Computational Physics

Approaches for constraining uncertainty and degeneracy in geometry reconstruction of molecules from simulated Coulomb explosion data

Kaili Tian [a,*], Ali Ramadhan [b], Marcel Nooijen [c], Stefan V. Pantazi [d], Reza Karimi [a], Joseph H. Sanderson [a]

[a] *Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
[b] *Massachusetts Institute of Technology, Cambridge, MA 02139, United States*
[c] *Department of Chemistry, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
[d] *Conestoga College Institute of Technology and Advanced Learning, Kitchener, ON N2G 4M4, Canada*

ARTICLE INFO

ABSTRACT

Coulomb explosion imaging (CEI) data, consisting of atomic ion momenta, resulting from a rapid (few fs) ionization event, is typically recorded as a first step in the generation of a single molecule geometry measurement or the production of a 'molecular movie' that watches the evolution of an excited state geometry in time. Deriving a reliable geometry from this data is a crucial and challenging step, as unique solutions may not exist. In this work, we start by simulating the geometries of a molecule, which is out of equilibrium, the asymmetric carbonyl sulfide (OCS) molecule. We generate momentum data resulting from six-fold ionization and investigate two methods to reconstruct geometries. The first method is a look-up table, which is simple in principle but shows limitations in terms of accuracy and ability to identify multiple 'degenerate' solutions. The second is a new method using nonlinear constrained optimization, which shows potential for scalability to higher dimension. We investigate the possibility of identifying problematic geometry regions where different degenerate geometries can not be distinguished.

1. Introduction

Coulomb explosion imaging (CEI) [1] is a technique for studying the structure of small molecules in the gas phase. Essentially, the molecule is ionized to induce fragmentation, during which the positively charged fragments repel each other with the Coulomb force. The momentum vector of each fragment is measured, in coincidence, by a sophisticated detection apparatus [2], which incorporates time and position sensitive detection [3]. Information about the initial geometry of the molecule is contained in the recorded momentum vectors of the ions, but there is no unique relationship between momentum and geometry. Recent experimental work [4] has shown that a "forward" method which takes a Monte Carlo approach to reproducing experimental data from the initial nuclear distribution, can match some significant features. In principle, however, it is possible to reconstruct the molecular structure from only the final momentum vectors of the fragment ions, by employing an inversion or "backward" approach, provide there is knowledge of the interaction potential governing the initial repulsion. Here we focus on the challenge of reproducing molecules that are out of equilibrium as

they respond to an excitation process [3], without prior knowledge of the nuclear distribution. This is however a non-linear problem, and the technique requires an optimization method in order to find the solution.

Légaré et al. [5] were the first to use few cycles, femtosecond laser pulses (sub 7fs) and CEI to report on the molecular structure of small molecule (D_2O, SO_2). They employed the inversion process using Coulomb's law and compared with *ab initio* potential energy surfaces to reconstruct the equilibrium geometries of D_2O and SO_2. Gagnon et al. [6] also used the inversion method and reported the reconstruction of the five-atom molecule dichloromethane (CH_2Cl_2) using a unique stochastic-based simulated annealing algorithm that globally optimizes the molecular spatial configuration. Their work suggested that the major uncertainty in the geometry reconstruction was due to the uncertainty in the measurement of the velocity vectors. However, they were only able to obtain molecular geometries, for five sets of measured velocity vectors, out of potentially hundreds of complete ion momentum measurements. Brichta et al. [7] proposed an inversion method for the reconstruction of small triatomic molecules using a simplex algorithm. They tested their method using the reconstruction of simulated

---

geometries of carbon dioxide (CO_2) and formaldehyde (CH_2O). The simplex algorithm was originally developed for solving linear problems [8], and further improvements were therefore needed to solve the nonlinear problem that constitutes the challenge of a molecular reconstruction, with its large numbers of variables and constraints [9]. Since the simplex algorithm is a local search method and relies heavily on the starting guess geometry, for some cases, it is hard to distinguish a local minimum from the global minimum. Nevertheless, Bocharova et al. [10] analyzing the results of their few cycles laser initiated multiple ionization experiment, obtained the average CO bond distance and bend angle for CO_2 data employing the simplex code [7]. An interesting geometry reconstruction effort using CEI was performed by Kunitski et al. [11], they employed a reconstruction method by using a "look-up table" to return geometries of the Eflimov state of $^4He_3$. The method works, by matching experimental momenta to a set of predetermined, simulated, values stored in a three-dimensional array, defined by atomic separations (R_1, R_2, R_3). However, like most experimental work, they did not give details of their method or show the look-up table. The basic idea of the method is still a linear process, while the problem remains a non-linear problem, where a small change in the initial molecular structure will result in a large difference in the final momentum. Sayler et al. [12] showed that multiple initial configurations can result in identical momenta for a triatomic molecular explosion, even when focusing on the limited case of fixed total energy release. In real Coulomb explosion experiments, the total released energy from a molecule would of course have a characteristic distribution, but these so called "degenerate" geometries could pose a fundamental obstacle to using the "backwards" approach to Coulomb explosion imaging in certain cases. Any successful reconstruction approach should have a process in place for identifying and if possible correcting errors that result from degeneracy.

This study focuses on the algorithm that involves the mapping from momenta to geometry structures and identifying degeneracy cases. Geometry reconstruction is performed for simulated carbonyl sulfide (OCS) in the OCS→O^{2+} + C^{2+} + S^{2+} concerted Coulombic explosion process, using two approaches: a look-up table method, and a nonlinear constrained optimization method. OCS is a common target in CEI. OCS is selected due to its asymmetry and linearity, while also exhibiting slight bends caused by vibrational motion in its ground state, making it of interest to experimentalists. The 6+ state is chosen because all events in this state arise from concerted breakup processes. As detailed in study [13], the 3+ and 4+ states involve a combination of both concerted and stepwise processes. Two simulated initial geometry datasets are employed. To make the methods applicable to any unknown geometry, the ground state nuclear distribution for OCS is not used; instead, possibilities that might arise from electronic and vibrational excitation are considered. Then the final momentum vectors are calculated by the Coulomb explosion model as described in Section 2, serving as our simulated experimental data. Each momentum vector is then fed into our reconstruction process, to find the reconstructed geometry. The known original geometries are used for comparing them with the reconstructed geometries and evaluating the ability of the two reconstruction methods. The evaluation is based on the number of events successfully reconstructed, speed, accuracy, and the ability to deal with degeneracy.

2. Preparation of simulated OCS geometries and momentum vectors

To simulate the Coulomb explosion of OCS^{6+}, similar simplifying assumptions are applied as in previous studies [5–8,10]. 1. All six electrons are instantly removed, and there is no charge redistribution afterward, or interaction between electrons and nuclei. 2. At t = 0, any structural deformation has already occurred. 3. ions behave as point-like charged particles. The point like ion approach means that the motion of the ions is governed only by their mutual Coulomb repulsion. Therefore,

only pure Coulomb potential is transferred into kinetic energy. 4. The impact of an external electric field, from the experimental setup, on the structures is not considered. 5. Every ion starts with zero initial momentum. Momentum is calculated until t=T (~1e-13s) when the momentum components reach their asymptotic values and are considered constant (exhibit minimal change within 0.0001e-22 kgm/s).

We aim to address two primary questions: Is a lookup table method superior to an optimization method, and how can we identify degenerate geometry results. We choose to do this within the restriction of the idealized dynamics described above, so that we can isolate these fundamental questions from influences which would distort our results. Furthermore, we have used these exact assumptions in previous work with experimental data which has given acceptable reconstructions [10], nevertheless, before we proceed with this study, we can briefly address the assumptions to identify which might be the most problematic and which are most reasonable. The validity of the assumptions we are making have varying levels of impact on the accuracy of the results. For the assumption 1, that the electrons are removed instantaneously, in reality they are removed during a laser pulse which may be sub 10 fs, between six and ten electrons could be removed depending on the ionization source [4,10], which is short compared to molecular dissociation timescale, but is not negligible, additionally, the nonzero effect of the laser field on the ion trajectories during this time should be considered for tunneling ionization events involving IR laser light. Assumption 3 is reasonable for a molecule ionized six times but because of the finite time to reach this state, some degree of bonding exists until around an overall charge state of 4+ is achieved. Assumption 4 is valid, as the impact of the external electric field, of the experiment is negligible on the timescale over which ions reach asymptotic momentum, having an influence of no more then $1/10^6$. Despite the initial vibrational motion of the molecule, assumption 5, of zero initial momentum is reasonable, ($1/10^3$ of the asymptotic momentum) in fact this initial momentum is small even for molecules undergoing isomerization and can reasonably be ignored [3], when determining the initial position of an ion. The final momenta measured are related to the initial positions of the atoms in the molecule rather than their initial momentum, and it is precisely because of this sensitivity to initial position within a molecule that Coulomb explosion imaging has merit, it can not determine the position in the experiment, where the molecule originates, to better than the micron level. In fact, for experimental purposes all ions are considered to have originated at the same physical point in space, to calculate their initial momentum with the errors which this introduces in momentum, no more than $1/10^6$.

$$\boldsymbol{r_{ij}} = \boldsymbol{r_j} - \boldsymbol{r_i}, \ \boldsymbol{i,j} \in \{\boldsymbol{O,C,S}\} \tag{1}$$

$$H(r_i, p_i', t) = \sum_{i=1}^{n} \frac{p_i'^2}{2m_i} + \frac{1}{4\pi\varepsilon_0} \sum_{\{i,j\},j\neq i} \frac{q_i q_j}{|r_i - r_j|} \tag{2}$$

$$\frac{dr_i}{dt} = \frac{\partial H}{\partial p_i'} = \frac{p_i'}{m_i} \tag{3}$$

$$\frac{dp_i'}{dt} = -\frac{\partial H}{\partial r_i} = \frac{1}{4\pi\varepsilon_0} \sum_{j,j\neq i} \frac{q_i q_j (r_i - r_j)}{|r_i - r_j|^3} \tag{4}$$

Consider the triatomic system as shown in Fig. 1 where $\mathbf{r}_O$, $\mathbf{r}_C$, and $\mathbf{r}_S$ represent the position vectors and $\mathbf{p'}_O$, $\mathbf{p'}_C$ and $\mathbf{p'}_S$ are the momentum vectors. Under the above assumptions, at t=0, we have $\mathbf{p'}_O = \mathbf{p'}_C = \mathbf{p'}_S = \mathbf{0}$, and initial geometry ($\mathbf{r}_O$, $\mathbf{r}_C$, $\mathbf{r}_S$). Since the CO and CS bond lengths, and bond angle θ_{OCS} are the key parameters of interest in our study, the initial geometry is described using (r_{co}, r_{cs}, θ_{ocs}) in the following discussions. In this way, the orientation of the geometry is no longer a factor since our focus is on studying the structure of each individual molecule. Furthermore, to ensure produce a unique momentum for each geometry, carbon is positioned at the origin, the oxygen along the -x-axis, then the position of sulfur can be easily calculated in the xy-plane.
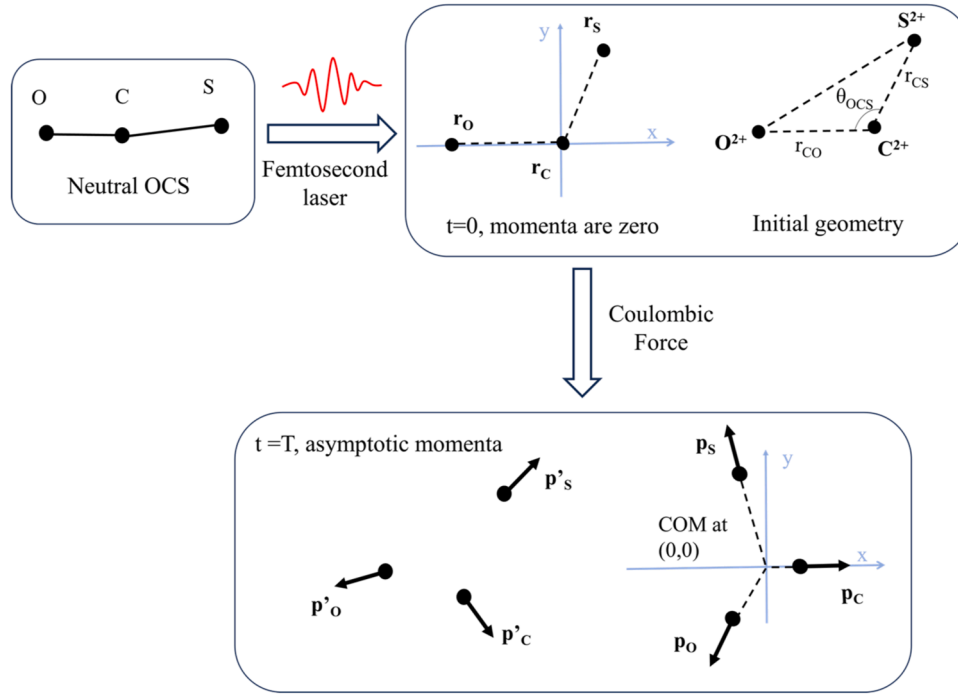
Fig. 1. Structural evolution of OCS. The initial geometry is defined at t=0, with carbon at the origin (0,0) and oxygen along the negative x-axis and sulfur in the first quadrant. The final asymptotic momentum is defined at time T with $\mathbf{p}_C$ along the positive x-axis, $\mathbf{p}_O$ in the third quadrant and $\mathbf{p}_S$ in the second quadrant.

At t=T, the final momentum ($\mathbf{p'}_O$, $\mathbf{p'}_C$, $\mathbf{p'}_S$) can be obtained from the Hamiltonian of the system, as detailed in equation (1-4). Additionally, to facilitate comparison with experimental data, the final momentum is adjusted to $p = (\mathbf{p}_O, \mathbf{p}_C, \mathbf{p}_S)$ as shown in Fig. 1. The center of mass is at the origin (0,0) with $\mathbf{p}_C$ along the positive x-axis, and structure remains in the xy-plane. The equations are solved by *ode45* in Matlab with an initial time step size 1e-18s.

Two datasets of initial geometries are generated: the first one has geometries with bond lengths r_{co}, r_{cs} and bond angle θ_{ocs} peaks at approximately 150pm,180pm and 160° respectively; and the second one with peaks at equilibrium structure (116pm,156pm,175°) [14]. Initially, the first dataset is the primary focus, as shown in Fig. 2, which contains more molecules that are stretched and bent compared to the equilibrium structure. The study [10] demonstrated that a small molecule such as CO_2 geometry can show significant stretching and bending in longer femtosecond pulses (30fs and longer). And it is important to consider geometries away from equilibrium, therefore we have chosen peak bond lengths exceeding equilibrium values and a peak bond angle smaller than the equilibrium value. We allow molecules to compress, stretch, and bend within the following ranges: $r_{co} = (0, 300\text{pm})$, $r_{cs} = (0, 350\text{pm})$ and $\theta_{ocs} = (140°, 180°)$. Then, the corresponding momentum vectors can be obtained by applying the Coulomb explosion algorithm.

The peaks of the distributions for r_{co}, r_{cs} and θ_{ocs} are at 149.740 pm, 177.641 pm, and 161.565° respectively. The average values are 150.033pm, 179.484pm, and 159.915°. In Fig. 3. Geometry plot with marginal distributions for original simulated geometries. The red, black and blue area represent oxygen, carbon and sulfur respectively., the nuclear distribution of each atom is displayed intuitively in 2D. To generate this plot, first, position carbon atom at the origin (0,0). Then, position the oxygen atom and the sulfur atom in such a way that θ_{ocs} is evenly divided by the y-axis and the oxygen atom is placed in the second quadrant, while the sulfur is placed in the first quadrant. Finally, adjust the molecule's coordinates to ensure that the center of mass is at the origin.
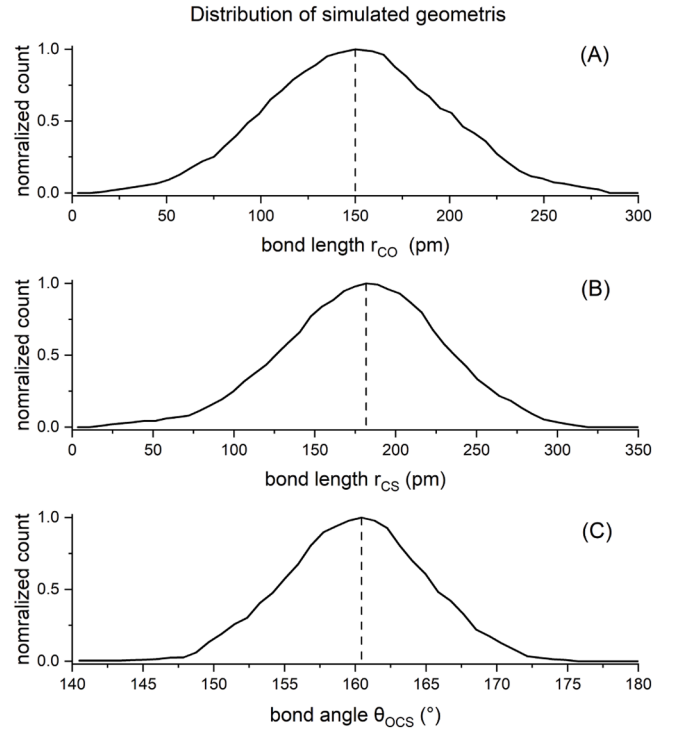
Fig. 2. Distributions of r_{co}, r_{cs} and θ_{ocs} from simulated OCS geometries.

3. Geometry reconstruction using a look-up table

The concept behind a look-up table is relatively straightforward: it involves simulating a certain number of Coulomb explosions with a range of molecular structures. This creates a mapping from molecular structures to momentum vectors. As a result of storing results from numerous simulations, a look-up table of asymptotic momentum vectors
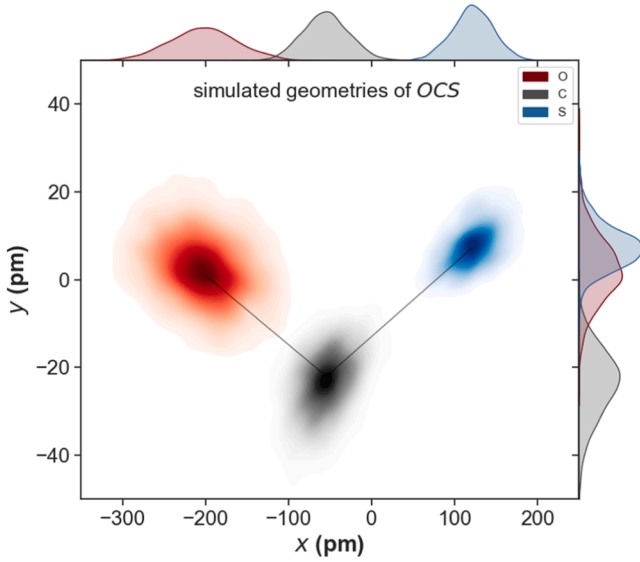
Fig. 3. Geometry plot with marginal distributions for original simulated geometries. The red, black and blue area represent oxygen, carbon and sulfur respectively.

is generated. This table can be correlated with any experimental measurement. To identify the structure corresponding to a particular set of observed momentum vectors, the table can be read in reverse. This involves searching for the momentum vectors that best match the observed set and subsequently deducing the corresponding structure. The degree of agreement is determined by calculating the squared norm between the observed and table vectors. The minimum value search algorithm used in this look-up table method is a linear search. This search goes through the entire table to find the entry with the minimum value of the squared norm.

The look-up table has two parts: the table of original geometry values $(r_{co}, r_{cs}, \theta_{ocs})$ and final momentum vectors $\boldsymbol{p} = (p_{ox}, p_{oy}, 0, p_{cx}, 0, 0, p_{sx}, p_{sy}, 0)$, expressed in the molecular frame. For effective comparison with this format, modifications are essential to the experimental momentum data. The sum of the momentum vectors of O, C and S from a single event has a very small number less than $5*10^{-23}$ kgm/s [15]. First, remove the motion of the center of mass, ensuring zero total momentum. Then, follow the same steps as mentioned in Section 2 to rotate momentum vectors into an xy-plane with the same orientation. This process reduces the momentum vector to 5-dimension without losing information of interest. To reconstruct the geometry corresponding to a set of molecular frame experimental momentum values the following steps are carried out.

(1) Generate a look-up table of 10,000,000 events, with $(r_{co}, r_{cs}, \theta_{ocs})$ uniformly distributed within the ranges (10pm to 500pm, 10pm to 500pm, 140° to 180°).
(2) Adjust the experimental momenta to match the format:

$\boldsymbol{p}_{exp} = (p_{e\_ox}, p_{e\_oy}, 0, p_{e\_cx}, 0, 0, p_{e\_sx}, p_{e\_sy}, 0)$, where $\boldsymbol{p}_{exp}$ represents the experimental data.

(3) For each experimental momentum, compare it with every reference momentum, $\boldsymbol{p}$ from the look-up table. Calculate the square norm of the difference as $\| p_{exp} - p \|^2$.
(4) Identify the smallest difference (error). If this error is below $1*10^{-46}$ (kgm/s)^2, indicating an error for momentum component P_i within $\pm 1*10^{-23}$ kgm/s, then recognize the corresponding geometry from the table as the reconstructed geometry.

(5) If no matching geometry can be found, the reconstruction for this $\boldsymbol{p}_{exp}$ is unsuccessful and the algorithm moves to the next one.

Here, $\boldsymbol{p}_{exp}$ represents the momentum vectors that generated in Section 2. For a total of 2000 momentum vectors, 1993 of them get reconstructed geometries, while 7 failed. Among the 1993 geometries, the peak values for r_{co}, r_{cs} and θ_{ocs} are 152.286pm, 179.715pm, and 161.390° respectively. The average values are 155.687pm, 194.824pm, and 162.113°. Compared to the simulated geometries, the average bond length r_{cs} of 194.824pm is much larger than the expected value of 179.484pm, similarly, the average bond length r_{co} and θ_{ocs} also exhibit deviations from the anticipated values. As shown in Fig. 4. Geometry plot with marginal distributions for reconstructed geometries by the look-up table method. The red, black and blue area represent oxygen, carbon and sulfur respectively., the shapes of the distributions for O, C and S are different compared to the original distributions in Fig. 3, with significant wings on the oxygen and sulfur distributions. Among these, only 376 geometries are reconstructed within a tolerance of (± 1pm, ± 1pm, $\pm 1°$) when compared to the original geometries. This means that the remaining 1617 reconstructed geometries either have low accuracy or failed to produce the correct answer. The estimated time of reconstruction for each momentum vector set is around 4.98s.

To understand the performance more quantitatively, the momentum difference M_d and the geometry difference G_d can be defined as:

$$M_d = \| p_{exp} - p_{solution} \|^2 = (p_{e\_O} - p_{s\_O})^2 + (p_{e\_C} - p_{s\_C})^2 + (p_{e\_S} - p_{s\_S})^2$$

$$G_d = \| G_{exp} - G_{solution} \|^2$$
$$= (r_{e\_CO} - r_{s\_CO})^2 + (r_{e\_CS} - r_{s\_CS})^2 + (r_{e\_SO} - p_{s\_SO})^2$$

M_d calculates the momentum difference in the same way as in the reconstruction process, while the G_d represents how far the reconstructed geometry deviates from the expected geometry. Fig. 5 displays the relation between M_d and G_d of the 1993 events. Even through the momentum difference being close to zero, there are many points with a large geometry difference. That is, the look-up table, which is based on the linear concept, is not a reliable method to accurately determining the geometry structures.
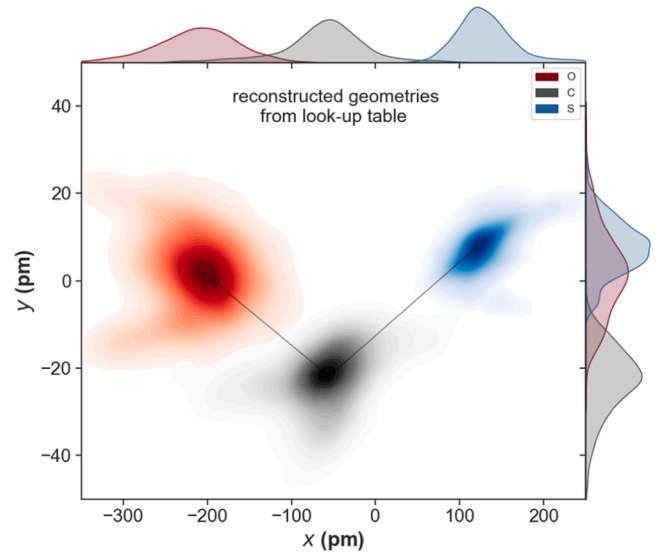
Fig. 4. Geometry plot with marginal distributions for reconstructed geometries by the look-up table method. The red, black and blue area represent oxygen, carbon and sulfur respectively.
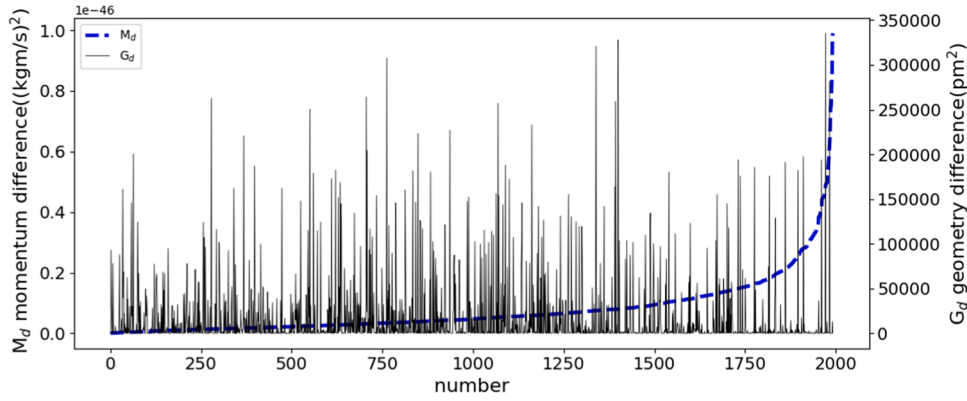
Fig. 5. Plot of Momentum difference (M$_d$) and geometry difference (G$_d$) vs. event IDs. The left y-axis shows momentum differences, the right y-axis shows geometry differences. The x-axis lists the event IDs in order of increasing momentum difference.

4. Geometry reconstruction using a constrained non-linear optimization method

Setting up the optimization problem is an essential step of the reconstruction process. The standard form of an optimization problem [16] is to minimize (or maximize) objective function $f(x)$, subject to constraint functions $g_i(x) \leq 0$, *where* $i \in \{1, ..., m\}$, and $h_j(x) = 0$, *where* $j \in \{1,...,n\}$. For our geometry reconstruction, it is regarded as a non-linear optimization problem. The goal is to minimize the objective function defined as $f(x) = |p(x) - p_{exp}|^2$. Here, $p(x)$ represents the momentum vectors iteratively calculated during the optimization process. The geometry structure x= (r$_{co}$, r$_{cs}$, θ_{ocs}) and momentum $p(x)$ satisfy the equality constraint functions $h_i(x)$, given by equations (1) - (4) in Section 2. And p_{exp} ($p_{experimental}$) represents the experimental momentum vectors. The inequality constrains, $g_i(x)$, define the box constraints (a $\leq$ r$_{co}$ $\leq$ b, c $\leq$ r$_{cs}$ $\leq$ d and e $\leq$ θ_{ocs} $\leq$ f). It is important to note that there is no restriction on dimensions of x and p. For larger molecules, more variables can be incorporated using similar steps. In our calculations, the 'fmincon' function in Matlab is utilized, employing the 'interior-point' method. This algorithm is well-suited for constrained nonlinear optimization.

The optimization algorithm begins with an initial starting point x$_0$= (r$_{co}$, r$_{cs}$, θ_{ocs}) and evaluates the objective function and constraints at this point. Once the initial point is valid, the algorithm proceeds with iterative updates to improve the solution. During each iteration, the algorithm updates the current value of the momentum difference and the current point (r$_{co}$, r$_{cs}$, θ_{ocs}) within the constrained region. Additionally, it calculates the gradient of the objective function with respective to (r$_{co}$, r$_{cs}$, θ_{ocs}), then determines the direction in which to search for the next point and computes an appropriate step size. The optimization process continues with iterations until convergence is achieved, with a convergence criterion set to a very small tolerance level of 1e-15, or until the maximum number of iterations (3000) is reached. To ensure the optimize the solutions and detect degenerate geometries, the entire process is repeated multiple times by a set of starting points within the constrained 3D space. These starting points can be uniformly distributed or follow any desired distribution, offering flexibility in the exploration of the solution space. In this case, there are a total of 125(5^3) starting points, which are uniformly selected from within the specified box range. To enhance the efficiency of the algorithm, iterations for each starting points are processed in a parallel fashion. If all starting points result in an identical solution, it indicates a unique solution, and the solution is considered fully optimized. However, if there are multiple solutions, it indicates there are multiple geometries corresponding to the same momentum vector. The algorithm automatically selects the solution with lowest value of $f(x)$ as the default, but it is important to note that all possible solutions are accessible for further analysis.

Computational steps:

(1) Set up the optimization problem:

Objective function: $f(x) = |p(x) - p_{exp}|^2$;
The Coulomb explosion model functions as described in Section 2;
Box constrains: 0 $\leq$ r$_{co}$ $\leq$ 500pm, 0 $\leq$ r$_{cs}$ $\leq$ 500pm and 140° $\leq$ θ_{ocs} $\leq$ 180°.

(2) Adjust the experimental momenta to match the format:

$p_{exp} = (p_{e\_ox}, p_{e\_oy}, 0, p_{e\_cx}, 0, 0, p_{e\_sx}, p_{e\_sy}, 0)$, where p_{exp} represents the experimental data.

(3) Initial guess points: uniformly select 125 points within the constrained space as the starting points.
(4) The preliminary result: the optimization algorithm begins at each starting point independently, each resulting in a solution.
(5) The final result: the results are classified based on the number of different solutions. Usually, results can be classified into one of three cases: a unique solution, a doubly degenerate solution, or a triply degenerate solution. For real experimental data, further analysis might be necessary for double degeneracy and triple degeneracy cases.

Here, p_{exp} represents the momentum vectors that generated in Section 2. For a total of 2000 momentum vectors, all of them produced reconstructed geometries as shown in Fig. 6. The peaks values for r$_{co}$, r$_{cs}$ and θ_{ocs} are 152.336pm, 178.814pm, and 161.65° respectively. The average values are 160.011pm, 175.540pm, and $\theta = 160.473$°. 1572 geometries are accurately reconstructed within a tolerance of ($\pm$0.001pm, $\pm$0.001pm, $\pm$0.001°) when compared to the original geometries. And 650 of them are identified as having a unique solution, 718 of them have double solutions, 632 of them have triple solutions. Therefore, 67.5% of them are related to the degeneracy problem, 32.5% are non-degeneracy cases. The estimated time of reconstruction for each momentum vector set is around 2.04 seconds.

5. Degeneracy region

In this section, we explore the possible existence of a degeneracy region, where degeneracy is more likely and requiring further analysis of the results. Starting with an example of degeneracy case as shown in Table 1.

For the test, a stretched and bent geometry is selected. The bond lengths r$_{co}$ and r$_{cs}$ are stretched to 127.5295 pm and 176.9708 pm, respectively. While the bond angle θ_{ocs} bends to 162.9222°. The momentum is calculated, resulting in an asymptotic momentum **p** of
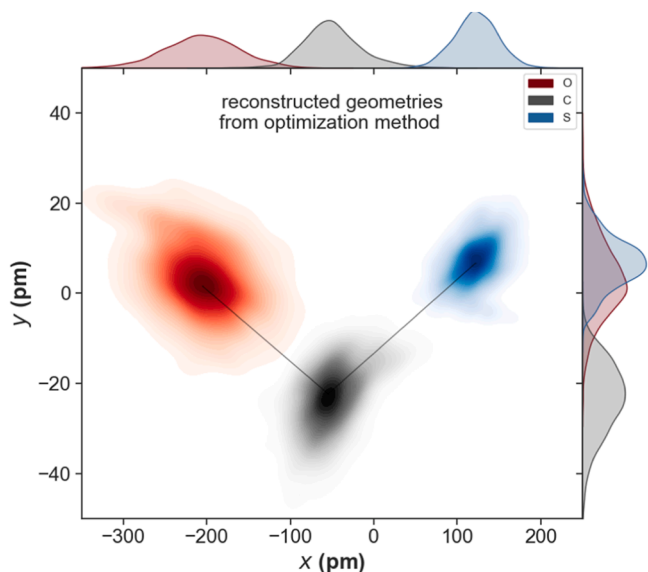
Fig. 6. Geometry plot with marginal distributions for reconstructed geometries by the optimization method. The red, black and blue area represent oxygen, carbon and sulfur respectively.

Table 1
Comparison of reconstruction results between the look-up table method and the constrained non-linear optimization method for a single test.

	Geometry (r_{co}, r_{cs}, θ_{ocs}) (pm, pm, degree)	Error $f(x)$ $(kgm/s)^2$
Original geometry	(127.5295, 176.970, 162.922)	
look-up method	(217.8787, 108.9899, 164.6464)	10^{-48}
optimization method	Geometry 1 = (127.530049, 176.969922, 162.91999)	$10^{-58.38}$
	Geometry 2 = (146.478598,151.771955, 161.696253)	$10^{-58.03}$
	Geometry 3 = (213.088342, 111.018173, 164.363248)	$10^{-57.95}$

(-3.765393e-24, -6.34491e-22, 0, 3.47457e-22, 0, 0, -3.43691e-22, 6.34491e-22, 0) kgm/s.

By employing the look-up table method, the reconstructed geometry is (217.8787 pm, 108.9899 pm, 164.6464°). The corresponding error value $f(x)$, representing the squared norm of the difference momentum, is 10^{-48}. However, when employing the constrained non-linear optimization process, three solutions are discovered as shown in Table 1. For a total of 125 starting points test, the most representative 3 calculations are selected to show the triple degeneracy. This indicates that for the given target momentum set, there exists three geometries, demonstrating a degeneracy of three. Notably, the error obtained from the optimization method is significantly smaller compared to the look-up table method. Additionally, the iteration plot. Fig. 7 shows that the optimization process converges after 60-80 iterations, reaching an error of 10^{-58}. In our previous study that proposed the simplex approach [7], only three initial points were used, and an error of 10^{-52} was achieved.

All geometries related to the degeneracy problem are identified within the space $100pm \leq r_{co}$ and $r_{cs} \leq 500pm$ and $140° \leq \theta \leq 180°$. As shown in Fig. 8(A), these geometries form a degeneracy region. This region is color-coded based on bond angle θ_{ocs}, where yellow represents more linear molecules, dark blue means more bent structures. A degeneracy region, in which geometries inside are considered degenerate cases, can be considered as a "danger" region for experimental results, such that if a result is found within this region it merits more attention. The degeneracy plot reveals that the degenerate cases are not randomly distributed in space. As Fig. 8 shows, the geometry of the "danger" region has a complicated structure. However, for simplicity, we could extract some general features such as a bond angle close to 160° is an indicator of degeneracy, and when one bond stretched much more than the other. Despite the complexity of the region, it would be possible for an AI to determine if an experimental result is inside, giving us an option to check the validity of the results. And, by plotting the reconstruction results in this multiple degeneracy region, we gain a visual understanding of how the degeneracy region correlates with the results presented in Section 4. In Fig. 8(B), blue points represent the 32.5% non-degeneracy cases, all of which are outside of the degeneracy region and the Red points in Fig. 8(C) represent the 67.5% of cases which are degenerate and located within the region.
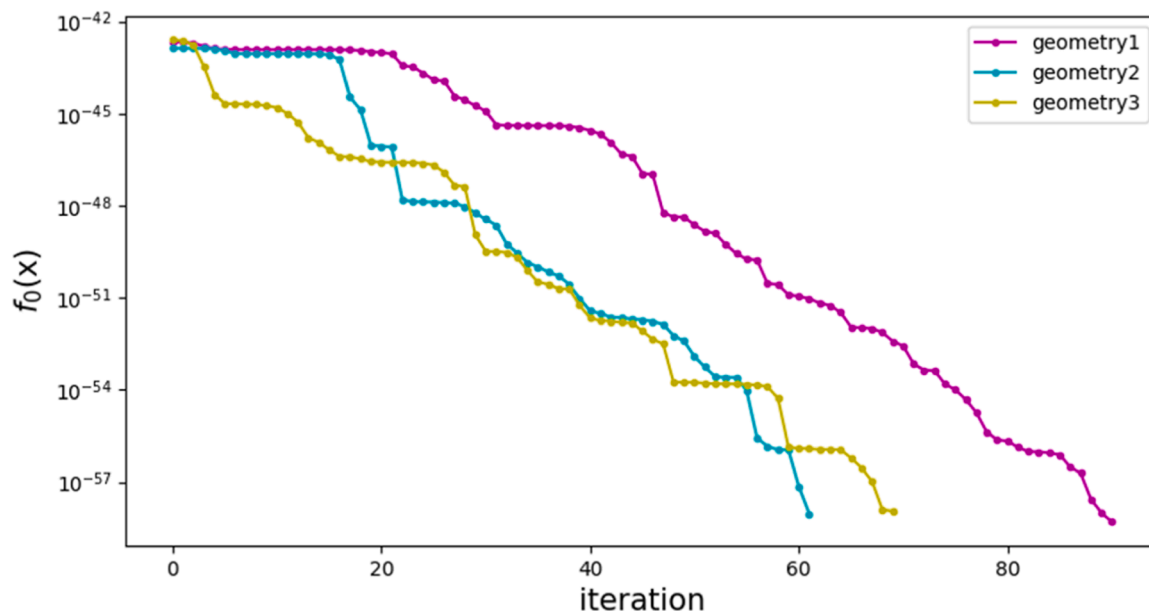
Fig. 7. Iteration plot (in log scale) for three initial starting geometries.
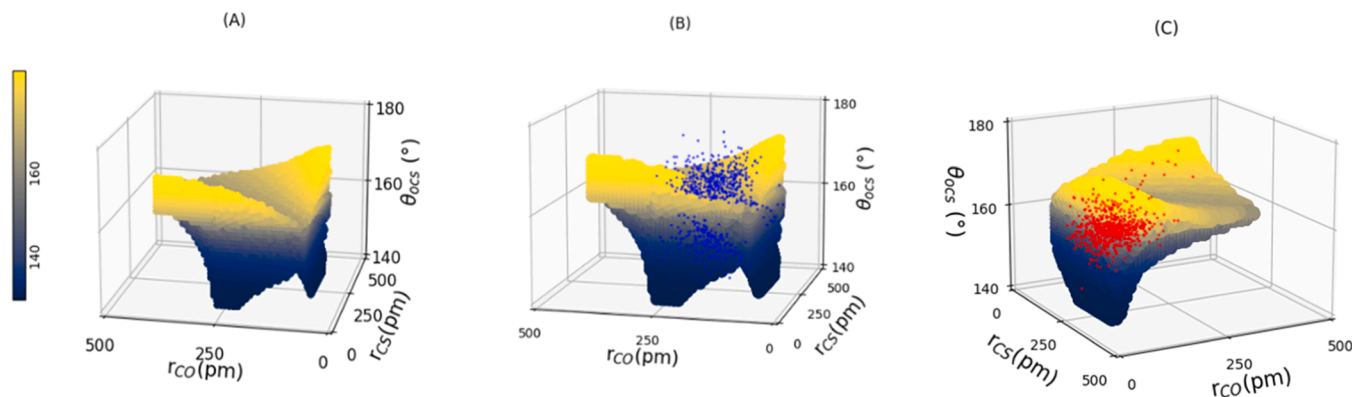
(A)   (B)   (C)

Fig. 8. Multiple degeneracy region. Yellow represents more linear molecules, dark blue means more bent structures. (A) the multiple degeneracy region. (B) blue points are non-degeneracy cases from Section 4. (C) red points are degeneracy cases in Section 4.

6. Comparison of look-up table method and constrained optimization method

We analyze the reconstruction results from Section 3 and 4. The table below presents a comparison of the reconstructed accuracy and speed between the look-up table method and the optimization method.

As Table 2 shows, the optimization method outperforms the look-up table method in terms of both accuracy and speed for the tests in Section 2 and 3. When considering the dataset of 2000 momentum vectors in Section 2, only 18.8% of the cases are successfully reconstructed within a tolerance of ($\pm$1pm, $\pm$1pm, $\pm$1°) using the look-up table method. The indication that a geometry is successful is typically an error less than10^{-46}, but seldom lower than 10^{-49}. However, with the optimization method, results are obtained for 100% the cases. Among them, 32.5% have a unique solution with a precision of ($\pm$0.001pm, $\pm$0.001pm, $\pm$0.001°), while the remaining 67.5% are degeneracy cases, where not only the solution found but other possible solutions are also discovered. Furthermore, the time required for each reconstruction using the look-up table method is more than twice that of the optimization method.

The precision of reconstruction is reliant on the resolution of the look-up table. To achieve greater precision, increasing the resolution is an option; however, that means a significant increase in storage space. For example, a look-up table which has a precision of 0.01 Å for bond lengths and 0.25° for bond angle needs over 3GB of storage space to store all the data in the range of r_{co} (10pm, 500pm), r_{cs} (10pm, 500pm), θ_{ocs} (140°, 180°). As the resolution increases, the size of the table expands exponentially. This exponential growth in table size also occurs as the target molecule gets larger, meaning more atoms are involved. Furthermore, a very large table size, in turn, would likely result in a less efficient reconstruction. It is important to note that the result found through the look-up table is limited to the values present in the table itself. We can only reconstruct geometries that exist in the table, and, since the data in the table is definitely discrete, it can not provide an exact solution to the reconstruction problem. The reconstruction speed is influenced by multiple factors, including the size of molecule, the size of look-up table and algorithms. For the test in this paper, the speed for the look-up table is 4.98s per reconstruction, while for the optimization method, it is 2s per reconstruction. The look-up table method requires

calculations for all geometries within a very large space and then works on the target momentum data set, whereas the optimization method operates only on the target momentum data set. Additionally, the table size exhibits exponential growth when increasing resolution, make it obviously time-consuming. So, the optimization method is faster than the look up table method. Overall, the constrained non-linear optimization method surpasses the look-up table method in aspects of accuracy, speed, precision, flexibility, and its ability to address the degeneracy cases and handle larger molecules.

7. Reconstruction for a second simulated dataset

Another dataset is generated with bond lengths r_{co}, r_{cs} and bond angle θ_{ocs} peaks at 115.710pm, 156.789pm, and θ_{ocs}=175.062°, close to the equilibrium structure. This dataset aims to capture the more typical behavior of molecules and avoids extreme cases of stretching, compression, or bend. This dataset is reconstructed using both the look-up table method and the optimization method. For a total of 200 momentum vectors, with the look-up table method, 57.6% of them get reconstructed geometries. However, only 21.5% of the cases are within the tolerance of ($\pm$1pm, $\pm$1pm, $\pm$1°). In contrast, when applying the constrained nonlinear optimization method, 100% of the geometries are reconstructed successfully, and all of them are within the stricter tolerance of ($\pm$0.001pm, $\pm$0.001pm, $\pm$0.001°).

Fig. 9 illustrates the r_{co}, r_{cs} and θ_{ocs} distributions obtained from both methods as well as the original geometries. And Table 3 shows the average values and peak values of them. The distributions obtained from the optimization method closely match the original geometry distributions, indicating high accuracy. However, the results from the look-up table method are different from the original geometries. The optimization method demonstrates its better accuracy in reconstructing geometries.

8. Conclusion

In our investigation, the comparison between the look-up table method and the constrained nonlinear optimization method have been conducted using simulated data from the asymmetric molecule OCS concerted explosion process. The look up table method is simple to implement, relies on a predefined table and exhibits some ability to perform reconstruction for non-degeneracy cases. However, it would encounter challenges when applied to larger molecules, as the required storage space increases exponentially with the number of atoms in a molecule, making it a less efficient method. The optimization method performs better. By providing a large set of starting points as the initial guess solutions and carrying out the iteration processes in parallel to identify the multiple convergent results, solutions can be found quickly. One other notable benefit of the optimization method is that it can

Table 2
Comparison of reconstruction accuracy and speed between two methods.

	Accuracy (pm, pm, degree)	Time per test	Identify degeneracy
Look-up table optimization method	($\pm$1, $\pm$1, $\pm$1) ($\pm$0.001, $\pm$0.001, $\pm$0.001)	4.98s 2.04s	No Yes
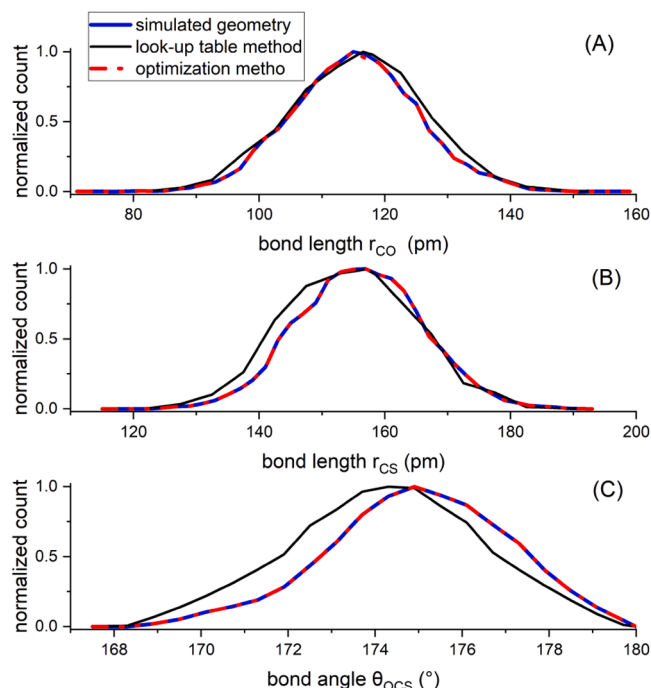
Fig. 9. Bond length and bond angle distribution for original geometries and reconstructed results

Table 3

Comparison of average values and peak values of distributions for original geometries and reconstructed results.

	Average geometry (unit in pm, degree)	Most likely geometry
Original geometry	$r_{co} = 115.238, r_{cs} = 156.046,$ $\theta = 174.989$	$r_{co} = 115.710, r_{cs} = 156.789,$ $\theta = 175.062$
Look-up table	$r_{co} = 116.685, r_{cs} = 155.300,$ $\theta = 174.300$	$r_{co} = 116.277, r_{cs} = 156.392,$ $\theta = 174.398$
Optimization method	$r_{co} = 115.238, r_{cs} = 156.046,$ $\theta = 174.989$	$r_{co} = 115.710, r_{cs} = 156.789,$ $\theta = 175.062$

indicate if the given momentum vector set is related to more than one geometric structure. In addition, all possible molecular structures found by this method are accessible for further analysis to determine the most appropriate solution. For the triatomic molecule OCS, a degeneracy region in 3D space is defined, where degeneracies are most probable. This allows for instant feedback and visualization of whether a derived geometry distribution is trustworthy throughout or if certain results should potentially be rejected. Alternatively, this could be used as a starting point of the automated (e.g., AI driven) filtering process which could impose conditions on geometry variation between nondegenerate values, such as smoothly varying parameters, in order to choose the most likely molecular structure. These advantages make the constrained optimization geometry reconstruction method a valuable approach for studying and understanding molecular systems using Coulomb explosion.

This study has demonstrated the abilities of two geometry reconstruction methods that map between momentum and geometry space based on the simple Coulomb explosion model. The model assumes that once electrons are removed and bonds break, ions move due to Coulomb repulsion. Ignoring affection from electrons, bonds, and external electric field. And the kinetic energy released does not only come from the Coulombic potential. Therefore, the reconstruction methods mainly limited to the concerted process and when the released kinetic energy is close to the expected Colomb potential energy. In future, the simulation model can be improved to reconstruct more accurate geometries.

Another future work will incorporate the optimization approach together with different models of ionization. Work to apply the optimization method to real experimental data for asymmetrical molecule OCS, and N_2O is under way.

CRediT authorship contribution statement

Kaili Tian: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis. **Ali Ramadhan:** Software, Methodology, Formal analysis. **Marcel Nooijen:** Writing – review & editing, Investigation, Conceptualization. **Stefan V. Pantazi:** Investigation, Conceptualization. **Reza Karimi:** Data curation. **Joseph H. Sanderson:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Z. Vager, R. Naaman, E.P. Kanter, Coulomb Explosion Imaging of Small Molecules, Science 244 (4903) (1989) 426–431, https://doi.org/10.1126/science.244.4903.426.

[2] R. Dörner, V. Mergel, O. Jagutzki, L. Spielberger, J. Ullrich, R. Moshammer, H. Schmidt-Böcking, Cold Target Recoil Ion Momentum Spectroscopy: a "momentum microscope" to view atomic collision dynamics, Physics Reports 330 (2-3) (2000) 95–192, https://doi.org/10.1016/s0370-1573(99)00109-x.

[3] H. Ibrahim, B. Wales, S. Beaulieu, B.E. Schmidt, N. Thiré, E.P. Fowe, É. Bisson, C. T. Hebeisen, V. Wanie, M. Giguére, J.-C. Kieffer, M. Spanner, A.D. Bandrauk, J. Sanderson, M.S. Schuurman, F. Légaré, Tabletop imaging of structural evolutions in chemical reactions demonstrated for the acetylene cation, Nature Communications 5 (1) (2014), https://doi.org/10.1038/ncomms5422.

[4] R. Boll, J. Schäfer, B. Richard, K. Fehre, Gregor Kastirke, Z. Jurek, M.S. Schöffler, Malik Muhammad Abdullah, N. Anders, T. Baumann, S. Eckart, B. Erk, A. De Fanis, R. Doerner, S. Grundmann, Patrik Grychtol, A. Hartung, M. Hofmann, M. Ilchen, Ludger Inhester, X-ray multiphoton-induced Coulomb explosion images complex single molecules, Nature Physics 18 (4) (2022) 423–428, https://doi.org/10.1038/s41567-022-01507-0.

[5] F. Légaré, K.F. Lee, I. Litvinyuk, P.W. Dooley, S.S. Wesolowski, P.R. Bunker, P. Dombi, F. Krausz, A.D. Bandrauk, D.M. Villeneuve, P.B. Corkum, Laser Coulomb-explosion imaging of small molecules, Physical Review A 71 (1) (2005), https://doi.org/10.1103/physreva.71.013415.

[6] J. Gagnon, K.F. Lee, D.M. Rayner, P.B. Corkum, V.R. Bhardwaj, Coincidence imaging of polyatomic molecules via laser-induced Coulomb explosion, Journal of Physics B 41 (21) (2008) 215104, https://doi.org/10.1088/0953-4075/41/21/215104.

[7] J.-P. Brichta, A.N. Seaman, J.H. Sanderson, Ultrafast imaging of polyatomic molecules with simplex algorithm, Computer Physics Communications 180 (2) (2009) 197–200, https://doi.org/10.1016/j.cpc.2008.09.006.

[8] George B. Dantzig, Original of the simplex method, Technical Report 87-5 (1987).

[9] K.A. Bhat, A. Ahmed, Simplex method and non-linear programming, International Journal of Computational Science and Mathematics 4 (2012) 299–303.

[10] I. Bocharova, R. Karimi, E.F. Penka, J.-P. Brichta, P. Lassonde, X. Fu, J. Kieffer, A. D. Bandrauk, I. Litvinyuk, J. Sanderson, François Légaré, Charge Resonance Enhanced Ionization of CO_2 Probed by Laser Coulomb Explosion Imaging, Physical Review Letters (6) (2011) 107, https://doi.org/10.1103/physrevlett.107.063201.

[11] M. Kunitski, S. Zeller, J. Voigtsberger, A. Kalinin, H. Schmidt, M.S. Schöffler, A. Czasch, W. Schöllkopf, R.E. Grisenti, T. Jahnke, D. Blume, R. Dörner, Observation of the Efimov state of the helium trimer, Science 348 (6234) (2015) 551–555, https://doi.org/10.1126/science.aaa5601.

[12] A.M. Sayler, E. Eckner, J.A. McKenna, B.D. Esry, K.D. Carnes, I. Ben-Itzhak, G. G. Paulus, Nonunique and nonuniform mapping in few-body Coulomb-explosion

imaging, Physical Review 97 (3) (2018), https://doi.org/10.1103/physreva.97.033412.

[13] B. Wales, É. Bisson, R. Karimi, S. Beaulieu, A. Ramadhan, M. Giguère, Z. Long, W.-K. Liu, J.-C. Kieffer, F. Légaré, J. Sanderson, Coulomb imaging of the concerted and stepwise break up processes of OCS ions in intense femtosecond laser radiation, Journal of Electron Spectroscopy and Related Phenomena 195 (2014) 332–336, https://doi.org/10.1016/j.elspec.2014.05.003.

[14] B. Wales, Tomonori Motojima, J. Matsumoto, Z. Long, W. Liu, H. Shiromaru, J Sanderson, Multiple ionization and complete fragmentation of OCS by impact

with highly charged ions Ar^{4+} and Ar^{8+} at 15 keV q^{-1}, Journal of Physics B 45 (4) (2012), https://doi.org/10.1088/0953-4075/45/4/045205, 045205–045205.

[15] R. Karimi, et al., N_2O ionization and dissociation dynamics in intense femtosecond laser radiation, probed by systematic pulse length variation from 7 to 500 fs, J. Chem. Phys. 138 (2013), 204311. https://doi-org.proxy.lib.uwaterloo.ca/10.1063/1.4804653.

[16] Sukanta Nayak, Fundamentals of Optimization Techniques with Algorithms, Academic Press, New York, 2020, pp. 135–143.