Inversion of Ultrafast X-ray Scattering with Dynamics Constraints

Martin Asenov
School of Informatics,
University of Edinburgh, UK

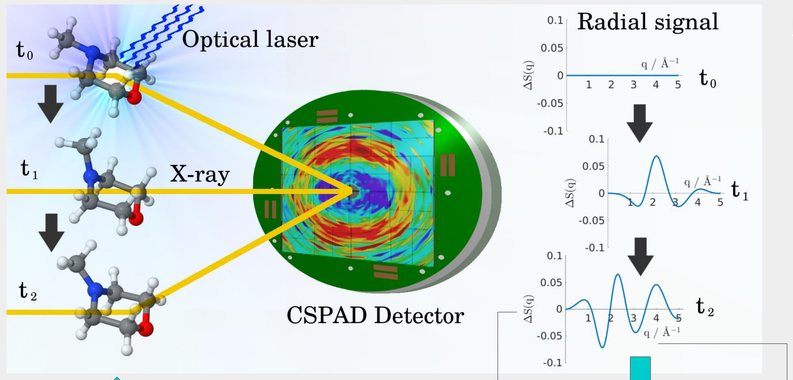
Nikola Zotev
EaStCHEM,
School of Chemistry,
University of Edinburgh, UK

Subramanian Ramamoorthy
School of Informatics,
University of Edinburgh, UK

Adam Kirrander
EaStCHEM and CSEC
School of Chemistry,
University of Edinburgh, UK

THE UNIVERSITY of EDINBURGH

Background



Can we recover the series of molecular geometries?

Ultrafast X-ray Scattering (UXS)

- Allows a direct time-resolved imaging of molecular motion, aka “molecular movies”.
- In this work: molecules are initially in thermal equilibrium in gas phase.
- Optical laser triggers the evolution of the molecular structure.
- X-ray pulses are used to probe the dynamics at a series of time points.
- The **time-series of images (aka the signal)** is collected on a detector.
- The radial component encodes the internal molecular motion via the **full set of atom-atom distances, {R_{AB}(t)}**.

The radial part of the signal is a function of the magnitude of the scattering vector, q, which measures the distance from the centre of the detector.

Here, we represent the signal at time t as a fractional signal change, ΔS(q, t), with respect to the signal at t_0.

The inversion problem

- In theory, there is a mathematical transformation that connects ΔS(q, t) to {R_{AB}(t)} for a given t.
- In practice, because of the finite size of the detector, there is not enough information in a single image.
- **Typical solution**: theoretical models are fitted to the experimental signal **on a frame-by-frame basis** to elucidate the dynamics.
- **Our approach: the inference of the atom-atom distances from the detector signal can be learned from quantum molecular dynamics trajectories, so that the time-dependence is accounted for.**

Methods

Moleclar dynamics and X-ray scattering

- We utilise simulated data to train, validate and test our model (separately for each molecule).
- Quantum molecular dynamics simulations based in CASSCF level of theory for 3 molecules.
- Simulation = a set of trajectories, taken to represent different plausible modes of motion (Quantum Mechanics ignored for now).

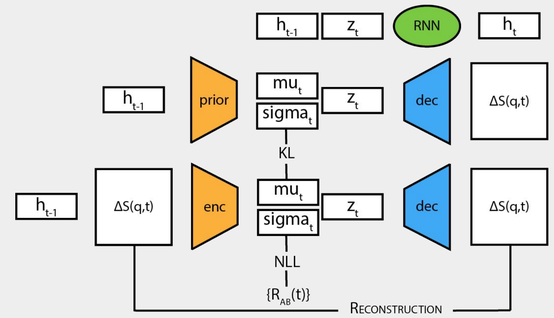
Molecule	NMM	CS_2	C_2H_4
Method	SHARC	SHARC [11]	AI-MCE [12]
Electronic states	X/3s/3p	S_0–S_3, T_1–T_4	S_0–S_2
Simulated time	1 ps	1 ps	150 fs
Time steps	201	201	151
# Train	85	753	800
# Val	11	94 (26)	100 (82)
# Test	11	95 (32)	100 (87)

- Simulation of the scattering signal via the Independent Atom Model (IAM) approximation (CS_2 dataset tested with more advanced methods as well):

$$I(q,t) = \sum_A |f_A(q)|^2 + \sum_A \sum_{B \neq A} f_A(q) f_B(q) \frac{\sin(qR_{AB}(t))}{qR_{AB}(t)}, \qquad \Delta S(q,t) = \frac{I(q,t) - I(t,t_0)}{I(q,t_0)}$$

Inversion with dynamics constraints

- Encode detector signal into a latent space z_t
- Include information about previous dynamics h_{t-1}
- Enforce part of that latent space to be atom-to-atom distances

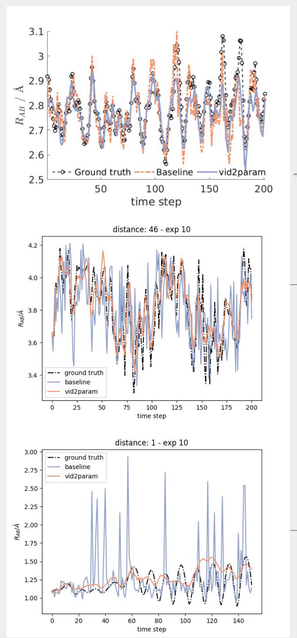


Metrics: Root-mean square deviation or sum of squared residuals

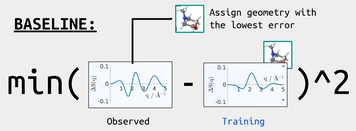
RMSD(-) Predicted / Ground truth

SSR(-) Analytically reconstructed / Ground truth

Results



BASELINE:

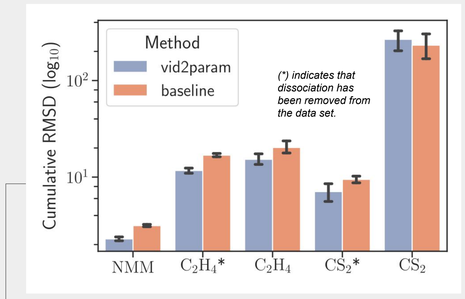


Interatomic distances (raw output):

- For heavy atoms: both baseline and vid2param do well
- Vid2param performs better on degenerate distances
- Vid2param performs better when light atoms such as hydrogen are involved

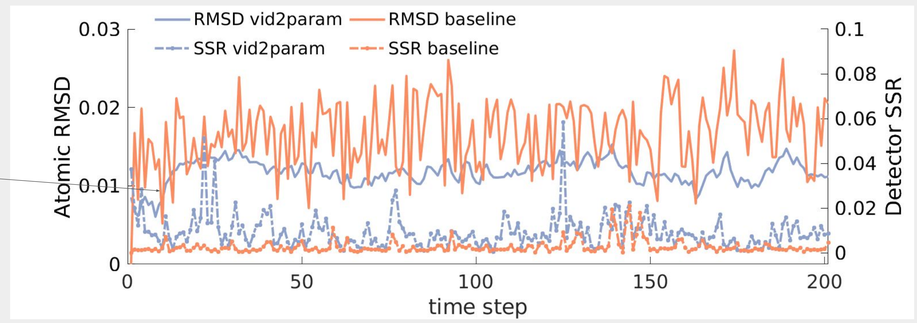
On the metrics Vid2param:

- Is better than baseline on cumulative RMSD (sum over all times and all trajectories)
- Suffers minor issues when the molecule is dissociating
- Has better RMSD throughout a typical trajectory
- Exhibits much smoother behaviour compared to the large variations with time seen in the baseline
- Captures time dependencies and constraints associated with movements of atoms



Analytic reconstruction of detector images

- SSR: second metric not based directly on the molecular geometry but on its scattering fingerprint
- Baseline seems to outperform vid2param on SSR
- Important note: we try to reconstruct only the visible part of the detector - the part of the signal that is not detected, i.e. at large values of q, still contains vital information.
- Better SSR in the visible part does not guarantee better SSR overall
- vid2param learns from previous time points, hence, producing a better estimate for the molecular structure (RMSD) regardless of the SSR



Impact

- UXS are state-of-the-art experiments performed at multibillion dollar facilities, which can help us answer fundamental questions in chemistry such as how molecules move, how bonds between atoms are broken or formed, and even how electrons may change their position inside molecules.
- We believe that the vid2param method addresses a major limitation in the current way of extracting information from UXS data.
- Immediate applications in other experimental techniques for studying ultrafast processes where direct (analytical) inversion is not possible.

Acknowledgements

M. Asenov is supported by the Engineering and Physical Sciences Research Council (EPSRC), as part of the CDT in RAS at Heriot-Watt University and The University of Edinburgh. S. Ramamoorthy acknowledges support from the Alan Turing Institute, through funding for the project: Safe AI for Surgical Assistance.

Datasets, expanded results and source code can be seen at https://sites.google.com/view/mlscattering/