nature portfolio

Peer Review File

Generative Modeling Enables Molecular Structure Retrieval from Coulomb Explosion Imaging

Corresponding Author: Dr Xiang Li

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This work presents a neural network model termed MOLEXA, which utilizes Machine learning technique to retrieve molecular structure from experimental Coulomb explosion imaging (CEI) data. The model is intelligently designed and employs a two-stage training strategy to reconstruct the molecular structures and estimate associated errors. Furthermore, when combined with time-resolved CEI techniques, MOLEXA will be capable of capturing 'snapshots' of a molecule at different instant time during a chemical reaction. Although the current version of MOLEXA can provide reconstruction of a molecule structure with fewer than ten atoms with reasonable reliabilities, it represents a significant and important step toward resolving the highly nonlinear inverse problem of inferring molecular configurations from the moment distributions of fragment ions produced in rapid Coulomb explosion processes.

This work holds considerable importance for the fields of ultrafast imaging in chemistry and physics, particularly in the context of Coulomb explosion techniques. The machine learning approaches applied are well-founded, and the data, analysis, and figures strongly support the reported conclusions and perspectives.

The study demonstrates clear novelty, and the manuscript is well organized with a logical flow. I recommend its publication in Nature Communications after minor revisions

The following comments and corrections should be addressed.
1. Under real experimental conditions, the complete detection of all fragment ions in coincidence become increasingly difficult, if not impossible. Reference 33 provides an example in which molecular structure was successfully reconstructed from only a subset of detected fragment ions – assuming complete ionization and fragmentation of the molecule. As the number of atoms in a molecule increases, manual reconstruction of molecular structure using such partial data becomes infeasible. Can the machine learning approach presented in this work be extended to reconstruct the molecular structure from partially detected fragment ions following a complete Coulomb explosion?
2. In Fig.2, molecular structure is illustrated using ball-and-stick models. In some cases, however, sticks between atoms are missing. please explain the reason for this omission.
3. Minor points:
Line 101, the word 'before' is duplicated; one should be removed.
Line 244, "FIG. 4C" should be corrected to "FIG. 3C"

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The manuscript "Generative Modeling Enables Molecular Structure Retrieval from Coulomb Explosion Imaging" described a much-expected effort to use a neutral network model for solving the problem of reconstructing molecular structure from Coulomb explosion imaging data.
The authors describe their Model and training method of choice, in particular the generation (calculation) of a large lower-

level training data-set and a smaller refined set calculated with a higher-level method. Testing was performed primarily on simulated data, but also applied to experimental Coulomb explosion imaging measurements.
This topic is very interesting and can be suitable for the broad audience of the journal. Nevertheless, there are major issues that need to be addressed and revised.

1. The main issue is the Model performance evaluation criteria:
The model is computed in the center of mass frame, so the RMSE and MAE can be misleading – in particular for small molecules. For example, consider the H2O structure described in figure 3: In the center of mass geometry the Oxygen position is automatically correct – giving the impression of successful reconstruction, while it is just a successful choice of the frame of reference. One would benefit much more from considering the error in bond lengths and angles. One would expect Coulomb explosion imaging to have better performance for bond-lengths compared with angles between bonds. And then too, it would be valuable to learn what is the Maximum Error – not the average one which can be small even if one out of 5 atoms is completely in the wrong position…
2. In the context of Coulomb explosion imaging of angles, it would be valuable to cite and discuss the https://doi.org/10.1103/PhysRevA.97.033412 work that shows the intrinsic non-unique mapping Coulomb explosion imaging on classical electrostatic potentials.

3. One example for the MAE criteria bias is the comparison of H2O and CF4: while inspection of the figure seems to indicate that reconstructed CF4 structure is better – the MAE shows better H2O (most likely due to the exact Oxygen position that reflects only the CM choice). A good test of this would be to look at CF4 vs CH4 which would appear better when using the MAE criteria to evaluate performance.

4. The problem of the performance criteria is even more pronounced when considering diatomic structures: On pg. 5, the authors claim 100% accuracy when considering diatomic structure. However, MAE of 0.4 a.u. might be not so different from a ball park guess with most diatomic bonds being somewhere between 2-4 a.u.? How does MOLEXA perform compared to a thumbs rule based on the kinetic energy release for the dication: R ~ 1.44/KER ? (or a more phenomenological ~1.44/1.2/KER, considering some screening from the remaining electrons) ? This comparison is important for the readers to understand the real potential of MOLEXA at least compared with thumb rule approximations….

5. Following up on point 1, one would also benefit from a discussion of which bonds are better predicted which are less ? e.g. C-C single bonds vs. C=C double bonds and C-H bonds in figure S2 etc…?

6. When considering the experimental Coulomb explosion Imaging: choosing the momentum centroids could be misleading: The ground state wavefunction has an intrinsic spread due to the QM uncertainty principle that samples different initial geometries. Could the authors use MOLEXA to map the experimentally measured distribution and then take the centroid geometry rather than take a centroid position in the momentum correlations that could misrepresent the most probable initial geometry due to the non-linear mapping (same as the shift between lambda0 and w0 in spectroscopy arising due to their non-linear relation)

7. Fig 4 shows that while MOLEXA does not seem to provide reliable information about bond lengths and angles - it can provide insight about the overall structure and identify gross structural rearrangements such as ring-opening and proton migration. In the community of Coulomb explosion imaging, there has been a long established effort to identify molecular isomerization. An early “classical” example is resolving acetylene and vinylidene , see for example https://doi.org/10.1103/PhysRevLett.81.3347 , another critical example is identification of chirality by Coulomb explosion imaging, see for example https://doi.org/10.1126/science.1246549 that could be cited in addition to the already cited Pitzer et al. – the MOLEXA ability of reconstructing the isomer state could be much more convincing than reporting on the maybe limited success in ACCURATE structure reconvery.

8. I think that if the authors focus on the ability of MOLEXA to correctly predict the topological structure (e.g. identify the molecular bonds in the system)

9. There are many seemingly technical, but crucial details that are not so well described in the main text. For example: the choice of molecular orientation or the exact nature of the approximations made in the Coulomb explosion simulations. I understand the desire to limit technical details for the broad audience, but it would be good to read in the main text if the simulated interaction potentials are merely electrostatic approximations – this is important as the readers should grasp the quality of the simulated data set as state of the art ab initio molecular dynamics simulations often are required to involve higher level electronic structure Coulomb explosion calculations that help extract time resolved structural information from experimental data, e.g. https://www.nature.com/articles/s41467-024-50759-2 and https://www.science.org/doi/full/10.1126/sciadv.abq8084

Additional minor comments:
- It is not clear for the readers what exactly was done and achieved in the cyclobutene test on pg.6: I understand that few structures were sampled from ref 38 time-resolved simulations – Coulomb explosion were calculated (using an unspecified approximation) and the simulated momentum distributions were FED to MOLEXA to reconstruct structure. Without taking anything away from the authors achievement, this is not exactly what one would naively understand from “By employing time-resolved CEI datasets, MOLEXA is able to provide ”snapshots” of a molecule at different instants during a chemical reaction.” as stated in the conclusions section. I mean that MOLEXA did not provide new information about a chemical reaction – maybe one can say that it “demonstrates” the potential of MOLEXA to provide such information.

- In figure 2, it is not very clear which are the correct positions and which are the simulated ones – specially in the panel showing low accuracy examples. Please revise.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Li et al. shows that it is possible to determine the structure of a molecule from the momentum vectors of the recoil ions produced in X-ray laser-induced Coulomb explosion based on neural networks.
The authors describe the two-level training protocol where one level comprises a data set obtained by classical trajectory calculation of the Coulomb explosion process and the second level comprises a data set obtained from an ab initio calculation. In terms of application to experimental data, results are shown for three different molecules (Fig. 2). The reconstructed molecular structures are fairly close to the known structures. Finally, to illustrate the use of the retrieval method to image molecules during time-dependent processes, it is applied to a simulated data set of different structures of cyclobutene as it undergoes ring-opening. Again, the reconstructed structures agree quite well with those from the simulation.
As the authors write, timed Coulomb explosion is a promising technique to measure the structure of molecules as they undergo various nuclear dynamics. Experimental data consists of time-resolved momentum distributions, and the challenge is to convert these to molecular structures, i.e. binding angles and distances. This is a notorious challenge for molecules beyond the diatomic case. The neural network approach, which I have not seen reported previously, introduced here is interesting and could be most useful in future studies.
There is certainly room for improvement of the structure accuracy of the method, but I consider the results presented and the prospects sufficiently interesting and promising to make a fit for Nature Communications.
There is, however, one point the authors should address. The results presented in Fig. 4 are, with due respect, just four different static structures of the same molecule that happen to represent the structures of cyclobutene at selected times during the ring-opening process. Since the neural network algorithm was already shown to work for static structure of similar size molecules (Fig. 2, panel 8 and 9) it is hardly surprising that it also works for the four cyclobutene structures. In an experiment, only a small fraction (< 10 %) of all molecules are photoexcited to undergo subsequent dynamics. The majority of the molecules are left in the ground state or, depending on the pump pulse intensity, may be singly ionized. Thus, the target in an experiment is mixed sample. Ideally, the authors should have considered this in their example to more convincingly demonstrate the potential of their novel structure retrieval algorithm to address time-dependent process. Is that possible? If not, the authors should at least include a discussion of this point.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
The manuscript proposes a new method, MOLEXA, based on diffusion models to reconstruct atomistic positions from Coulomb Explosion Imaging data. The approach consists of several components: an embedding module for the ion momenta, a transformer-based architecture to compute subsequent conditioning features, and a denoising model employing a similar transformer architecture for structure reconstruction. In addition, a confidence model is trained to estimate the uncertainties of the reconstructed geometries. For training and validation, the authors develop a new dataset that combines both inexpensive classical simulations and more accurate ab initio calculations, which is also made available online. Although the model is trained only on molecules containing up to seven atoms, it can also be applied to slightly larger systems (up to nine atoms). Moreover, the method is evaluated on some experimental data.

The introduction is exceptionally well written, clearly explaining the task, background, and proposed approach.

Some architectural and model choices are unclear from the manuscript and require more clarifications:
- As described in Fig. 2 and Algorithm 12, the diffusion sampler performs only five denoising steps. It is not clear why such a small number of steps was chosen. Providing ablation studies or a more detailed justification for this design choice would be helpful.
- The proposed memory transformer module could be motivated more clearly, for example through ablation studies comparing it against simple skip connections or learnable residual gates.
- The noise schedule is unclear. Is it just a straight application of the EDM schedule?

The uncertainty model requires some clarifications as well:
- Why 200 bins/classes and not a continuous uncertainty/error score?
- The uncertainty model seems to underestimate the error of reconstructed samples, particularly for samples with high deviations. Especially in those cases, a correct uncertainty value might be important. Could this behavior be a consequence of training with a cross-entropy loss where some categories are sparsely populated? Potentially, a weighted cross-entropy loss or continuous error could help mitigate this issue.
- Could the uncertainty model be used to select the best reconstruction of a molecule from multiple samples generated by the model? An ablation study investigating whether this leads to improved reconstruction quality would be interesting. For

example, scatter plots similar to those shown in the fourth row of Figure 2, but evaluated for individual molecules.
- Figure 2, third row: How are the bins selected, given that the uncertainty model is trained to predict 200 bins? Additionally, are the bar heights representing the mean values, and the error bars the standard deviation?

It is generally unclear how the reconstructed structures are selected, given that the diffusion model can generate multiple candidate structures per molecule. Is only a single reconstruction generated and reported for each molecule? It would be interesting to assess how consistently the model reconstructs structures for a given molecule.

The model does not generalize that well to the unseen larger molecules with 8 or 9 atoms. Could a reason for this be that the ionization process for larger molecules differs from that of smaller ones, e.g. maybe the larger molecules break into smaller fragments first? If so, training data in this regime may be crucial for the model to accurately reconstruct such molecules. However, this also implies that scaling the method to even larger systems would again require corresponding training data, which might be expensive to obtain.

Additional questions and suggestions:
- A runtime evaluation might help the work, as the inference time of the model is most likely much faster than any existing method for the reconstruction
- It may be useful to compare not only absolute positions of the reconstructed structures, but also structural properties such as bond lengths and bond angles. Otherwise, a small difference in the alignment could result in large differences.
- In the tables in the supplementary material, it is unclear how the reported errors are computed.
- How are the ground truth structures in figure 3 generated?
- Are there specific types or classes of molecules for which the model performs particularly well or poorly?

(Remarks on code availability)
The code enables both training the model and performing inference using the trained model or pre-trained weights. In addition, the training, validation, and test sets are provided. However, the packages used and their corresponding versions should be documented somewhere.

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have satisfactorily addressed all of my concerns and have made substantial improvements to the manuscript. Having also reviewed the full rebuttal to the other reviewers, I believe that the authors have thoughtfully considered all comments and critiques. the authors have corrected errors, refined the content of main text ans supplementary materials, and enhanced the clarity of the figures. with these revisions, the manuscript is now well-structured and clearly presented for readers. Therefore, I recommend its acceptance for publication in NC in its current form.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The authors made substantial revisions that improved their manuscript and addressed the concerns and comments of the reviewers. Specifically, MOLEXA performance for diatomic systems shown to exceed the typically applied thumb rules for interpreting KER in terms of interatomic distances emphasizes the relevance of the presented work to a broad audience.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I thank the authors for their additional experiments and clarifications. All my concerns have been addressed, and I think the

paper improved significantly.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

We thank the editor and reviewers for their valuable time and insightful comments which are crucial for further improving our paper. We have revised it accordingly. The following is a point-by-point response.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This work presents a neural network model termed MOLEXA, which utilizes Machine learning technique to retrieve molecular structure from experimental Coulomb explosion imaging (CEI) data. The model is intelligently designed and employs a two-stage training strategy to reconstruct the molecular structures and estimate associated errors. Furthermore, when combined with time-resolved CEI techniques, MOLEXA will be capable of capturing 'snapshots' of a molecule at different instant time during a chemical reaction. Although the current version of MOLEXA can provide reconstruction of a molecule structure with fewer than ten atoms with reasonable reliabilities, it represents a significant and important step toward resolving the highly nonlinear inverse problem of inferring molecular configurations from the moment distributions of fragment ions produced in rapid Coulomb explosion processes.

This work holds considerable importance for the fields of ultrafast imaging in chemistry and physics, particularly in the context of Coulomb explosion techniques. The machine learning approaches applied are well-founded, and the data, analysis, and figures strongly support the reported conclusions and perspectives.

We appreciate the reviewer's acknowledgement of the importance of our study.

The study demonstrates clear novelty, and the manuscript is well organized with a logical flow. I recommend its publication in Nature Communications after minor revisions

The following comments and corrections should be addressed.
1. Under real experimental conditions, the complete detection of all fragment ions in coincidence become increasingly difficult, if not impossible. Reference 33 provides an example in which molecular structure was successfully reconstructed from only a subset of detected fragment ions – assuming complete ionization and fragmentation of the molecule. As the number of atoms in a molecule increases, manual reconstruction of molecular structure using such partial data becomes infeasible. Can the machine learning approach presented in this work be extended to reconstruct the molecular structure from partially detected fragment ions following a complete Coulomb explosion?

The reviewer pointed out one of the major limitations of MOLEXA. Currently the model is incapable of reconstructing molecular structures from incomplete sets of detected fragment ions. We agree with the reviewer that prediction from partial coincidence sets is a very promising research direction. We hope to report our progress on this challenge in future work. To emphasize this limitation, we have added the following to the main text (line 265-274):

The other limitation is that the input data must include all ions produced by Coulomb explosion of the molecule. But in a CEI experiment, the complete coincident ion detection from each explosion is a relatively low-probability event, especially for larger systems. Our current way to circumvent this limitation is by distilling the required full-coincidence momentum vectors from the data accumulated over many explosion events [21-24]. This approach is used to prepare the input data of the experimentally studied molecules discussed in the next section. Further development of the neural network is needed to make it capable of molecular structure prediction from partial coincidence sets, which would be particularly beneficial for single-event reconstruction.

2. In Fig.2, molecular structure is illustrated using ball-and-stick models. In some cases, however, sticks between atoms are missing. please explain the reason for this omission.

We thank the reviewer for notifying us of this missing piece of information. The model only predicts the atomic positions but cannot provide bonding information. The bonding is instead automatically decided by the visualization software (PyMOL) based on the atomic distances. The following was added to the caption of Fig. 2:

The ball-and-stick models were plotted with PyMOL which considers two atoms bonded if their distance is smaller than a tolerance-expanded sum of their covalent sizes.

3. Minor points:
Line 101, the word ‘before’ is duplicated; one should be removed.
Line 244, “FIG. 4C” should be corrected to “FIG. 3C”

We thank the reviewer for pointing them out. They have been corrected.

Reviewer #2 (Remarks to the Author):

The manuscript “Generative Modeling Enables Molecular Structure Retrieval from Coulomb Explosion Imaging” described a much-expected effort to use a neutral network model for solving the problem of reconstructing molecular structure from Coulomb explosion imaging data.

The authors describe their Model and training method of choice, in particular the generation (calculation) of a large lower-level training data-set and a smaller refined set calculated with a higher-level method. Testing was performed primarily on simulated data, but also applied to experimental Coulomb explosion imaging measurements.
This topic is very interesting and can be suitable for the broad audience of the journal. Nevertheless, there are major issues that need to be addressed and revised.

We appreciate the reviewer's acknowledgement of our work's potential broad interest.

1. The main issue is the Model performance evaluation criteria:
The model is computed in the center of mass frame, so the RMSE and MAE can be misleading – in particular for small molecules. For example, consider the H2O structure described in figure 3: In the center of mass geometry the Oxygen position is automatically correct – giving the impression of successful reconstruction, while it is just a successful choice of the frame of reference. One would benefit much more from considering the error in bond lengths and angles. One would expect Coulomb explosion imaging to have better performance for bond-lengths compared with angles between bonds. And then too, it would be valuable to learn what is the Maximum Error – not the average one which can be small even if one out of 5 atoms is completely in the wrong position…

We agree with the reviewer that it would be beneficial to present the prediction errors in bond lengths and angles. The model performance has been re-evaluated with the mean and maximum distance errors of all atomic pairs, and the mean and maximum angle errors of all directional atomic triplets . The following changes have been made to the manuscript:

1. The mean and maximum distance error (DE) and angle error (AE) evaluated on the test datasets are now reported as (line 195-198 and 202-204):

   Using the test dataset of molecules with less than eight atoms, the mean absolute error (MAE) is 0.52 a.u., the mean (maximum) distance error (DE) of all atomic pairs in a molecule is 0.98 (2.11) a.u., and the mean (maximum) angle error (AE) of all directional triplets in a molecule is 13.97 (38.39) degrees.

   For these molecules larger than those included in the training dataset, the MAE is 0.66 a.u., the mean (maximum) DE is 1.16 (3.43) a.u., and the mean (maximum) AE is 14.12 (57.91) degrees.

2. In the top two rows of Fig. 2, the RMSE and MAE values previously displayed for exemplary reconstructions were replaced with the mean DE and AE values. And their maximum DE and AE values are now summarized in Supplementary Tables 2-3.

3. The previous accuracy and MAE plots in the bottom two rows of Fig. 2 have been relocated to Supplementary Fig. 4. They were replaced with the distributions of mean DE and AE values as a function of the predicted uncertainties.

4. The distributions of maximum DE and AE as a function of the predicted uncertainties were added to Supplementary Fig. 5.

2. In the context of Coulomb explosion imaging of angles, it would be valuable to cite and discuss the https://doi.org/10.1103/PhysRevA.97.033412 work that shows the intrinsic non-unique mapping Coulomb explosion imaging on classical electrostatic potentials.

We thank the reviewer for recommending this work. It is now cited as Ref.[41]. The following discussion was added in order to tell the readers that the performance of our current model could be affected by the non-uniquiness problem and suggest a potential route to mitigate this issue in the future development (line 260-265):

The model has two other major limitations. One originates from the non-unique mapping between ion momentum vectors and real-space molecular geometries as revealed by Ref.[41], which can increase the overall prediction error. In future developments, the non-uniqueness issue can be mitigated by a model that makes use of multiple coincidence channels as input data, in contrast to a single channel used by the current model.

3. One example for the MAE criteria bias is the comparison of H2O and CF4: while inspection of the figure seems to indicate that reconstructed CF4 structure is better – the MAE shows better H2O (most likely due to the exact Oxygen position that reflects only the CM choice). A good test of this would be to look at CF4 vs CH4 which would appear better when using the MAE criteria to evaluate performance.

We agree with the reviewer that MAE is not as good a metric to compare the H_2O and CF_4 reconstructions. To address this, the mean and maximum DE and AE values are now reported for the reconstructions of all three experimentally studied molecules as the following (line 327-329, 333-335, and 345-347):

H_2O: *The reconstructed molecular geometry (opaque) is plotted on top of the ground truth (semi-transparent), with the corresponding MAE being 0.296 a.u. The mean (maximum) DE and AE are 0.674 (1.199) a.u. and 18.459 (27.689) degrees, respectively.*

CF_4: *The reconstructed position-space geometry has a MAE of 0.238 a.u., a mean (maximum) DE of 0.66 (1.117) a.u., and a mean (maximum) AE of 5.943 (17.173) degrees.*

C_2H_6O: *The retrieved molecular geometry plotted together with the ground truth at the right has a MAE of 0.429 a.u. The mean (maximum) DE and AE are 1.024 (2.007) a.u.and 9.011 (32.319) degrees, respectively.*

4. The problem of the performance criteria is even more pronounced when considering diatomic structures: On pg. 5, the authors claim 100% accuracy when considering diatomic structure. However, MAE of 0.4 a.u. might be not so different from a ball park guess with most diatomic bonds being somewhere between 2-4 a.u.? How does MOLEXA perform compared to a thumbs rule based on the kinetic energy release for the dication: R ~ 1.44/KER ? (or a more phenomenological ~1.44/1.2/KER, considering some screening from the remaining electrons) ? This comparison is important for the readers to understand the real potential of MOLEXA at least compared with thumb rule approximations….

We agree with the reviewer that it is important to compare MOLEXA predictions with classical calculations based on the thumb rule and have added the following text (line 241-246):

For diatomics, the mean DE is 0.155 a.u. As a reference, the mean DE's from calculating the bond length of the same set of diatomic molecules with the classical 1/KER model and the optimized empirical 0.797/KER model are 1.27 and 0.49 a.u., respectively. (KER stands for the kinetic energy release from Coulomb explosion. The empirical factor 0.797 was determined by minimizing the error of the c/KER model on the test data.)

We would like to note that the optimized empirical factor (0.797) is less than 1 because the molecule takes time to be charged up, which results in smaller KER than the instantaneous charge-up model 1/KER. A related experimental example can be found in [*Phys.Rev.Research 4, 013029 (2022)*].

The mean DE and AE errors from using these two classical models have also been plotted in the bottom left corner of Fig. 2 to facilitate their comparison with MOLEXA prediction errors.

5. Following up on point 1, one would also benefit from a discussion of which bonds are better predicted which are less ? e.g. C-C single bonds vs. C=C double bonds and C-H bonds in figure S2 etc…?

We did not identify the performance dependence on the bond types. As discussed in the paper, the main dependence we found is on the number of atoms in a molecule. In addition, we investigated the dependence on other factors and found that prediction error also tended to increase with atomic distances. The following text was added along with the new Supplementary Fig. 8 showing the dependence of maximum DE and AE values on the maximum atomic distance of a molecule (line 255-257):

In addition to the dependence on the number of atoms, predictions for molecules with larger atomic distances tend to result in larger errors as indicated by Supplementary Fig. 8.

6. When considering the experimental Coulomb explosion Imaging: choosing the momentum centroids could be misleading: The ground state wavefunction has an intrinsic spread due to the QM uncertainty principle that samples different initial geometries. Could the authors use MOLEXA to map the experimentally measured distribution and then take the centroid geometry rather than take a centroid position in the momentum correlations that could misrepresent the most probable initial geometry due to the non-linear mapping (same as the shift between lambda0 and w0 in spectroscopy arising due to their non-linear relation)

We thank the reviewer for raising this point and agree that, due to the nonlinear mapping between ion momentum and real-space molecular geometry, averaging in momentum space does not strictly correspond to averaging in real space. In principle, the equilibrium structure is best estimated by reconstructing molecular geometries from individual single-shot, full-coincidence Coulomb explosion events and then averaging the predicted structures in real space. At present, however, single-shot full-coincidence detection of all ionic fragments remains experimentally challenging. In the current work and as the reviewer wrote, we obtain the input full-coincidence momentum vectors by averaging the data over many explosion events. They are then used to estimate the equilibrium molecular geometry. The following discussion was added in order to emphasize to the readers that our approach is an approximate way to obtaining the equilibrium geometry (line 306-318):

In real molecules, the nuclear ground state exhibits a spatial distribution due to the uncertainty principle, such that different nuclear geometries are sampled even in the absence of excitation. Because the mapping between ion momentum and molecular geometry is nonlinear, averaging in momentum space does not strictly correspond to averaging in real space. In principle, the equilibrium molecular structure is best estimated by reconstructing geometries from individual single-shot, full-coincidence Coulomb explosion events and averaging the predicted structures in real space. But single-shot full-coincidence detection remains challenging for today's CEI experiments. In the following, the structures are instead predicted from the coincident momentum vectors (Supplementary Figs. 9 - 14) obtained by averaging over the single-shot data. This procedure yields an approximate equilibrium geometry from the ensemble-averaged observables, rather than the full nuclear probability distribution or any single instantaneous molecular configuration.

7. Fig 4 shows that while MOLEXA does not seem to provide reliable information about bond lengths and angles - it can provide insight about the overall structure and identify gross structural rearrangements such as ring-opening and proton migration. In the community of Coulomb explosion imaging, there has been a long established effort to identify molecular isomerization. An early "classical" example is resolving acetylene and

vinylidene , see for example https://doi.org/10.1103/PhysRevLett.81.3347 , another critical example is identification of chirality by Coulomb explosion imaging, see for example https://doi.org/10.1126/science.1246549 that could be cited in addition to the already cited Pitzer et al. – the MOLEXA ability of reconstructing the isomer state could be much more convincing than reporting on the maybe limited success in ACCURATE structure reconvery.

We thank the reviewer for suggesting these two pioneering studies and have added them as Refs.[6-7]. We also agree with the reviewer that the science cases in them are among the best examples for demonstrating the capability of MOLEXA and hope to make use of them in our future work.

8. I think that if the authors focus on the ability of MOLEXA to correctly predict the topological structure (e.g. identify the molecular bonds in the system)

We share this view with the reviewer and have added the texts below (line 366-368 and 381-384):

The reconstructions show that the model can provide insight about the overall structure and identify gross structural rearrangements such as proton migration and ring opening

It allows for inverting momentum-space datasets to position space, providing the structure of a molecule right before its explosion by an X-ray pulse and showing particular effectiveness in reconstructing the overall structure of molecules.

9. There are many seemingly technical, but crucial details that are not so well described in the main text. For example: the choice of molecular orientation or the exact nature of the approximations made in the Coulomb explosion simulations. I understand the desire to limit technical details for the broad audience, but it would be good to read in the main text if the simulated interaction potentials are merely electrostatic approximations – this is important as the readers should grasp the quality of the simulated data set as state of the art ab initio molecular dynamics simulations often are required to involve higher level electronic structure Coulomb explosion calculations that help extract time resolved structural information from experimental data, e.g. https://www.nature.com/articles/s41467-024-50759-2 and https://www.science.org/doi/full/10.1126/sciadv.abq8084

We thank the reviewer for pointing this out and have added more details to the main text. They include

1. Moving the following molecular orientation discussion from Methods to the main text (line 172-181)

Specifically, for each molecule, the emission direction of the heaviest ion fragment is defined as the x axis. Then the cosine similarities of the momenta of the other ion fragments relative to the x direction are calculated. The ion momentum with the smallest cosine similarity is chosen to define the y axis through Gram-Schmidt orthonormalization so that this momentum vector would fall within the first quadrant of the x-y plane. Both the momenta of all ion fragments and the initial molecular structure, after being centered at the origin, are transformed into this coordinate frame. For diatomics, the coordinate system is fully determined once the emission direction of the heavier ion fragment is defined as the x axis.

2. Providing a brief explanation of the nature of the ab initio simulation (line 154-156)

   One level involves the ab initio calculation of the XFEL-induced Coulomb explosion of molecules, which tracks the quantum transition probabilities across all participating electronic configurations while treating the nuclei as moving in a classical force field [40].

3. Moving the experimental details from Methods to the main text (line 292-305)

   The experiments were performed in multiple CEI beamtimes using the COLTRIMS (Cold Target Recoil Ion Momentum Spectroscopy) Reaction Microscope[42] at the Small Quantum Systems (SQS) instrument. The beamtimes and x-ray pulse parameters are summarized in Table 4. In all experiments, the molecular samples were delivered to the interaction region through a supersonic expansion followed by three skimmers and an adjustable collimator. The distance from the nozzle to the interaction region is about 54 cm. The pressure in the main chamber was maintained at 1 × 10−11 mbar. The focal spot size of the X-ray beam was about 1.5 –3 μm and the X-ray pulse duration was less than 25 fs based on the 250pC electron bunch charge. The ion fragments produced in the interaction region were guided by a homogeneous electric field to a time- and position-sensitive detector. The lab-frame momentum vectors of the ion fragments were then reconstructed from the detector readouts, which were subsequently transformed into the molecular frame with the alignment procedure discussed in the Training subsection and Supplementary Methods.

4. Several details related to the neural network, including model architecture, inference time and uncertainty estimation, are now added to the main text. (line 121-123, 133-134, 143-146, 204-210, 216-219, and 224-229)

Additional minor comments:
- It is not clear for the readers what exactly was done and achieved in the cyclobutene test on pg.6: I understand that few structures were sampled from ref 38 time-resolved simulations – Coulomb explosion were calculated (using an unspecified approximation) and the simulated momentum distributions were FED to MOLEXA to reconstruct structure. Without taking anything away from the authors achievement, this is not exactly what one

would naively understand from “By employing time-resolved CEI datasets, MOLEXA is able to provide ”snapshots” of a molecule at different instants during a chemical reaction.” as stated in the conclusions section. I mean that MOLEXA did not provide new information about a chemical reaction – maybe one can say that it “demonstrates” the potential of MOLEXA to provide such information.

We agree with the reviewer that our previous way of presenting the cyclobutene example can give readers a false impression of MOLEXA’s capability. As written in the response to point 8, the following text is added to the cyclobutene discussion to clarify our message (line 366-368):

The reconstructions show that the model can provide insight about the overall structure and identify gross structural rearrangements such as proton migration and ring opening.

We have also taken the reviewer’s advice and revised the related text in the Conclusions section as the following (line 387-390):

MOLEXA has the potential to provide ”snapshots” of a molecule at different instants during a chemical reaction, which can enable the use of the CEI technique for direct reconstruction of molecular dynamics in position space as they unfold on their natural time scales.

- In figure 2, it is not very clear which are the correct positions and which are the simulated ones – specially in the panel showing low accuracy examples. Please revise.

We thank the reviewer for bringing out this issue. To make the predicted and ground-truth structures more differentiable, we have increased the transparency of the latter in Fig. 2. This change was also applied to the ball-and-stick models in Fig. 3 and Supplementary Fig. 15.

Reviewer #3 (Remarks to the Author):

Li et al. shows that it is possible to determine the structure of a molecule from the momentum vectors of the recoil ions produced in X-ray laser-induced Coulomb explosion based on neural networks.
The authors describe the two-level training protocol where one level comprises a data set obtained by classical trajectory calculation of the Coulomb explosion process and the second level comprises a data set obtained from an ab initio calculation. In terms of application to experimental data, results are shown for three different molecules (Fig. 2). The reconstructed molecular structures are fairly close to the known structures. Finally, to illustrate the use of the retrieval method to image molecules during time-dependent processes, it is applied to a simulated data set of different structures of cyclobutene as it

undergoes ring-opening. Again, the reconstructed structures agree quite well with those from the simulation.
As the authors write, timed Coulomb explosion is a promising technique to measure the structure of molecules as they undergo various nuclear dynamics. Experimental data consists of time-resolved momentum distributions, and the challenge is to convert these to molecular structures, i.e. binding angles and distances. This is a notorious challenge for molecules beyond the diatomic case. The neural network approach, which I have not seen reported previously, introduced here is interesting and could be most useful in future studies.
There is certainly room for improvement of the structure accuracy of the method, but I consider the results presented and the prospects sufficiently interesting and promising to make a fit for Nature Communications.

We are thankful for the reviewer's recognition of our work's potential.

There is, however, one point the authors should address. The results presented in Fig. 4 are, with due respect, just four different static structures of the same molecule that happen to represent the structures of cyclobutene at selected times during the ring-opening process. Since the neural network algorithm was already shown to work for static structure of similar size molecules (Fig. 2, panel 8 and 9) it is hardly surprising that it also works for the four cyclobutene structures. In an experiment, only a small fraction (< 10 %) of all molecules are photoexcited to undergo subsequent dynamics. The majority of the molecules are left in the ground state or, depending on the pump pulse intensity, may be singly ionized. Thus, the target in an experiment is mixed sample. Ideally, the authors should have considered this in their example to more convincingly demonstrate the potential of their novel structure retrieval algorithm to address time-dependent process. Is that possible? If not, the authors should at least include a discussion of this point.

We thank the reviewer for raising this point. It made us realize our previous way of presenting Fig. 4 could give readers a false impression that the neural network could be directly applied to a time-resolved CEI experiment. Currently, the model can only make reconstructions if the input momentum vectors originate from a single initial molecular geometry and hence cannot deal with mixed structures. To address this point, we added the following to the main text (line 371-378):

It is worth noting that the reconstructions in Fig. 4 are idealized scenarios where the input momentum vectors correspond to a single well-defined molecular geometry. This can work if the input coincidence data is assembled from a single explosion event. For time-resolved CEI experiments, the accumulated data for each time delay typically results from a mixture of geometries corresponding to multiple quantum states. Classification of the experimental data

into individual states, based on, e.g., ion charge state characteristics and kinetic energies[18], is hence required prior to applying MOLEXA for reconstruction.

Reviewer #4 (Remarks to the Author):

The manuscript proposes a new method, MOLEXA, based on diffusion models to reconstruct atomistic positions from Coulomb Explosion Imaging data. The approach consists of several components: an embedding module for the ion momenta, a transformer-based architecture to compute subsequent conditioning features, and a denoising model employing a similar transformer architecture for structure reconstruction. In addition, a confidence model is trained to estimate the uncertainties of the reconstructed geometries. For training and validation, the authors develop a new dataset that combines both inexpensive classical simulations and more accurate ab initio calculations, which is also made available online. Although the model is trained only on molecules containing up to seven atoms, it can also be applied to slightly larger systems (up to nine atoms). Moreover, the method is evaluated on some experimental data.

The introduction is exceptionally well written, clearly explaining the task, background, and proposed approach.

We thank the reviewer for providing this encouraging feedback for the introduction.

Some architectural and model choices are unclear from the manuscript and require more clarifications:
- As described in Fig. 2 and Algorithm 12, the diffusion sampler performs only five denoising steps. It is not clear why such a small number of steps was chosen. Providing ablation studies or a more detailed justification for this design choice would be helpful.

We agree with the reviewer and have added Supplementary Fig. 18 which shows the dependence of prediction errors on the number of denoising (sampling) steps. It can be seen that with 5 denoising steps, the prediction errors are already getting close to the minimum values that are achievable with less than 100 denoising steps. The following justification is added (page 51, line 737-742):

The prediction errors evaluated as a function of the number of sampling steps (NSS) are displayed in Supplementary Fig. 18. The overall minimum is reached at the NSS of 33. Since five sampling steps can already get to an error within 1% of the minimum value (achievable with less than 100 sampling steps), it was chosen as a good compromise between prediction accuracy and inference time.

- The proposed memory transformer module could be motivated more clearly, for example through ablation studies comparing it against simple skip connections or learnable residual gates.

We agree that a clearer justification of the memory transformer module is important. We carried out an ablation study in which the memory mechanism was replaced by skip connections while keeping all other components unchanged. The results show that the original model consistently outperforms the "skip connection" variant, reducing the mean (maximum) atomic distance errors by 3.6% (4.0%) and the mean (maximum) angle errors by 1.3% (4.4%). We added the following to the main text (line 121-123):

In comparison to using skip connections, the memory mechanism is found to suppress the mean (maximum) atomic distance and angle errors by 3.6% (4%) and 1.3% (4.4%), respectively.

- The noise schedule is unclear. Is it just a straight application of the EDM schedule?

Yes, it is mostly an application of the EDM schedule from Ref.[34], with the detailed parameters provided in Algorithm 12. We added the following to the main text (line 133-134):

During inference, this structure is iteratively refined by a diffusion sampler with a noise schedule adapted from Ref.[34].

The uncertainty model requires some clarifications as well:
- Why 200 bins/classes and not a continuous uncertainty/error score?

Indeed, the model first predicts the probabilities of the prediction error falling into each of the 200 bins/classes. Then the uncertainty is calculated by taking a probability-weighted sum of the 200 bin values. This uncertainty produced this way is continuous. To clarify this, we added the following to the main text (line 143-146):

It pre-defines the uncertainty bins r = [0, 0.05, 0.1, . . . , 9.95] and estimates the probability that the prediction error falls within each of these bins. For each coordinate of the predicted atomic positions, the uncertainty is then calculated as the probability-weighted sum of the bin values.

- The uncertainty model seems to underestimate the error of reconstructed samples, particularly for samples with high deviations. Especially in those cases, a correct uncertainty value might be important. Could this behavior be a consequence of training with a cross-entropy loss where some categories are sparsely populated? Potentially, a weighted cross-entropy loss or continuous error could help mitigate this issue.

We thank the reviewer for raising this issue. It led us to spot a mistake in the accuracy and MAE plots in the previous Fig. 2. The predicted uncertainty was designed to match the MAE. Previous plots accidentally used the square root of the uncertainties and RMSE (typically larger than MAE), which made the prediction error look larger than the predicted uncertainty. This mistake has been corrected and the accuracy and MAE plots in the previous Fig.2 have been relocated to Supplementary Fig. 4. It shows a more reasonable correspondence between the MAE's and the predicted uncertainties.

We agree with the reviewer, even with the corrected plot, the model can underestimate the prediction errors, especially for samples with large prediction errors. As the reviewer wrote, it could be caused by some uncertainty bins being sparsely populated and could potentially be mitigated with a weighted cross-entropy loss. To emphasize this limitation, we added the following to the main text (line 224-229):

It can be observed from these heatmaps that the predicted uncertainties can underestimate the reconstruction errors, especially for larger molecules with larger prediction errors. We attribute this to a training bias caused by the Uncertainty Estimation Module being trained on reconstructions most of which have relatively low errors. Such bias can potentially be mitigated with a weighted cross-entropy loss in future developments.

- Could the uncertainty model be used to select the best reconstruction of a molecule from multiple samples generated by the model? An ablation study investigating whether this leads to improved reconstruction quality would be interesting. For example, scatter plots similar to those shown in the fourth row of Figure 2, but evaluated for individual molecules.

We thank the reviewer for making this suggestion. We randomly selected 1000 molecules, performed 1000 independent reconstructions for each molecule. The distributions of the correlation coefficient between reconstruction errors and predicted uncertainties are plotted in Supplementary Fig. 6. They show no clear positive correlation, although their means are slightly positive. It makes us conclude that the benefit of using uncertainty values to select reconstructions is marginal. The following was added to the main text to discuss this idea (line 235-239):

It needs to be noted that although an uncertainty estimate can assist the assessment of the trustworthiness of a prediction, it is not effective in selecting the most accurate reconstruction out of many predictions for a molecule. This is indicated by the small correlation coefficients between uncertainty estimate and prediction errors plotted in Supplementary Fig. 6.

- Figure 2, third row: How are the bins selected, given that the uncertainty model is trained to predict 200 bins? Additionally, are the bar heights representing the mean values, and the error bars the standard deviation?

The predicted uncertainty is a continuous value given by the probability-weighted sum of the 200 bin values. Each of the bins in the third row of Fig. 2 has a size of 0.1 a.u. The bar height represents the percentage of the model predictions with a MAE less than 0.6 a.u. out of all predictions which have an uncertainty falling into that bin. The error bar is derived from the standard deviations. We added the text below to the caption of Supplementary Fig. 4:

At each uncertainty bin of size 0.1 a.u., the bar height is the percentage of predictions with a MAE below 0.6~a.u out of all predictions which have an uncertainty belonging to that bin. The error bar is calculated from standard deviations, assuming a Poisson distribution for the number of counts falling into each bin.

It is generally unclear how the reconstructed structures are selected, given that the diffusion model can generate multiple candidate structures per molecule. Is only a single reconstruction generated and reported for each molecule? It would be interesting to assess how consistently the model reconstructs structures for a given molecule.

We thank the reviewer for this suggestion. Each reconstruction is indeed from a single model prediction. We found little discrepancies between independent predictions. Quantitatively, this is reflected by the distribution of standard deviations of model predictions in Supplementary Fig. 2, which reveals that the distribution is mostly below 0.075 a.u. We added the text below to discuss this point (line 206-210):

Each of the reconstructed molecular structures is from a single model prediction. Based on the standard deviations of the model predictions plotted in Supplementary Fig. 2, the expected discrepancy between the corresponding atomic coordinates of two independent reconstructions for a single molecule is about 0.07 a.u.

The model does not generalize that well to the unseen larger molecules with 8 or 9 atoms. Could a reason for this be that the ionization process for larger molecules differs from that of smaller ones, e.g. maybe the larger molecules break into smaller fragments first? If so, training data in this regime may be crucial for the model to accurately reconstruct such molecules. However, this also implies that scaling the method to even larger systems would again require corresponding training data, which might be expensive to obtain.

As the reviewer wrote, the charge-up and fragmentation processes for larger molecules could be different from small ones. And we agree that training with larger molecules is crucial to improve its performance. We are working on ways to more efficiently produce

training datasets for larger molecules and hope to report our progress in this respect in a future work.

Additional questions and suggestions:
- A runtime evaluation might help the work, as the inference time of the model is most likely much faster than any existing method for the reconstruction

The runtime evaluation has been added to Supplementary Fig. 1, with the corresponding text added as the following (line 204-206):

The inference time distribution evaluated on all test molecules is plotted in Supplementary Fig. 1 and has a mean of 59.8 ms.

- It may be useful to compare not only absolute positions of the reconstructed structures, but also structural properties such as bond lengths and bond angles. Otherwise, a small difference in the alignment could result in large differences.

We agree with the reviewer and have added evaluation metrics related to bond lengths and angles. The model was re-evaluated based on the distance errors of all atomic pairs and angle errors of all directional triplets in a molecule. Related changes in the manuscript include

1. The mean and maximum distance error (DE) and angle error (AE) evaluated on the test datasets are now reported as (line 195-198 and 202-204):

   Using the test dataset of molecules with less than eight atoms, the mean absolute error (MAE) is 0.52 a.u., the mean (maximum) distance error (DE) of all atomic pairs in a molecule is 0.98 (2.11) a.u., and the mean (maximum) angle error (AE) of all directional triplets in a molecule is 13.97 (38.39) degrees.

   For these molecules larger than those included in the training dataset, the MAE is 0.66 a.u., the mean (maximum) DE is 1.16 (3.43) a.u., and the mean (maximum) AE is 14.12 (57.91) degrees.

2. In the top two rows of Fig. 2, the RMSE and MAE values previously displayed for exemplary reconstructions were replaced with the mean DE and AE values. And their maximum DE and AE values are now summarized in Supplementary Tables 2-3.

3. The previous accuracy and MAE plots in the bottom two rows of Fig. 2 have been relocated to Supplementary Fig. 4. They were replaced with the distributions of mean DE and AE as a function of the predicted uncertainties.

4. The distributions of maximum DE and AE as a function of the predicted uncertainties were added to Supplementary Fig. 5.

- In the tables in the supplementary material, it is unclear how the reported errors are computed.

The errors were the uncertainties directly predicted by the model after taking the probability-weighted sum of the 200 bin values. To clarify it, we added the following to the main text (line 216-219):

The Uncertainty Estimation Module produces uncertainties for all coordinates of all atoms in a molecule. The overall uncertainty displayed here (Fig. 2) *was obtained by taking an average of the uncertainties of all atomic positions of a molecule.*

The following text is added to the Uncertainty Estimation Module subsection in Supplementary Methods (page 55, line 754-758):

For each predicted coordinate, the uncertainty is then calculated as the probability-weighted sum of the bin values. Such values are directly used in Supplementary Tables 5 - 7. The uncertainty used in Fig. 2 and Supplementary Figs. 4- 5 was calculated by taking an average of the uncertainty values predicted for all atomic coordinates in a molecule.

- How are the ground truth structures in figure 3 generated?

They are from the NIST Computational Chemistry Comparison and Benchmark Database. The text below has been added to the caption of Fig. 3:

The ground truth structures are from the NIST Computational Chemistry Comparison and Benchmark Database [47].

- Are there specific types or classes of molecules for which the model performs particularly well or poorly?

The main dependence we found is on the number of atoms as discussed in the main text. We also investigated other potentially influencing factors and found that the model also tended to have larger prediction errors for molecules with larger atomic distances. The following text was added along with the new Supplementary Fig. 8 showing the dependence of maximum DE and AE on the maximum atomic distance of a molecule (line 255-257):

In addition to the dependence on the number of atoms, predictions for molecules with larger atomic distances tend to result in larger errors as indicated by Supplementary Fig. 8.

Reviewer #4 (Remarks on code availability):

The code enables both training the model and performing inference using the trained model or pre-trained weights. In addition, the training, validation, and test sets are provided. However, the packages used and their corresponding versions should be documented somewhere.

We thank the reviewer for this suggestion and have listed the following dependencies in the environment.yml file of the MOELXA repository on GitHub:

dependencies:
- *python=3.12.2*
- *pytorch=2.3.0*
- *pytorch-cuda=12.1*
- *cudnn=8.9.2*
- *fastai=2.7.15*
- *fastcore=1.5.33*
- *fastprogress=1.0.3*
- *numpy=1.26.4*
- *pandas=2.2.1*
- *matplotlib=3.8.4*
- *gdown=5.2.0*
- *torch-geometric==2.5.3*
- *torcheval==0.0.7*
- *einops==0.8.0*

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank the reviewer for helping review our manuscript.