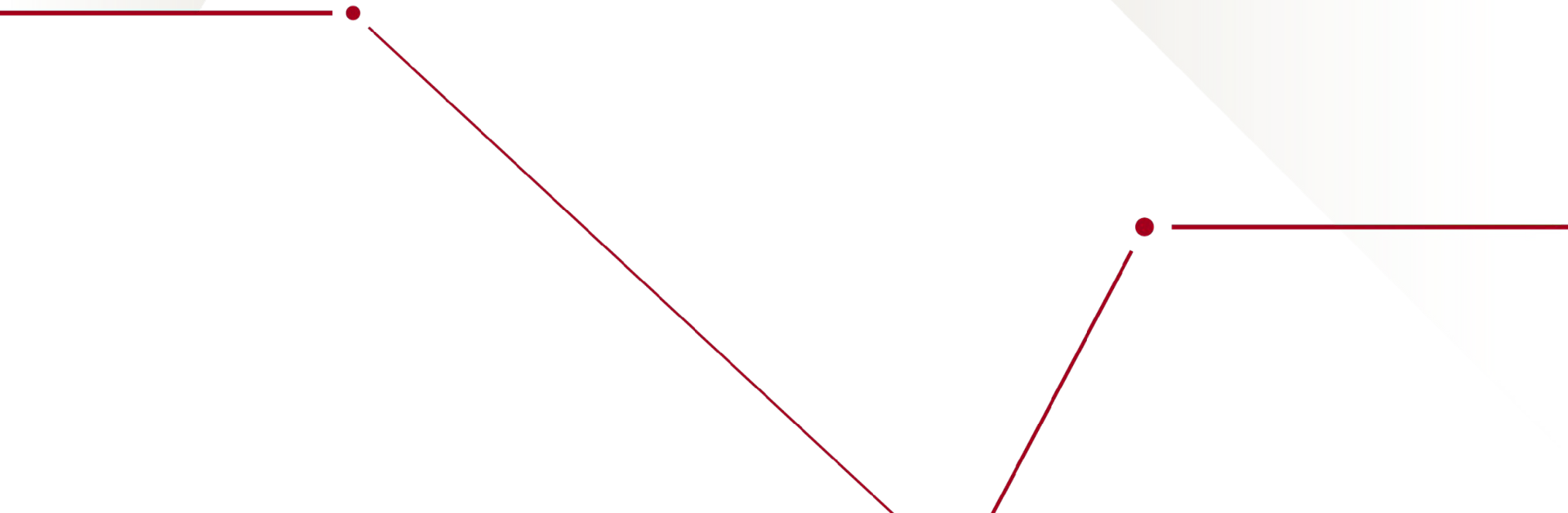
Metropolis Hastings Algorithm

- Unbiased search for **Θ**
- **Θ** are preferentially selected based on their agreement with data
- Spend more time searching regions of high likelihood
  - If **Θ** is twice as likely as **Θ'** we spend twice as much time sampling the area around **Θ**

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

$$= \min\left[\frac{P(C|\boldsymbol{\Theta}')P(\boldsymbol{\Theta}')}{P(C|\boldsymbol{\Theta})P(\boldsymbol{\Theta})}, 1\right]$$

Application to simulated NO_2 and measured N_2O



Retrieving the MF Geometry Probability Distribution
Algorithm Input

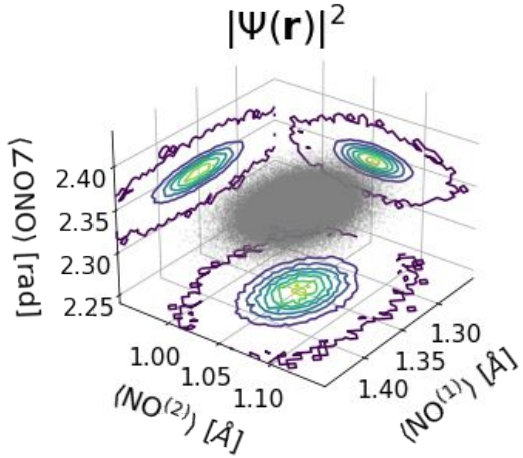
PULSE

SLAC

Simulated NO_2

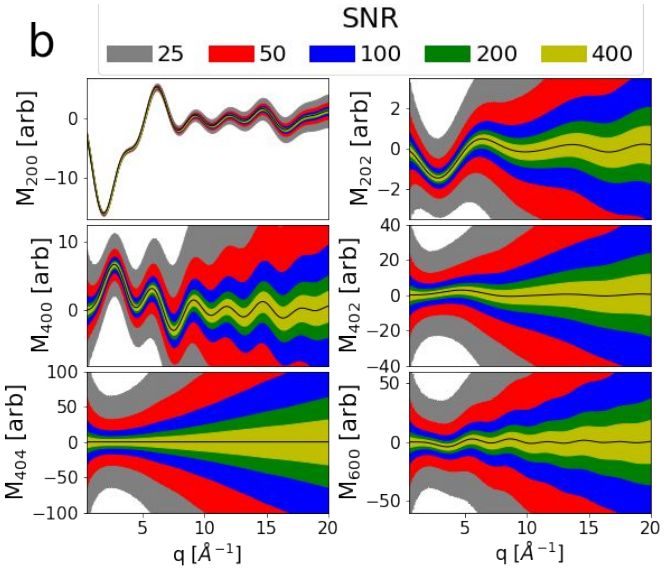
Measured N_2O

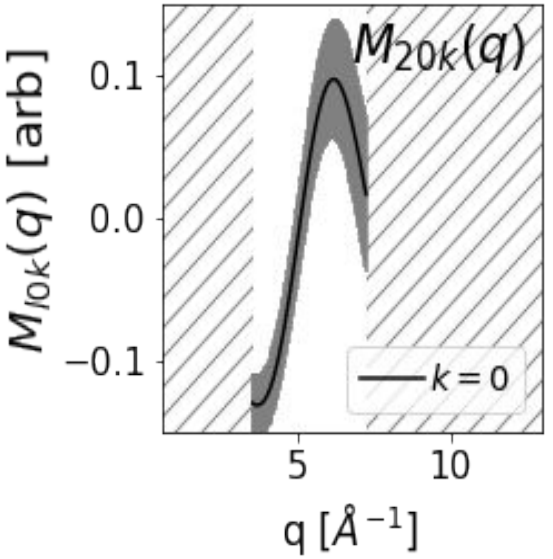
$|\Psi(\boldsymbol{r}^{(mf)})|^2$



?

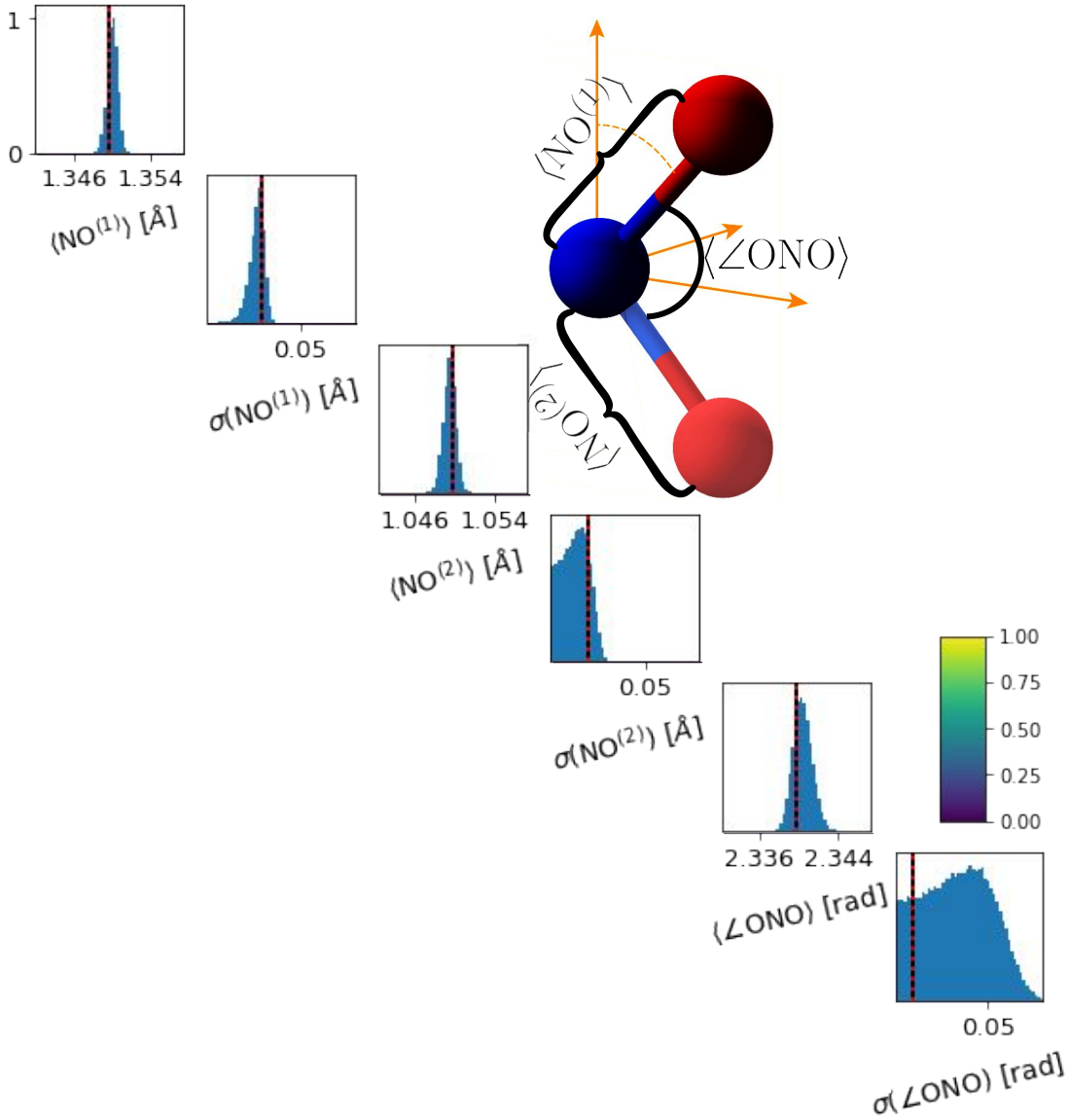
Data: $C_{lmk}(q)$





Only 3.5 Å^{-1}!

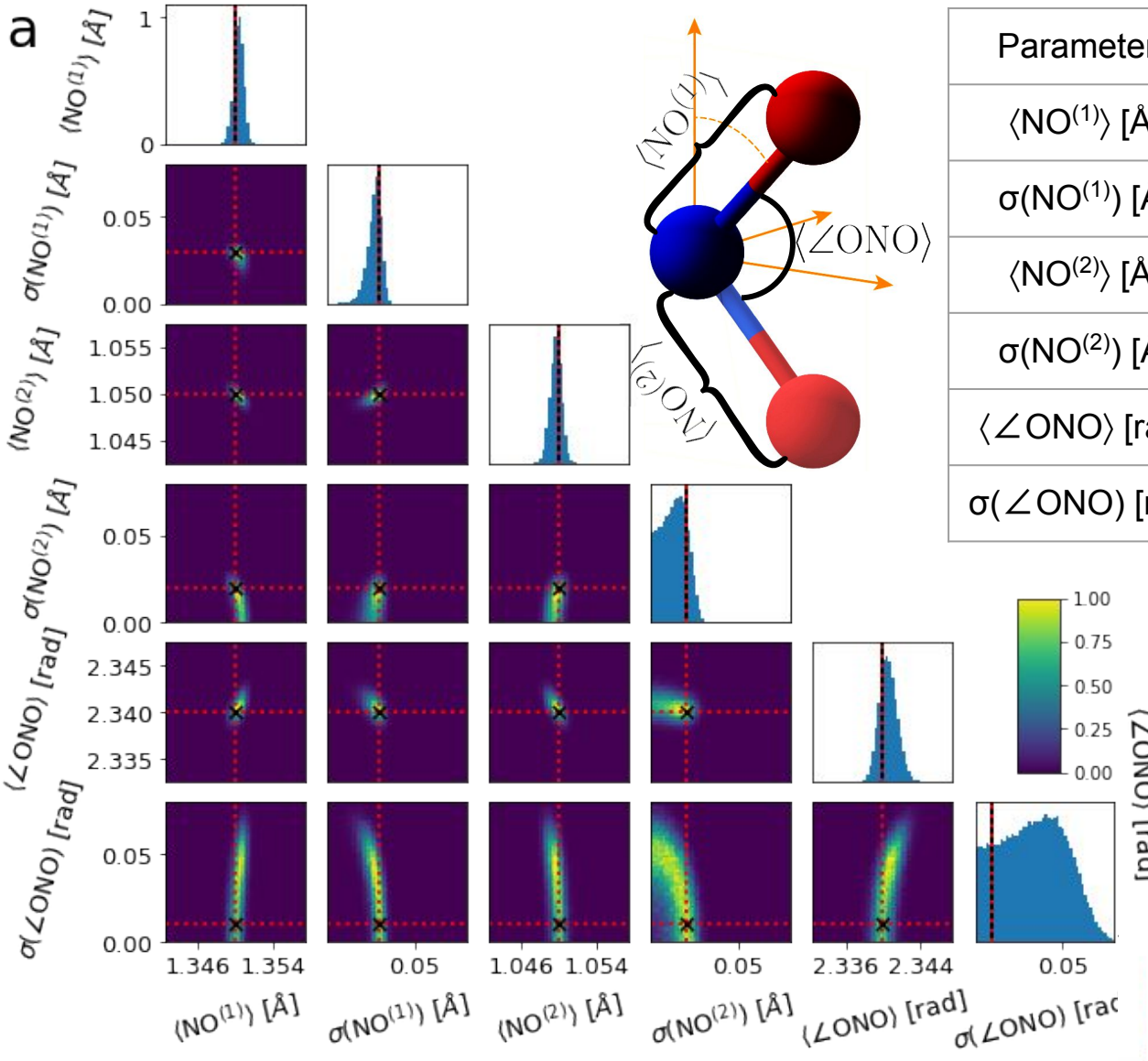
Retrieving the MF Geometry Probability Distribution
Results: P(θ|C)



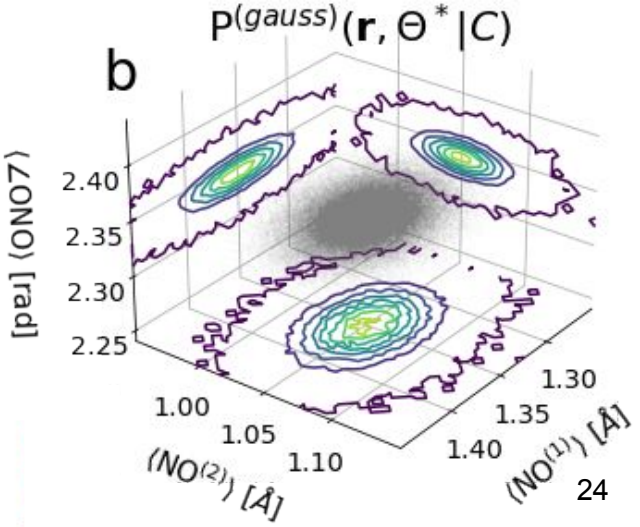
Retrieving the MF Geometry Probability Distribution
Results: P(θ|C)

PULSE

SLAC



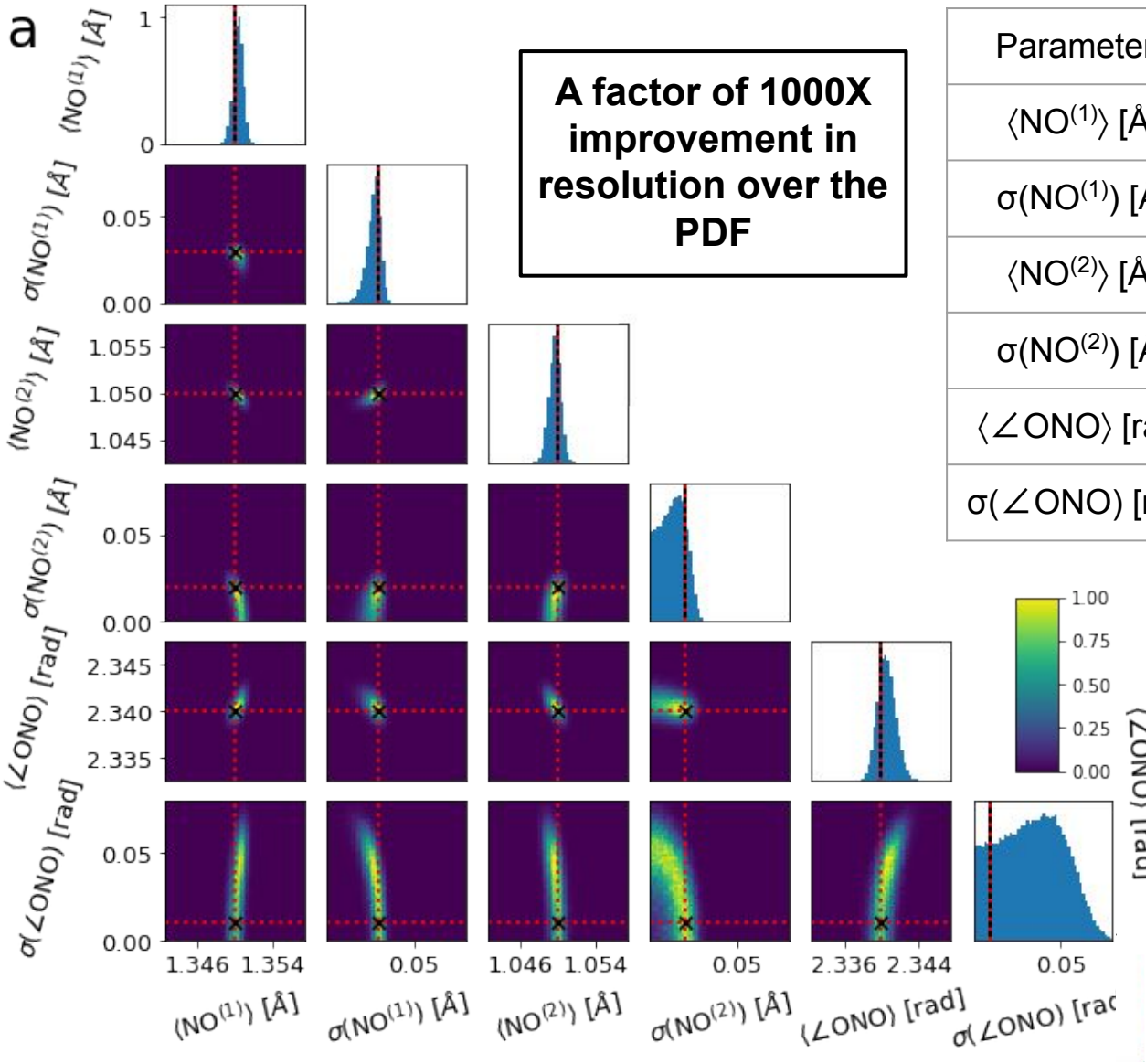
Parameters	Input	Retrieved	Resolution
⟨NO(1)⟩ [Å]	1.35	1.35	0.00029
σ(NO(1)) [Å]	0.03	0.03	0.0019
⟨NO(2)⟩ [Å]	1.05	1.05	0.00029
σ(NO(2)) [Å]	0.02	0.02	0.0054
⟨∠ONO⟩ [rad]	2.34	2.34	0.00047
σ(∠ONO) [rad]	0.01	0.0101	0.015



Retrieving the MF Geometry Probability Distribution Results: P(θ|C)

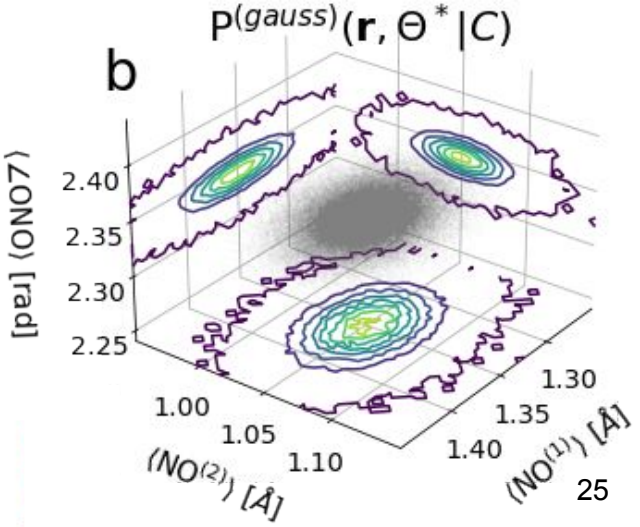
PULSE

SLAC

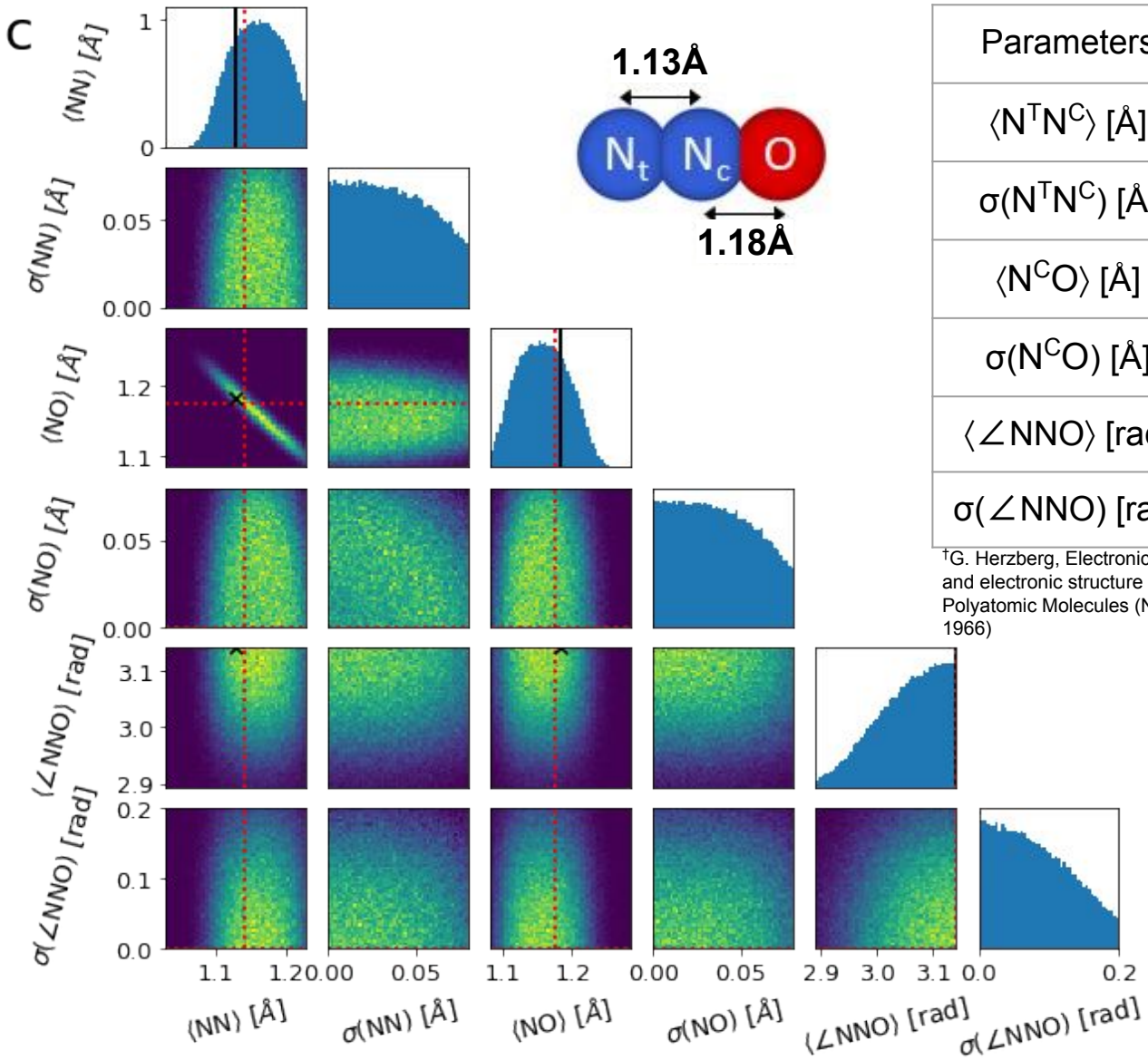


A factor of 1000X improvement in resolution over the PDF

Parameters	Input	Retrieved	Resolution
$\langle NO^{(1)} \rangle$ [Å]	1.35	1.35	0.00029
$\sigma(NO^{(1)})$ [Å]	0.03	0.03	0.0019
$\langle NO^{(2)} \rangle$ [Å]	1.05	1.05	0.00029
$\sigma(NO^{(2)})$ [Å]	0.02	0.02	0.0054
$\langle \angle ONO \rangle$ [rad]	2.34	2.34	0.00047
$\sigma(\angle ONO)$ [rad]	0.01	0.0101	0.015

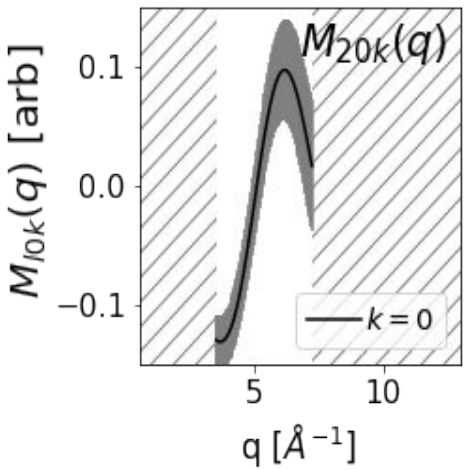


Retrieving the MF Geometry Probability Distribution Results: P(θ|C)

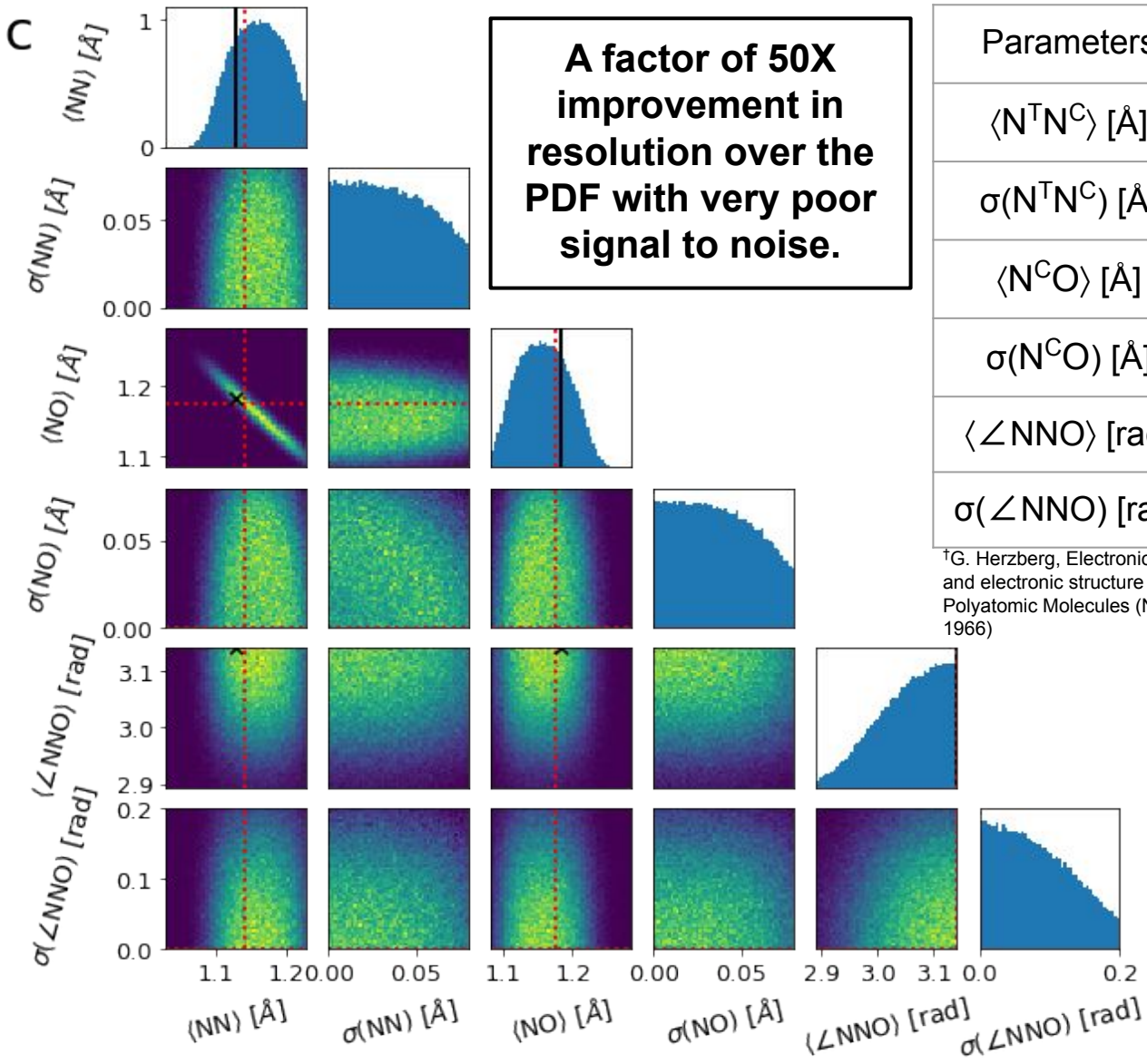


Parameters	Lit.†	Retrieved	Resolution
$\langle N^T N^C \rangle$ [Å]	1.128	1.143	0.039
$\sigma(N^T N^C)$ [Å]		0.081	0.028
$\langle N^C O \rangle$ [Å]	1.184	1.175	0.036
$\sigma(N^C O)$ [Å]		$3.08*10^{-8}$	0.027
$\langle \angle NNO \rangle$ [rad]	3.142	3.142	0.061
$\sigma(\angle NNO)$ [rad]		$5.5*10^{-12}$	0.062

†G. Herzberg, Electronic Spectra and electronic structure of Polyatomic Molecules (Nostrand, 1966)



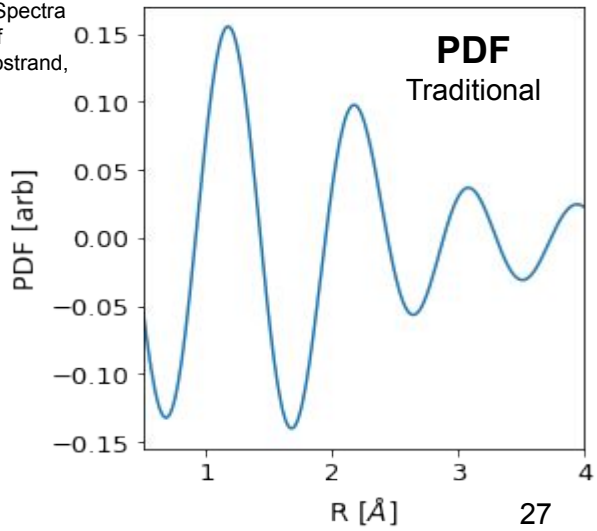
Retrieving the MF Geometry Probability Distribution Results: P(θ|C)



A factor of 50X improvement in resolution over the PDF with very poor signal to noise.

Parameters	Lit.†	Retrieved	Resolution
$\langle N^T N^C \rangle$ [Å]	1.128	1.143	0.039
$\sigma(N^T N^C)$ [Å]		0.081	0.028
$\langle N^C O \rangle$ [Å]	1.184	1.175	0.036
$\sigma(N^C O)$ [Å]		$3.08*10^{-8}$	0.027
$\langle \angle NNO \rangle$ [rad]	3.142	3.142	0.061
$\sigma(\angle NNO)$ [rad]		$5.5*10^{-12}$	0.062

†G. Herzberg, Electronic Spectra and electronic structure of Polyatomic Molecules (Nostrand, 1966)



Retrieving Excited State Dynamics (Isotropic)

$$\langle I(\mathbf{q},t)\rangle = \mathcal{I}\Bigg(\sum_{\mu} |f_\mu(q)|^2 + \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\Bigg\{ f_\mu(q) f_\nu^*(q) \sum_l 4\pi i^l \times \sum_{m,k} (-1)^k \underbrace{Y_l^m\left(\theta_q^{(\mathrm{lf})}, \phi_q^{(\mathrm{lf})}\right)}_{\text{Lab Frame}} \langle \Psi(t)| \underbrace{D_{mk}^l\left(\phi_1^{(\mathrm{lf})}, \theta_1^{(\mathrm{lf})}, \chi_1^{(\mathrm{lf})}\right)}_{\text{Ensemble Anisotropy}} \underbrace{j_l(q\Delta R_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right)}_{\text{Molecular Frame Structure}} |\Psi(t)\rangle \Bigg\}\Bigg)$$

Use only the isotropic component

- Lose explicit dependence on MF angles
  - l=0, m=0, k=0
- Can apply to current datasets
- One may expect similar results
  - Resolution: Order 1 - 10mÅ

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

Retrieving Excited State Dynamics (Anisotropic)

$$\langle I(\mathbf{q})\rangle^{(2)}_{\text{sep}}(t,\tau) \approx \mathcal{I}\Bigg(\sum_{\mu} |f_\mu(q)|^2 + \sum_{\mu,\nu:\mu\neq\nu} \text{Re}\Bigg\{ f_\mu(q) f^*_\nu(q) \sum_l \frac{32\pi^3 i^l}{2l+1} \sum_{m_1,m_2} (-1)^{m_1} Y_l^{m_2}\left(\theta_q^{(\text{lf})}, \phi_q^{(\text{lf})}\right)$$

$$\times \sum_{n,n'} \tilde{\mathcal{A}}^{(2)l}_{m_2 m_1}(n,n';\tau) \left\langle \psi^{n'}_{\text{el-vib}}(t) \middle| j_l(q\Delta R_{\mu\nu}) Y_l^{-m_1}\left(\theta^{(\text{mf})}_{\mu\nu}, \phi^{(\text{mf})}_{\mu\nu}\right) \middle| \psi^{n}_{\text{el-vib}}(t) \right\rangle \Bigg\} \Bigg).$$

- Separation of rotational and vibrational time scales
- Independent atom approximation
- **Measurement:** Lab frame anisotropy
- **Simulation:** Ensemble anisotropy
- **Result:** Molecular frame geometry
  - Explicit dependence on MF geometric angles
  - Each pair-wise contribution is labelled

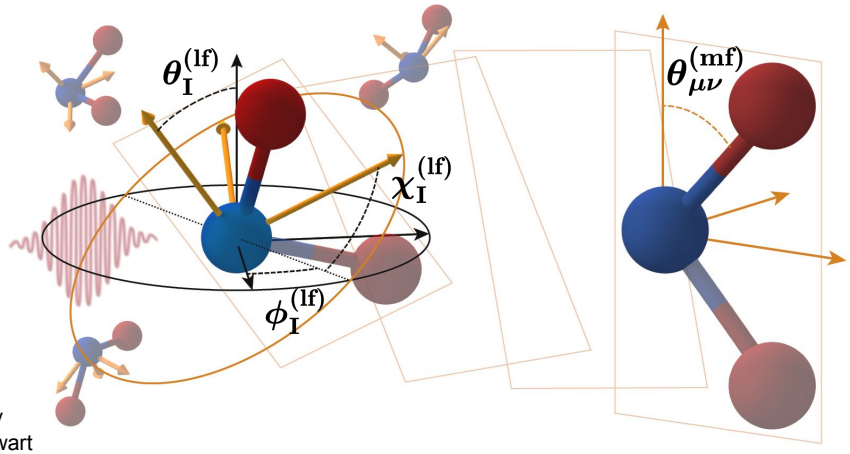


Illustration by Gregory Stewart

Rotational dynamics is of order a few to 10 picoseconds and vibrational or isomerization dynamics of interest are often on the femtosecond timescale. If sufficient rotation occurs outside of the ground state geometry one can use $C_{000}(q,t)$ which is independent of anisotropy but does not have an explicit dependence on the molecular frame angles.

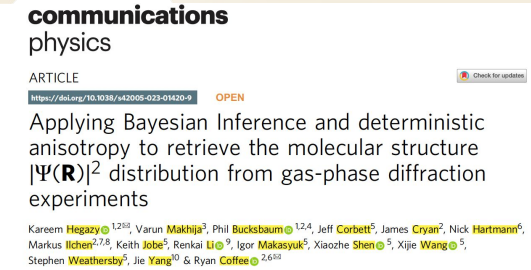
Summary

Khegazy@slac.stanford.edu

PULSE

SLAC

- Directly probe $|\Psi(\mathbf{R}^{(mf)})|^2$ in a high dimensional space
  - Retrieve distributions of geometry parameters in high dimensions
  - **MHA: efficient, unconstrained, and unbiased search**
- Generally applicable to current experiments
  - **Choice of model (Normal, …) allows novel measurements like width**
  - Can use induced anisotropy from excitation dipole and $C_{000}(q,t)$
- High resolution (roughly 100-1000 times better than PDF)
  - **Simulation resolutions distances/angles: Order 1 - 0.1 pm / ~1 mrad**
  - **Data with LMK=[2,0,0] and q=[3.5,7] Å^{-1} distances/angles: ~4 pm / 6 mrad**
  - High signal to noise is more important than q range
- **Potentially turn ultrafast gas phase diffraction into a discovery oriented technique without requiring excited state simulations**
  - This analysis only requires knowledge of the ground state geometry and ensemble anisotropy simulations which are much more tractible than excited state dynamics


communications physics

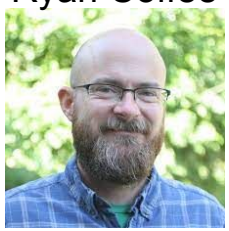
ARTICLE

https://doi.org/10.1038/s42005-023-01420-9 OPEN

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 Makhija[3], Phil Bucksbaum[1,2,4], Jeff Corbett[5], James Cryan[2], Nick Hartmann[6], Markus Ilchen[2,7,8], Keith Jobe[5], Renkai Li[9], Igor Makasyuk[5], Xiaozhe Shen[5], Xijie Wang[5], Stephen Weathersby[5], Jie Yang[10] & Ryan Coffee[2,6✉]

Ryan Coffee



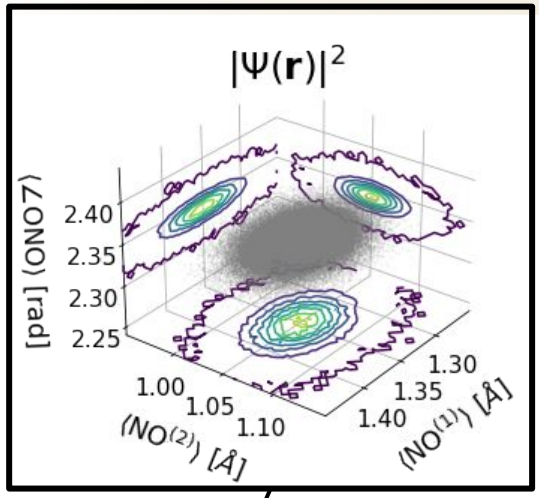
Varun Makhija

SLAC UED Team

U.S. DEPARTMENT OF ENERGY | Office of Science

Backup Slides

Method Overview



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

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

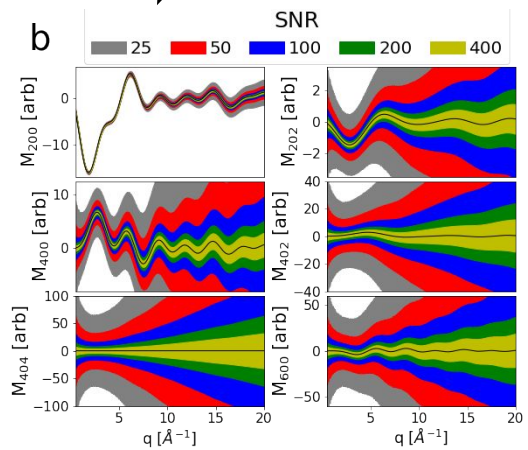
$$\mathbf{\Theta}^{(\delta)} = \left[\langle \mathrm{NO}^{(1)}\rangle, \langle \mathrm{NO}^{(2)}\rangle, \langle \angle \mathrm{ONO}\rangle\right]$$

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

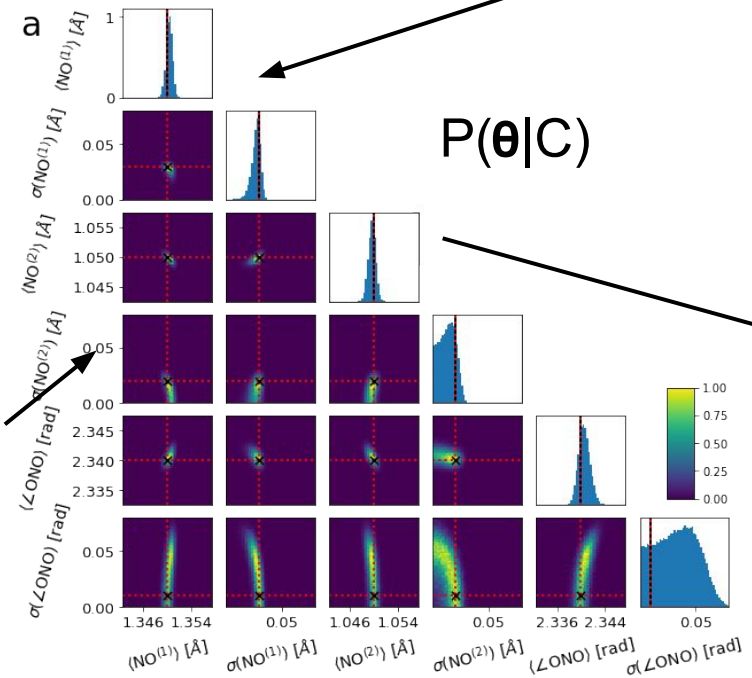
$$\mathbf{\Theta}^{(\mathcal{N})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \sigma\left(\mathrm{NO}^{(1)}\right), \langle \mathrm{NO}^{(2)}\rangle, \sigma\left(\mathrm{NO}^{(2)}\right), \langle \angle \mathrm{ONO}\rangle, \sigma(\angle \mathrm{ONO})\right]$$

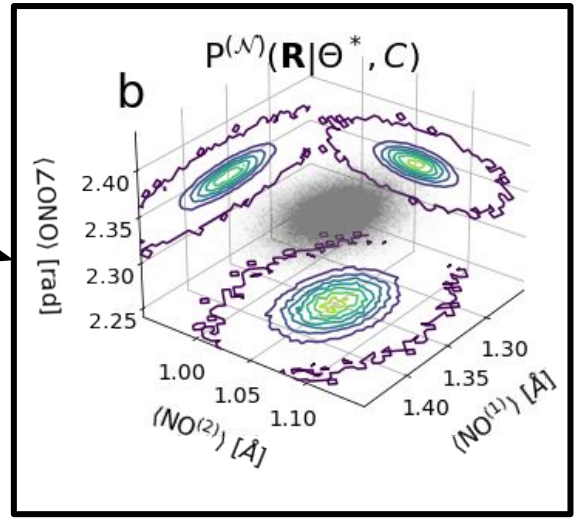
Model of $|\Psi(\mathbf{R}^{(\mathrm{mf})})|^2$

Data



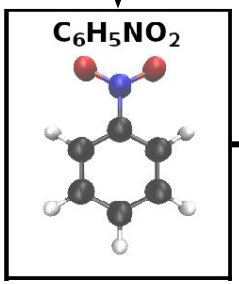
P(θ|C)

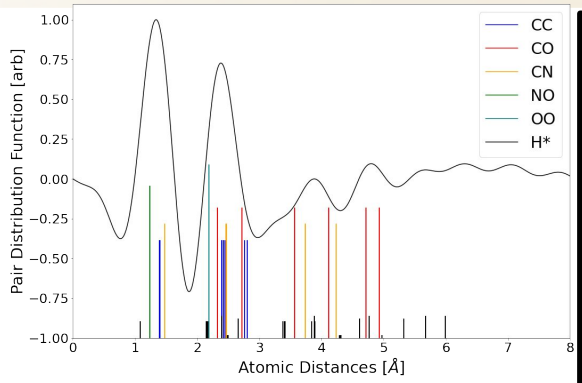


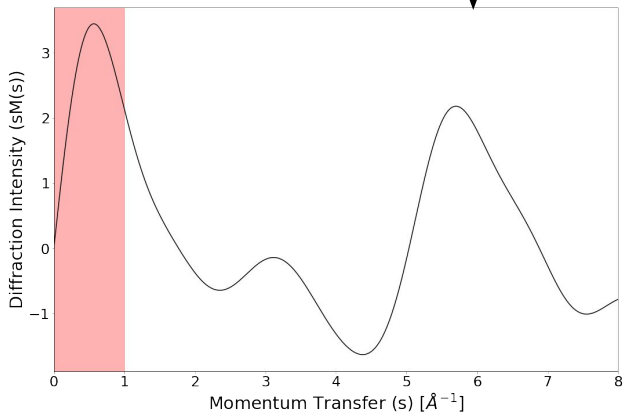


Applying Bayesian Inferencing
Bayesian Inferencing

SLAC





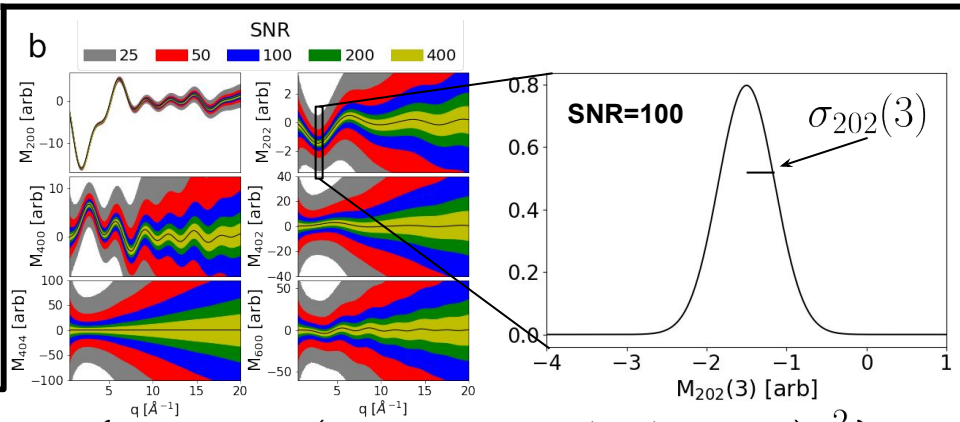


Select **Θ'** nearby **Θ** and calculate expected $C_{lmk}(q)$

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

$$\Theta^{(\text{gauss})} = \left[\langle \text{NO}^{(1)}\rangle, \sigma\left(\text{NO}^{(1)}\right), \langle \text{NO}^{(2)}\rangle, \sigma\left(\text{NO}^{(2)}\right), \langle \angle\text{ONO}\rangle, \sigma\left(\angle\text{ONO}\right)\right]$$

$$C_{lmk}^{(\text{calc})}(q,\Theta) = \mathcal{I}\sum_{\mu,\nu,\mu\neq\nu}\text{Re}\left\{f_\mu(q)f_\nu^*(q)(-1)^{k-m}i^l 8\pi^2 \times\sqrt{\frac{4\pi}{(2l+1)}}\underbrace{\int j_l(q\Delta r_{\mu\nu})Y_l^{-k}\left(\theta_{\mu\nu}^{(\text{mf})},\phi_{\mu\nu}^{(\text{mf})}\right)P(\mathbf{r},\Theta|C)d\mathbf{r}}_{\text{Molecular Frame Geometry}}\right\}$$



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

Bayesian Inferencing
Reframing the Problem

SLAC

What we have: Molecular frame representations for the nuclear geometry with explicit dependence on pair-wise distances and angles

What we want: Invert the integral equation for $|\Psi(\mathbf{R}^{(mf)})|^2$

How we do it: Approximate $|\Psi(\mathbf{R}^{(mf)})|^2$ with a chosen distribution and solve for the model pa… distribution P(**θ**|…

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

Bayesian Inference

$$C_{lmk}^{(\mathrm{calc})}(q) = \mathcal{I} \sum \mathrm{Re}\left\{ f_\mu(q) f_\nu^*(q) (-1)^k \frac{32\pi^3 i^l}{2l+1} \left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right) P(\mathbf{R}|\boldsymbol{\Theta}, C) d\mathbf{R} \right\}$$

r Frame Structure

Curse of Dimensionality
- Evaluate ~20 Equations
- Order 100 terms
- Embedded in a k*(3N-6)

When finding **θ** one can make novel measurements

Delta Distributio…

Normal Distribution: $P^{(\mathcal{N})}(\mathbf{R}|\,\boldsymbol{\Theta}, C) = \frac{1}{\sqrt{2\pi}^{N_{\mathrm{dof}}} \prod_{i=0}^{i<N_{\mathrm{dof}}} \boldsymbol{\Theta}_{2i+1}^{(\mathcal{N})}} \exp\left\{ \frac{-1}{2} \sum_{i=0}^{i<N_{\mathrm{dof}}} \left(\frac{\boldsymbol{\Theta}_{2i}^{(\mathcal{N})} - \mathbf{R}_i}{\boldsymbol{\Theta}_{2i+1}^{(\mathcal{N})}} \right)^2 \right\}$

$\boldsymbol{\Theta}^{(\mathcal{N})} = \left[\langle \mathrm{NO}^{(1)} \rangle, \sigma\left(\mathrm{NO}^{(1)}\right), \langle \mathrm{NO}^{(2)} \rangle, \sigma\left(\mathrm{NO}^{(2)}\right), \langle \angle \mathrm{ONO} \rangle, \sigma(\angle \mathrm{ONO}) \right]$

Accessing the MF via Deterministic Anisotropy

SLAC

$$\langle I(\mathbf{q})\rangle(t) = \mathcal{I}\Bigg(\sum_{\mu} |f_{\mu}^{*}(q)|^2 + \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\Bigg\{ f_{\mu}(q) f_{\nu}^{*}(q) \sum_{l} i^{l} 8\pi^2 \sqrt{4\pi\,(2l+1)}$$

$$\times \sum_{m,k} (-1)^{k-m} \underbrace{Y_{l}^{-m}\left(\theta_{q}^{(\mathrm{lf})}, \phi_{q}^{(\mathrm{lf})}\right)}_{\text{Lab Frame}} \langle \Psi(t)| \underbrace{D_{mk}^{l}\left(\phi_{\mathrm{I}}^{(\mathrm{lf})}, \theta_{\mathrm{I}}^{(\mathrm{lf})}, \chi_{\mathrm{I}}^{(\mathrm{lf})}\right)}_{\text{Ensemble Anisotropy}} \underbrace{j_{l}(q\Delta r_{\mu\nu}) Y_{l}^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right)}_{\text{Molecular Frame Geometry}} |\Psi(t)\rangle \Bigg\} \Bigg)$$

Independent atom approximation

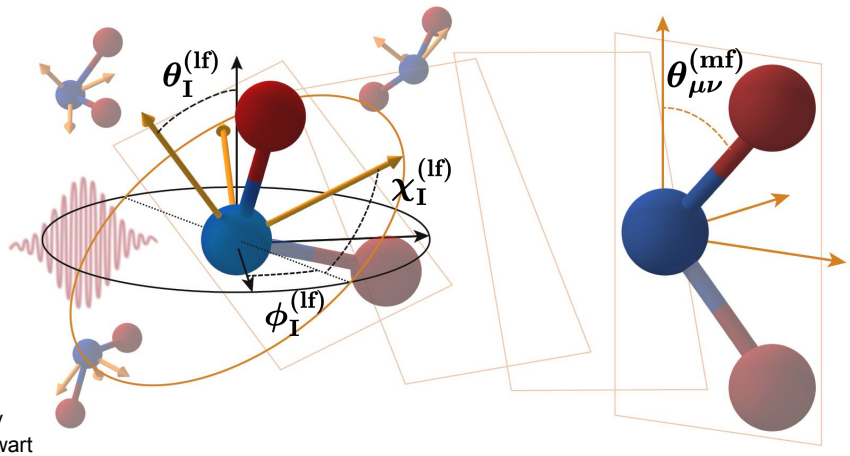
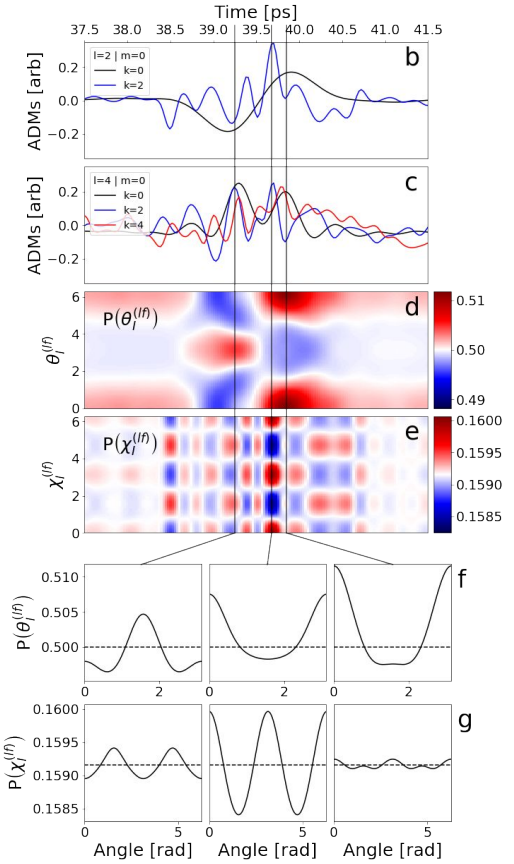


Illustration by
Gregory Stewart



Anisotropy Reveals the Molecular Frame (MF)

- Anisotropy provides constraints on molecular frame (MF) degrees of freedom.
- Combining many measurements (constraints) allows one to retrieve the MF.
- Application in photo-electron spectroscopy
  - V. Makhija, et. al., (2016), arXiv:1611.06476 [physics.atom-ph]
  - C. Marceau, et. al., Phys. Rev. Lett. 119, 083401 (2017)
  - M. Gregory, et. al., (2020), arXiv:2012.04561 [physics.chem-ph]
- Use a stretched NO_2: an asymmetric top

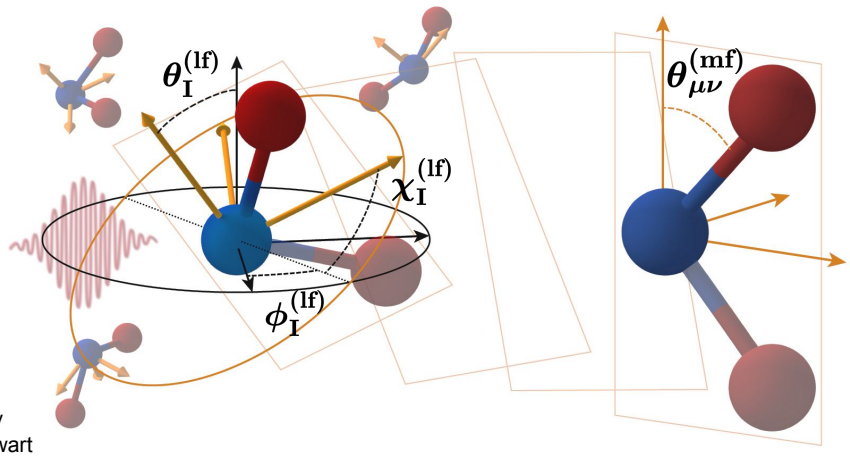
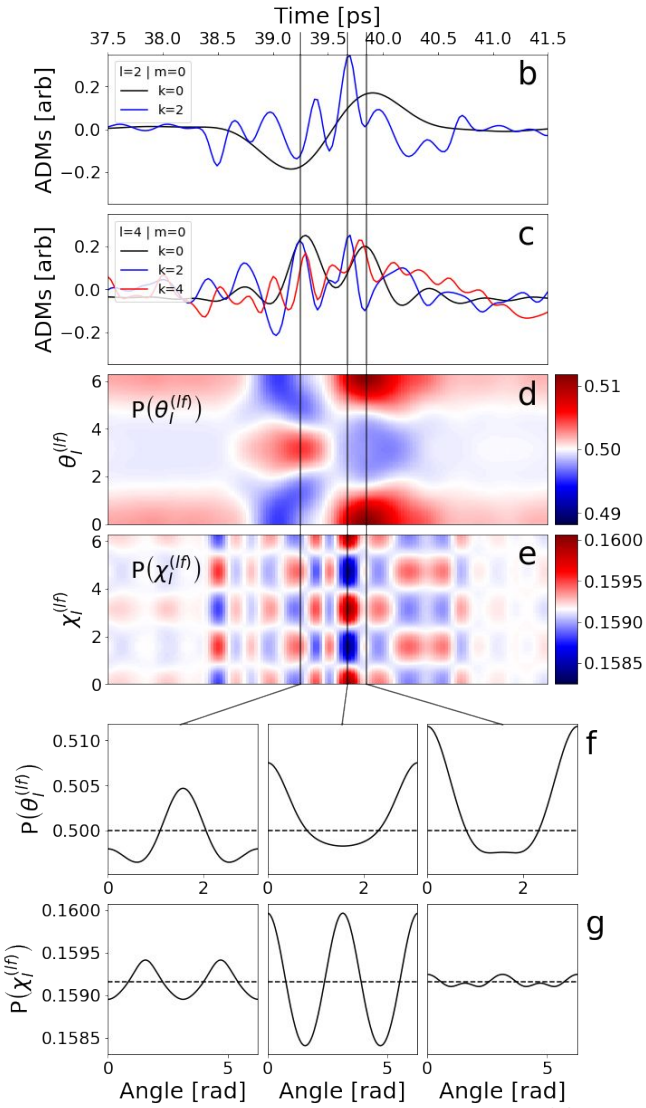


Illustration by Gregory Stewart



NO_2 rotational wavepacket anisotropy signature

Accessing the MF via Deterministic Anisotropy

SLAC

$$\langle I(\mathbf{q})\rangle(t) = \mathcal{I}\Bigg(\sum_{\mu} |f_\mu^*(q)|^2 + \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\Bigg\{ f_\mu(q) f_\nu^*(q) \sum_l i^l 8\pi^2 \sqrt{4\pi\,(2l+1)} \times \sum_{m,k} (-1)^{k-m} \underbrace{Y_l^{-m}\left(\theta_q^{(\mathrm{lf})}, \phi_q^{(\mathrm{lf})}\right)}_{\text{Lab Frame}} \langle \Psi(t) | \underbrace{D_{mk}^l\left(\phi_{\mathrm{I}}^{(\mathrm{lf})}, \theta_{\mathrm{I}}^{(\mathrm{lf})}, \chi_{\mathrm{I}}^{(\mathrm{lf})}\right)}_{\text{Ensemble Anisotropy}} \underbrace{j_l(q\Delta r_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right)}_{\text{Molecular Frame Geometry}} | \Psi(t) \rangle \Bigg\} \Bigg)$$

- Independent atom approximation
- **Measurement:** Lab frame anisotropy
- **Simulation:** Ensemble anisotropy
- **Result:** Molecular frame geometry
  - Explicit dependence on MF geometric angles
  - Each pair-wise contribution is labelled

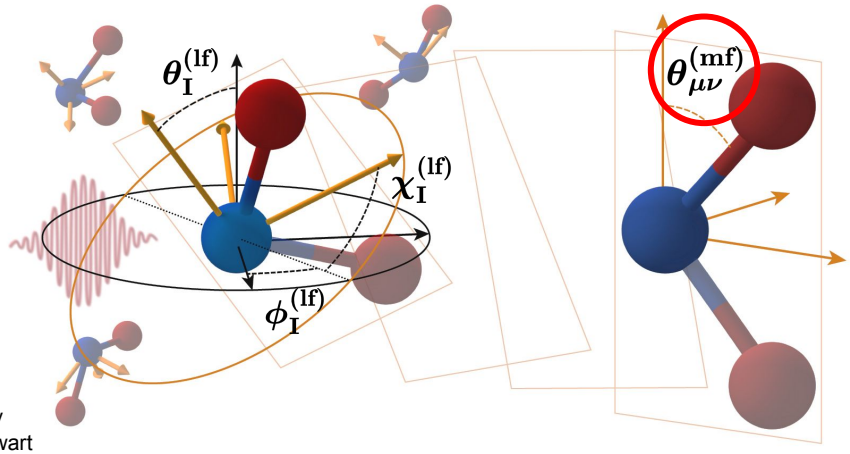


Illustration by Gregory Stewart

Explicit dependence on molecular frame angles provides an opportunity to develop methods that extract these angles, and hence a unique geometry, from data alone.

Retrieving the MF Geometry Probability Distribution
Delta Distribution and Systematic Error

SLAC

Delta Distribution Systematic Error

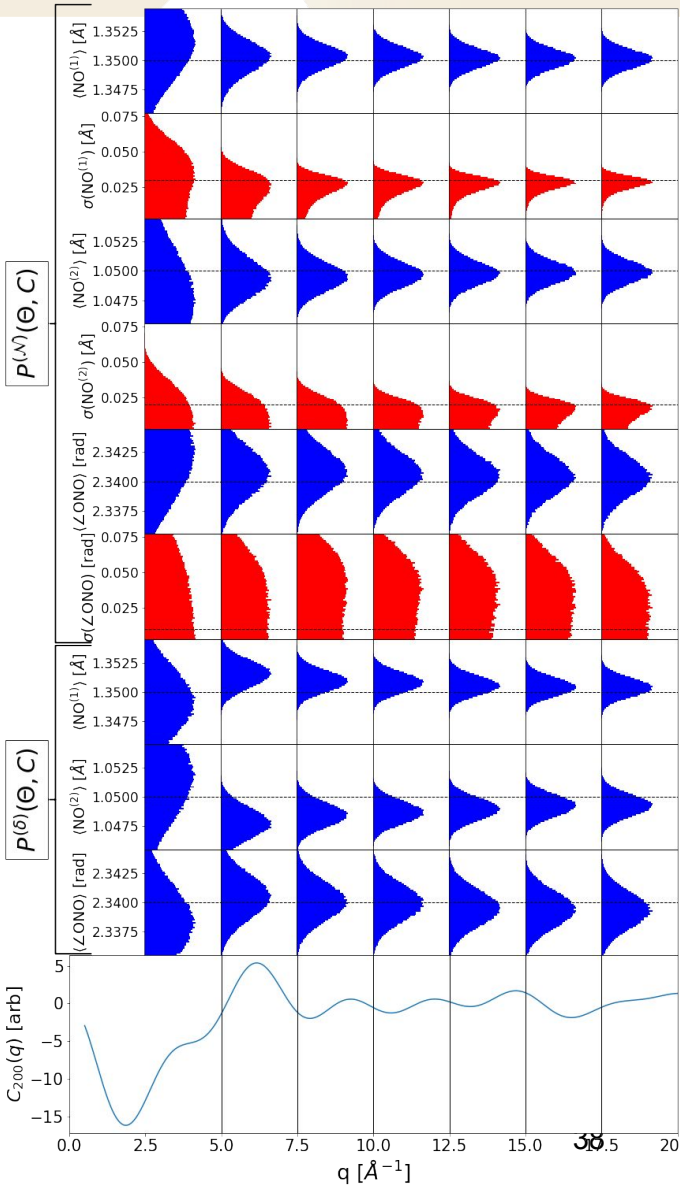
- Assumes signal from ensembles and a single geometry are comparable
- Accuracy is roughly 1000 times worse
- At ≤picometer resolution, the correct value can be 2-3 standard deviations of P(**θ**|C)

Normal Distribution Mitigates this Systematic

- Normal distribution P(**θ**|C) distributions are closely centered around the expected value
- The 1d mode does not change with q range

Why use the delta distribution?

- It is of order 100 times faster
- Provides sufficient accuracy for debugging analysis



Convergence: Autocorrelation Time (τ)

SLAC

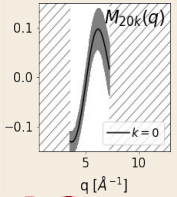
Criteria for each chain (1000 chains are used)

- At least 100 τ in length
- $\Delta\tau/\tau$ < 0.01

Autocorrelation Time

- Number of steps needed for geometries to be uncorrelated
- Expectations: $E_{p(\theta)}\left[f(\theta)\right] \approx \frac{1}{N}\sum_{n}^{N} f\left(\theta^{(n)}\right)$ $\qquad \mathrm{SEM} = \sqrt{\frac{\mathrm{Var}_{p(\theta)}[f(\theta)]}{N}}$
- Correlation Effects: $\mathrm{SEM} = \sqrt{\frac{\tau}{N}\mathrm{Var}_{p(\theta)}[f(\theta)]}$

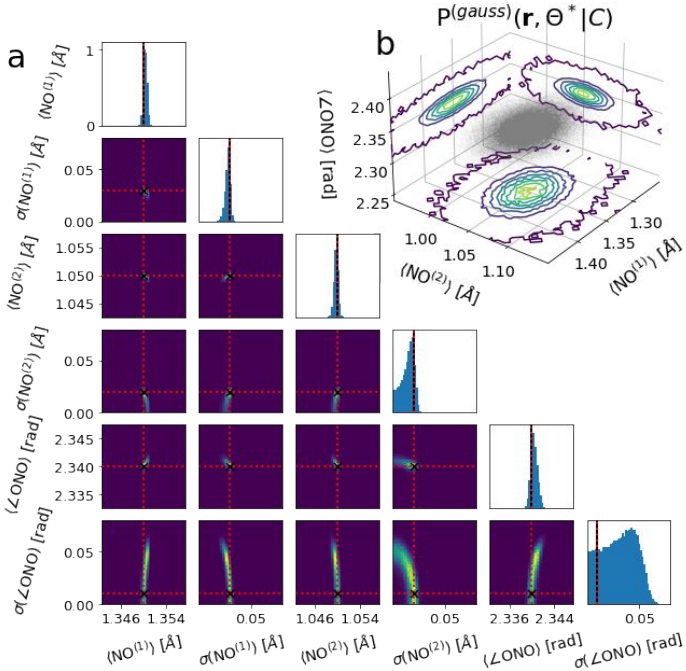
Retrieving the MF Geometry Probability Distribution Results



SLAC

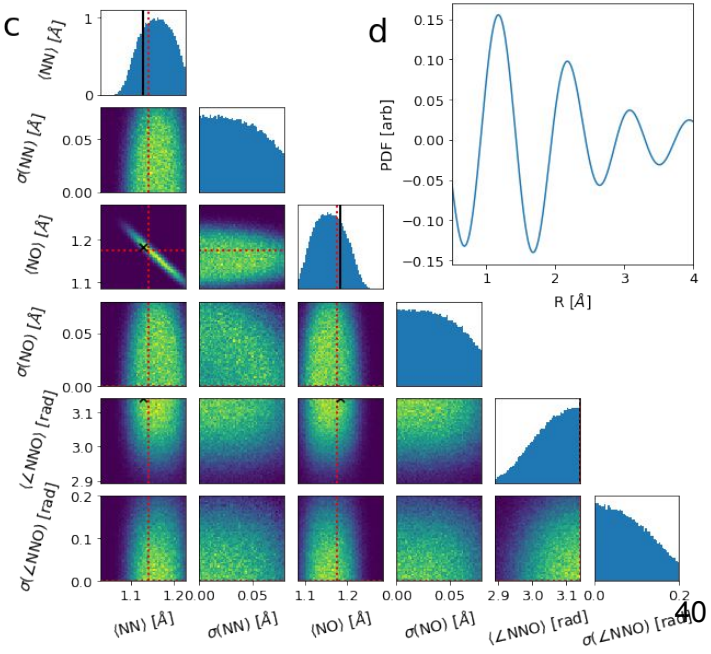
Simulated NO_2

	Input	Θ^*	σ^Θ
$\langle NO^{(1)} \rangle$ [Å]	1.35	1.3500	0.00029
$\sigma\left(NO^{(1)}\right)$ [Å]	0.03	0.03000	0.0019
$\langle NO^{(2)} \rangle$ [Å]	1.05	1.0500	0.00029
$\sigma\left(NO^{(2)}\right)$ [Å]	0.02	0.02000	0.0054
$\langle \angle ONO \rangle$ [rad]	2.34	2.340	0.00047
$\sigma\left(\angle ONO\right)$ [rad]	0.01	0.01010	0.015



Measured N_2O

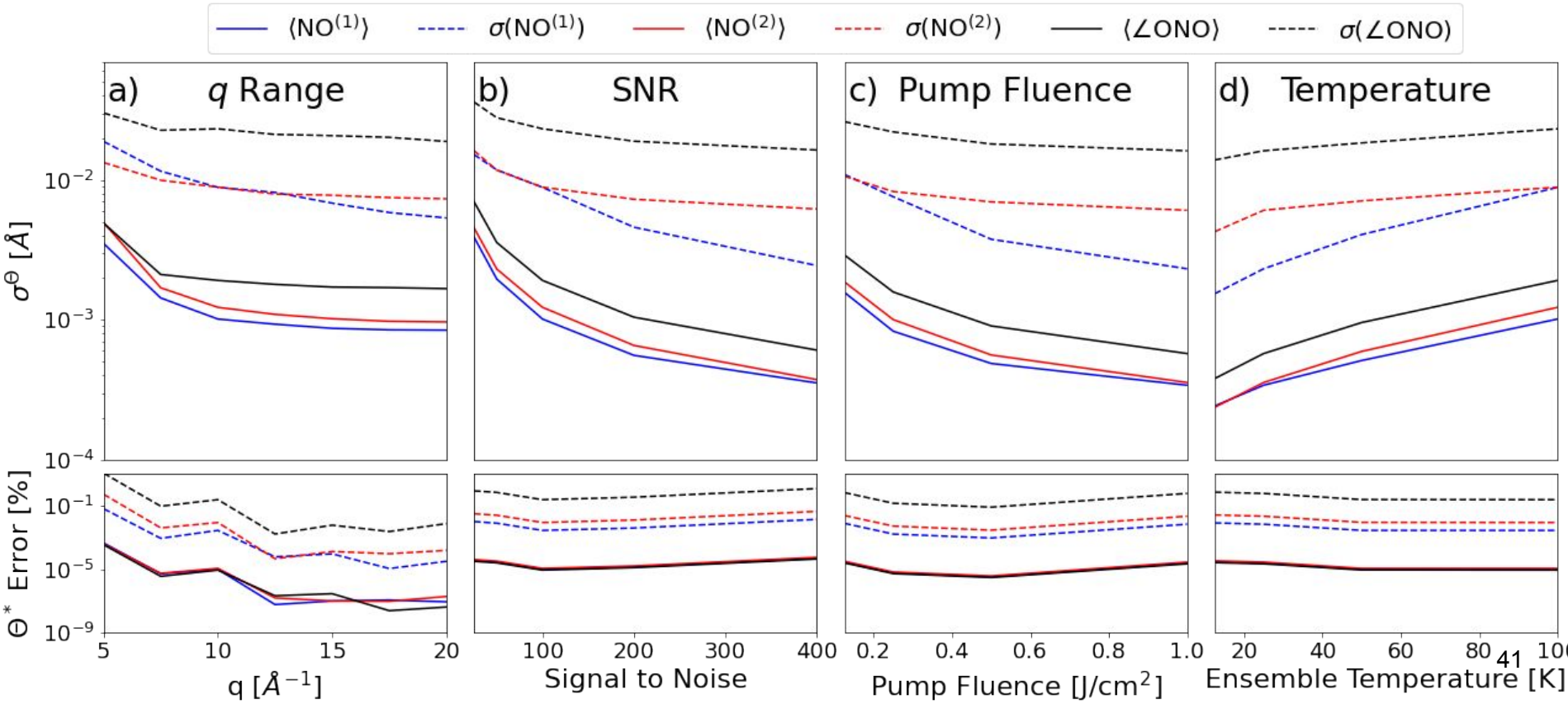
	$\Theta^*_{\text{Literature}}$	Θ^*	σ^Θ
$\langle N^T N^C \rangle$ [Å]	1.128	1.142	0.039
σ (NN) [Å]		0.081	0.028
$\langle NO \rangle$ [Å]	1.184	1.175	0.036
σ (NO) [Å]		3.08×10^{-8}	0.027
$\langle \angle NNO \rangle$ [rad]	3.142	3.142	0.061
σ ($\angle$NNO) [rad]		5.5×10^{-12}	0.062



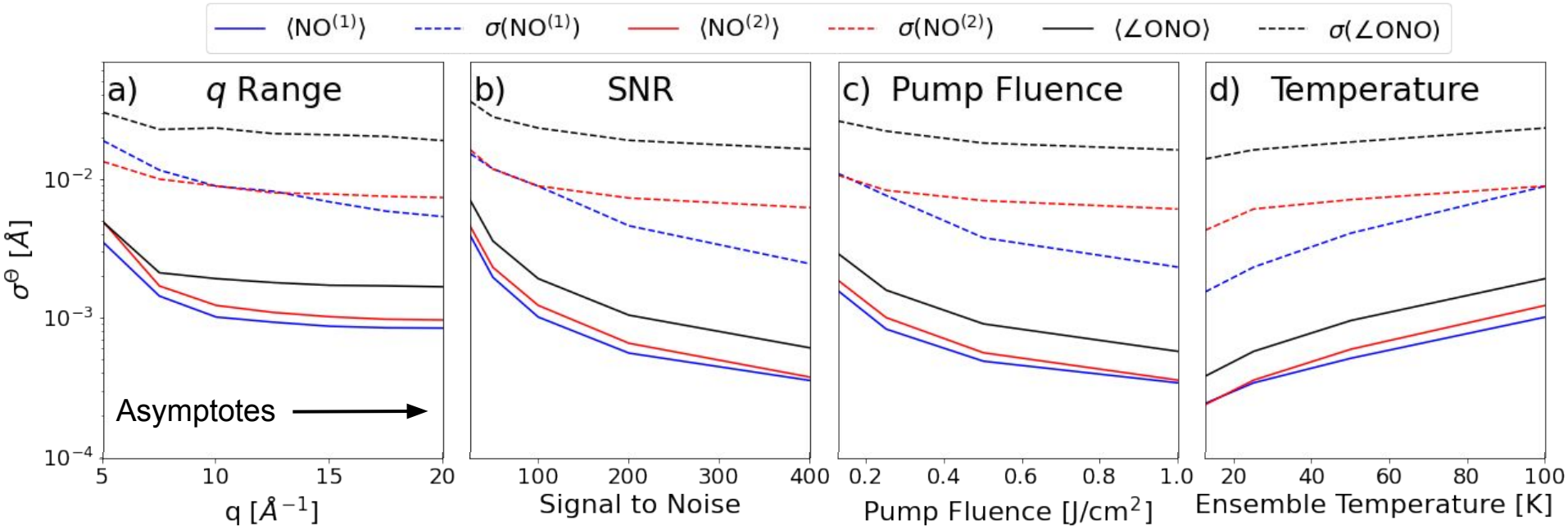
Retrieving the MF Geometry Probability Distribution
Experimental Parameters

SLAC

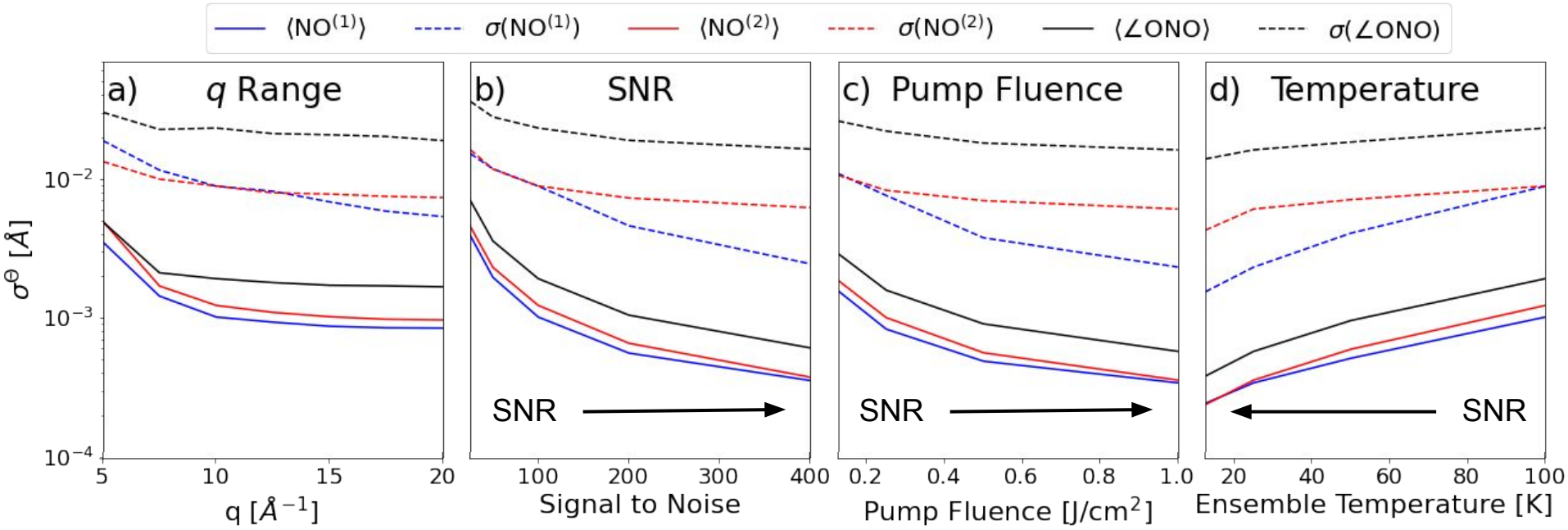
a) q Range	b) SNR	c) Pump Fluence	d) Temperature
q : 0.5 - [5, 20] Å^{-1} SNR: 100 Fluence: 1 J/cm^2 Temp: 100 K	q : 0.5 - 10 Å^{-1} SNR: [25, 400] Fluence: 1 J/cm^2 Temp: 100 K	q : 0.5 - 10 Å^{-1} SNR: 100 Fluence: [0.12, 1] J/cm^2 Temp: 100 K	q : 0.5 - 10 Å^{-1} SNR: 100 Fluence: 1 J/cm^2 Temp: [12.5, 100] K



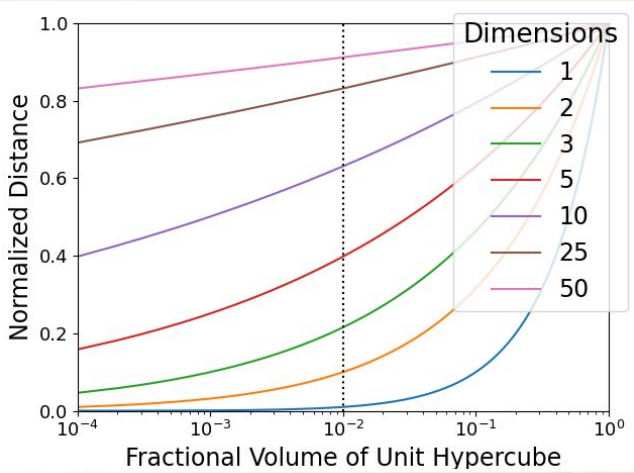
Retrieving the MF Geometry Probability Distribution
Experimental Parameters



Retrieving the MF Geometry Probability Distribution
Experimental Parameters



Curse of Dimensionality

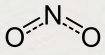


Assumptions

1. Grid Search
2. Ignore hydrogens
3. Know atom pair-wise distances within 1 Å

Degrees of freedom: $N_{dof} = 3N_{atoms} - 6$
Number of samples: N_S
Number of sample per dimension: N_S/D
Sampled volume given SS step size: $V_S|SS$

NO_2



N_{dof}= 3

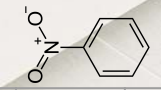
N_S	N_S/D
10^6	100
10^7	215
10^8	464

Cyclohexadiene

N_{dof}= 12

N_S	N_S/D	$V_S\|0.1$	$V_S\|0.25$
10^6	3.16	10^{-5}%	5.96%
10^7	3.83	10^{-4}%	59.6%
10^8	4.64	10^{-3}%	>100%

Nitrobenzene



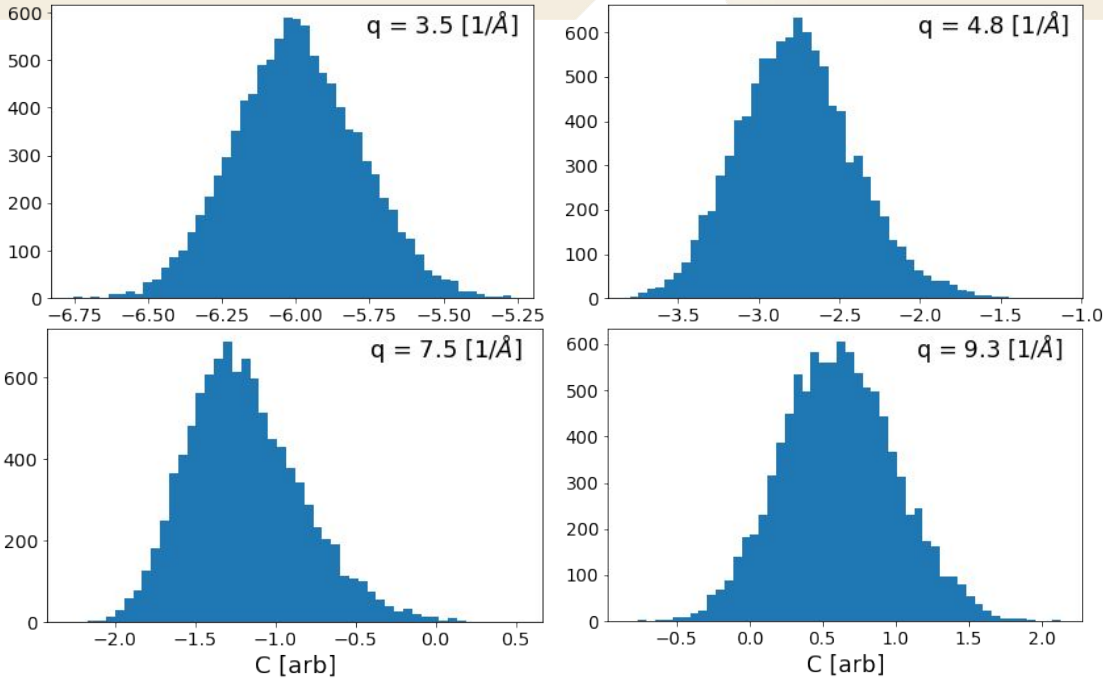
N_{dof}= 21

N_S	N_S/D	$V_S\|0.1$	$V_S\|0.25$
10^6	1.93	10^{-14}%	2.3×10^{-5}%
10^7	2.15	10^{-13}%	2.3×10^{-4}%
10^8	2.40	10^{-12}%	2.3×10^{-3}%

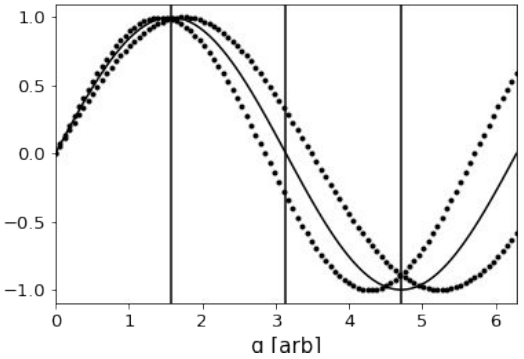
Systematic Errors in Retrieving Geometric Parameters

SLAC

- Systematic errors are caused by non-gaussian $C_{lmk}(q)$ distributions
- Gaussian distributions of geometries lead to non-gaussian distributions due to sinusoids.



$$P(C|\mathbf{r}^{(mf)}) = \prod_{lmk,q} \frac{1}{\sigma_{lmk}(q)\sqrt{2\pi}} \exp\left\{\frac{-1}{2}\left(\frac{C_{lmk}^{(data)}(q) - C_{lmk}^{(calc)}\left(q, \mathbf{r}^{(mf)}\right)}{\sigma_{lmk}(q)}\right)^2\right\}$$



Bayesian Inferencing
Reframing the Problem

SLAC

What we have: Molecular frame representations for the nuclear geometry with explicit dependence on pair-wise distances and angles

$$C_{lmk}(q) = \mathcal{I} \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\left\{ f_\mu(q) f_\nu^*(q) (-1)^{k-m} i^l 8\pi^2 \times \sqrt{\frac{4\pi}{(2l+1)}} \underbrace{\left\langle j_l(q\Delta r_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right)\right\rangle}_{\text{Molecular Frame Geometry}} \right\}$$

$$\longrightarrow \quad C_{lmk}(q) = \mathcal{I} \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\left\{ f_\mu(q) f_\nu^*(q) (-1)^{k-m} i^l 8\pi^2 \times \sqrt{\frac{4\pi}{(2l+1)}} \underbrace{\int j_l(q\Delta r_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right) \left|\Psi(\mathbf{r}^{(\mathrm{mf})})\right|^2 d\mathbf{r}^{(\mathrm{mf})}}_{\text{Molecular Frame Geometry}} \right\}$$

What we want: Invert the integral equation for $|\Psi(\boldsymbol{r}^{(\mathrm{mf})})|^2$

How we do it: Approximate $|\Psi(\boldsymbol{r}^{(\mathrm{mf})})|^2$ with a chosen distribution and solve for the model parameter ($\boldsymbol{\theta}$) distribution P($\boldsymbol{\theta}$|C)

$$P(\mathbf{r}, \boldsymbol{\Theta}|\, C) \approx |\Psi(\mathbf{r})|^2$$

$$P(\boldsymbol{\Theta}|C) = \int P(\mathbf{r}, \boldsymbol{\Theta}|\, C)\, d\mathbf{r}$$

Delta Distribution:

$$P^{(\delta)}(\mathbf{r}, \boldsymbol{\Theta}|\, C) = \delta\left(\boldsymbol{\Theta}^{(\mathrm{delta})} - \mathbf{r}\right)$$

$$\boldsymbol{\Theta}^{(\mathrm{delta})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \langle \mathrm{NO}^{(2)}\rangle, \langle \angle \mathrm{ONO}\rangle\right]$$

Normal Distribution:

$$P^{(\mathcal{N})}(\mathbf{r}, \boldsymbol{\Theta}|\, C) = \frac{1}{\sqrt{2\pi}^{N_{dof}} \prod_{i=0}^{i<N_{dof}} \Theta_{2i+1}^{(\mathrm{gauss})}} \exp\left\{ \frac{-1}{2} \sum_{i=0}^{i<N_{dof}} \left(\frac{\Theta_{2i}^{(\mathrm{gauss})} - \mathbf{r}_i}{\Theta_{2i+1}^{(\mathrm{gauss})}} \right)^2 \right\}$$

$$\boldsymbol{\Theta}^{(\mathrm{gauss})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \sigma\left(\mathrm{NO}^{(1)}\right), \langle \mathrm{NO}^{(2)}\rangle, \sigma\left(\mathrm{NO}^{(2)}\right), \langle \angle \mathrm{ONO}\rangle, \sigma(\angle \mathrm{ONO})\right]$$

Bayesian Inferencing
Reframing the Problem

SLAC

What we have: Molecular frame representations for the nuclear geometry with explicit dependence on pair-wise distances and angles

$$C_{lmk}(q) = \mathcal{I} \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\left\{ f_\mu(q) f_\nu^*(q) (-1)^{k-m} i^l 8\pi^2 \times \sqrt{\frac{4\pi}{(2l+1)}} \underbrace{\left\langle j_l(q\Delta r_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right)\right\rangle}_{\text{Molecular Frame Geometry}} \right\}$$

$$\rightarrow \quad C_{lmk}(q) = \mathcal{I} \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\left\{ f_\mu(q) f_\nu^*(q) (-1)^{k-m} i^l 8\pi^2 \times \sqrt{\frac{4\pi}{(2l+1)}} \underbrace{\int j_l(q\Delta r_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right) \left|\Psi(\mathbf{r}^{(\mathrm{mf})})\right|^2 d\mathbf{r}^{(\mathrm{mf})}}_{\text{Molecular Frame Geometry}} \right\}$$

What we want: Invert the integral equation for $|\Psi(\boldsymbol{r}^{(\mathrm{mf})})|^2$

How we do it: Approximate $|\Psi(\boldsymbol{r}^{(\mathrm{mf})})|^2$ with a chosen distribution and solve for the model parameter ($\boldsymbol{\theta}$) distribution P($\boldsymbol{\theta}$|C)

$$P(\mathbf{r}, \boldsymbol{\Theta}|\, C) \approx |\Psi(\mathbf{r})|^2$$

$$P(\boldsymbol{\Theta}|C) = \int P(\mathbf{r}, \boldsymbol{\Theta}|\, C)\, d\mathbf{r}$$

When finding $\boldsymbol{\theta}$ one can make novel measurements

Delta Distribution:

$$P^{(\delta)}(\mathbf{r}, \boldsymbol{\Theta}|\, C) = \delta\left(\boldsymbol{\Theta}^{(\mathrm{delta})} - \mathbf{r}\right)$$

$$\boldsymbol{\Theta}^{(\mathrm{delta})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \langle \mathrm{NO}^{(2)}\rangle, \langle \angle \mathrm{ONO}\rangle\right]$$

Normal Distribution:

$$P^{(\mathcal{N})}(\mathbf{r}, \boldsymbol{\Theta}|\, C) = \frac{1}{\sqrt{2\pi}^{N_{dof}} \prod_{i=0}^{i<N_{dof}} \Theta_{2i+1}^{(\mathrm{gauss})}} \exp\left\{ \frac{-1}{2} \sum_{i=0}^{i<N_{dof}} \left(\frac{\Theta_{2i}^{(\mathrm{gauss})} - \mathbf{r}_i}{\Theta_{2i+1}^{(\mathrm{gauss})}} \right)^2 \right\}$$

$$\boldsymbol{\Theta}^{(\mathrm{gauss})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \sigma\left(\mathrm{NO}^{(1)}\right), \langle \mathrm{NO}^{(2)}\rangle, \sigma\left(\mathrm{NO}^{(2)}\right), \langle \angle \mathrm{ONO}\rangle, \sigma(\angle \mathrm{ONO})\right]$$

Bayesian Inferencing
Reframing the Problem

SLAC

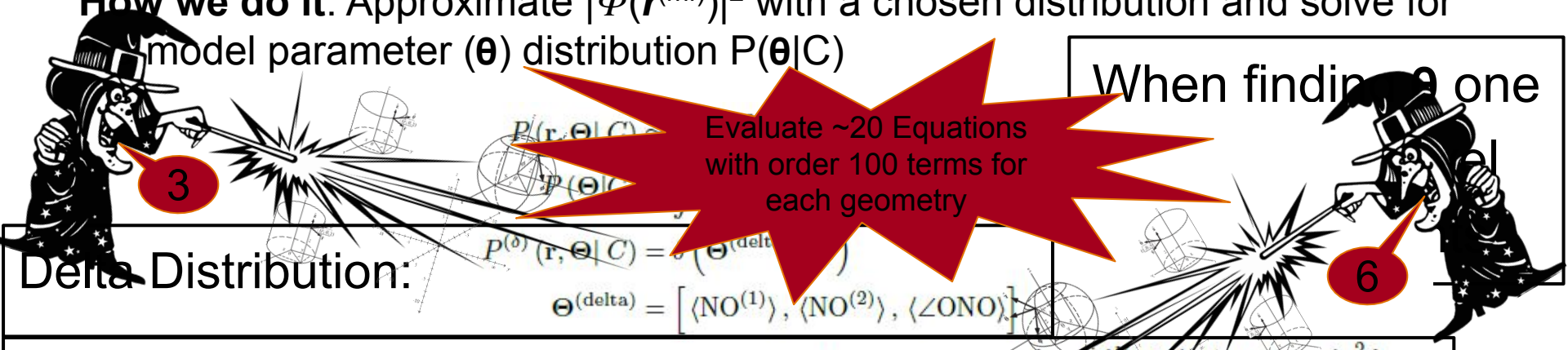
What we have: Molecular frame representations for the nuclear geometry with explicit dependence on pair-wise distances and angles

$$C_{lmk}(q) = \mathcal{I} \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\left\{ f_\mu(q) f_\nu^*(q) (-1)^{k-m} i^l 8\pi^2 \times \sqrt{\frac{4\pi}{(2l+1)}} \underbrace{\left\langle j_l(q\Delta r_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right)\right\rangle}_{\text{Molecular Frame Geometry}} \right\}$$

$$\longrightarrow \quad C_{lmk}(q) = \mathcal{I} \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\left\{ f_\mu(q) f_\nu^*(q) (-1)^{k-m} i^l 8\pi^2 \times \sqrt{\frac{4\pi}{(2l+1)}} \underbrace{\int j_l(q\Delta r_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right) |\Psi(\mathbf{r}^{(\mathrm{mf})})|^2 d\mathbf{r}^{(\mathrm{mf})}}_{\text{Molecular Frame Geometry}} \right\}$$

What we want: Invert the integral equation for $|\Psi(\boldsymbol{r}^{(\mathrm{mf})})|^2$

How we do it: Approximate $|\Psi(\boldsymbol{r}^{(\mathrm{mf})})|^2$ with a chosen distribution and solve for model parameter ($\boldsymbol{\theta}$) distribution P($\boldsymbol{\theta}$|C)



Delta Distribution:

$$P^{(\delta)}(\mathbf{r}, \Theta|\, C) = \delta\left(\Theta^{(\mathrm{delta})}\right)$$

$$\Theta^{(\mathrm{delta})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \langle \mathrm{NO}^{(2)}\rangle, \langle \angle \mathrm{ONO}\rangle\right]$$

Normal Distribution:

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

$$\Theta^{(\mathrm{gauss})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \sigma\left(\mathrm{NO}^{(1)}\right), \langle \mathrm{NO}^{(2)}\rangle, \sigma\left(\mathrm{NO}^{(2)}\right), \langle \angle \mathrm{ONO}\rangle, \sigma\left(\angle \mathrm{ONO}\right)\right]$$

Bayesian Inferencing
Reframing the Problem

SLAC

What we have: Molecular frame representations for the nuclear geometry with explicit dependence on pair-wise distances and angles

$$C_{lmk}(q) = \mathcal{I} \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\left\{ f_\mu(q) f_\nu^*(q) (-1)^{k-m} i^l 8\pi^2 \times \sqrt{\frac{4\pi}{(2l+1)}} \int j_l(q\Delta r_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right) \right\}$$

Metropolis Hastings Algorithm

Wh **nvert the integral equation for** $|\Psi(\boldsymbol{r}^{(\mathrm{mf})})|^2$

H **we d** **o it**: Ap**roximate** $|\Psi(\boldsymbol{r}^{(\mathrm{mf})})|^2$ with a chosen distributio**n** **mode** **paran** **ter** ($\boldsymbol{\theta}$) distribution P($\boldsymbol{\theta}$|C)

Metropolis Hastings Algorithm

Evaluate ~20 Equations with order 100 terms for each geometry

3

6

Wh**n** find**i**** **9 one **el**

Delta Distribution:

$$P^{(\delta)}(\mathbf{r}, \Theta|\, C) = \delta\left(\Theta^{(\mathrm{delta})}\right)$$

$$\Theta^{(\mathrm{delta})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \langle \mathrm{NO}^{(2)}\rangle, \langle \angle \mathrm{ONO}\rangle\right]$$

Normal Distribution:

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

$$\Theta^{(\mathrm{gauss})} = \left[\langle \mathrm{NO}^{(1)}\rangle, \sigma\left(\mathrm{NO}^{(1)}\right), \langle \mathrm{NO}^{(2)}\rangle, \sigma\left(\mathrm{NO}^{(2)}\right), \langle \angle \mathrm{ONO}\rangle, \sigma\left(\angle \mathrm{ONO}\right)\right]$$

Bayesian Inferencing
Metropolis Hastings Algorithm

SLAC

- Inverts the system of equations to solve for the joint P(**θ**|C) distribution
- **Unbiased** sampling method designed for high dimensional spaces

$$C_{lmk}(q) = \mathcal{I} \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\Big\{ f_\mu(q) f_\nu^*(q) (-1)^{k-m} i^l 8\pi^2 \times \sqrt{\frac{4\pi}{(2l+1)}} \underbrace{\int j_l(q\Delta r_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right) \boxed{|\Psi(\mathbf{r})|^2} d\mathbf{r}}_{\text{Molecular Frame Geometry}} \Big\}$$

Bayesian Inferencing

$$C_{lmk}^{(\mathrm{calc})}(q,\Theta) = \mathcal{I} \sum_{\mu,\nu:\mu\neq\nu} \mathrm{Re}\Big\{ f_\mu(q) f_\nu^*(q) (-1)^{k-m} i^l 8\pi^2 \times \sqrt{\frac{4\pi}{(2l+1)}} \underbrace{\int j_l(q\Delta r_{\mu\nu}) Y_l^{-k}\left(\theta_{\mu\nu}^{(\mathrm{mf})}, \phi_{\mu\nu}^{(\mathrm{mf})}\right) \boxed{P(\mathbf{r},\Theta|C)} d\mathbf{r}}_{\text{Molecular Frame Geometry}} \Big\}$$

Method

1. Select neighboring **θ**
2. Add new **θ** to P(**θ**|C) with probability $P(C|\boldsymbol{\theta})_{\mathrm{New}}/P(C|\boldsymbol{\theta})_{\mathrm{Prev}}$

$$P\left(C|\boldsymbol{\Theta}\right) = e^{L(\boldsymbol{\Theta})} \left[\prod_{lmk,q} \frac{1}{\sigma_{lmk}(q)\sqrt{2\pi}}\right] \exp\left\{\frac{-1}{2} \sum_{lmk,q} \left(\frac{C_{lmk}(q) - C_{lmk}^{(\mathrm{calc})}(q,\boldsymbol{\Theta})}{\sigma_{lmk}(q)}\right)^2\right\}$$

SLAC National Accelerator Laboratory