Determining electronic damage to biomolecular structures in x-ray free-electron-laser imaging experiments

E. K. Curwood, H. M. Quiney, and K. A. Nugent

ARC Centre of Excellence for Coherent X-ray Science, School of Physics, The University of Melbourne, Melbourne, Victoria, 3010, Australia

(Received 17 August 2012; revised manuscript received 14 December 2012; published 15 May 2013)

The use of femtosecond pulses produced by x-ray free-electron-laser (XFEL) sources to image the structures of biomolecules involves a competition between the elastic scattering of photons to form a diffraction pattern and the damage initiated by inelastic collisions with the target. Since the electron density of the biomolecule changes rapidly throughout its interaction with a femtosecond XFEL pulse, the diffraction process measured in "diffract and destroy" experiments is, at best, partially coherent. It has been established that a detailed knowledge of these electrodynamical processes may be used to ameliorate the effects of damage in diffractive imaging experiments. It is shown here that, subject to conventional assumptions about the nature of the interactions, it is possible to characterize the extent of electronic damage in biomolecular imaging experiments using XFELs and to use this information transferably across similar systems. We develop a physical model of the interaction of a coherent x-ray pulse with a molecular system that describes the dynamical electronic response of the molecule. The resulting insights open a way forward for the measurement of atomic processes in such systems.

DOI: 10.1103/PhysRevA.87.053407 PACS number(s): 33.80.Wz, 82.53.Ps, 33.90.+h, 42.25.Fx

I. INTRODUCTION

The determination of the structure of a biomolecule is a vital step in the process of understanding its function. Biomolecular structures drive the process of rational drug design and the development of pharmaceuticals for the treatment or prevention of disease. Of particular interest are membrane proteins, which typically sit astride the bilipid membranes that form barriers between cells and their environment. These biomolecules control the passage of ions and small molecules and regulate cellular function. X-ray crystallography is a widely used technique that allows for near-atomic resolution of biomolecular samples but it is difficult or impossible to form high-quality crystals of membrane proteins for analysis. Any advance in our ability to determine membrane protein structures without forming large crystals is likely to be a significant driver in the development of new pharmaceuticals.

Coherent diffractive imaging (CDI) [1] has been proposed as an effective replacement for crystallography in the determination of biomolecular structures because it does not require high-quality crystals. In CDI, a finite, noncrystalline sample is illuminated by a coherent light source and the scattered photons are collected by an area detector in the far field. The Fourier-transform mapping between the wave field leaving the vicinity of the scatterer and the wave field in the far-field diffraction plane is utilized to enable direct determination of a two-dimensional projection of the structure. The resolution attainable using CDI is limited by the largest angle to which scattered photons can be measured. Only the intensity of the wave field can be measured so the imaging process necessarily involves the restoration of the phase of the complex scattered wave field, typically using an iterative projective algorithm [2,3].

The interaction between x rays and the principal constituents of biological materials, involving the "low-Z" elements carbon, nitrogen, and oxygen, is weak. An x-ray source of immense brightness, such as an x-ray free-electron laser (XFEL), is required to cause sufficient high-angle scatter to enable the determination of molecular structures to atomic resolution. The use of femtosecond pulses produced by x-ray free-electron sources to determine the structures of biomolecules has been the subject of active research since the publication of a detailed theoretical study by Neutze *et al*. [4]. It was recognized in that article that the electrodynamical processes driven by the interaction of such an intense pulse with matter would inevitably destroy the sample. This has led to the "diffract and destroy" paradigm [5] involving multiple sample copies that guides current experimental design. The time that one might expect the nuclear positions to remain in their equilibrium configuration in such an experiment was also estimated in [4], setting limits on the coherence of the desired molecular structural information. This temporal window, of approximately 10 fs, enables diffraction measurements to reveal the dominant characteristics of the underlying electron distribution, which is localized to atomic positions. Extensive studies have also been performed on the use of tampers to delay the onset of the Coulomb explosion and to extend the time over which diffraction data can be collected [6–8].

Implicit in all these approaches are the assumptions that damage processes lead to irretrievable and, ultimately, catastrophic loss of structural information and that one must somehow design the experiment so that all forms of damage are reduced or eliminated while maximizing the diffracted signal by increasing the incident flux. These conflicts between the incidence of inelastic and elastic processes possess, however, the characteristics of a competition that cannot be won by brute force alone. The conventional atomic models by which the rates of these processes are estimated indicate that the rate of photoionization in first-row atoms exceeds that of Thomson scattering by a factor of 10 at 10 keV [9]. An experiment designed to determine a molecular structure by scattering x rays from the electron density of the target is actually dominated by photoabsorption events that trigger a secondary cascade of Auger decay and electron recapture processes that lead to further collisional ionization; photon scattering is actually one of the least-favored events in the

experiment. Nevertheless, it has already been shown that molecular structures can be determined from diffraction data under the proposed interaction conditions if due regard is given to these electrodynamic processes in the reconstruction algorithm [10].

Recent simulations have also demonstrated that the signal can be separated from the isotropic background due to the stochastic nature of the damage processes under realistic interaction conditions [11]. It has also recently been demonstrated experimentally [12,13] that the use of femtosecond x-ray pulses with nanocrystalline biomolecular samples can outrun both the electronic damage and the radiation-induced structural disorder that it causes because of the persistence of the ordered nuclear structure over the duration of the pulse. Conventional crystallographic structure determination is more robust than coherent diffractive imaging methods for aperiodic structures because it utilizes only a finite number of Bragg reflections as data that are used to fit trial molecular structures. Structure determination using CDI, however, utilizes a continuous diffraction pattern and depends for its success on a highly coherent wave field and the clear identification of the zeros of that wave field. Even if the pulse is short enough that there is no disorder to the nuclear structure, electronic damage undermines both of these fundamental assumptions when applied to very small nanocrystals or individual molecules. Electronic processes occur on attosecond time scales, so that the coupled system involving the electron density and the radiation field evolves more rapidly than the nuclear distribution.

Here we consider proposals to determine molecular structures using the diffract and destroy approach applied to single molecules. We assume that sufficient data have been collected to sample the diffraction pattern to a specified resolution and that a suitable classification, orientation, and averaging scheme has been implemented in order to construct a three-dimensional data set from two-dimensional projections of randomly oriented molecules [14,15]. The full diffraction volume can then be phased using iterative methods [3] and a complete molecular structure obtained if detailed knowledge of the electrodynamical processes is included in the recovery scheme [10]. We further assume no reliance on molecular replacement strategies, so that structural information is obtained directly from the scattering data.

In this paper we show that the recovery of molecular structures from such partially coherent scattering data can be made without reference to any specific model subject to a number of general assumptions about the nature of the scattering. It is shown that the general characteristics of the electrodynamical processes enable information obtained from the diffraction data involving a known structure to be transferred to a general description of the degree of partial coherence induced by the time-varying electron density of any molecule of similar mass and composition. This allows the three-dimensional structures of unknown molecules to be determined without relying solely on electrodynamical simulations in spite of the extensive electronic damage that they endure. We also show that quantitative information about the rates of these electrodynamical processes can be inferred from the diffraction data obtained in an XFEL imaging experiment.

II. CONSTRUCTION OF A SCATTERING MODEL

A. The electronic processes

Structural analysis is based on establishing a relationship between the electronic structure of a single biomolecule and the scattered far-field intensity under femtosecond XFEL pulse illumination within the diffract and destroy approach. The interaction of the pulse with the molecule causes inelastic events, the most dominant of which is photoionization. While any electron in the atom is a candidate for interaction, photons of the energies produced by the XFEL are almost certain to interact with the electrons of the inner shells of first-row atoms and produce a core hole.

Values for the photoionization cross section of atoms or ions may be obtained from online databases [16]. Using these cross sections, the rate for a photoionization event is easily calculated as $R_{\rm ph} = \sigma_{\rm ph} n_{\rm phot}$ where $n_{\rm phot}$ is the photon flux through the material and $\sigma_{\rm ph}$ is the cross section.

A secondary electronic process occurs in ionized atoms with $1s$ holes. Core-hole states occupy energetically unfavorable electronic configurations and in a very short time ($\sim$10 fs for carbon [17]), the atom relaxes. In the case of x rays of energy 8–10 keV incident on low-Z materials, the vast majority of relaxation (approximately 97% [18]) occurs via an Auger process. The rate of Auger emission is given by $R_{\rm Auger} = 1/\tau_{\rm Auger}$ where $\tau_{\rm Auger}$ is the Auger lifetime; these are available in published tables [17]. We also expect some electron recapture to occur late in an exposure as the molecule becomes more heavily ionized, but we expect this process to have little impact on diffraction data.

The occurrence of these processes renders the electronic occupancies of the atoms which constitute the biomolecule time-dependent quantities. Within an atomic superposition model, the calculation of the time-varying occupancies involves the solution of a set of coupled, linear differential equations. These equations can be written in the form [18]

$$\frac{dN_{i,j}}{dt} = \sum_{k\neq i, l\neq j}^{n} (R_{kl\rightarrow ij} N_{k,l} - R_{ij\rightarrow kl} N_{i,j}), \qquad (1)$$

where (i,j) denotes the state of an atom, i refers to the number of electrons in the $1s$ orbital and j refers to the number of electrons in the $2s$ and $2p$ orbitals. The number of atoms in this state is denoted $N_{i,j}$ and $R_{kl\rightarrow ij}$ is the rate for transitions from the state (k,l) to (i,j).

B. The shell electron density

Rather than rely on tabulated form factors for ground-state atoms we have adopted a simple electronic structure model that readily accommodates the electronic state of each atom without superfluous computation. The shell orbital electron density was constructed using Slater's rules [19]. This approach employs a screened hydrogenic approximation to describe the orbital wave function of an electron in an atomic orbital. The functional form of these orbitals is given by

$$\psi_{nlm}(r) = N_n r^{n-1} \exp(-\zeta_n r) Y_l^m(\theta,\phi), \qquad (2)$$

where n is the principal quantum number of the orbital, l is the azimuthal quantum number, m is the magnetic quantum

number, and N_n is the normalization constant. The parameter ζ_n represents both the effective nuclear charge and the principal quantum number. For the first three shells this is defined as $\zeta_n = (Z - s)/n$ where Z is the nuclear charge of the atom and s is a semiempirical shielding constant, known as Slater's number [19]. The effects of orbital relaxation and consequent modification of the effective exponents due to the variable occupancies of different electronic states may be readily incorporated in this model by extending the definition of s to include highly excited inner-shell core-hole states. The function $Y_l^m(\theta,\phi)$ is a spherical harmonic which gives the angular dependence of the shape of the electronic shell. In these simulations, all angular dependence of the wave function is ignored, since the scattering is assumed to take place from spherical centers of electron density. If one wishes, more complicated descriptions of the electronic wave function could be used, such as the Hartree-Fock or Hartree-Slater models with no change to the essential workings of the model. In the same way the vibrational modes of the system, while neglected here, may be easily incorporated as nuclear distribution functions.

Noting that the electron density is given as $\rho_n = \psi_n^* \psi_n$, the normalized orbital wave function yields an orbital density of the form

$$\rho_n(r) = \frac{(2\zeta_n)^{2n+1}}{(2n)!} r^{2n-2} \exp(-2\zeta_n r). \tag{3}$$

In this model no distinction is made between orbitals of the same n and ζ_n and differing l, which is also reflected in more sophisticated electronic structure models.

C. The orbital form factor

The average scattered power in the far field for an individual atom is proportional to the Fourier transform of the electron density, referred to as the form factor. At the incident x-ray energies of interest here it can be safely assumed that the photon energy does not change during the scattering process. In this model the only inelastic processes considered are absorption events occurring on localized atomic positions. The analysis presented here is analogous to the natural orbital method of electronic structure theory [20] and shares the same theoretical foundation.

We make the assumption that individual atomic electron densities can be expanded in an orbital density basis, so that

$$\rho_{\text{atom}}(\mathbf{r}) = \sum_\gamma a_\gamma \rho_\gamma(\mathbf{r}), \tag{4}$$

where a_γ is the occupancy of the orbital labeled by the shell symbol γ, which generally takes values corresponding to $1s$, $2s$, or $2p$ for carbon, nitrogen, or oxygen, extended to $3s$ or $3p$ for phosphorus and sulfur; these five elements and hydrogen are the elemental constituents of almost all biological substances. The atomic form factor can also be expanded in terms of an orbital form-factor basis, so that

$$f_{\text{atom}}(\mathbf{q}) = \sum_\gamma a_\gamma f_\gamma(\mathbf{q}), \tag{5}$$

where $\mathbf{q}$ represents a point in a three-dimensional diffraction volume. To determine an analytic expression for the orbital form factor we take the Fourier transform of the orbital electron density defined in Eq. (3), yielding

$$f_{\gamma=n}(q) = \frac{(2\zeta_n)^{2n+1}}{q(2n)} \text{Im}\left\{\frac{(2\zeta + iq)^{2n}}{(4\zeta^2 + q^2)^{2n}}\right\}. \tag{6}$$

This formulation of the orbital form factor assumes that molecular scattering is dominated by a superposition of spherical atomic scatterers, which is why $\mathbf{q}$ in Eq. (5) is replaced by q in Eq. (6).

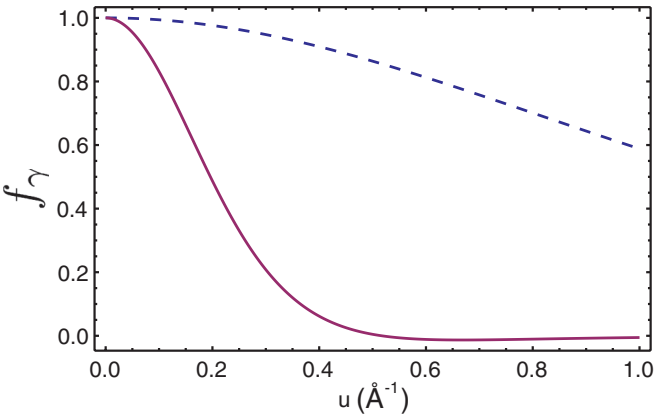

FIG. 1. (Color online) Plots of the orbital form factors for carbon, $Z = 6$, with increasing spatial frequency u, where $q = 2\pi u$. The $2s$ and $2p$ orbital (solid line) decays more rapidly, decaying rapidly for $u > 0.2$ Å^{-1}. The $1s$ orbital (dashed line) contains almost all of the high-resolution information ($u > 0.5$ Å^{-1}) corresponding to 2 Å resolution.

The orbital form factors for carbon are plotted in Fig. 1. It is evident that the high-resolution information required for atomic-resolution imaging is primarily provided by the $1s$ orbital and, therefore, inner-shell photoionization must have a dramatic effect on the likelihood of successful reconstruction. These calculations are in qualitative agreement with those of Hau-Riege [21], but employ a less computationally expensive approach.

D. Time-dependent atomic form factor

The electron occupancies of an atom in an illuminated molecule change with time over the pulse. The simulations here aim to represent an average over many pulses; the experiment detailed in [5] requires the accumulation of many repeated exposures in order to obtain atomic resolution in three dimensions. This large amount of data, when constructed into a three-dimensional volume using a classification and orientation scheme, allows us to make the assumption, given the assertion in Eq. (5), that all changes in the scattering properties of each atomic type are confined to the occupancies $a_{Z\gamma}(t)$. That is, the individual atomic orbital occupancies [Eq. (4)] are averaged over all atoms of the same type, smoothing out the stochastic nature of the damage mechanisms and making the occupancy a continuous function of time. This "average atom" approximation is appropriate for experiments, such as diffract and destroy molecular imaging, that produce very large data sets, which when combined into a three-dimensional diffraction volume will resemble a large-scale ensemble average of the randomly fluctuating electronic

state of the molecule. The stochastic fluctuations from the mean inherent to the processes can be accommodated by subtracting an isotropic q-dependent background term from the measured intensities [11]. Furthermore, we assume that the general forms of the orbital wave functions of the electrons may be included in the Slater s factor, but these are small effects compared to the variation in orbital occupancy. We may, therefore, write an expression for the time-dependent atomic form factor for species Z, as

$$f_Z(q,t) = \sum_\gamma a_{Z,\gamma}(t) f_{Z,\gamma}(q). \tag{7}$$

Equation (7) can be regarded as an extension of the time-dependent form factor presented by Hau-Riege *et al*. [6]. The principal difference is that our model recognizes the differential depletion of different orbitals, while the model of Hau-Riege *et al*. averages this depletion over all orbitals; the depopulation of orbitals at different rates makes the variation in the form factor with time q dependent.

The orbital occupancies $a_{Z\gamma}(t)$ can be calculated by an appropriate summation of the time-dependent state values found by solving the rate equation (1). For example the occupancy of the $\gamma = 1s$ orbital can be calculated as

$$a_{Z,\gamma=1s}(t) = \sum_{i,j} \frac{N^Z_{(i,j)}(t)}{N^Z_{(i,j)}(0)} i, \tag{8}$$

for the case of the $1s$ orbital, and similarly for j in the case of the $2s$ and $2p$ orbitals.

Figure 2 compares the form factor of neutral carbon calculated using this method to that obtained from crystallographic tables [22]; the two are seen to be in good agreement.

E. Structure factors

A structure factor describes the x-ray scattering properties of a complex molecule comprised of many atoms. When using the model described here it is convenient to separate the atoms into groups of their respective elements. The structure factor for a system of atoms in a molecule can therefore be written as

$$F(\mathbf{q},t) = \sum_Z \sum_{m_Z} f_Z(q,t) \exp\left(i\mathbf{q}\cdot\mathbf{R}_{m_Z}\right), \tag{9}$$

where m_Z is the mth atom of element Z, located at position $\mathbf{R}_{m_Z}$, with an atomic form factor $f_Z(q)$. The vector $\mathbf{q}$ represents a point in a three-dimensional diffraction volume, and as the form factor is assumed to be spherically symmetric, we set $q = |\mathbf{q}|$. The time dependence of the form factors follows Eq. (7). It is assumed that the atomic centers at $\mathbf{R}_{m_Z}$ are stationary throughout the pulse, this is considered a valid assumption if the pulses are shorter than 10 fs [4].

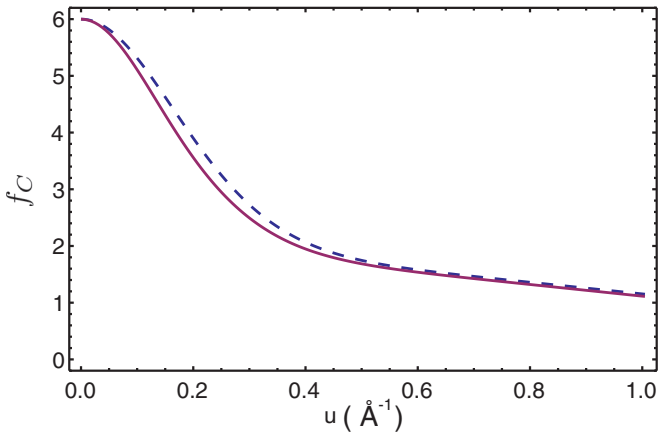

FIG. 2. (Color online) Plots of the atomic form factor for carbon, $Z = 6$. The dashed line represents the form factor calculated using the analysis presented here, and the solid line represents the tabulated values [22]. For $u > 0.4$ Å^{-1} the form factor is almost entirely due to contributions from the $1s$ orbital density.

F. Calculation of intensity

The intensity is proportional to the structure factor multiplied by its complex conjugate, $I \propto F^*F$. The intensity measured at the detector is assumed to be the time average of the instantaneous intensities resulting from the time-dependent structure factor over the life of the pulse. For a square pulse of duration T we write

$$I(\mathbf{q}) = \sum_{Z_1,Z_2} \frac{1}{T} \int_0^T f^*_{Z_1}(q,t) f_{Z_2}(q,t) dt \times \sum_{m_{Z_1},m_{Z_2}} \exp\left[i\mathbf{q}\cdot\left(\mathbf{R}_{m_{Z_2}} - \mathbf{R}_{m_{Z_1}}\right)\right], \tag{10}$$

providing an expression for the intensity expected from a molecule with a time-varying electron density. An example of the diffraction patterns simulated using this formulation are shown in Fig. 3. The molecule chosen as a diffraction target is bacteriorhodopsin, a light-harvesting molecule consisting of 2039 nonhydrogen atoms, including 1391 carbon atoms.

It should be noted that the intensity defined by Eq. (10) separates the unchanged structural components of the molecule (the positions of the atoms $\mathbf{R}_{m_Z}$) from the time-dependent components of the diffraction. Given the average-atom approximation [Eq. (7)], we now separate these two components explicitly to yield

$$I(\mathbf{q}) = \sum_{Z_1,Z_2} T_{Z_1}(\mathbf{q}) A_{Z_1,Z_2}(q) T^*_{Z_2}(\mathbf{q}), \tag{11}$$

where we define $T_Z(\mathbf{q})$ to contain the structural information through the relation

$$T_Z(\mathbf{q}) = \sum_{m_Z} \exp\left(-i\mathbf{q}\cdot\mathbf{R}_{m_Z}\right), \tag{12}$$

which is the Fourier transform of a series of Dirac δ functions centered on the atomic nuclei of all atoms of species Z. Following [10], we obtain $A_{Z_1,Z_2}(q)$ from $A_{Z_1,Z_2}(q,q)$, where

$$\begin{aligned} A_{Z_1,Z_2}(q_1,q_2) &= \frac{1}{T}\int_0^T f_{Z_1}(q_1,t) f^*_{Z_2}(q_2,t) dt \\ &= \sum_{\gamma_1,\gamma_2} \frac{1}{T}\int_0^T a_{Z_1,\gamma_1}(t) a_{Z_2,\gamma_2}(t) dt \\ &\quad \times f_{Z_1,\gamma_1}(q_1) f_{Z_2,\gamma_2}(q_2) \\ &= \sum_{\gamma_1,\gamma_2} P_{Z_1\gamma_1,Z_2\gamma_2} f_{Z_1,\gamma_1}(q_1) f_{Z_2,\gamma_2}(q_2), \end{aligned} \tag{13}$$

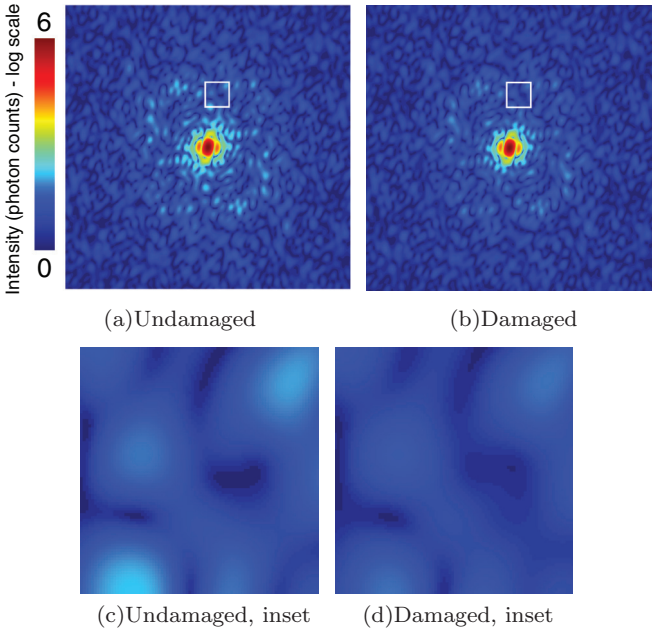

FIG. 3. (Color online) A two-dimensional (2D) projection of the simulated far-field diffracted intensity of bacteriorhodopsin on a logarithmic scale, calculated according to Eq. (11), for the (a) undamaged and (b) damaged cases. The insets (c) and (d) provide a closeup of a region corresponding to ∼6 Å resolution. The change in contrast between damaged and undamaged cases is evident: there is an approximately ∼7% loss in contrast between damaged and undamaged cases at 6 Å resolution; this number varies with q. The amount of damage corresponds to an incident fluence of 5×10^{12} photons/(100 nm)^2 with a photon energy of 10 keV. The edge of the array corresponds to a resolution of 1.085 Å.

where the form factors $f_{Z,\gamma}(q)$ are real-valued quantities defined by Eq. (6). The elements $A_{Z_1,Z_2}(q_1,q_2)$ form a matrix $\mathbf{A}$, which contains all of the dynamical information about the system within the atomic scattering model. It is derived from the elements of the time-averaged orbital population matrix $\mathbf{P}$, whose elements are $P_{Z_1\gamma_1,Z_2\gamma_2}$. The intensity is, as a consequence, the diagonal part of the mutual optical intensity $J(\mathbf{q}_1,\mathbf{q}_2)$, which is defined by

$$J(\mathbf{q}_1,\mathbf{q}_2) = \sum_Z \sum_{Z'} T_Z(\mathbf{q}_1) A_{Z,Z'}(q_1,q_2) T_{Z'}(\mathbf{q}_2). \qquad (14)$$

This function describes the coherence properties of the electromagnetic wave scattered by the time-dependent electron density.

G. Summary of the scattering model

In formulating Eq. (11) some quite general assumptions about the electrodynamic processes have been made. These assumptions form an electrodynamical model of the scattering process and are reiterated here in concise form.

It is assumed that the positions of the atoms are fixed throughout their interactions with the x-ray field. This assumption is considered reliable if the pulse duration is less than ∼10 fs, and rules out any scattering interaction during the "Coulomb explosion" of the molecule. The localized, stable position of atomic centers enables the treatment of the atomic postions as Dirac δ functions. Consequently, the contribution of the positions of atoms in the far field is expressed as the Fourier transform of a set of δ functions centered around the atomic positions $\mathbf{R}_m$ [see Eq. (12)]. This is readily extended to include vibrational amplitudes caused by thermal motion provided the characteristic lengths of the associated probability distributions are not too large.

It is assumed that the atomic electron densities may be expanded as a set of orbital occupancies and that the electron densities of the orbitals do not depend strongly on the degree of ionization, so that the variability in the electronic state of the molecule through the pulse is expressed in terms of a time-dependent orbital occupancies. Any scatter from the diffuse distribution of recaptured electrons is neglected. We also note that the high-angle scatter that corresponds to the high-resolution information in the detector plane is largely dependent on core-shell electrons [21].

All scattering interactions between the molecule and the x-ray field are assumed to involve interactions with a superposition of atomic electron densities. The primary inelastic interaction expected for objects consisting of biological elements at the wavelengths typical of XFEL illumination at atomic resolution is photoabsorption; Compton scattering is neglected. Our expression for the total scattered intensity [Eq. (10)] contains within it the signature of partial spatial coherence because the fluctuations in the electron density render the complex structure factor time dependent. The total scattered intensity may be regarded as the weighted superposition of intensities formed by scattering from the instantaneous electron density. At the beginning of the pulse, the molecule is presumed to be in its ground electronic state, but at the end of the pulse it is left in a highly excited nonequilibrium electronic state. The structure factors for each electronic state sampled by the scattering process are nontrivially related by a succession of electronic processes. The total scattered intensity cannot, as a consequence, be regarded as being proportional to the complex square of a single structure factor derived from an electron density whose spatial extent matches that of the target molecule.

The partially coherent scatter resulting from a damage-affected molecule invalidates the main assumption of CDI, which is the full coherence of the wave field leaving the sample; there is no longer a simple mapping between the detected intensity and the electron density. In general, if the spatial coherence length of the wave field leaving the object is at least twice as large as the largest spatial dimension of the object, then the field may be considered fully coherent with respect to the object, to a good approximation [23]. However, coherent imaging techniques are employed regularly with sources of partially coherent light [24,25] using a modal decomposition method [26]. It is evident that in the case of biomolecular imaging at XFELs, the effect of the illumination is to create disturbances in the electron density of the molecule, through photoionization, Auger relaxation, and other events. The spatial extent of these disturbances, when projected onto a plane perpendicular to the propagation direction of the pulse, is small compared to the mean diameter of the molecule. This can induce a coherence length that is smaller than the dimensions of the scattered field leaving the molecule. The field leaving the object may be considered to be a certain type of partially

coherent field produced by a *quasihomogeneous secondary source* [[27], Sec. 5.3.2], provided that the likelihood of photoionization is similar for all elements of the same species in the molecule.

III. SOLVING FOR THE MODES

A. The molecule as a secondary source

Even if the illumination of the sample is fully coherent, the mutual optical intensity of the scattered wave, Eq. (14), may exhibit partial spatial coherence due to the effects of time-averaged electrodynamical processes; the problem of propagating partially coherent light fields from entirely static scatterers is mathematically analogous to the problem of propagating fully coherent light fields from dynamic scatterers.

It is convenient to write the mutual optical intensity, Eq. (14), as an equivalent modal expansion in the manner of Wolf [26],

$$J(\mathbf{q}_1,\mathbf{q}_2) = \sum_k \eta_k \psi_k(\mathbf{q}_1)\psi_k^*(\mathbf{q}_2). \tag{15}$$

The diffracted intensity is obtained by setting $\mathbf{q}_1 = \mathbf{q}_2$, so that

$$I(\mathbf{q}) = \sum_k \eta_k \psi_k(\mathbf{q})\psi_k^*(\mathbf{q}). \tag{16}$$

The functions $\psi_k(\mathbf{q})$ represent mutually incoherent optical modes that satisfy $\langle\psi_j|\psi_k\rangle = \delta_{jk}$, and η_k represents the occupancy of the kth mode.

If the scattered light can be described as being emitted from a planar, secondary, quasihomogeneous source, then the degree of coherence can be approximated by a Gaussian function based on a few parameters: the size of the molecule, the wavelength of illumination, and the relative elemental composition. The degree of coherence measured for any one such object holds for all such objects within any electrodynamical model based on atomic scattering, photoabsorption, and Auger emission and secondary ionization events determined by a mean-field model of the molecule-ion potential. The degree of partial coherence induced by damage can be calculated by estimating the rates of the physical processes occurring due to the illumination. We propose a simpler method in which the matrix **A** is determined from experimental data using a known structure of a similar size and composition to the target molecule as a calibrator. The damage-induced partial coherence can then be used to update an iterative phase recovery algorithm for unknown molecules by rescaling the intensity to compensate for the effect of damage. This is analogous to using the known structure of a Young's double slit to measure the coherence of a source prior to imaging with partially coherent diffractive methods [24,28]. We now extend the theoretical framework of Quiney and Nugent [10] to show how such a measurement could be performed.

B. Derivation of the eigenvalue equation

In terms of the parameters of the electronic structure model, the mutual optical intensity within the molecular volume is defined by

$$J(\mathbf{r}_1,\mathbf{r}_2) = \sum_{Z_1\gamma_1,Z_2\gamma_2} \rho_{Z_1\gamma_1}(\mathbf{r}_1) A_{Z_1\gamma_1,Z_2\gamma_2} \rho_{Z_2,\gamma_2}(\mathbf{r}_2), \tag{17}$$

where $\rho_{Z\gamma}(\mathbf{r})$ is the orbital density of an electron in an atom of element type Z and, as before, the orbital label is denoted by γ. The matrix of average atomic populations **P**, whose elements are denoted P_{Z_1,Z_2}, may be obtained from elemental and orbital components; following Eq. (13), $P_{Z_1,Z_2} = \sum_{\gamma_1,\gamma_2} P_{Z_1\gamma_1,Z_2\gamma_2}$. As mentioned in Sec. II B, we apply a nodeless hydrogenic spherically symmetric approximation to our orbital densities, and therefore the $2s$ and $2p$ orbitals are indistinguishable in our model.

We expand the mutual optical intensity in terms of a set of orthonormal modes ψ_k, weighted by the modal occupancy η_k, using Mercer's theorem [29]

$$\sum_{Z_1\gamma_1,Z_2\gamma_2} \rho_{Z_1\gamma_1}(\mathbf{r}_1) P_{Z_1\gamma_1,Z_2\gamma_2} \rho_{Z_2\gamma_2}(\mathbf{r}_2) = \sum_k \eta_k \psi_k(\mathbf{r}_1)\psi_k^*(\mathbf{r}_2). \tag{18}$$

To simplify we multiply both sides by an arbitrary mode $\psi_m(\mathbf{r}_2)$ and integrate over all space, so that

$$\sum_{Z_1\gamma_1,Z_2\gamma_2} \rho_{Z_1\gamma_1}(\mathbf{r}_1) P_{Z_1\gamma_1,Z_2\gamma_2} \int \rho_{Z_2,\gamma_2}(\mathbf{r}_2)\psi_m(\mathrm{r}_2)\,d\mathbf{r}_2 = \sum_k \eta_k \psi_k(\mathbf{r}_1) \int \psi_k(\mathbf{r}_2)\psi_m(\mathbf{r}_2)\,d\mathbf{r}_2. \tag{19}$$

Orthonormality of the modes requires that $\langle\psi_k|\psi_{k'}\rangle = \delta_{kk'}$ where $\delta_{kk'}$ is the Kronecker δ. The integral on the right-hand side of Eq. (19) then vanishes except for the case $k = m$. By analogy with the natural orbital method [20], we expand the modes in terms of a complete shell-orbital density basis, so that $\psi_m(\mathbf{r}) = \sum_{Z_3\gamma_3} c^m_{Z_3\gamma_3} \rho_{Z_3\gamma_3}(\mathbf{r})$. We also define $S_{Z_1\gamma_1,Z_2\gamma_2}$, the squared orbital density, to be

$$S_{Z_1\gamma_1,Z_2\gamma_2} = \int \rho_{Z_1\gamma_1}(\mathbf{r})\rho_{Z_2\gamma_2}(\mathbf{r})d\mathbf{r}. \tag{20}$$

This enables us to rewrite the integral on the left-hand side of Eq. (19) as

$$\int \rho_{Z_2\gamma 2}(\mathbf{r}_2)\psi_m(\mathbf{r}_2)d\mathbf{r}_2 = \sum_{Z_3\gamma_3} c^m_{Z_3\gamma_3} S_{Z_2\gamma_2,Z_3\gamma_3}. \tag{21}$$

Multiplying Eq. (19) by an arbitrary shell density $\rho_{Z_4\gamma_4}(\mathbf{r}_1)$ and integrating with respect to $\mathbf{r}_1$ yields

$$\sum_{Z_1\gamma_1,Z_2\gamma_2,Z_3\gamma_3} S_{Z_4\gamma_4,Z_1\gamma_1} P_{Z_1\gamma_1,Z_2\gamma_2} S_{Z_2\gamma_2,Z_4\gamma_4} c^m_{Z_3\gamma_3} = \eta_m \sum_{Z_1\gamma_1} c^m_{Z_1\gamma_1} S_{Z_4\gamma_4,Z_1\gamma_1}. \tag{22}$$

This may be rewritten as a matrix equation

$$\mathbf{SPSc}_m = \eta_m \mathbf{Sc}_m, \tag{23}$$

where **S** is a matrix whose elements consist of the orbital density values $S_{Z_1\gamma_1,Z_2\gamma_2}$. The other elements in the equation, η_m and $\mathbf{c}_m$, represent the modal occupancy and the expansion of the mode in terms of orbital densities. Expressing the left-hand side of Eq. (23) in terms of the matrix $\mathbf{J} = \mathbf{SPS}$, we arrive at the form of the generalized eigenvalue equation,

$$\mathbf{JC} = \eta\mathbf{SC}, \tag{24}$$

where η is a diagonal matrix whose elements are the modal occupancies, and $\mathbf{C}$ is a matrix whose columns are the eigenvectors containing the expansion coefficients of the modes. It is more convenient to write the orbital density expansion of the modes in terms of its Fourier transform, involving the orbital form factors

$$\tilde{\psi}_k(q) = \sum_{Z\gamma} c^k_{Z\gamma} f_{Z\gamma}(q). \tag{25}$$

The structure of a large molecule can be easily incorporated into the expansion of the mode in terms of orbital densities as

$$\psi_k(\mathbf{r}) = \sum_{Z\gamma} c^k_{Z\gamma} \sum_{m_Z} \rho_{Z\gamma}\left(\mathbf{r} - \mathbf{R}^Z_{m_Z}\right), \tag{26}$$

where $\mathbf{R}^Z_{m_Z}$ is a vector defining the location of the mth atom of type Z in the molecule. In the far field, this is given by its Fourier transform,

$$\tilde{\psi}_k(\mathbf{q}) = \sum_{Z} T_Z(\mathbf{q}) \sum_{\gamma} c^k_{Z\gamma} f_{Z\gamma}(q), \tag{27}$$

where T_Z is the structure vector defined in Eq. (12).

C. Orbital density matrix

Given the elements of $\mathbf{S}$ that have been defined as the integral of orbital densities, shown in Eq. (20), an analytical expression for the elements of this matrix can be found using expressions for the orbital density given in Eq. (3). Using the integral identity $\int_0^\infty x^n \exp(-\alpha x)dx = n!/\alpha^{n+1}$, the elements of $\mathbf{S}$ become

$$S_{Z_1\gamma_1,Z_2\gamma_2} = \frac{(2\zeta_{\gamma_1})^{2\gamma_1+1}}{(2\gamma_1)!} \frac{(2\zeta_{\gamma_2})^{2\gamma_2+1}}{(2\gamma_2)!} \frac{(2\gamma_1+2\gamma_2-4)!}{(2\zeta_{\gamma_1}+2\zeta_{\gamma_2})^{2\gamma_1+2\gamma_2-3}}. \tag{28}$$

Here we apply the tight-binding approximation [30,31], which assumes that the atomic wave functions vanish at distances corresponding to the nearest-neighbor distance, to the matrix $\mathbf{S}$. Assuming that the electron densities of different atomic species never overlap allows us to set $S_{Z_1\gamma_1,Z_2\gamma_2} = 0$ when $Z_1 \neq Z_2$, for all γ_1 and γ_2.

D. Resultant modes

The eigenvectors $\mathbf{c}_k$ that form the expansion of modes in terms of orbital densities [Eq. (26)] have values that represent a normalized occupancy of that orbital density. Determining these modes involves solving the eigenvalue equation [Eq. (24)] for a given damage matrix $\mathbf{A}$. For a large complex molecule such as bacteriorhodopsin the modes are difficult to represent. To gain physical insight into the modes we present the modes for a simple test molecule, 3-hydroxypyridine, a heterocyclic molecule with chemical formula C_5NH_3OH. The hydrogen atoms contribute negligible scattering, so this is considered a seven-atom molecule, consisting of three different elements of interest. These elements of interest have two distinct orbitals given spherical symmetry, the $1s$ and the $2s$ and $2p$ orbitals. To these two we add a third orbital to account for electrons lost to the continuum due to photoionization events. Keeping track of these electrons ensures that the transformation from static scatterer to damaged scatterer is unitary. The orbital density of the continuum states is set to zero when calculating diffraction, reflecting the negligible contribution of free electrons to x-ray diffraction. The calculation of the modes $\psi_k(\mathbf{r})$ for 3-hydroxypyridine illuminated by a square pulse of fluence 5×10^{12} photons/(100 nm)^2 is given in Fig. 4.

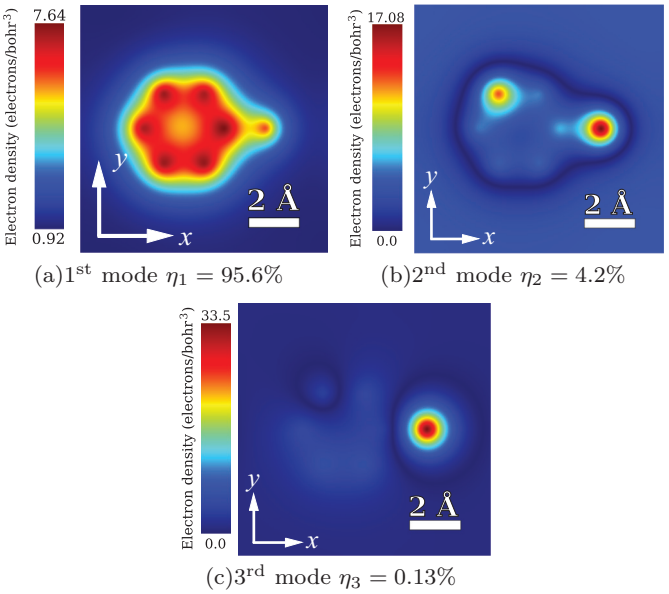

FIG. 4. (Color online) The first three modes and respective occupancies for 3-hydroxypyridine, illuminated by a uniform pulse of duration 5 fs with fluence of 5×10^{12} photons/(100 nm)^2. The modes are represented in Hartree atomic units of electron density.

One may make a physical interpretation of the modes. The differing rates of atomic processes for different atomic species leads to a differential change in the average populations of, say, carbon and oxygen. In the modal analysis this appears like a polarization, although there is, of course, no actual electronic transport involved. This general behavior is reflected in the modal decomposition of bacteriorhodposin in which the Z-dependent rates of photoionization and Auger recombination cause differential depletion of electron density.

IV. MEASUREMENT OF THE MODES

The measurement or characterization of the damage to a sample given an XFEL pulse is performed completely by determining the matrix $\mathbf{A}$. This can be achieved by measuring the modes ψ_k and modal coefficients η_k that characterize the damage. Figure 5 shows the variation of the time-averaged occupancy of the carbon orbitals, that is, $\langle a_{Z\gamma}\rangle$ for $Z =$ carbon and $\gamma = 1s$, $2s$, and $2p$ for incident photon fluences over a 5 fs pulse.

As the level of incident photon flux increases we see the values of the orbital occupancies decay, indicating the damage is affecting the sample. It is observed that the occupancy of the 1s orbital decreases, corresponding to an increase in the number of core hole vacancies in carbon for pulses with large X-ray fluxes. The variability of occupancies and, hence, modes, with incident flux means one set of modes cannot be used to describe all damage conditions. This makes impossible

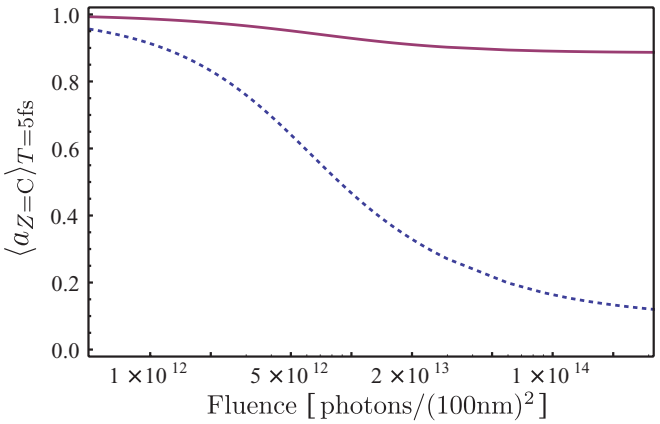

FIG. 5. (Color online) The time-averaged occupancy for the $1s$ orbital (dashed line) and the $2s$ and $2p$ orbitals (solid line) of carbon for increasing incident photon fluence, and hence damage. Photon energy was set to 10 keV.

a measurement of **A** by measuring occupancies in the manner described in [28]. Our measurement of the damage must now include a determination of both the form and occupancies of the damage modes.

This measurement of the modes relies on their completeness and orthonormality. This property can be ensured by enforcing the unitarity of the matrix **A**, which can be accomplished by keeping track of all electrons lost during exposure using continuum states. Given these conditions, we can expand a single mode that forms part of a complete description of a damage scenario, ψ_k^0, as an expansion in terms of a set of approximate, "trial" modes ψ'_m, or

$$\tilde{\psi}_k^0(\mathbf{q}) = \sum_m b_m \tilde{\psi}'_m(\mathbf{q}) \tag{29}$$

where we are considering modes in the far field, and where b_m are auxiliary expansion coefficients. These trial modes $\tilde{\psi}'_m$ are similar in form to the target modes $\tilde{\psi}_k^0$, but are calculated using an initial estimate of the damage matrix **A**.

A new expression for the diffracted intensity is therefore obtained by substituting Eq. (29) into Eq. (16), yielding

$$I(\mathbf{q}) = \sum_k \eta_k^0 \sum_{mm'} b_m b^*_{m'} \tilde{\psi}'_m(\mathbf{q}) \tilde{\psi}^{*\prime}_{m'}(\mathbf{q}). \tag{30}$$

We require an expression for the damage matrix in terms of the trial modes ψ'_m. To obtain this, we equate two of our expressions [Eqs. (11) and (16)] for the intensity,

$$\sum_{Z_1,Z_2} T_{Z_1}(\mathbf{q}) A_{Z_1,Z_2}(q) T^*_{Z_2}(\mathbf{q}) = \sum_k \eta_k \tilde{\psi}_k^0(\mathbf{q}) \tilde{\psi}_k^{0*}(\mathbf{q}). \tag{31}$$

Expressing our exact modes in terms of a trial mode basis [Eq. (29)] and writing our trial modes as an explicit expansion in terms of an orbital form-factor basis set [Eq. (27)] enables Eq. (31) to be rewritten. After some simplification we obtain an expression for the elements of **P**, which is

$$P_{Z_1\gamma_1,Z_2\gamma_2} = \sum_k \sum_{b_1,b_2} \eta_k b_{m_1} b_{m_2} c^{m_1}_{Z_1,\gamma_1} c^{m_2}_{Z_2,\gamma_2}. \tag{32}$$

This is the principal result of this section. Equation (32) indicates that the task of measuring the effects of damage processes becomes one of determining the auxiliary coefficients b_m and the modal occupancies η_k, given an arbitrary set of trial modes defined, after Eq. (27), by the expansion coefficients $c^m_{Z,\gamma}$ and a known structure $T(\mathbf{q})$. We therefore endeavor to take a simulated intensity measurement and to determine these values by a fitting procedure, given our assumed structure and modes.

A. Fitting modes to intensities

We now fit a set of modes to the simulated far-field diffraction expected from illumination of the protein bacteriorhodopsin. To perform this fitting with respect to η and b_m we require some objective function marking the deviation in our fit. We select the metric

$$\mathcal{E} = \sum_i (I_i - I_{0,i})^2, \tag{33}$$

where I_i is the ith pixel in the current guess of the diffracted intensity, and $I_{0,i}$ is the ith pixel in the simulated intensity measurement corresponding to experiment. Using this objective function and its derivative allows the use of standard conjugate gradient techniques to fit a measured intensity to an arbitrary damage matrix via a nonlinear least-squares method; for these methods convergence is defined as determination of the solution to within a tolerable error metric; indeed the presence of noise in these simulations precludes a pointwise solution. As a guide to convergence we define a second metric ρ, the average ratio of the fitted intensity I' to the input simulated intensity I^0, that is, $\rho = (1/N)\sum_i (I'_i/I^0_i)$ for $I^0_i \neq 0$, and where N is the number of nonzero elements in I^0. A perfect fit would have a ratio equal to unity, with a standard deviation close to machine error.

The first trial example of the fit procedure was initialized as follows: a diffracted intensity corresponding to an incident flux of 1.5×10^{11} [photons/(100 nm)^2]/fs was calculated. To fit to this damage-effected intensity, the trial modes $\psi'_m(\mathbf{q})$ and the initial modal occupancies η_k were chosen to correspond to an incident flux of 3.0×10^{11} [photons/(100 nm)^2]/fs, precisely double the incident flux used to calculate the intensity distribution. The coefficients $b_{m,k}$ were chosen such that $b_{m,k} = 1$ when $m = k$, and $b_{m,k} = 0.01$ when $m \neq k$. Setting the $m = k$ coefficients to unity and the cross terms $m \neq k$ to zero corresponds to the ideal scenario where $\psi_k^0(\mathbf{q}) = \psi'_m(\mathbf{q})$. In other words the trial modes are indistinguishable from the exact modes. This will generally not be the case, so it is expedient to set the cross terms to some small, nonzero number rather than zero, reflecting their small, yet non-negligible, contribution in likely fits. Initially, this fit began with an average ratio ρ of 0.976, with standard deviation σ of 0.003. This shows the general decrease in intensity as the damage is increased. After 600 iterations, the final ratio was $\rho = 1.0002$ with $\sigma = 3\times10^{-5}$.

A second fit was attempted, this time initializing the procedure using eigenvectors and modal occupancies for a minimal amount of damage, corresponding to an incident flux of 4×10^5 [photons/(100 nm)^2]/fs. This fit started with an initial ratio $\rho = 1.6$ with $\sigma = 0.5$. After 200 iterations the intensity converged, leaving $\rho = 0.9826$ with $\sigma = 0.0003$. The value of the objective function with each iteration for both fits is given in Fig. 6.

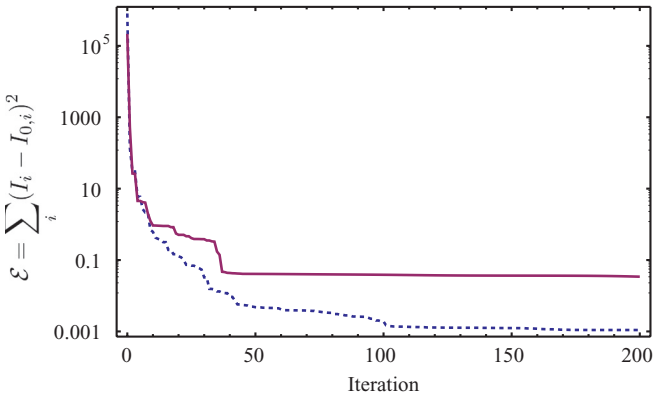

FIG. 6. (Color online) The value of the objective function $\mathcal{E}$ on a logarithmic scale, for increasing routine iteration, for the case of intialization with double incident flux modes and occupancies (solid line) and for the case of minimal incident flux mode and occupancy intialization (dashed line). The routine performs most of the minimization within the first 50 iterations.

It is evident that the routine converges even from initial modes and modal occupancies that belong to incident fluences far different from the intensity being fitted. As the routine performs better when the initial damage estimate is closer to the correct result, it may be beneficial to measure the XFEL flux and perform preliminary simulations of the damage processes prior to fitting.

B. The recovered damage matrix

To quantify the deviation of the fitted population matrix **P** we define the metric

$$d_{Z_1,Z_2} = \frac{1}{n_\gamma}\sqrt{\sum_{\gamma_1\gamma_2}\left(P'_{Z_1\gamma_1 Z_2\gamma_2} - P^0_{Z_1\gamma_1 Z_2\gamma_2}\right)^2}, \qquad (34)$$

where n_γ is the number of orbitals, $\mathbf{P}'$ is the fitted population matrix, and $\mathbf{P}^0$ is the population matrix for the damage scenario used to simulate the intensity. These calculations were performed for cases where $Z_1 = Z_2$; the results appear in Table I.

The fit with modes and occupancies was created with an incident photon flux exactly four times that used to simulate the intensity. This approach was adopted to reflect a likely scenario in which neither the details of the model nor the experimental interaction parameters are known precisely, but for which an order-of-magnitude estimate is likely to suffice to capture the relative kinetic behavior of each element. This approach enabled the recovery of the elements of **P** for carbon, nitrogen, and oxygen to within $\approx$1% precision.

TABLE I. The percentage deviation (for $Z_1 = Z_2$) in the elements of **P** for three elements of biological interest, at the start of the fitting procedure, d_{ZZ}^{initial}, and at the end, d_{ZZ}^{final}. The initial values were generated by a simulation corresponding to a photon flux four times that used to generate the diffraction data.

	d_{ZZ}^{initial} (%)	d_{ZZ}^{final} (%)
Carbon	18.9	0.785
Nitrogen	25.6	1.27
Oxygen	29.5	1.53

It is important to recall that **A** (or **P** from which it is derived) completely characterizes the time-varying nature of the electron densities. All dynamical information during the pulse is encoded in this quantity and, therefore, this matrix is all that is needed to incorporate the damage for an unknown structure. This information can be included in a subsequent single-molecule phase reconstruction, for instance, by updating the modulus constraint using the method proposed by Quiney and Nugent [10].

V. RECOVERY OF CROSS SECTIONS

The recovery of the damage inflicted on the illuminated molecule, contained in the time-averaged population matrix **P** can be extended to infer an effective photoionization cross section for carbon. A measurement of the cross section performed in this way will enable certain assumptions about the damage processes as they exist in large biomolecules to be tested directly. This measurement could quantify precisely what physical mechanisms are dominating the damage process, as well as determining the applicability of established rates to large molecular environments under XFEL illumination.

A nonlinear optimization is used to fit a cross section to the elements of the population matrix **P**. Following the approach adopted in previous sections we define an objective function denoting the difference between our guess of **P** using our assumed cross section and the measured **P** that comes from the fitting of modes and occupancies. Restricting ourselves to the case of $Z_1 = Z_2 = 6$, corresponding to carbon, we write

$$\mathcal{E}(\sigma_{\text{ph}}) = \sum_{\gamma_1,\gamma_2}\left[P_{\gamma_1,\gamma_2}(\sigma_{\text{ph}}) - P^0_{\gamma_1,\gamma_2}\right]^2, \qquad (35)$$

and seek the derivative of $\mathcal{E}$ with respect to σ_{ph}. We make the assumption that the photoionization cross section remains largely constant with respect to time and is the same for atoms with one core hole; ionization from $n = 2$ orbitals is neglected entirely. To obtain the derivative with respect to the cross section, we must find the derivative of the population matrix with respect to the cross section. An easy way is to solve the equation

$$\frac{d}{dt}\frac{dP_{\gamma_1,\gamma_2}(q)}{d\sigma_{\gamma_1}} = a_{\gamma_1}(t)\frac{da_{\gamma_2}(t)}{d\sigma_{\gamma_1}} + a_{\gamma_2}(t)\frac{da_{\gamma_1}(t)}{d\sigma_{\gamma_1}} \qquad (36)$$

over the interval $0 \leqslant t \leqslant T$ with the initial condition

$$\left.\frac{dP_{\gamma_1,\gamma_2}(q)}{d\sigma_{\gamma_1}}\right|_{t=0} = 0. \qquad (37)$$

The solution at $t = T$ is then equal to the value of the integral. Therefore, one must calculate not only $a_{\gamma_1}(t)$ over the length of the pulse, but also its derivative with respect to the rate, which is also time dependent. This derivative is dependent on the rate and, hence, the photon flux, which is a time-dependent quantity, so this calculation must be repeated for each guess of the cross section. It also means that recovery of the cross section requires an accurate measurement of the incident flux that interacted with the molecule. Such measurements are now possible at XFEL facilities. The integration itself can

be accomplished using a standard fourth-order Runge-Kutta integrator.

Fitting our photoabsorption cross sections in this way yields a value of 2.09 ± 0.09 cm^2/g when the modal fit for $A_{\gamma_1,\gamma_2}(q)$ is initialized assuming pulse fluences two orders of magnitude less than that used to calculate the intensity, and a value of 2.03 ± 0.07 cm^2/g when the modal fit for $A_{\gamma_1,\gamma_2}(q)$ is intialized assuming pulse fluences one order of magnitude less than that used to calculate the intensity. This compares favorably with the value used in initial simulations of 2.06 cm^2/g at 10 keV incident photon energy taken from the tables of Henke *et al*. [16]. The uncertainty of the fitted value may be reduced by collecting more signal in diffraction regions corresponding to atomic resolution; these fits were performed assuming the ability to measure ten photons out to atomic resolution. In these simulations tabulated Auger decay rates were assumed; however, it should be possible to expand the fit to measure them, and any of the rates of any other applicable processes, as well.

VI. CONCLUSION

We have proposed a measurement scheme for femtosecond x-ray diffraction experiments that removes the reliance on detailed simulation of the molecular electrodynamics for their structural interpretation. It is based on the assumption that inner-shell processes, such as Thomson scattering, *K*-shell photoionization, and Auger decay are primarily atomic processes, with characteristic rates that are largely independent of chemical environment. The measurement of the diffraction from a biomolecule containing, for example, carbon, nitrogen, and oxygen will contain the signatures of these electronic processes in a manner that is transferable between systems of similar chemical composition. The characterization of the electronic component of the mean scattering amplitude under given experimental conditions using experimental data is equally valid in any biomolecule of similar composition. Rather than rely on electrodynamical simulations, which carry with them considerable uncertainty regarding the validity of atomic model parameters under XFEL interaction conditions, one can instead extract the relevant electronic parameters from a calibration experiment on a molecule of known structure and similar chemical composition to the target. This approach retains the convenience of an atomic scattering model while recognizing that the rates of each process may be modified significantly when the atoms are embedded in a rapidly evolving, highly ionized biomolecular system.

The foundation of the atomic model adopted here involves approximations that can be justified only by comparison with experimental data for single-molecular diffraction which are not yet available. The most obvious potential failing of this approach is that the details of the biomolecular structure play no explicit role in the electrodynamics, either within existing simulations of the interaction or within the proposed method of experimental analysis. In addition to electron recapture and collisional ionization processes, the dominant effects not included here most likely arise because of the formation of a large positive charge distributed over the molecule and the consequent decrease in the kinetic energy of the photoelectrons that are ejected. This may impart a position dependence within the molecule of the electrodynamical properties of atoms of a given type, each of which experiences an electric field that depends primarily on the distance of the atom from the center of charge. This, in turn, may influence the rates of each of the electronic processes that the atom undergoes. Also untreated in the present model is the effect, if any, of the ejected photoelectrons on the measured diffraction pattern on the proposed time scale of the interaction. Since these effects are all electronic, however, it is reasonable to assume that molecules of similar chemical composition *and* physical dimensions may possess similar average scattering properties.

The procedure outlined here offers a scheme by which this electronic information may be transferred between biomolecular systems to facilitate the determination of unknown biomolecular structures using *a priori* information about their electrodynamic behavior under specified interaction conditions. This reduces the reliance on modeling of the electronic processes and on molecular replacement strategies in structure determination. It also offers a way to measure the effective rates of fundamental electrodynamical processes in complex biomolecular systems.

[1] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, Nature (London) **400**, 342 (1999).

[2] R. W. Gerchberg and W. O. Saxton, Optik (Stuttgart) **35**, 237 (1972).

[3] J. R. Fienup, Appl. Opt. **21**, 2758 (1982).

[4] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, Nature (London) **406**, 752 (2000).

[5] K. J. Gaffney and H. N. Chapman, Science **316**, 1444 (2007).

[6] S. P. Hau-Riege, R. A. London, H. N. Chapman, A. Szoke, and N. Timneanu, Phys. Rev. Lett. **98**, 198302 (2007).

[7] Z. Jurek and G. Faigel, Eur. Phys. J. D **50**, 35 (2008).

[8] B. Ziaja, H. N. Chapman, R. Santra, T. Laarman, E. Weckert, C. Bostedt, and T. Möller, Phys. Rev. A **84**, 033201 (2011).

[9] M. J. Berger, J. H. Hubbell, S. M. Seltzer, J. Chang, J. S. Coursey, R. Sukumar, D. S. Zucker, and K. Olsen, XCOM: Photon cross-section database. Online at http://www.nist.gov/pml/data/xcom/ (2010).

[10] H. M. Quiney and K. A. Nugent, Nat. Phys. **7**, 142 (2011).

[11] U. Lorenz, N.M. Kabachnik, E. Weckert, and I. A. Vartanyants, Phys. Rev. E **86**, 051911 (2012).

[12] H. N. Chapman *et al*., Nature (London) **470**, 73 (2011).

[13] S. Boutet *et al*., Science **337**, 362 (2012).

[14] R. Fung, V. Shneerson, D. K. Saldin, and A. Ourmazd, Nat. Phys. **5**, 64 (2009).

[15] Ne-TeDuane Loh and V. Elser, Phys. Rev. E **80**, 026705 (2009).

[16] B. L. Henke, E. M. Gullikson, and J. C. Davis, At. Data Nucl. Data Tables **54**, 181 (1993); see also http://henke.lbl.gov/optical_constants/.

[17] E. J. McGuire, Phys. Rev. **185**, 1 (1969).

[18] S. P. Hau-Riege, R. A. London, and A. Szoke, Phys. Rev. E **69**, 051906 (2004).
[19] J. C. Slater, Phys. Rev. **36**, 57 (1930).
[20] E. R. Davidson, Rev. Mod. Phys. **44**, 451 (1972).
[21] S. P. Hau-Riege, Phys. Rev. A **76**, 042511 (2007).
[22] D. Waasmaier and A. Kirfel, Acta Crystallogr., Sect. A: Found. Crystallogr. **51**, 416 (1995).
[23] J. C. H. Spence, U. Weierstall, and M. Howells, Ultramicroscopy **101**, 149 (2004).
[24] L. W. Whitehead, G. J. Williams, H. M. Quiney, D. J. Vine, R. A. Dilanian, S. Flewett, K. A. Nugent, A. G. Peele, E. Balaur, and I. McNulty, Phys. Rev. Lett. **103**, 243902 (2009).
[25] B. Abbey, L. W. Whitehead, H. M. Quiney, D. J. Vine, G. A. Cadenazzi, C. A. Henderson, K. A. Nugent, E. Balaur, C. T. Putkunz, A. G. Peele, G. J. Williams, and I. McNulty, Nature Photonics **5**, 420 (2011).
[26] E. Wolf, J. Opt. Soc. Am. **72**, 343 (1982).
[27] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995).
[28] S. Flewett, H. M. Quiney, C. Q. Tran, and K. A. Nugent, Opt. Lett. **34**, 2198 (2009).
[29] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, USA, 1953).
[30] F. Bloch, Z. Phys. **52**, 555 (1928).
[31] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).