



Recovery of undamaged electron-density maps in the presence of damage-induced partial coherence in single-particle imaging

Alexander Kozlov,^{a,*} Timur E. Gureyev,^{a,b,c,d} David M. Paganin,^d Andrew V. Martin,^{a,e} Carl Caleman^{f,g} and Harry M. Quiney^a

Received 4 March 2020
Accepted 25 September 2020

Keywords: radiation damage; XFELs; partial coherence; single-particle imaging.

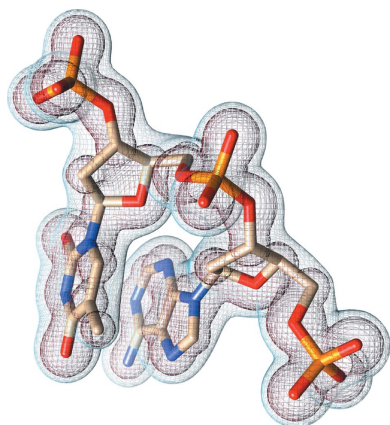
^aARC Centre of Excellence in Advanced Molecular Imaging, School of Physics, University of Melbourne, Parkville, Victoria 3010, Australia, ^bFaculty of Health Sciences, University of Sydney, Sydney, NSW 2006, Australia, ^cSchool of Science and Technology, University of New England, Armidale, NSW 2351, Australia, ^dSchool of Physics and Astronomy, Monash University, Clayton, Victoria 3800, Australia, ^eSchool of Physics, RMIT University, Melbourne, Victoria 3000, Australia, ^fDepartment of Physics and Astronomy, Uppsala University, PO Box 516, SE-751 20 Uppsala, Sweden, and ^gCenter for Free-Electron Laser Science, DESY, Notkestraße 85, DE-22607 Hamburg, Germany.
*Correspondence e-mail: okozlov@unimelb.edu.au

Resolving the electronic structure of single biological molecules in their native state was among the primary motivations behind X-ray free-electron lasers. The ultra-short pulses they produce can outrun the atomic motion induced by radiation damage, but the electronic structure of the sample is still significantly modified from its original state. This paper explores the decoherence of the scattered signal induced by temporal evolution of the electronic structure in the sample molecule. It is shown that the undamaged electron density of a single-molecule sample can often be retrieved using only the two most occupied modes from the coherent mode decomposition of the partially coherent diffraction fluence.

1. Introduction

Single-particle imaging (SPI) using X-ray free-electron lasers (XFELs) seeks to achieve atomic resolution ($\sim 1 \text{ \AA}$) in a reconstructed three-dimensional (3D) electron density of a non-crystalline sample, such as single biological molecules, obtained from a large number of two-dimensional (2D) X-ray diffraction patterns (Sayre *et al.*, 1998; Chapman *et al.*, 2006; Aquila *et al.*, 2015). These 2D diffraction patterns are collected in the far field (Fresnel number much smaller than unity) from randomly oriented copies of the sample, each of which is destroyed by the ultra-high X-ray flux (Neutze *et al.*, 2000). Elastic scattering in light elements of biological importance at X-ray photon energies $\hbar\omega \simeq 10 \text{ keV}$, needed to achieve atomic resolution, is relatively weak. In fact inelastic interactions such as photoionization dominate over the Thompson scattering at these X-ray energies. Photoionization starts a cascade of secondary ionization processes which together turn the sample into a highly charged molecular ion that expands rapidly due to Coulomb repulsion (Hau-Riege *et al.*, 2004; Fortmann-Grote *et al.*, 2017).

While radiation damage is a recognized limitation for SPI (Aquila *et al.*, 2015), a range of technical challenges currently constrain the achievable resolution (Chapman, 2019). These challenges include sample delivery, background scattering by a carrier medium, sample reproducibility and purity, low number of photons scattered by individual samples, and various sources of detection noise. For a more detailed discussion we refer the reader to Chapman (2019) and



references therein. In this work we focus on the following question, raised by Aquila *et al.* (2015): ‘Will it be possible to recover undamaged electron-density maps based on damaged sample data?’.

Alongside atomic motion, radiation damage induces electronic structure change in the sample during the pulse (Hau-Riege *et al.*, 2007a; Son *et al.*, 2011). The changes in electron structure, caused by photoionization (Neutze *et al.*, 2000) and a cascade of secondary ionization events during the pulse, occur on a sub-femtosecond timescale (Seibert *et al.*, 2010). In nanocrystalline samples the scattering from the damaged structure is suppressed by a breakdown in crystalline periodicity and a loss of bound electrons (Barty *et al.*, 2012; Coleman *et al.*, 2015b; Nass *et al.*, 2020). In single-molecule samples a reduction in the contrast of speckles in continuous diffraction fluence is caused by the damage-induced thermal motion of atoms (Martin *et al.*, 2015). If a molecule is confined to a buffer, for example a droplet, the atomic motion will be suppressed compared with a molecule in a vacuum (Hau-Riege *et al.*, 2007b). Temporal variation of the electronic structure leads to a loss of coherence in the scattered X-ray photons (Quiney & Nugent, 2011; Gorobtsov *et al.*, 2015; Martin & Quiney, 2016; Coleman *et al.*, 2020) and significantly modifies the atomic structure factors (Hau-Riege *et al.*, 2007a; Son *et al.*, 2011). This partial coherence can be described in terms of coherent modes (Mandel & Wolf, 1995; Quiney & Nugent, 2011), so that the partially coherent scattered intensity is represented as the sum of two or more coherent modes with a certain number of photons occupying each of them. Hence, electron-density maps recovered from a single coherent mode depict a damaged electron structure. Even if atomic motion is insignificant, the depicted electron structure is modified from its initial state. Below we use the term ‘damaged electron density’ in this sense. This damaged electron density tends towards a pulse-averaged distribution in the limit of low incident fluence (Martin & Quiney, 2016). An application of phase retrieval algorithms which assume full coherence and retrieve the phase of a single mode will therefore yield such a damaged electron density. Generalized phase retrieval algorithms capable of retrieving multiple coherent modes have been reported in the literature (Whitehead *et al.*, 2009; Quiney, 2010; Sala *et al.*, 2019).

In this paper we investigate the possibility of utilizing a sample electrodynamics theory combined with experimental measurements of partially coherent diffraction data to overcome the damage-induced electronic structure changes in the retrieved electron density of the sample. We demonstrate that a linear combination of retrieved coherent modes with modal weights obtained from the theory yields a reconstruction of local atomic electron densities with significantly suppressed radiation damage effects. This reconstruction requires input from a theoretical ionization-dynamics model which relies only on knowledge of the chemical composition of the sample without any structural information. Addition of a second mode significantly improves the reconstruction of undamaged electron density compared with that obtained with a single coherent mode, conventionally used under an assumption of

full coherence in the scattered X-ray intensity. The coherence of the scattered signal as a function of the pulse length and fluence is investigated for a model system of DNA origami. These artificial structures have recently been proposed for testing SPI using XFELs (Xavier & Chandrasekaran, 2018) and cryo-electron microscopy (Kopatz *et al.*, 2019). For typical contemporary XFEL pulse parameters, as many as 8% of photons can occupy modes other than the primary one. For the theory described below these photons serve as a source of useful structural information rather than decoherence, as would be the case for a conventional approach with an assumption of full coherence. We further discuss the inclusion of additional modes, beyond the first two, into the reconstruction of undamaged electron density.

2. Modal decomposition

In the context of SPI we consider the problem of reconstruction of the 3D electron-density distribution in a sample from the 3D intensity distribution of scattered X-rays. The latter is obtained by merging many 2D diffraction patterns after solving for sample orientations in each pattern. We assume that the incident X-ray pulse is a quasi-monochromatic plane wave propagating along the optic axis z with intensity distribution $I(x, y, 0, t) = I_{\text{in}}(t)$, which is spatially uniform in the object plane $z = 0$ but can vary in time and has a finite duration. Here, (x, y) denote coordinates in planes perpendicular to the optic axis, and t denotes time. The phase distribution of the incident pulse, $\phi(x, y, 0, t) \simeq \exp(-i\omega t)$, is assumed to be approximately flat over the pulse duration T within the area occupied by the sample in the object plane, where $\omega = 2\pi c/\lambda$, c is the speed of light in a vacuum and λ is the mean wavelength. We also assume that the scattering is weak, so that the standard first Born approximation can be applied to describe the diffracted intensity in a remote detector plane $z = R$. Under these assumptions, the integrated intensity of diffracted X-rays $W(\mathbf{q})$ in the far (Fraunhofer) field, $R \gg d^2/\lambda$, where d is the diameter of the sample, obtained by integrating the diffracted intensity $I_R(\mathbf{q}, t)$ over a single exposure (pulse length) time, can be expressed as

$$W(\mathbf{q}) = \int dt I_R(\mathbf{q}, t) = \frac{r_e^2}{R^2} \int dt I_{\text{in}}(t) |F(\mathbf{q}, t)|^2 + W_B(\mathbf{q}), \quad (1)$$

$$F(\mathbf{q}, t) = \int d\mathbf{r} \rho(\mathbf{r}, t) \exp(-i\mathbf{q} \cdot \mathbf{r}), \quad (2)$$

where the molecular form factor $F(\mathbf{q}, t)$ is the Fourier transform of the time-dependent molecular electron density $\rho(\mathbf{r}, t)$, r_e is the classical electron radius, and $W_B(\mathbf{q})$ is the background integrated intensity that contains both the Compton scattering term and other incoherent contributions (Lorenz *et al.*, 2012; Slowik *et al.*, 2014; Gorobtsov *et al.*, 2015). Below we shall assume that background subtraction has been carried out and set $W_B(\mathbf{q}) = 0$. In contrast with previous publications (Lunin *et al.*, 2015; Gureyev *et al.*, 2018), here we include plasma effects and consider entire atomic configurations (Son *et al.*, 2011) rather than individual atomic orbitals (Quiney, 2010; Lunin *et*

al., 2015) in the damage model during illumination by the incident X-ray pulse. Note that because of the stochastic character of radiation damage processes and the variability of the repeated illumination pulses, the right-hand side of equation (1) is a stochastic distribution; this is taken into account below.

We adopt an independent-atom model for the total molecular density (Quiney & Nugent, 2011; Son *et al.*, 2011), so that the total molecular electron density $\rho(\mathbf{r}, t)$ is given by the sum of the atomic electron densities $\rho_\alpha^{(\text{at})}(\mathbf{r})$ of all atoms in the molecule,

$$\rho(\mathbf{r}, t) = \sum_{\alpha m_Z} a_{\alpha m_Z}(t) \rho_{\alpha m_Z}^{(\text{at})}(\mathbf{r} - \mathbf{R}_{m_Z}). \quad (3)$$

Here $\alpha = \{Z, \gamma\}$, and $a_{\alpha m_Z}(t)$ represents the occupancy of the electronic configuration γ in the m th atom of chemical element Z , located at a position \mathbf{R}_{m_Z} in the molecule. For example, $\gamma = 1$ in carbon enumerates the configuration $1s^2 2s^2 2p^2$, $\gamma = 2$ stands for $1s^2 2s^2 2p^1$, and so on for all 27 configurations. This treatment of $a_{\alpha m_Z}(t)$ as entire configuration occupancies is in contrast to the original model (Quiney, 2010), which treated these quantities as the number of electrons in an orbital γ of atom Z . This allows us to account for the change in atomic orbitals between different configurations, which can be significant (Son *et al.*, 2011), especially for heavier elements (Kozlov & Quiney, 2019).

Substituting the electron density (3) into (1), we can write the intensity integrated over the X-ray pulse duration as

$$W(\mathbf{q}) = \sum_{\alpha\beta m_Z n_Z'} \Pi_{\alpha\beta m_Z n_Z'} f_\alpha^*(\mathbf{q}) f_\beta(\mathbf{q}) \exp[-i\mathbf{q} \cdot (\mathbf{R}_{n_Z'} - \mathbf{R}_{m_Z})], \quad (4)$$

$$\Pi_{\alpha\beta m_Z n_Z'} = \frac{r_e^2}{R^2} \int dt I_{\text{in}}(t) a_{\alpha m_Z}(t) a_{\beta n_Z'}(t), \quad (5)$$

where $f_\alpha(\mathbf{q}) \exp[-i\mathbf{q} \cdot \mathbf{R}_{m_Z}]$ is the Fourier transform of the atomic electron density $\rho_{\alpha m_Z}(\mathbf{r} - \mathbf{R}_{m_Z})$, so that $f_\alpha(\mathbf{q})$ is an atomic form factor. Here the temporal pulse variability and radiation damage effects are contained in the stochastic ‘damage matrix’ $\Pi_{\alpha\beta m_Z n_Z'}$. Note that the quantities $a_{\alpha m_Z}(t)$ themselves depend on the incident pulse intensity $I_{\text{in}}(t)$, as the time-dependent intensity drives the electron-density dynamics of atoms in the molecule. The XFEL pulse is assumed to be sufficiently short that the atomic positions \mathbf{R}_{m_Z} of atoms do not change during the pulse.

The diffracted intensity in equation (4) is a statistical function of repeated measurements. The mean value over the ensemble is equal to

$$\overline{W}(\mathbf{q}) = \sum_{\alpha\beta} P_{\alpha\beta} \hat{\rho}_\alpha^*(\mathbf{q}) \hat{\rho}_\beta(\mathbf{q}), \quad (6)$$

where $\hat{\rho}_\alpha(\mathbf{q}) = f_\alpha(\mathbf{q}) T_Z(\mathbf{q})$ and the structure function $T_Z(\mathbf{q}) = \sum_{m_Z} \exp[-i\mathbf{q} \cdot \mathbf{R}_{m_Z}]$ has been introduced. Note that $\hat{\rho}_\alpha(\mathbf{q})$ is the Fourier transform of the electron density of all atoms Z in the sample, each of which has an electronic configuration denoted by γ . Within the framework of the independent-atom model, the mean damage is independent of the position of

each atom in the sample (Hau-Riege *et al.*, 2004; Lorenz *et al.*, 2012), so that

$$P_{\alpha\beta} = \overline{\Pi}_{\alpha\beta m_Z n_Z'} = \frac{r_e^2}{R^2} \int dt \overline{I_{\text{in}}(t) a_\alpha(t) a_\beta(t)}, \quad (7)$$

where the atomic position in $a_\alpha(t)$ has been dropped, and the overline denotes the average value over an ensemble of damage scenarios. The independent-atom model assumes that, on average, each atom of chemical element Z , illuminated by incident X-rays, has the same charge density $\sum_\alpha a_\alpha(t) \rho_\alpha(\mathbf{r})$ at any time t .

The average integrated intensity (6) can be rewritten in diagonal form,

$$\overline{W}(\mathbf{q}) = \text{Tr}[\hat{\rho}^\dagger(\mathbf{q}) \mathbf{P} \hat{\rho}(\mathbf{q})] = \sum_k \eta_k |\langle \mathbf{v}_k, \hat{\rho}(\mathbf{q}) \rangle|^2, \quad (8)$$

where the vector $\hat{\rho}(\mathbf{q})$ has components $\hat{\rho}_\alpha(\mathbf{q})$, and \mathbf{v}_k is an eigenvector of the matrix $\mathbf{P}\mathbf{S}$, so that $\mathbf{P}\mathbf{S}\mathbf{v}_k = \eta_k \mathbf{v}_k$, and $(\mathbf{S})_{\alpha\beta} = \int d\mathbf{q} \hat{\rho}_\alpha^*(\mathbf{q}) \hat{\rho}_\beta(\mathbf{q})$ is the overlap matrix (see Appendix A for details). In the above equation the eigenvalues η_k are associated with the occupancies of coherent modes,

$$\psi_k(\mathbf{q}) = \langle \mathbf{v}_k, \hat{\rho}(\mathbf{q}) \rangle = \sum_\alpha v_{k\alpha} \hat{\rho}_\alpha(\mathbf{q}), \quad (9)$$

which are orthonormal, so that equation (8) takes the form of a modal decomposition (Mandel & Wolf, 1995; Quiney, 2010; Quiney & Nugent, 2011; Lorenz *et al.*, 2012; Lunin *et al.*, 2015) that reflects the coherence properties of the mean integrated intensity $\overline{W}(\mathbf{q})$. The mode occupancy η_k is proportional to the total number of photons \overline{N}_k in the mode $\psi_k(\mathbf{q})$. Indeed, the mode occupancy can be written as $\eta_k = \omega \overline{N}_k / \kappa$ (in atomic units, used henceforth in this article, the reduced Planck constant \hbar is set to unity and $\omega = 2\pi\nu$), where κ is the quantum efficiency of the detector (Goodman, 1985; Mandel & Wolf, 1995; Gureyev *et al.*, 2017).

Now consider the problem of extracting the sample electron density from the coherent modes. This can be achieved using a set of vectors $\mathbf{u}_k = \mathbf{S}\mathbf{v}_k$ which are bi-orthogonal to vectors \mathbf{v}_k : $\langle \mathbf{u}_k, \mathbf{v}_{k'} \rangle = \delta_{kk'}$. The vectors \mathbf{u}_k are eigenvectors of the matrix $\mathbf{S}\mathbf{P}$ so that

$$\mathbf{S}\mathbf{P}\mathbf{u}_k = \eta_k \mathbf{u}_k. \quad (10)$$

Expanding the vector $\hat{\rho}(\mathbf{q})$ we obtain

$$\hat{\rho}(\mathbf{q}) = \sum_k \mathbf{u}_k \psi_k(\mathbf{q}). \quad (11)$$

The undamaged electron density can be expressed as $\rho_1(\mathbf{r}) = \sum_{Z, m_Z} \rho_{Z\gamma=1}(\mathbf{r} - \mathbf{R}_{m_Z})$, where the index $\gamma = 1$ indicates the initial undamaged configuration of each atom. The Fourier transform of the undamaged electron density $\hat{\rho}_1(\mathbf{q}) = \sum_Z T_Z(\mathbf{q}) f_{Z\gamma=1}(\mathbf{q}) = \sum_{\alpha \in J_1} \hat{\rho}_\alpha(\mathbf{q})$, where $J_1 = \{Z, \gamma = 1\}$ is a subset of all electronic configurations that includes only those that correspond to undamaged atoms. Therefore,

$$\hat{\rho}_1(\mathbf{q}) = \sum_k \left(\sum_{\alpha \in J_1} u_{k\alpha} \right) \psi_k(\mathbf{q}). \quad (12)$$

Equation (12) expresses the undamaged electron density in the sample via the coherent modes that can be retrieved from experimental data, collected in the presence of radiation damage, using phase retrieval algorithms (Whitehead *et al.*, 2009; Quiney, 2010; Sala *et al.*, 2019) and the eigenvectors of the matrix \mathbf{SP} obtained from the theoretical model. Importantly, the first summation in equation (12) can always be truncated at a finite $k = M$, which corresponds to the singular-value decomposition (SVD) of the matrix \mathbf{SP} , $\mathbf{SPx} = \sum_k \eta_k \langle \mathbf{x}, \mathbf{v}_k \rangle \mathbf{u}_k$ (Golub & Van Loan, 1989). However, if the sum in equation (12) is truncated so that only $M < N$ modes are included, where N is the total number of modes, then

$$\sum_{k=1}^M \sum_{\alpha \in J_1, \beta} u_{k\alpha} v_{k\beta} = K_{\alpha\beta}^{(M)} \neq \delta_{\alpha\beta}. \quad (13)$$

This leads to imperfect reconstruction of the undamaged electron densities using equation (12). We introduce the normalization factor $\Omega^{(M)} = \sum_{\beta} K_{\alpha\beta}^{(M)} \neq 1$ for $M < N$, which rapidly converges to unity when several modes with the highest eigenvalues η_k are included in the summation. For $M = 2$ (while $N \simeq 10^3$), $\Omega^{(M)}$ differs from unity by no more than 40% for all the simulation parameters in Table 1 that we considered. To correct the reconstructed electron density, we introduce $\Omega^{(M)}$ into equation (12) and obtain

$$\hat{\rho}_1(\mathbf{q}) = \frac{1}{\Omega^{(M)}} \sum_{k=1}^M \left(\sum_{\alpha \in J_1} u_{k\alpha} \right) \psi_k(\mathbf{q}), \quad (14)$$

for the norm-adjusted undamaged electron-density reconstruction from the truncated sum over coherent modes. If the radiation damage is negligible, the integrated intensity $\overline{W}(\mathbf{q})$ is fully coherent and equation (14) becomes exact with just a single mode occupied by all scattered photons.

In summary, equation (14) demonstrates that the electronic structure corresponding to a sample that is unaffected by radiation damage can potentially be retrieved from only the first two coherent modes $\psi_k(\mathbf{q})$ and their occupancies η_k . All coefficients in equation (14) depend only on the chemical composition of the sample and can be obtained via a radiation damage simulation, while the coherent modes $\psi_k(\mathbf{q})$ contain information about the sample structure. These modes can in principle be retrieved from the 3D diffraction fluence, collected in the presence of radiation damage, and *a priori* information about the sample (Quiney, 2010; Quiney & Nugent, 2011; Curwood *et al.*, 2013). A different algorithm for solving a generalized phase retrieval problem (Thibault & Menzel, 2013) has been successfully used to retrieve ten coherent modes from the diffraction data collected in an XFEL imaging experiment (Sala *et al.*, 2019).

3. Radiation-damage dynamics model

Finding the undamaged electron density (14) from the coherent modes $\psi_k(\mathbf{q})$ given by equation (14) requires the expansion coefficients $u_{k\alpha}$ given by equation (10). The latter form eigenvectors \mathbf{u}_k of the matrix \mathbf{SP} , so finding them requires

Table 1

Percentages of the total number of photons \overline{N} scattered by the sample that occupy coherent modes $\psi_k(\mathbf{q})$.

Only values that correspond to the first ($k = 0$, top row for each pulse length) and second ($k = 1$, bottom row for each pulse length) coherent modes are depicted for a range of XFEL pulse lengths and fluences with an incident XFEL photon energy of 8 keV.

Pulse length (fs)	k	Pulse fluence ($\times 10^4$ J cm $^{-2}$)				
		1	10	100	318	1000
10	0	99.98	99.46	96.52	94.66	94.65
	1	0.020	0.53	3.34	5.07	5.07
15	0	99.96	99.30	95.49	93.44	93.62
	1	0.032	0.69	4.33	6.25	6.00
20	0	99.96	99.15	94.69	92.63	93.02
	1	0.044	0.84	5.10	7.03	6.57
25	0	99.94	99.01	94.07	92.07	92.61
	1	0.055	0.96	5.70	7.57	6.96
30	0	99.93	98.89	93.56	91.65	92.31
	1	0.066	1.08	6.18	7.97	7.24

knowledge of both the overlap matrix $(\mathbf{S})_{\alpha\beta} = \int d\mathbf{q} \hat{\rho}_\alpha^*(\mathbf{q}) \hat{\rho}_\beta(\mathbf{q})$ and the damage matrix \mathbf{P} . According to equation (7), the damage matrix depends on the electron-density occupancies $a_\alpha(t)$, which are determined by the ionization dynamics of the sample. This section describes finding these occupancies and modelling the ionization dynamics of the molecule imaged by an XFEL.

In the independent-atom model that was adopted in the derivation of the coherent mode expansion (6), the molecular electron density is approximated by those of individual atoms. Adopting the average-over-configuration description of electron configurations in atoms (Lindgren & Morrison, 1986; Son *et al.*, 2011; Kozlov & Quiney, 2019) and using Parseval's theorem, the overlap matrix can be written as

$$(\mathbf{S})_{Z\gamma Z'\gamma'} = 4\pi N_Z^2 \delta_{ZZ'} \int d\mathbf{q} q^2 f_{Z\gamma}(q) f_{Z'\gamma'}(q), \quad (15)$$

where N_Z is the number of atoms of element Z . The atomic scattering form factor $f_{Z\gamma}(q)$, introduced in equation (4), depends only on the absolute value of \mathbf{q} for a spherically symmetric atomic electron density. The overlap matrix (15) depends only on the atomic scattering form factors and is independent of the underlying molecular structure (Appendix A). The evolution of the atomic electron configuration occupancies can be described by a system of coupled rate equations (Hau-Riege *et al.*, 2004) which account for various radiation damage processes:

$$\frac{d}{dt} a_{Z\gamma}(t) = \sum_{\gamma'} R_{Z,\gamma' \rightarrow \gamma}(t) a_{Z\gamma'}(t) - R_{Z,\gamma \rightarrow \gamma'}(t) a_{Z\gamma}(t). \quad (16)$$

Here we have expanded the index $\alpha = \{Z\gamma\}$ to reflect the changes in occupancy of electronic configuration γ of element Z . The transition rate $R_{Z,\gamma' \rightarrow \gamma}(t)$ from configuration γ' to configuration γ in general can be time-dependent. For biomolecules illuminated by XFEL pulses longer than 10 fs it has been demonstrated (Gorobtsov *et al.*, 2015) that photoelectrons and secondary electrons, produced by electron

impact ionization and the Auger effect, make a substantial contribution to the dynamics of configuration occupancies $a_{Z\gamma}(t)$. We expanded the *AC4DC* atomic toolkit (Kozlov & Quiney, 2019; Kozlov *et al.*, 2020), which solves the atomic rate equations (16) with photoionization, Auger effect and fluorescence processes, to include electron impact ionization and three-body electron recombination. Here we modified the continuum model (Hau-Riege *et al.*, 2004; Gorobtsov *et al.*, 2015) to treat secondary electrons adiabatically as a Maxwellian plasma separately from the photoelectrons, which can collide with the former. The latter are described by two additional equations for their number and energy densities:

$$\frac{dN_s(t)}{dt} = \sum_{Z\gamma} n_Z a_{Z\gamma}(t) \sum_{\gamma'} R_{Z,\gamma \rightarrow \gamma'}^{(s)}, \quad (17)$$

$$\frac{dE_s(t)}{dt} = N_p(t) N_s(t) R_{\text{coll}} + \sum_{Z\gamma} n_Z a_{Z\gamma}(t) \sum_{\gamma'} e R_{Z,\gamma \rightarrow \gamma'}^{(s)}, \quad (18)$$

$$\frac{dN_p(t)}{dt} = -N_p(t) R_{\text{esc}} + \sum_{Z\gamma} n_Z a_{Z\gamma}(t) \sum_{\gamma'} R_{Z,\gamma \rightarrow \gamma'}^{(p)}, \quad (19)$$

$$\begin{aligned} \frac{dE_p(t)}{dt} = & -E_p(t) R_{\text{esc}} - N_p(t) N_s(t) R_{\text{coll}} \\ & + \sum_{Z\gamma} n_Z a_{Z\gamma}(t) \sum_{\gamma'} e R_{Z,\gamma \rightarrow \gamma'}^{(p)}. \end{aligned} \quad (20)$$

Here N_s , E_s and N_p , E_p are the number and energy density of the secondary and photoelectrons, respectively, n_Z is the number density of element Z , $R_{Z,\gamma \rightarrow \gamma'}^{(s)}$ include those of the above-mentioned atomic processes that release or capture secondary electrons, $e R_{Z,\gamma \rightarrow \gamma'}^{(s)}$ are the rates of the same processes weighted by their energy intake or release (see Appendix B for details), and $R_{Z,\gamma \rightarrow \gamma'}^{(p)}$ and $e R_{Z,\gamma \rightarrow \gamma'}^{(p)}$ indicate the respective rates for photoelectrons. The temperature of secondary electrons is $T(t) = 2E_s(t)/3N_s(t)$ (Boltzmann's constant is equal to unity in atomic units). The photoelectron energy distribution is approximated by a delta function $\delta[\epsilon_p - \bar{\epsilon}_p(t)]$, where the average photoelectron energy is $\bar{\epsilon}(t) = E_p(t)/N_p(t)$. This leads to the loss of a major part of the high-energy tail of the total electron partition function (Hau-Riege, 2013), where electrons can acquire energies in excess of the incident photon energy ω . To capture this effect, a more sophisticated model (Leonov *et al.*, 2014) is required. In our model, collisions between photoelectrons are ignored, but their collisions with secondary electrons are included via the rate R_{coll} in the weak-beam approximation (Kunc, 1989) where the number density of secondary electrons exceeds that of photoelectrons by at least an order of magnitude. This model assumes that the finite thermalization time of photoelectrons is longer than the pulse length. This is a valid assumption for X-ray photon energies in excess of 2 keV and typical XFEL pulse parameters (Hau-Riege, 2013). We do not consider Coulomb trapping of photoelectrons at the later stages of the pulse, which is justified if the incident X-rays are sufficiently energetic and photoelectron scattering leaves them with a substantial portion of their initial kinetic energy (Gorobtsov *et*

al., 2015). Secondary electrons, which include electrons produced via Auger decay and impact ionization, are assumed to be trapped at all times during the pulse. This results in a spike in the secondary electron temperature at the early stages of the pulse, when only a few energetic secondary electrons are present. These electrons are assumed to undergo rapid thermalization and are trapped by the net positive charge of the sample that is the result of escaping photoelectrons, described by the rate R_{esc} , which depends on the sample size (see Appendix B for more details).

To estimate the accuracy of our atomic and plasma simulation model, *AC4DC*, we compare its predictions with the non-local thermodynamic equilibrium (*NLTE*) approach (Scott, 2001). The latter was initially developed for the description of laboratory plasmas and later adapted to model sample dynamics illuminated by an XFEL pulse (Caleman *et al.*, 2015a; Jönsson *et al.*, 2015; Beyerlein *et al.*, 2018; Jönsson *et al.*, 2018). Two major differences between our approach and *NLTE* are the atomic structure description (*AC4DC* calculates electron transition rates for all possible ionic states) and the different dynamics of photoelectron thermalization (weak-beam approximation in *AC4DC*, instantaneous thermalization in *NLTE*). However, our model does not include continuum lowering that modifies the ionization potentials of atoms due to high-density electron plasma in the sample; this process is included in *NLTE*.

We compared average ionizations per atom of water and lysozyme under an XFEL pulse fluence ranging from 10^5 to 10^6 J cm⁻² and a pulse length of 20 fs with an X-ray photon energy of 6 to 9 keV (Jönsson *et al.*, 2018). For the shortest pulse length available in the *NLTE* implementation (Jönsson *et al.*, 2018) the difference in prediction between the two models for the average charge per atom in lysozyme and water is under 10%. The secondary electron temperature predicted by *AC4DC* was higher for all but the weakest incident fluences compared with the temperature predicted by *NLTE*. The reason for this discrepancy is an amplification of Auger decay in ions, included in *AC4DC*, and continuum lowering, accounted for in *NLTE*. Finally, the accuracy of the rate equation model presented above depends on the relative scales of the photoelectron thermalization time and the XFEL pulse length. As the photoelectron thermalization time increases for higher X-ray photon energies, the approximation of the free-electron energy distribution that *AC4DC* relies on becomes more accurate.

4. Results

At present, atomic resolution using XFEL sources remains a challenge. Imaging and reconstructing pre-designed DNA origami structures was proposed to guide the experimental and data analysis techniques required for the success of SPI (Xavier & Chandrasekaran, 2018). This approach was recently demonstrated in cryo-electron microscopy, where DNA origami was encapsulated into a capsid of simian vacuolating virus 40 (Kopatz *et al.*, 2019). These structures have a number of features that make them an attractive target for an SPI

experiment, such as high reproducibility and homogeneity, structural robustness under various conditions, a chemical composition similar to that of most proteins, and a high degree of control over their structure (Xavier & Chandrasekaran, 2018; Kopatz *et al.*, 2019). While DNA origami structures may appear semi-periodic in small-angle scattering, they can be designed to resemble single molecules at wide angles. Here we apply the theoretical framework described above and implemented in *AC4DC* to investigate the coherence properties of scattering by DNA origami. First, we use information about the structure of the sample to solve the forward problem and simulate the first two coherent modes. Next, we use only the chemical composition (without the structural information) of the sample and the XFEL pulse parameters to construct the overlap matrix **S** and damage matrix **P**. This enables us to find the coefficients in equation (14) and obtain an approximation to the undamaged electron-density map. We consider incident X-ray photons at 8 keV in XFEL pulses of Gaussian shape with FWHMs ranging from 10 to 30 fs and fluences between 10^4 and 10^7 J cm⁻². In our simulations we consider DNA origami that is spherical in shape with a radius of 20 nm, made up of equal proportions of adenine, guanine, cytosine and thymine nucleotides with a total number ratio of chemical elements of 39 (C):15 (N):24 (O):4 (P). The DNA strands occupy a volume equal to that of a cylinder with a radius of 1 nm and a length of 10 nm per 32 base pairs, with an additional 0.5 nm added to the radius to account for spacing between the strands (Rothemund, 2005).

Table 1 shows the fraction of photons in the first and second coherent modes given by equation (9) for the XFEL pulse parameters listed above.

We observed the largest decoherence for a 30 fs pulse with a fluence of 3.2×10^7 J cm⁻² (corresponding to a peak intensity of 1×10^{21} W cm⁻²). For these pulse parameters 91.7% of photons occupy the first mode and 8.0% occupy the second,

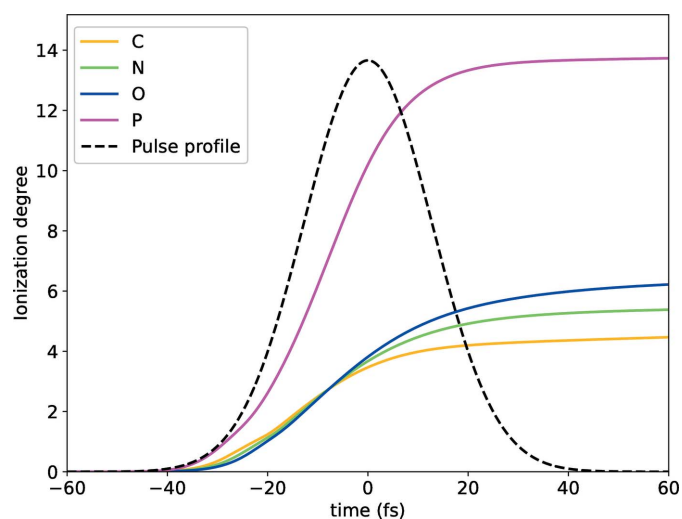


Figure 1
The average loss of bound electrons (ionization degree) for carbon, nitrogen, oxygen and phosphorus atoms in DNA origami, illuminated by a 30 fs XFEL pulse with a fluence of 3.2×10^7 J cm⁻².

with the remaining 0.3% occupying all other modes combined. Fig. 1 shows the ionization dynamics obtained by solving the system of equations (16) for atoms of the DNA origami exposed to a 30 fs XFEL pulse with a fluence of 3.2×10^7 J cm⁻².

As is evident from Fig. 1, by the time the pulse reaches its peak almost half the electrons are stripped from their atoms due to ionization. The reduction in the number of bound electrons can be observed in Fig. 2 for an electron-density map (dark-red mesh) obtained from a single mode. In the diffraction data this reduction in scattering power is somewhat compensated by an increase in the number of incident photons for a higher incident fluence. The effective atomic radial densities, the Fourier transform of which gives pulse-averaged atomic form factors, are plotted in Fig. 3 in red for the dominant mode where they were scaled by a factor $\Omega^{(M)}$ [see equation (13) and the subsequent description]. The radial atomic electron density as depicted by the dominant coherent mode (which in the low-damage limit represents a pulse-averaged structure) has a relatively higher depletion of valence-shell electrons compared with the *K* shell, although the photoionization cross section is much higher for the latter. This occurs due to an enhancement of *K*-shell hole recombination in Auger decay in ions (Son *et al.*, 2011) and a stronger suppression of the contribution to the total wide-angle scattering from atoms that have lost a *K*-shell electron compared with atoms that have lost an electron from a valence shell.

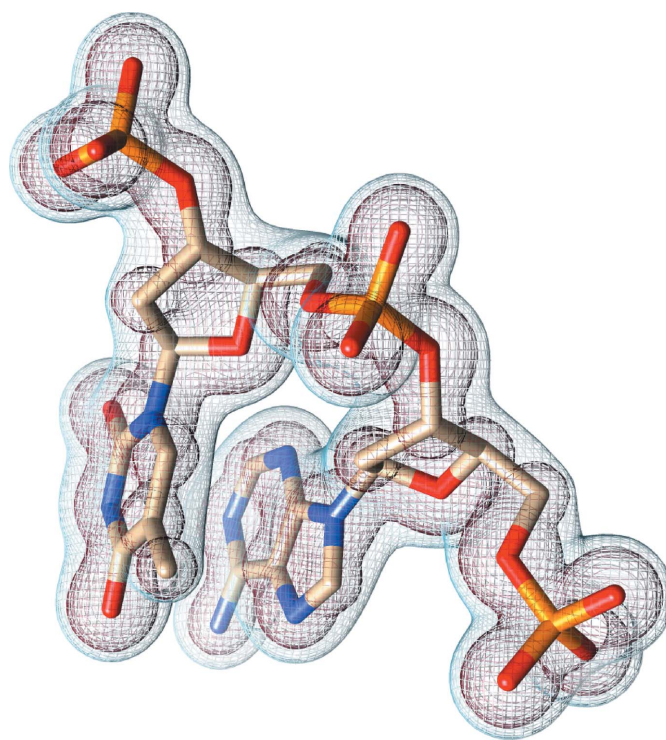


Figure 2
A simulation of the reconstructed electron density of a DNA fragment obtained from equation (14) using one (dark-red mesh) and two (pale-blue mesh) coherent modes. Values of constant electron density in both depicted surfaces are equal.

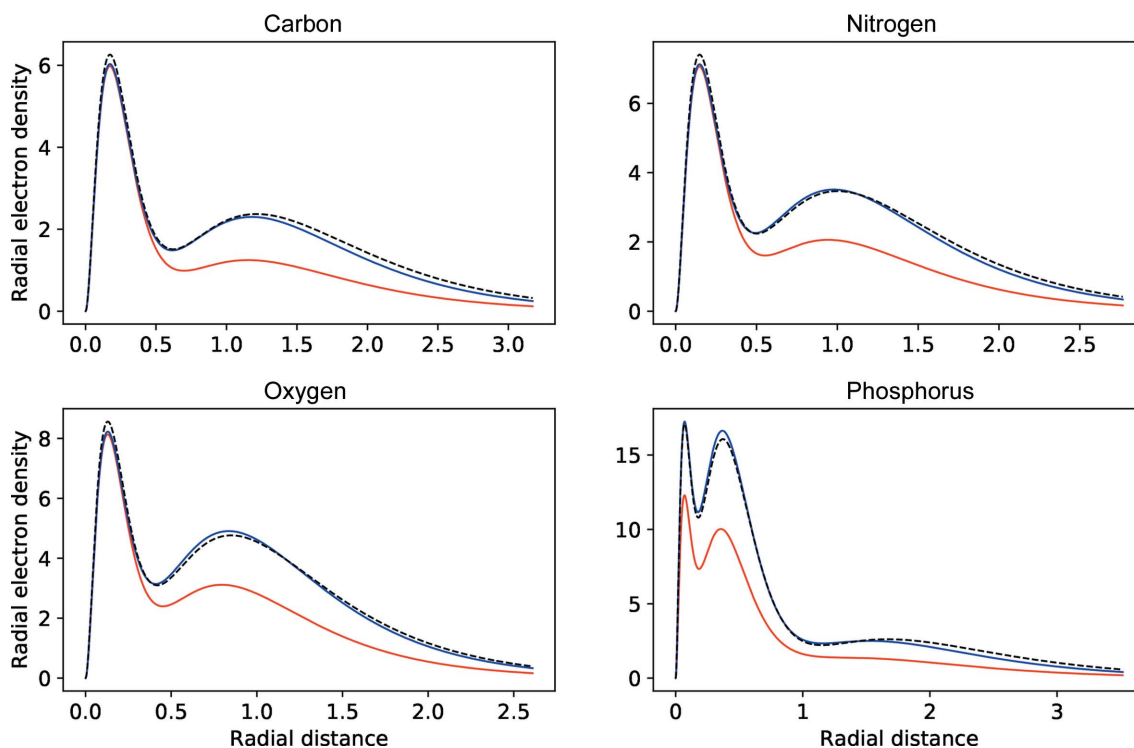


Figure 3 Radial atomic electron densities reconstructed using expansion (14) with two modes (blue lines) and one mode (red lines). Undamaged densities (black dashed lines) are shown for comparison. Radial distances from the atomic nucleus are expressed in Bohr radii, $a_B = 0.529 \text{ \AA}$, and the radial electron densities are in electrons per a_B .

5. Discussion and conclusions

Incorporating modes beyond the dominant mode, which can be obtained from partially coherent diffraction images, enables us to obtain a reconstruction of the sample electron density (14) that is improved compared with the pulse-averaged density affected by radiation damage. This approach assumes that the motion of atoms in a sample during an XFEL pulse is small compared with their size, which may be achieved if the imaged molecule resides in a droplet illuminated with a short pulse (Hau-Riege *et al.*, 2007b) but can be quite large otherwise (Fortmann-Grote *et al.*, 2017), especially for pulses well in excess of 30 fs (Martin *et al.*, 2015; Coleman *et al.*, 2015b). The possibility of extracting coherent modes beyond the dominant one depends on the degree of coherence and the signal-to-noise ratio (SNR) in the average integrated intensity $\overline{W}(\mathbf{q})$. Indeed, since the number of photons in the second mode does not rise beyond 8% in our simulations, we expect at least an inversely proportional reduction in the SNR in the retrieved second mode compared with the first dominant mode. However, a sufficiently high number of diffraction patterns can provide the required SNR and will be achievable in the near future with the progress of XFEL technology (Sobolev *et al.*, 2020). Our results, shown in Table 1, indicate that for higher-fluence pulses rather substantial radiation damage is imprinted into the electron density of the sample. In such XFEL regimes the atomic form factors need to be adjusted using radiation damage models to prevent structure reconstruction algorithms from confusing neutral atoms with

ions of different chemical elements. For higher pulse lengths and fluences we observe a relative reduction in the degree of coherence. For short pulses, higher coherence is observed due to the relative reduction in radiation damage associated with the contribution from secondary ionization processes (Hau-Riege *et al.*, 2004).

The impact of radiation damage on the reconstructed electronic structure of a fragment of DNA origami can be observed in Fig. 2. The electron density, depicted in dark red, was reconstructed using equation (14) from only the first coherent mode. The depletion of valence electrons observed in this figure is clearly visible when compared with the electron density reconstructed from both the first and second modes (light blue). Both reconstructions in Fig. 2 represent surfaces of constant electron density, the values of which are the same for both surfaces. The radial atomic electron densities plotted in Fig. 3 offer a deeper insight into the qualitative advantage of multi-mode electron-density reconstruction in the presence of radiation damage. The red and blue lines represent the 1D radial atomic electron densities used in the 3D electron-density surfaces of the corresponding colour in Fig. 2 within the framework of the independent-atom model. The two-mode reconstructions (blue) accurately reproduce the undamaged radial electron densities, while the one-mode reconstructions (red) clearly show the depletion of valence electrons for all atoms except phosphorus, for which a more dramatic depletion is observed for all electrons including those in the *K* shell (first peak in radial density). Note that, in general, a single coherent mode reconstruction becomes

dependent on the incident pulse parameters in the presence of significant radiation damage. This will introduce an additional complication into subsequent structure refinement. A two-mode reconstruction, on the other hand, is more robust to the choice of XFEL pulse parameters and yields a close approximation to the undamaged electron structure.

It is well known that elements as heavy as sulfur and phosphorus are ionized substantially faster by XFEL pulses than lighter elements of biological significance. The results in Fig. 1 indicate that phosphorus starts undergoing a rapid ionization several femtoseconds earlier than carbon, nitrogen and oxygen. Therefore, the radial electron density of phosphorus in Fig. 3, reconstructed using a single mode (red line), misses a larger fraction of initial electrons than lighter elements. Addition of the second mode into the reconstruction (blue line in Fig. 3) corrects phosphorus as well as the lighter elements. Therefore, the second coherent mode has an amplified contribution from heavier elements and a reduced contribution from lighter atoms compared with the first mode. In the context of single-particle imaging, this may have implications for imaging techniques that rely on anomalous diffraction from heavy elements, and this will be explored in future work.

APPENDIX A Coherent mode equations

In order to find the expression for coherent modes in (9) it is useful to consider the properties of the overlap matrix,

$$S_{\alpha\beta} = \int d\mathbf{q} \hat{\rho}_\alpha^*(\mathbf{q}) \hat{\rho}_\beta(\mathbf{q}) = \int d\mathbf{q} T_Z^*(\mathbf{q}) f_{Z\gamma}^*(q) f_{Z'\gamma'}(q) T_{Z'}(\mathbf{q}). \quad (21)$$

The product of the structure factor $T_Z(\mathbf{q})$ and atomic form factor $f_{Z\gamma}(q)$ is the Fourier transform of the sum of electron densities of atoms Z in a configuration γ ,

$$\hat{\rho}_\alpha(\mathbf{q}) = T_Z(\mathbf{q}) f_{Z\gamma}(q), \quad (22)$$

$$\rho_\alpha(\mathbf{r}) = \sum_{m_Z} \rho_\alpha^{(\text{at})}(\mathbf{r} - \mathbf{R}_{m_Z}). \quad (23)$$

Using Parseval's theorem, the overlap matrix can be rewritten as

$$\begin{aligned} S_{\alpha\beta} &= \sum_{m_Z n_{Z'}} \int d\mathbf{r} \rho_\alpha^{(\text{at})}(\mathbf{r} - \mathbf{R}_{m_Z}) \rho_\beta^{(\text{at})}(\mathbf{r} - \mathbf{R}_{n_{Z'}}) \\ &\simeq N_Z^2 \delta_{ZZ'} \int d\mathbf{r} \rho_\alpha^{(\text{at})}(\mathbf{r}) \rho_\beta^{(\text{at})}(\mathbf{r}). \end{aligned} \quad (24)$$

The last part of this expression was obtained by neglecting atomic density overlap. An assumption of radial symmetry of the atomic electron density $\rho_\alpha^{(\text{at})}(\mathbf{r})$, followed by application of Parseval's theorem, yields equation (15) for the overlap matrix. Any such matrix is Gramian (Horn & Johnson, 2013), and hence positive semi-definite, since $\langle \mathbf{S}\mathbf{x}, \mathbf{x} \rangle = \sum_{\alpha\beta} S_{\alpha\beta} x_\alpha x_\beta = \int d\mathbf{q} |\sum_\alpha x_\alpha \hat{\rho}_\alpha(\mathbf{q})|^2$ for any vector \mathbf{x} . Moreover, the overlap

matrix considered in the present paper is invertible, and therefore positive definite. Indeed, as we use a tight-binding model in which each atom of the sample is considered independent [equation (15)], the overlap matrix $(\mathbf{S})_{ZZ'}$ is always equal to zero when $Z \neq Z'$. This means that \mathbf{S} has a block-diagonal structure, with each block corresponding to a different chemical element Z . Thus, its determinant can be zero only if the determinant of one of these individual blocks is zero. The latter condition would mean that the overlap matrix for different configurations of one atom is degenerate, which would correspond to a linear dependence of the atomic electron densities $\rho_{Z\gamma}(\mathbf{r})$. This is never the case for a set of atomic electron densities that correspond to different orbital occupancies (electron configurations), since the difference in the Coulomb interaction between electron configurations lifts degeneracy and therefore prevents linear dependence. Note that, in the atomic model (Kozlov & Quiney, 2019) used in this paper, we average over the total angular momentum and its projections, treating all these states in the ensemble $\{Z\gamma\}$ as one state with an effective weight.

Let η_k and \mathbf{g}_k be the eigenvalues and the orthonormal eigenvectors of the symmetric matrix $\mathbf{Q} = \mathbf{S}^{1/2} \mathbf{P} \mathbf{S}^{1/2}$. Then for any vector \mathbf{x} we have $\mathbf{Q}\mathbf{x} = \sum_k \eta_k \langle \mathbf{x}, \mathbf{g}_k \rangle \mathbf{g}_k$. As $\mathbf{P} = \mathbf{S}^{-1/2} \mathbf{Q} \mathbf{S}^{-1/2}$, we obtain

$$\mathbf{P}\mathbf{x} = \sum_k \eta_k \langle \mathbf{x}, \mathbf{S}^{-1/2} \mathbf{g}_k \rangle \mathbf{S}^{-1/2} \mathbf{g}_k. \quad (25)$$

Like the matrix \mathbf{S} , the matrix \mathbf{P} is also positive semi-definite, but unlike \mathbf{S} , \mathbf{P} can be highly degenerate. In particular, in the absence of radiation damage, the rank of \mathbf{P} is equal to one, *i.e.* only one of the eigenvalues η_k is non-zero (in that case there is only one coherent mode, *i.e.* the diffracted wave is spatially fully coherent). Substituting equation (25) into the expression $\langle \mathbf{P}\boldsymbol{\rho}(\mathbf{q}), \boldsymbol{\rho}(\mathbf{q}) \rangle$ in equation (6), where $\boldsymbol{\rho}(\mathbf{q})$ is the vector with elements $\rho_\alpha(\mathbf{q})$, we obtain the following representation of the mean diffracted intensity in terms of coherent modes and their mean 'occupation numbers' (Mandel & Wolf, 1995; Quiney, 2010; Quiney & Nugent, 2011; Lorenz *et al.*, 2012):

$$\bar{W}(\mathbf{q}) = \sum_k \eta_k |\psi_k(\mathbf{q})|^2, \quad (26)$$

$$\psi_k(\mathbf{q}) = \langle \mathbf{S}^{-1/2} \mathbf{g}_k, \boldsymbol{\rho}(\mathbf{q}) \rangle. \quad (27)$$

The functions $\psi_k(\mathbf{q})$ are orthonormal in the sense that

$$\begin{aligned} &\int d\mathbf{q} \psi_k^*(\mathbf{q}) \psi_{k'}(\mathbf{q}) \\ &= \sum_{\alpha\beta} \int d\mathbf{q} (\mathbf{S}^{-1/2} \mathbf{g}_k)_\alpha \rho_\alpha(\mathbf{q}) (\mathbf{S}^{-1/2} \mathbf{g}_{k'})_\beta \rho_\beta(\mathbf{q}) \\ &= \langle \mathbf{S} \mathbf{S}^{-1/2} \mathbf{g}_k, \mathbf{S}^{-1/2} \mathbf{g}_{k'} \rangle \\ &= \delta_{kk'}, \end{aligned} \quad (28)$$

where we have used the orthonormality of vectors \mathbf{g}_k . It is also easy to verify that the vectors $\mathbf{v}_k = \mathbf{S}^{-1/2} \mathbf{g}_k$ are eigenvectors of the matrix $\mathbf{P}\mathbf{S}$, *i.e.* $\mathbf{P}\mathbf{S}\mathbf{v}_k = \eta_k \mathbf{v}_k$: the proof consists of substituting $\mathbf{g}_k = \mathbf{S}^{1/2} \mathbf{v}_k$ into the eigenvector equation $\mathbf{S}^{1/2} \mathbf{P} \mathbf{S}^{1/2} \mathbf{g}_k = \eta_k \mathbf{g}_k$ and multiplying both sides by $\mathbf{S}^{-1/2}$ from the left.

Therefore, the coherent modes (9) are the scalar products of the electron-density vector and the eigenvectors of the matrix **PS**.

APPENDIX B

Atomic and electron plasma rate equations

This appendix describes the details of the rate-equation model for the description of atomic and electron plasma dynamics in an XFEL pulse. The rates in the electron plasma equations (17)–(20) are written as

$$R_{Z,\gamma\rightarrow\gamma'}^{(s)} = R_{Z,\gamma\rightarrow\gamma'}^{(\text{Aug})} + N_s R_{Z,\gamma\rightarrow\gamma'}^{(\text{sEII})} + N_p R_{Z,\gamma\rightarrow\gamma'}^{(\text{pEII})} - N_s^2 R_{Z,\gamma\rightarrow\gamma'}^{(\text{sTBR})}, \quad (29)$$

$$eR_{Z,\gamma\rightarrow\gamma'}^{(s)} = \varepsilon_{Z,\gamma\gamma'} \left[R_{Z,\gamma\rightarrow\gamma'}^{(\text{Aug})} - N_s R_{Z,\gamma\rightarrow\gamma'}^{(\text{sEII})} + N_s^2 R_{Z,\gamma\rightarrow\gamma'}^{(\text{sTBR})} \right] + N_p W_{Z,\gamma\rightarrow\gamma'}^{(\text{pEII})}, \quad (30)$$

$$R_{Z,\gamma\rightarrow\gamma'}^{(p)} = R_{Z,\gamma\rightarrow\gamma'}^{(\text{PhI})}, \quad (31)$$

$$eR_{Z,\gamma\rightarrow\gamma'}^{(p)} = R_{Z,\gamma\rightarrow\gamma'}^{(\text{PhI})} (\omega - \varepsilon_{Z,\gamma\gamma'}) - N_p \left[\varepsilon_{Z,\gamma\gamma'} R_{Z,\gamma\rightarrow\gamma'}^{(\text{pEII})} + W_{Z,\gamma\rightarrow\gamma'}^{(\text{pEII})} \right], \quad (32)$$

where $R_{Z,\gamma\rightarrow\gamma'}^{(\text{Aug})}$ is the Auger rate for element Z , which transitions from the initial electronic configuration γ to the final configuration γ' , $R_{Z,\gamma\rightarrow\gamma'}^{(\text{PhI})}$ is the photoionization rate, ω is the photon energy and $\varepsilon_{\gamma\gamma'}$ is the ionization potential. The electron impact ionization (EII) and three-body recombination (TBR) rates are given by the following expressions:

$$R_{Z,\gamma\rightarrow\gamma'}^{(\text{sEII})} = 4\pi \int_0^\infty dp p^3 f(p) \sigma_{Z,\gamma\rightarrow\gamma'}^{(\text{EII})}(p), \quad (33)$$

$$R_{Z,\gamma\rightarrow\gamma'}^{(\text{sTBR})} = 4\pi [2\pi/T(t)]^{3/2} \exp(\varepsilon_{Z,\gamma\gamma'}) R_{Z,\gamma\rightarrow\gamma'}^{(\text{sEII})}, \quad (34)$$

$$R_{Z,\gamma\rightarrow\gamma'}^{(\text{pEII})} = \bar{v}_p(t) \sigma_{Z,\gamma\rightarrow\gamma'}^{(\text{EII})} [\bar{v}_p(t)], \quad (35)$$

$$W_{Z,\gamma\rightarrow\gamma'}^{(\text{pEII})} = \bar{v}_p(t) \int_0^{[\bar{v}(t) - \varepsilon_{Z,\gamma\gamma'}]^{1/2}} dW W \frac{d\sigma_{Z,\gamma\rightarrow\gamma'}^{(\text{EII})} [\bar{v}_p(t), W]}{dW}, \quad (36)$$

where $f(p) = [1/2\pi T(t)]^{3/2} \exp[-p^2/2T(t)]$ is the Maxwellian partition function for secondary electrons (normalized to unity), $\bar{v}_p(t) = [2\varepsilon(t)]^{1/2} = E_p(t)/N_p(t)$ is the average photoelectron speed, and $W_{Z,\gamma\rightarrow\gamma'}^{(\text{pEII})}$ represents the rate of kinetic energy transfer from the photoelectron to the secondary electron it creates when ionizing an element Z in the initial configuration γ . The three-body recombination rate $R_{Z,\gamma\rightarrow\gamma'}^{(\text{sTBR})}$ was obtained using the principle of detailed balance (Landau & Lifshitz, 2002; Hau-Riege, 2011). For the electron impact ionization cross sections $\sigma_{Z,\gamma\rightarrow\gamma'}^{(\text{EII})}(p)$ and the differential cross section $d\sigma_{Z,\gamma\rightarrow\gamma'}^{(\text{EII})}(p, W)/dW$ we used the binary-encounter Bethe model (Kim & Rudd, 1994).

The escape rate of photoelectrons from the sample is accounted for by the rate $R_{\text{esc}} = 3\bar{v}_p(t)/2r$ in equations (19) and (20), where r is the radius of a spherical sample and $\bar{v}_p(t) =$

$[2\bar{\varepsilon}_p(t)]^{1/2}$. In the derivation of R_{esc} it was assumed that the photoelectrons are distributed uniformly in the sample and have a random, but uniformly distributed, direction of their speed $\bar{v}_p(t)$.

Prior to numerical solution of the atomic (16) and plasma rate equations (17)–(20), all atomic configurations γ are calculated for all atoms Z . The atomic rates and the respective blocks of the overlap matrix are calculated and stored prior to solving the rate equations. During solution of the rate equations, at every time step the integrals in (33) and (36) are calculated using 13-point Gaussian quadrature. The two remaining rates, secondary electron recombination (34) and photoelectron impact ionization (35), are calculated at each time step as well, using the atomic parameter stored at the previous step.

Acknowledgements

Alexander Kozlov is grateful to Dr Adrian Mancuso and Dr Andrew Morgan for useful discussions.

Funding information

Funding for this research was provided by: the Australian Research Council through the Centre of Excellence for Advanced Molecular Imaging, the Swedish Foundation for Strategic Research, the Swedish Research Council (2018-00740), the Carl Trygger Foundation, and the Helmholtz Association through the Center for Free-Electron Laser Science at DESY.

References

- Aquila, A., Barty, A., Bostedt, C., Boutet, S., Carini, G., dePonte, D., Drell, P., Doniach, S., Downing, K. H., Earnest, T., Elmlund, H., Elser, V., Gühr, M., Hajdu, J., Hastings, J., Hau-Riege, S. P., Huang, Z., Lattman, E. E., Maia, F. R. N. C., Marchesini, S., Ourmazd, A., Pellegrini, C., Santra, R., Schlichting, I., Schroer, C., Spence, J. C. H., Vartanyants, I. A., Wakatsuki, S., Weis, W. I. & Williams, G. J. (2015). *Struct. Dyn.* **2**, 041701.
- Barty, A., Caleman, C., Aquila, A., Timneanu, N., Lomb, L., White, T. A., Andreasson, J., Arnlund, D., Bajt, S., Barends, T. R. M., Barthelmess, M., Bogan, M. J., Bostedt, C., Bozek, J. D., Coffee, R., Coppola, N., Davidsson, J., DePonte, D. P., Doak, R. B., Ekeberg, T., Elser, V., Epp, S. W., Erk, B., Fleckenstein, H., Foucar, L., Fromme, P., Graafsma, H., Gumprecht, L., Hajdu, J., Hampton, C. Y., Hartmann, R., Hartmann, A., Hauser, G., Hirsemann, H., Holl, P., Hunter, M. S., Johansson, L., Kassemeyer, S., Kimmel, N., Kirian, R. A., Liang, M., Maia, F. R. N. C., Malmerberg, E., Marchesini, S., Martin, A. Y., Nass, K., Neutze, R., Reich, C., Rolles, D., Rudek, B., Rudenko, A., Scott, H., Schlichting, I., Schulz, J., Seibert, M. M., Shoeman, R. L., Sierra, R. G., Soltau, H., Spence, J. C. H., Stellato, F., Stern, S., Strüder, L., Ullrich, J., Wang, X., Weidenspointner, G., Weierstall, U., Wunderer, C. B. & Chapman, H. N. (2012). *Nat. Photon.* **6**, 35–40.
- Beyerlein, K. R., Jönsson, H. O., Alonso-Mori, R., Aquila, A., Bajt, S., Barty, A., Bean, R., Koglin, J. E., Messerschmidt, M., Ragazzon, D., Sokaras, D., Williams, G. J., Hau-Riege, S., Boutet, S., Chapman, H. N., Timneanu, N. & Caleman, C. (2018). *Proc. Natl Acad. Sci. USA*, **115**, 5652–5657.
- Caleman, C., Hultdt, G., Maia, F. R. N. C., Ortiz, C., Parak, F. G., Hajdu, J., van der Spoel, D., Chapman, N. C. & Timneanu, N. (2015a). *ACS Nano*, **25**, 139–146.

- Caleman, C., Jares, F. Jr, Grånäs, O. & Martin, A. V. (2020). *Crystals*, **10**, 585.
- Caleman, C., Timneanu, N., Martin, A. V., Jönsson, H. O., Aquila, A., Barty, A., Scott, H. A., White, T. A. & Chapman, H. N. (2015*b*). *Opt. Express*, **23**, 1213.
- Chapman, H. N. (2019). *Annu. Rev. Biochem.* **88**, 35–58.
- Chapman, H. N., Barty, A., Marchesini, S., Noy, A., Hau-Riege, S. P., Cui, C., Howells, M. R., Rosen, R., He, H., Spence, J. C. H., Weierstall, U., Beetz, T., Jacobsen, C. & Shapiro, D. (2006). *J. Opt. Soc. Am. A*, **23**, 1179–1200.
- Curwood, E. K., Quiney, H. M. & Nugent, K. A. (2013). *Phys. Rev. A*, **87**, 053407.
- Fortmann-Grote, C., Buzmakov, A., Jurek, Z., Loh, N.-T. D., Samoylova, L., Santra, R., Schneidmiller, E. A., Tschentscher, T., Yakubov, S., Yoon, C. H., Yurkov, M. V., Ziaja-Motyka, B. & Mancuso, A. P. (2017). *IUCrJ*, **4**, 560–568.
- Golub, G. H. & Van Loan, C. F. (1989). *Matrix Computations*, 2nd ed. Baltimore, Maryland, USA: Johns Hopkins University Press.
- Goodman, J. W. (1985). *Statistical Optics*. Chichester: Wiley.
- Gorobtsov, O. Yu., Lorenz, U., Kabachnik, N. M. & Vartanyants, I. A. (2015). *Phys. Rev. E*, **91**, 062712.
- Gureyev, T. E., Kozlov, A., Nesterets, Ya. I., Paganin, D. M. & Quiney, H. M. (2017). *Sci. Rep.* **7**, 4542.
- Gureyev, T. E., Kozlov, A., Nesterets, Y. I., Paganin, D. M., Martin, A. V. & Quiney, H. M. (2018). *IUCrJ*, **5**, 716–726.
- Hau-Riege, S. P. (2011). *High-Intensity X-rays – Interaction with Matter*. Weinheim: Wiley-VCH.
- Hau-Riege, S. P. (2013). *Phys. Rev. E*, **87**, 053102.
- Hau-Riege, S. P., Boutet, S., Barty, A., Bajt, S., Bogan, M. J., Frank, M., Andreasson, J., Iwan, B., Seibert, M. M., Hajdu, J., Sakdinawat, A., Schulz, J., Treusch, R. & Chapman, H. N. (2007*a*). *Phys. Rev. Lett.* **104**, 064801.
- Hau-Riege, S. P., London, R. A., Chapman, H. N., Szoke, A. & Timneanu, N. (2007*b*). *Phys. Rev. Lett.* **98**, 198302.
- Hau-Riege, S. P., London, R. A. & Szoke, A. (2004). *Phys. Rev. E*, **69**, 051906.
- Horn, R. A. & Johnson, C. R. (2013). *Matrix Anal.* 2nd ed. Cambridge University Press.
- Jönsson, H. O., Östlin, C., Scott, H. A., Chapman, H. N., Aplin, S. J., Timneanu, N. & Caleman, C. (2018). *High Energy Density Physics*, **26**, 93–98.
- Jönsson, H. O., Timneanu, N., Östlin, C., Scott, H. A. & Caleman, C. (2015). *J. Synchrotron Rad.* **22**, 256–266.
- Kim, Y.-K. & Rudd, M. E. (1994). *Phys. Rev. A*, **50**, 3954–3967.
- Kopatz, I., Zalk, R., Levi-Kalisman, Y. & Zlotkin-Rivkin, E. (2019). *Nanoscale*, **11**, 10160.
- Kozlov, A., Martin, A. V. & Quiney, H. M. (2020). *Crystals*, **10**, 478.
- Kozlov, A. & Quiney, H. M. (2019). *Phys. Scr.* **94**, 075404.
- Kunc, J. A. (1989). *Phys. Rev. A*, **40**, 1507–1514.
- Landau, L. D. & Lifshitz, E. M. (2002). *Physical Kinetics*. Moscow: Fizmatlit.
- Leonov, A., Ksenzov, D., Benediktovitch, A., Feranchuk, I. & Pietsch, U. (2014). *IUCrJ*, **1**, 402–417.
- Lindgren, I. & Morrison, J. (1986). *Atomic Many-Body Theory*. Heidelberg: Springer-Verlag.
- Lorenz, U., Kabachnik, N. M., Weckert, E. & Vartanyants, I. A. (2012). *Phys. Rev. E*, **86**, 051911.
- Lunin, V. Y., Grum-Grzhimailo, A. N., Gryzlova, E. V., Sinitsyn, D. O., Petrova, T. E., Lunina, N. L., Balabaev, N. K., Tereshkina, K. B., Stepanov, A. S. & Krupyanskiy, Y. F. (2015). *Acta Cryst. D71*, 293–303.
- Mandel, L. & Wolf, E. (1995). *Optical Coherence and Quantum Optics*. Cambridge University Press.
- Martin, A. V., Corso, J. K., Caleman, C., Timneanu, N. & Quiney, H. M. (2015). *IUCrJ*, **2**, 661–674.
- Martin, A. V. & Quiney, H. M. (2016). *J. Phys. B At. Mol. Opt. Phys.* **49**, 244001.
- Nass, K., Gorel, A., Abdullah, M. M. V., Martin, A., Kloos, M., Marinelli, A., Aquila, A., Barends, T. R. M., Decker, F. J., Bruce Doak, R., Foucar, L., Hartmann, E., Hilpert, M., Hunter, M. S., Jurek, Z., Koglin, J. E., Kozlov, A., Lutman, A. A., Kovacs, G. N., Roome, C. M., Shoeman, R. L., Santra, R., Quiney, H. M., Ziaja, B., Boutet, S. & Schlichting, I. (2020). *Nat. Commun.* **11**, 1814.
- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature*, **406**, 752–757.
- Quiney, H. M. (2010). *J. Mod. Opt.* **57**, 1109–1149.
- Quiney, H. M. & Nugent, K. A. (2011). *Nat. Phys.* **7**, 142–146.
- Rothmund, P. W. K. (2005). *ICCAD-2005. IEEE/ACM International Conference on Computer-Aided Design*, pp. 471–478.
- Sala, S., Daurer, B. J., Odstrcil, M., Capotondi, F., Pedersoli, E., Hantke, M. F., Manfreda, M., Duane Loh, D. N., Thibault, P. & Maia, F. R. N. C. (2019). arXiv:1901.08970v2.
- Sayre, D., Chapman, H. N. & Miao, J. (1998). *Acta Cryst. A54*, 232–239.
- Scott, H. A. (2001). *J. Quant. Spectrosc. Radiat. Transfer*, **71**, 689–701.
- Seibert, M. S., Boutet, S., Svenda, M., Ekeberg, T., Maia, F. R. N. C., Bogan, M. J., Timneanu, N., Barty, A., Hau-Riege, S., Caleman, C., Frank, M., Benner, H., Lee, J. Y., Marchesini, S., Shaevitz, J. W., Fletcher, D. A., Bajt, S., Andersson, I., Chapman, H. N. & Hajdu, J. (2010). *J. Phys. B: At. Mol. Opt. Phys.* **43**, 194015.
- Slowik, J. M., Son, S.-K., Dixit, G., Jurek, Z. & Santra, R. (2014). *New J. Phys.* **16**, 073042.
- Sobolev, E., Zolotarev, S., Giewekemeyer, K., Bielecki, J., Okamoto, K., Reddy, H. K. N., Andreasson, J., Ayyer, K., Barak, I., Bari, S., Barty, A., Bean, R., Bobkov, S., Chapman, H. N., Chojnowski, G., Daurer, B. J., Dörner, K., Ekeberg, T., Flückiger, L., Galzitskaya, O., Gelisio, L., Hauf, S., Hogue, B. G., Horke, D. A., Hosseinizadeh, A., Ilyin, V., Jung, C., Kim, C., Kim, Y., Kirian, R. A., Kirkwood, H., Kulyk, O., Küpper, J., Letrun, R., Loh, N. D., Lorenzen, K., Messerschmidt, M., Mühlig, K., Ourmazd, A., Raab, N., Rode, A. V., Rose, M., Round, A., Sato, T., Schubert, R., Schwander, P., Sellberg, J. A., Sikorski, M., Silenzi, A., Song, C., Spence, J. C. H., Stern, S., Sztuk-Dambietz, J., Teslyuk, A., Timneanu, N., Trebbin, M., Uetrecht, C., Weinhausen, B., Williams, G. J., Xavier, P. L., Xu, C., Vartanyants, I. A., Lamzin, V. S., Mancuso, A. & Maia, F. R. N. C. (2020). *Commun. Phys.* **3**, 97.
- Son, S.-K., Young, L. & Santra, R. (2011). *Phys. Rev. A*, **83**, 033402.
- Thibault, P. & Menzel, A. (2013). *Nature*, **494**, 68–71.
- Whitehead, L. W., Williams, G. J., Quiney, H. M., Vine, D. J., Dilanian, R. A., Flewett, S., Nugent, K. A., Peele, A. G., Balaur, E. & McNulty, I. (2009). *Phys. Rev. Lett.* **103**, 243902.
- Xavier, P. L. & Chandrasekaran, A. R. (2018). *Nanotechnology*, **29**, 062001.