Large-scale atomistic calculations of clusters in intense x-ray pulses

Phay J Ho and Chris Knight

1 Argonne National Laboratory, 9700 S Cass Ave., Argonne, IL 60439, United States of America
2 Leadership Computing Facility, Argonne National Laboratory, 9700 S Cass Ave., Argonne, IL 60439, United States of America

E-mail: pho@anl.gov and knightc@anl.gov

Received 20 September 2016, revised 21 March 2017
Accepted for publication 29 March 2017
Published 28 April 2017

Abstract

We present the methodology of our recently developed Monte Carlo/molecular-dynamics method for studying the fundamental ultrafast dynamics induced by high-fluence, high-intensity x-ray free electron laser (XFEL) pulses in clusters. The quantum nature of the initiating ionization process is accounted for by a Monte Carlo method to calculate probabilities of electronic transitions, including photo absorption, inner-shell relaxation, photon scattering, electron collision and recombination dynamics, and thus track the transient electronic configurations explicitly. The freed electrons, atoms and ions are followed by classical particle trajectories using a molecular dynamics algorithm. Our calculations reveal the role of electron–ion recombination processes that lead to the development of nonuniform spatial charge density profiles in x-ray excited clusters over femtosecond timescales at XFEL intensities exceeding $10^{20}$ W cm$^{-2}$. The Ar cluster fluorescence spectrum is found to be very different from the Ar atom spectrum, in which recombination processes enable additional pathways to reach the required EC for fluorescence transitions. In the high-intensity limit, recombination dynamics can play an important role in the calculated scattering response even for a 2 fs pulse. We demonstrate that our numerical codes and algorithms can make efficient use of the computational power of massively parallel supercomputers to investigate the intense-field dynamics in systems with increasing complexity and size at the ultrafast timescale and in nonlinear x-ray interaction regimes. In particular, picosecond trajectories of XFEL clusters with attosecond time resolution containing millions of particles can be efficiently computed on upwards of 262 144 processes.

Keywords: light–matter interaction, x-ray free-electron laser, multi-photon ionization and excitation, laser–cluster interaction, high-performance computing

(Some figures may appear in colour only in the online journal)

1. Introduction

High-brightness x-ray free-electron laser (XFEL) pulses [1–3], currently available at the Linac Coherent Light Source (LCLS) and Spring-8 Angstrom Compact Free Electron Laser (SACLA) facilities, provide researchers with unprecedented tools to follow the dynamics of atoms and electrons with atomic spatial resolution [4–6]. When combined with femtosecond pulse durations, the extraordinary spatial resolution and elemental specificity from these short wavelength pulses can potentially be used to create real-time ‘movies’ and track the ultrafast processes at play in chemical reactions, advanced
materials, and biological systems. However, the unprecedented intensities of the XFEL pulses also exhibit a strong interaction with the sample inducing complex dynamics [7]. These dynamics start predominantly with inner-shell electrons and inner-shell vacancy decay cascades and on the longer time scale include many pathways for electronic to nuclear relaxation. Extended nanometer-sized samples are rapidly transformed into a highly excited state. The correlation between sample excitation and measurement response is important to many XFEL applications [8–11] and central to the LCLS Single Particle Imaging Initiative roadmap [12]. The ultimate resolution and contrast that can be obtained from experiments with XFEL pulses depends sensitively on the detailed XFEL induced dynamics in the nanosized sample.

XFELs have long been envisioned as a means to perform femtosecond serial coherent diffraction imaging (CDI) on a single non-periodic entities, by making use of the very high number of photons in a single pulse. Application of this ‘probe before destroy’ approach, which has worked well in the field of single shot nanocrystallography and for a large sample in CDI [13–16], remains challenging both for small samples and for achieving atomic scale resolution [17, 18]. To realize the limitation of the approach, it is thus important to understand the physics of x-ray interactions with nanosized systems in the high-intensity hard x-ray limit required for imaging (>10^{20} W cm^{-2}, 3–8 keV and few femtosecond pulses) [12]. Here we present a combined Monte Carlo/molecular-dynamics (MC/MD) computational model to facilitate theoretical studies of clusters exposed to XFEL pulses and, as a first step, we employ Ar clusters as our model system to characterize the response in the above high-intensity regime.

Traditionally, van der Waals clusters have been exploited as testbeds as intense lasers have evolved from optical to x-ray wavelengths [19–22]. Since the composition and structure of the atomic clusters can be controlled, they can serve as a testing ground for new regimes of intense laser-matter interaction [23, 24]. These cited studies have provided a large body of knowledge on the strong wavelength dependence of initial cluster excitation to the subsequent nanoplasma dynamics from recombination to disintegration (coulomb explosion versus hydrodynamic expansion) in light pulses with intensity reaching 10^{15} W cm^{-2}. The resulting recombination and expansion dynamics in clusters exposed to intense X-ray radiation can have strong effects on the charge state distribution and ion and electron energy distributions [27, 36, 38, 45, 46]. Apart from the electron and ion observables, fluorescence spectroscopy has been used to reveal the transient states in cluster excitation [33] and interesting recombination dynamics in Ar clusters exposed to intense VUV pulse, in which delocalized electrons recombine with multiply charged Ar ions to produce excited ions that fluoresce from the electronic transition 4d \rightarrow 3p [41].

It is important to note that typical x-ray excitation mechanisms are initiated through inner-shell ionization events as opposed to thermal excitation or ionization of higher lying valence electrons [43, 47]. In the case of intense XFEL pulses, multiphoton absorption via a sequence of single-photon absorption (photoionization + resonant excitation) events in atoms and molecules produce a multitude of transient electron states with lifetimes of femtoseconds or shorter, which are subsequently followed by sequences of inner-shell electronic relaxation and photoabsorption processes [5, 48–56]. An important feature in this non-perturbative regime is that the ionization rate in a femtosecond, intense x-ray pulse can be comparable to the Auger rate, such that the normal K-shell cascades via Auger and fluorescence can be interrupted (lifetime of Ar K-shell is \sim 1 fs).

In this case, the double core-hole state, [5, 57] in addition to the single core-hole state, is relevant. The need to follow a large number of electronic configurations (EC) has been confirmed by the rate equation approach [48, 52, 56, 57].

Beyond isolated atoms and molecules, it is not trivial to experimentally infer the single cluster response due to both particle size and spatial intensity averaging [26]. However, recently developed coincident spectroscopic and scattering allow unprecedented tracking of isolated nanoparticle dynamics [7, 27, 28, 45, 58] to pinpoint the fluence dependence of cluster dynamics via simultaneous spectroscopic and scattering measurement analysis. In this technique, the recorded scattering pattern is used to infer the exposed intensity and cluster size, such that the ion observables from a single cluster in different pulse parameters can be obtained. Recently, with this coincident technique, the peak of the measured charge state distributions were shown to shift from low charge states at low peak fluences to the highest charge states at (>10^{15} W cm^{-2}), revealing the strong pulse intensity dependence of the cluster recombination and expansion dynamics [27, 45]. Motivated by the potential of this single particle imaging technique, it is thus important to explore its applicability at extreme intensities required for SPI.

Here we employed the MC/MD method to study Ar clusters exposed to intense 2 fs, hard x-ray pulses with intensities reaching beyond 10^{20} W cm^{-2}. This model uses a particle approach to follow the x-ray induced ionization, nanoplasma formation, and sample relaxation. Since the seminal work by Ditmire [59] on simulating clusters interacting with intense optical field, many particle approaches have been developed for studying microscopic details of x-ray/matter interaction [8, 21, 34, 60–75]. An important feature of the MC/MD method is that it uses a Monte Carlo method to account for transient electron dynamics induced by XFEL interaction [52–55, 76] and the subsequent electronic relaxation [61, 67] and electron–ion recombination processes quantum mechanically.

The strength of this method is that various types of observables can be computed within the same model and they can studied systematically as a function of x-ray beam parameters, sizes and composition. Previously, we revealed the important role of compton scattering found in low Z systems is substantially reduced in high Z clusters for single-particle imaging and showed that recovery of the original nanostucture with 3 Å resolution is possible with femtosecond XFEL pulses [77]. Here, we depict the roles of electron–ion recombination processes in nanoclusters exposed to 2 fs hard x-ray pulses at extreme intensities >10^{20} W cm^{-2} and their signatures in ion charge state distributions, fluorescence spectrum and scattering data. We found that recombination processes rapidly transform a small charged cluster into a core/shell structure with nonuniform spatial charge distribution within femtoseconds. Also, we
examine the feasibility of implementing the MC/MD method on high-performance computing (HPC) platforms for investigation of complex and larger systems, potentially paving the way for simulating ultrafast, collective phenomena in clusters approaching the experimental system sizes [7]. We examine the performance of our algorithms on Mira, a 10-petaflops IBM Blue Gene/Q supercomputer with 786,432 processors at the Argonne Leadership Computing Facility (ALCF).

This paper is organized as follows. In section 2, we describe the hybrid quantum Monte Carlo/classical molecular dynamics method, hereafter referred to as MC/MD. In section 3, we demonstrate the validity of this method and discuss the role of the electron–ion recombination process. We follow that with a brief discussion on the achieved parallel scalability of MC/MD methodology and comment on observed performance bottlenecks. Finally, a summary of our results and an outlook is presented in section 4.

2. Method

In order to obtain an atomistic view of the dynamical x-ray damage processes on individual atoms and the subsequent structural distortion on the lattice throughout the x-ray pulse, we use a combined MC/MD algorithm, whereby the classical degrees of freedom (coordinates and velocities of all active particles, including atoms, ions and unbound electrons) are propagated via MD, while the quantum degrees of freedom (electronic transitions) are sampled according to MC (see figure 1). The method treats the electronic and nuclear excitations on equal footing. At each step in the simulation, all particle coordinates are updated in the usual manner as a function of forces derived from interaction potentials. The equation of motion for all particles are derived from Newtonian mechanics

\[ m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i = -\nabla U_{\text{total}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N), \]  

where \( m_i \) and \( \mathbf{r}_i \) are the mass and position of the \( i \)th particle, respectively, and \( U_{\text{total}} \) is the total potential energy resulting from particle–particle interactions. For a van der Waals cluster, \( U_{\text{total}} \) is the sum of the long-range coulombic potential \( U_{\text{coul}} \) and the short-range Lennard–Jones potential \( U_{\text{LJ}} \). In particular,

\[ U_{\text{coul}} = \sum_{i<j} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|^2} - \frac{b_j r_{ij}}{|\mathbf{r}_i - \mathbf{r}_j|^2}, \]

where \( q_i \) is the charge of the \( i \)th particle and the \( b_j \) is a parameter that depends on the types of interacting pairs. For electron pairs and ion pairs, \( b_j = 0 \). For the interaction between electron and ion with charge \( Z \), \( b_j \) is chosen such at \( -Z/b \) is less than the ionization potential of the deepest bound electron. This is to ensure that the potential can support recombination and at the same time prevent numerical instability resulting from charged classical particles approaching too close. The Lennard–Jones potential is included for atom pairs and it has the form of

\[ U_{\text{LJ}} = \sum_{i<j} \epsilon \left[ \left( \frac{r_m}{|\mathbf{r}_i - \mathbf{r}_j|} \right)^{12} - 2 \left( \frac{r_m}{|\mathbf{r}_i - \mathbf{r}_j|} \right)^6 \right], \]

where \( r_m \) is the distance at which the potential has a minimum value of \( \epsilon \).

At each MD step, an MC procedure is used to incorporate quantum effects via cross-sections and evaluate probabilities for electron ionization, resonant excitation, inner-shell relaxation, electron–atom impact ionization and three-body recombination to occur. The overall transition rate, \( \Gamma \), between different EC I and J is given by

\[ \Gamma_{IJ} = \Gamma_{P}^{P} + \Gamma_{R}^{A} + \Gamma_{R}^{F} + \Gamma_{R}^{E} + \Gamma_{R}^{RC} + \Gamma_{R}^{RE}. \]

Starting from the ground state of the neutral atom, we include the contribution from photoionization \( \Gamma_{R}^{P} \), Auger decay \( \Gamma_{R}^{A} \), fluorescence \( \Gamma_{R}^{F} \), electron–impact ionization \( \Gamma_{R}^{E} \), recombination \( \Gamma_{R}^{RC} \) and resonant excitation \( \Gamma_{R}^{RE} \). In particular, all these
transition rates are calculated using the Hartree–Fock–Slater (HFS) model [78], which has been shown to provide adequate description for multiphoton processes for atoms in intense XFEL pulses [50, 76, 79], including resonant excitation phenomena [55, 56]. The HFS formulation and numerical method are described in Appendix A. Currently, the electronic excitation from Compton scattering is not included as its cross section is typically many orders of magnitude smaller than the photo-ionization process in the considered x-ray photon energy range. Also, molecular effects and environment impacts are not included.

2.1. Treatment of electron-impact ionization

We employ the binary-encounter-Bethe model [80] to compute the electron-impact ionization cross sections. This model has been in previous x-ray studies [61, 81] and it provides a parameterized formula to calculate the cross sections as a function of incident electron energy, \( T \), for all occupied orbitals associated with different EC. For an atom/ion transitioning between ECs \( I \) and \( J \), the ionization cross section of an electron in the \( j \)th orbital is given by

\[
\sigma_{ij}(E) = \frac{S}{t + (u + 1)/n} \left[ \frac{Q \ln t}{2} \left( 1 - \frac{1}{t^2} \right) + (2 - Q) \left( 1 - \frac{1}{t} - \frac{\ln t}{t+1} \right) \right],
\]

where

\[
Q = \frac{2}{N} \int \frac{B_{ij}(\omega)}{B_{ij}^{(s)}} \frac{df}{dW} \, dW
\]

\( t = T/B_{ij}^{(f)}, u = U_{ij}^{(f)}/B_{ij}^{(f)}, S = 4\pi a_0^2 N_{ij}^{(f)} (R_j/B_{ij}^{(f)})^2, a_0 \) is the Bohr radius, \( R \) is Rydberg constant in unit of energy, \( B_{ij}^{(f)} \) is binding energy of the bound electron, \( U_{ij}^{(f)} \) is the orbital kinetic energy, \( N_{ij}^{(f)} \) is the number of electrons in that orbital. For atoms, \( n \) is set to the value of the principal quantum number if it is \( \geq 3 \), otherwise, \( n = 1 \). The rate of electron impact ionization is then given by \( \Gamma_{ij}^{(I)} = \sigma_{ij}(E) n_e v_e \), where \( n_e \) is the number density of electrons and \( v_e \) is the velocity of the incident electron. The product \( n_e v_e \) is the incident electron flux.

The occurrence of electron-impact ionization is determined by first constructing a cylindrical volume with the area of the cross sectional plate corresponding to the sum of all the \((e, 2e)\) cross sections of the occupied orbitals with ionization energy less than the incident electron energy. The velocity vector of the incident electron is normal to these cross sections. As the electron moves this plate draws out a volume. If the cylindrical volume contains an atom or ion, an electron impact ionization can take place [67]. A random number is then generated to select one occupied orbital, in which its probability is weighted by its \((e, 2e)\) cross section.

2.2. Treatment of electron–ion recombination

In an electron–ion recombination process, an electron recombines to the \( j \)th subshell of an ion and gives away its excess energy to a second electron. The generalized cross section of this process that changes the EC of the ion from \( I \) to \( J \) is given by [82]

\[
\sigma_{1j}^{RC}(\varepsilon) = 3\pi^2 \frac{B_{1j}^{(i)}}{g_{1j}^{(i)}} \varepsilon + B_{1j}^{(f)} - B_{1j}^{(i)},
\]

where \( \varepsilon \) is the sum of the energy of the two electrons participating in the recombination process, \( B_{1j}^{(f)} \) and \( B_{1j}^{(i)} \) are the binding energies of the \( j \)th orbital, and \( g_{1j}^{(i)} \) and \( g_{1j}^{(f)} \) are the statistical weights of the ion before and after recombination, respectively. Here the cross section has units of \((\text{area}^2 \text{time})^{-1}\) and is obtained from the detailed balance equation, such that three-body recombination cross section is the reverse process of the electron impact ionization. The recombination rate can be obtained as

\[
\Gamma_{1j}^{RC} = \sigma_{1j}^{RC} n_1^2 v_1 v_2,
\]

where \( v_1 \) and \( v_2 \) are the velocities of the two participating electrons, \( n_e \) is the number density of the electron present in the cluster. Since \( n_e v_1 \) and \( n_e v_2 \) are the flux of the two electrons, by assuming that these two electrons are independent, \( n_e(t) \) can be expressed as \( n_e(t) = \frac{1}{v_1} N_1 \), where \( V_1 = \frac{4\pi}{3} [R_x^3 - R_x^2 R_y + R_x R_y^2] \) and \( \bar{R}_x \) are the positions of the electron and ion, respectively. Our working definition for the number density makes use of the information of three participating particles rather than using the average number density of the cluster, which can underestimate the recombination dynamics in samples that have spatially nonuniform charge distributions and are far from thermal equilibrium. Since equation (8) can account for the recombination of electrons into different subshells, it thus allows the tracking of EC for individual atoms and ions. We assume that recombination is most probable when the two electrons are close to an ion. With this assumption, the probability of recombination of a given ion is considered only for the two electrons with the shortest distances to the ion. Also, the closest electron must have a total energy less than the orbital energy of the orbital selected for recombination in order to avoid artificial enhancements due to \( |\bar{R}_x - \bar{R}_y| \approx 0 \).

A second method to compute recombination rates is based on energy and distance criteria. If the energy of the electron in the considered ion is decreasing and less than the orbital energy of subshell \( j \) with vacancy and the distance between the electron and ion is less than 1 Bohr radius, then the electron can be considered to have recombined with the ion with excess energy transferred to the second nearest electron. A similar working definition developed by Ackad and coworkers has been shown to capture the recombination rate of clusters in EUV regime [83].

To examine the effectiveness of our working definition given in equation (8), we consider the recombination dynamics of a preionized \( \text{Ar}_{1415} \) cluster of singly charged ions. Initially, the cluster is assumed to have a Maxwell–Boltzmann electron temperature distribution of 13 eV. Figure 2 shows the number of recombination events as a function of time obtained from the model by Ackad and coworkers [83] is similar to the standard three-body recombination model given in [84], which assumes no cluster.
expansion. It is clear from figure 2 that the MC/MD calculation, which employed 10 replicas, gives a good agreement with these previous calculations and thus demonstrates the appropriateness of our three-body recombination model. We point out that for the MC/MD calculations presented in section 3, both of these methods give similar recombination rate (see appendix C) and ion yield data with variations less than 10%.

### 2.3. Model of x-ray scattering response

The observed scattering response can be characterized as a sum of the instantaneous scattering patterns weighted by the pulse intensity, $j_k(\tau, t)$ with FWHM duration $\tau$. In the high scattering limit, the scattering signals expressed in terms of the total differential cross section of the cluster can be regarded as the sum of the coherent, free-electron and Compton (inelastic) scattering [85–88]

$$\frac{d\sigma_{\text{total}}}{d\Omega}(q) = \frac{d\sigma_{\text{coh}}}{d\Omega}(q) + \frac{d\sigma_{\text{in}}}{d\Omega}(q) + \frac{d\sigma_{\text{comp}}}{d\Omega}(q),$$  

where coherent scattering can be expressed as

$$\frac{d\sigma_{\text{coh}}}{d\Omega}(q) = \frac{d\sigma_{\text{coh}}}{d\Omega} \frac{1}{F} \int_{-\infty}^{+\infty} dt j_k(\tau, t) F_b(q, t)^2,$$

with $d\sigma_{\text{coh}}/d\Omega$ being the Thomson scattering cross section and $F = \int_{-\infty}^{+\infty} dt j_k(\tau, t)$ the fluence of an XFEL pulse. Here $F_b(q, t)$ is the time-dependent form factor of the bound electrons of the target cluster. Using an independent atom model, it is given by

$$F_b(q, t) = \sum_{j=1}^{N_a} f_j(q, C_j(t)) e^{i q \cdot R_j(t)},$$

where $N_a$ is the total number of atoms/ions, $C_j(t)$ and $f_j(q, C_j(t))$ are the EC and the atomic form factor of the $j$th atom/ion respectively. The momentum transfer vector is defined as $q = 4\pi \sin(\theta/2)/\lambda$, where $\lambda$ is the wavelength of the XFEL pulse and $\theta$ is the scattering angle defined as the angle between the incoming and scattered XFEL beams.

Note that the atomic form factor, $f_j(q, C_j(t))$ for all participating EC $C_j(t)$ is calculated using the HFS model discussed in appendix A.

The free-electron contribution is approximated as

$$\frac{d\sigma_{\text{in}}}{d\Omega}(q) = \frac{d\sigma_{\text{coh}}}{d\Omega} \frac{1}{F} \int_{-\infty}^{+\infty} dt j_k(\tau, t) N_i(t),$$  

where $N_i(t)$ is the number of delocalized electrons within the focal region of the x-ray pulse, and $d\sigma_{\text{coh}}/d\Omega$ is the Klein–Nishina scattering cross section [89]. For an x-ray photon energy that is much less the electron rest mass energy, $d\sigma_{\text{coh}}/d\Omega$ can be approximated by $d\sigma_{\text{in}}/d\Omega$.

The contribution from Compton scattering processes can be given in terms of the inelastic scattering function, $S(q, t)$ [75, 85, 88]:

$$\frac{d\sigma_{\text{comp}}}{d\Omega}(q) = \frac{d\sigma_{\text{coh}}}{d\Omega} \frac{1}{F} \int_{-\infty}^{+\infty} dt j_k(\tau, t) S(q, t),$$

with

$$S(q, t) = \sum_{j=1}^{N_i} s_j(q, C_j(t)),$$

and $s_j(q, C_j(t))$ is the inelastic scattering function of the $j$th atom/ion with EC $C_j(t)$. For biomolecules, the Compton scattering can limit the resolution as its contribution can overtake the contribution from the coherent scattering at high scattering angles [75]. In contrast, Compton scattering is less important in high-Z clusters [77].

### 2.4. Initial cluster configuration and numerical propagation

In typical MC/MD calculations, the initial cluster configuration consists of a collection of neutral atoms, and all the electrons reside with the parent ions and are assumed to be inactive. Starting from a crystalline structure, individual atoms are relaxed under the influence of the lattice potential, but in the absence of an x-ray pulse, to reach a desired temperature, in which their velocities follow a Boltzmann distribution. Since an MC method is used, each MC/MD calculation will require multiple realizations to compute average observables. The phase space of all atoms at different times after equilibration can serve as initial configurations of the cluster for different realizations.

Once the initial configuration is chosen, standard integration algorithms, such as Velocity-Verlet, can be used to integrate Newton’s equations of motion and advance the MC/MD dynamics forward in time. The choice of the time step is important for capturing the many-body correlated dynamics from photoionization, inner-shell and collision events. For hard x-ray simulations, the time step is chosen to be of the order of 1 as, which is small enough such that the distance traveled by the fastest electrons is on the order of typical interatomic separations and it is sufficient to capture the fastest electronic processes. After the EC for each atom in the system has been updated, the algorithm proceeds to the next step of the MD simulation with evaluation of the interaction potential and propagating particles forward in time.
With this MC/MD model, rich datasets are generated that include coherent x-ray scattering patterns, ion, electron and fluorescence spectra. If only scattering patterns are needed, calculations can be terminated three pulse lengths after the peak of the pulse. On the other hand, if fluorescence, electrons and ion observables are of interest, calculations might need to be carried out for many pulse lengths until a steady state is reached. Since the fastest electrons usually escape during the pulse, an adaptive timestep scheme can potentially be used for these calculations, such that the chosen time step will adequately characterize the electron dynamics in the cluster volume. At the end of the calculations, the results from all realizations are averaged.

3. Results and discussions

3.1. Method validation

Methodologies similar to MC/MD methods have been successfully used to model the interactions of $C_60$ [90] and small Ar clusters [35] exposed to intense XFEL pulses, and they were recently extended to study heterogeneous systems [68]. In particular, figure 3 shows that our calculated kinetic energy distribution of the ionized electrons from Ar10000 clusters exposed to an intense 5 keV, 10 fs XFEL pulse has an excellent agreement with the experimental data from SACLA [35], and this thus establishes the validity of our MC/MD method. We note that figure 3 does not include the data from the photoelectrons and Auger electrons from KLL/KLM/KMM since they are not measured experimentally. These electrons have keVs of energy, and they mostly escape the small Ar cluster without having much collisional interactions. Rather, the distribution mostly results from the low Auger electrons via LLM and LMM Auger cascades and the subsequent secondary electrons from collisional ionization events by these Auger electrons, as described in [35]. Before being emitted, these low energy electrons are first trapped by the coulombic field of the charged cluster and contribute to the formation of the nanoplasma. Thus, inclusion of x-ray inner-shell dynamics of many MC is essential to capture the width of the distribution.

We note that the calculated distribution includes a spatial intensity averaging effect, in which 1000 independent cluster calculations were performed with the initial positions of individual clusters randomly distributed within the interaction volume defined by the intersection of the XFEL beam (1 μm focal size) with the cluster jet (2 mm diameter). In the focal region, a maximum fluence close to 50 $\mu$J $\mu$m$^{-2}$ is obtained. But, the majority of clusters lie outside of the x-ray focus and are exposed to much lower fluence (access to smaller number of ECs). The high number of MC/MD runs serves not only for the spatial intensity averaging, but more importantly it is needed to sample the large number of EC potentially accessed in the high-fluence region.

3.2. The role of electron–ion recombination

Due to the spatial averaging effect, it is not trivial to extract dynamical signatures associated with interactions between the high-intensity XFEL pulse and cluster. But, with the recent advance of the single-shot single-particle imaging technique [27], one can now correlate the measured signals of individual clusters with known exposure intensity and follow the induced dynamics with atomic resolution both in time and space. We investigate the role of electron–ion recombination processes in XFEL excitation of single Ar1415 clusters (26 885 nuclei + electrons) in the high-intensity limit. The XFEL parameter sets analyzed fall within the SPI parameters sets (10$^{12}$ photons, fs pulse duration, 3–8 keV photon energy, <1 μm focus) which have been have been identified as advantageous to record images with 3 Å resolution [12]. For the results presented in this section, 100 replicas are used to capture the response for each set of pulse parameters.

We first examine the ultrafast evolution of the transient electronic dynamics of an Ar1415 cluster exposed to a 5 keV, 2 fs pulse with a peak fluence of 4 × 10$^{12}$ photons $\mu$m$^{-2}$. This peak fluence can be obtained by focusing a 3.7 mJ XFEL pulse into 1 $\mu$m$^2$ sectional area of intersecting target beam and gives an intensity of 1.85 × 10$^{20}$ W cm$^{-2}$. A tighter focus of 100 nm produces a 4 × 10$^{14}$ photons $\mu$m$^{-2}$, which corresponds to an intensity of 1.85 × 10$^{22}$ W cm$^{-2}$. We note that the exposed fluence is close to 80 times higher than that shown in figure 3, giving rise to a more complex electronic excitation and relaxation landscape. The complexity is found already in the early ionization stage. In particular, only about 42% of the created K-shell core hole is filled by Auger (30.7% for K-LL, 5.0% for K-LM and 0.3% for K-MM) and fluorescence (5% for $K_\alpha$, 0.5% for $K_\beta$) events. The rest of the population with a single K-shell hole undergoes further photoionization to produce hollow Ar atoms with empty K-shell electrons (47% branching ratio) and an additional vacancy in the L or M shell (11% branching ratio). Our calculation shows that similar branching ratios are found in the early ionization stage of isolated Ar atoms exposed to the same pulse parameters.

However, unlike the case of an isolated atom, the Ar cluster is also charged by electron-impact ionization (EI) and electron–ion recombination events as depicted in figure 4. We
see that multiple photon absorption processes lead to rapid charging of the cluster, rearrangement of a large number of electrons, and displacement of ions/atoms within femtoseconds. For \(-4 \text{ fs} \leq t \leq 4 \text{ fs}\), figure 4(a) illustrates that the energetic photoelectrons and Auger electrons escape the cluster volume while an increasingly large Coulombic field is being developed. This field generates a large attractive force that traps slow Auger and secondary electrons even after the pulse is gone. At the same time, the field provides a repulsive force that leads to particle volume expansion and full cluster disintegration. This cluster expansion mechanism, however, is strongly dependent on the nature of the participating quantum electronic transitions at the microscopic level. On average, our MC/MD model predicts that each atom absorbs between 5 and 6 photons, undergoes about 9 Auger decays, 4 electron impact ionization events, more than 13 recombination events and emits less than 1 fluorescence photon, as shown in figure 4(b). We note that this combination of events gives rise to complicated ionization/relaxation pathways involving more than 1000 different EC. We will discuss later in this section that these transient EC have direct implications on the calculated fluorescence.

3.2.1. Ion yield data. By exploiting the atomistic capability of the MC/MD method, we can track and compare subgroups of ions/atoms and electrons based on their initial type, EC, position, and other properties. Figure 5 examines the time-dependent average charge state for atoms originating from different geometric shells of \( \text{Ar}_{1415} \) which initially had an icosahedral structure. Due to the x-ray pulse’s high transparency, multiphoton absorption processes charge the system uniformly within the cluster volume before reaching the peak intensity of the pulse. The developed Coulombic field traps electrons within the clusters. Surprisingly, shortly after the peak of the pulse, the average charge states of the first 5 shells drop significantly. For example, the average
charge state of the ions from the 4th geometric shell reduces to $1^+$ after peaking at $5^+$. This charge state reduction provides a clear signature of electron–ion recombination. Interestingly, while these inner cluster shells undergo recombination, ions from the two outermost shells continue to be charged. As a result, an excited cluster transient with a highly charged surface and a core with low charge is formed as shown in figure 5(e). This nonuniform spatial charge distribution means that disintegration of the cluster occurs with fast ablation of surface layers, which have large Coulomb explosion energy, followed by a slow expansion of the cluster core. The resulting final ion yield data shows a broad charge state distribution in figure 5(g).

To highlight the importance of electron ion recombination processes in intense x-ray interactions with Ar clusters, we perform a ‘No RC’ calculation, i.e. a calculation without recombination, using the same pulse parameters and initial cluster conditions. Figure 5(b) shows that the cluster in the ‘No RC’ calculation exhibits a very different time-dependent charging profile. In particular, an opposite space charge distribution is found with the ions from the outer shell having a lower average charge than those from the core. Also, the obtained ion distribution is much narrower and peaks at $13^+$ to gives an average charge of 12.6, which is about $5^+$ higher than that from the RC calculation. To further support the observed features, we examine the cluster dynamics with a 2 fs, 8 keV pulse with the same pulse energy, but with a peak fluence of $2.6 \times 10^{12}$-photons $\mu$m$^{-2}$. Due to the lower absorption cross section at 8 keV, the degree of cluster ionization is less, resulting in a slower disintegration timescale, but a similar

Figure 5. (top panel) Time-dependent average charge states for atoms originating from different shells, (middle panel) positions and charge states of atoms/ions (electrons are not shown) at $t = 8$ fs, and (bottom panel) yield of ion at $t = 1$ ps calculated for an Ar$_{1415}$ cluster in a 2 fs, 3.7 mJ $\mu$m$^{-2}$ pulse. These three quantities are shown for three cases: 5 keV pulse with recombination (5 keV, RC), without recombination (5 keV, no RC) and 8 keV pulse with recombination (8 keV, RC).
nonuniform spatial charge distribution due to electron ion recombination processes is evident (see figures 5(c), (f) and (i)). We note that the effect of recombination in producing highly charged layers and neutral cores has been demonstrated in the infrared, XUV and soft x-ray regimes [24, 38, 45, 91, 92]. However, the obtained timescales of these electronic structure transformations in the Ar cluster exposed to an intense XFEL (>10^{12} photons μm^{-2}) are on the order of tens of femtoseconds, which are much shorter than the picosecond timescales observed previously in XUV [36] or x-ray pulses [35]. At higher fluence levels, one would expect that the distribution will peak at or near the highest charge states, as shown by [27, 45] (see appendix D for the distribution calculated for 10^{14} photons μm^{-2}).

We note that our analysis of the ion charge state does not distinguish between delocalized and localized electrons. In the calculation by Arbeiter and coworkers [36], a large fraction of localized electrons, i.e. classical electrons that are bound by an ion, were found in clusters exposed to an XUV pulse. These electrons were then removed from the calculation. To estimate the number of localized electrons produced by hard x-ray pulses, we performed the same analysis used in [36] on our RC and No RC calculations. We found that less than 0.1% of the ionized electrons are localized during and after the pulse due to the fact that the generated electrons from the hard x-ray pulse are much more energetic than those generated in the XUV field.

3.2.2. Fluorescence spectra. The power of the MC/MD method is that all electron, ion and photon observables can be computed simultaneously within the same calculation. In particular, the scattering data probe the short-time dynamics during the laser pulse, whereas the electron/ion and fluorescence spectra contain dynamical information over a much longer timescale. In particular, fluorescence spectra have been shown to be a useful spectroscopic tool for identifying transient electronic states [33, 41, 42] in the cluster excitation dynamics. Figure 6(a) shows the fluorescence spectrum obtained for an Ar cluster exposed to a 2 fs, 5 keV pulse with a fluence of 3.7 mJ μm^{-2}. For comparison, the fluorescence spectra for the Ar cluster without recombination dynamics and an isolated Ar atom exposed to the same pulse parameters are also shown in figure 6. To highlight the interesting features, the fluorescence spectra depict only a limited range of energies focusing on Kα, Kα satellites, hypersatellite (Kα^h) and its satellites (Kα^h(L^{-1}) and Kα^h(L^{-2}) shown in figure 6(a)). We see that the Ar cluster spectrum obtained when including recombination is very different from the spectra for an isolated atom and the cluster without recombination. Interestingly, the Ar cluster spectrum depicts stronger fluorescence lines for Kα and Kα^h. Moreover, their satellites (Kα(L^{-1}), Kα^h(L^{-1}) and others), which are negligible in the atomic spectrum, become very pronounced in the Ar cluster spectrum. Also, there are more Kα^h events than Kα events. Our analysis shows that recombination processes are an important contribution to these fluorescence features. In the case of Kα^h, which are generated from the decay of doubly charged Ar ions with hollow K shell, our analysis show that there are additional pathways for producing these Ar^{2+} beyond sequential photoionization pathways. In particular, those ions with EC 2s^22p^63s^33p^6, 2s^22p^63s^23p^6, 2s^22p^63s^23p^5 and 2s^22p^63s^23p^5 can undergo recombination to produce Ar^{1+} with double K-shell holes by filling the vacancy in the L or M shell. In fact, these recombination processes account for close to 40% of the pathways to produce this particular ion. We note that the fluorescence decay of 1s2s^12p^63s^23p^6 and 1s2s^22p^63s^23p^6 can give rise to Kα(L^{-1}) line, and there are recombination pathways leading to these two EC. This further highlights the capability of our MC/MD method in examining fluorescence pathways in high-intensity x-ray interaction with nanoclusters by tracking EC explicitly.

3.2.3. Scattering patterns. In addition to the ion yield data and fluorescence data, the MC/MD method can also provide scattering data. Having these rich data sets are useful in understanding the recently developed single-shot single-particle imaging [27, 45]. In this technique, the recorded scattering pattern is used to infer the exposed intensity and cluster size, such that the ion observables from a single cluster in different pulse parameters can be obtained. In order to
understand the applicability of this intensity and size assignment method, we compared the scattering patterns (RC data) of the same Ar cluster exposed to different fluences and photon energies. To show the impact of pulse fluence on the cluster electron density profile, we compute the differential cross sections rather than the scattered photon number. Figures 7(a) and (c) show that the scattering data obtained with 2 fs, 5 keV pulses at the fluence levels of $4.1 \times 10^{12}$ and $4.1 \times 10^{14}$ photons $\mu$m$^{-2}$ are very different, in which a 100-fold increase in fluence results in lower fringe visibility and reduces the differential cross section by more than an order of magnitude. This is due to the fact that a higher fluence pulse leads to an increase of delocalized electrons and loss of electrons in the cluster volume, which results in a smaller scattering cross section. Similar trends are also shown in figures 7(b) and (d) for 2 fs, 8 keV pulses at fluence levels $2.6 \times 10^{12}$ and $2.6 \times 10^{14}$. The sensitivity of the scattering signals with respect to the pulse fluence suggests a different intensity assignment scheme may be needed. On the other hand, the location of the first minimum in the interference fringes remains the same, which suggests that it is a robust feature to infer the field-free particle size, at least for the maximum fluence level considered.

In addition, figure 7 also compares the differential scattering cross sections calculated with and without recombination processes. We see that the ‘No RC’ scattering data do not deviate substantially from those data with recombination included for the fluence level close $10^{12}$ photons $\mu$m$^{-2}$, unlike the ion yield data. This is due to the fact the calculated scattering is a pulse-weighted quantity and the recombination process does not play an important role during the pulse, as shown in figure 4(b). However, for the 5 keV pulse at $4.1 \times 10^{12}$ photons $\mu$m$^{-2}$, the RC and No RC data are very different, suggesting that substantial recombination dynamics is induced during the pulse due to very high number of delocalized electrons being created.

3.3. Algorithms and performance

The MC/MD method presents a simpler framework for treating the direct coupling of electronic and nuclear degrees of freedoms of many-electron systems explicitly in ultrafast timescales, in comparison to the full quantum treatment, which is impossible to solve. Despite its simplicity, computation with the MC/MD method can still be challenging as both MC and MD computations must be evaluated at each step in the simulation. In each MC step, the work needed to determine the quantum probability of photoabsorption and inner-shell ionization scales linearly with $N_a$, the number of atoms (ions). On the other hand, the effort to determine collisional ionization and recombination scales with $N_a^2$, the number of ion–electron pairs. We point out that this scaling is achieved since recombination of a given ion is considered only for the two electrons with the shortest distances to the ion. On the other hand, the effort to determine collisional ionization and recombination scales with $N_t$, the number of active particles in the system.

Given the randomness inherent to an MC algorithm, it is
necessary to simulate statistically independent replicas for each system, such that important transient ECs can be sampled and statistically converged physical observables computed. Typically, calculations involving high Z elements, resonant excitations or high-intensity pulses will require a larger number of replicas, \( N_r \), to sample the larger number of relevant ECs. All of these factors combine to create a need for an efficient and scalable implementation for computation on both small- and large-scale computing resources.

The high number of force calculations suggests that MC/MD calculations will benefit from using an efficient MD implementation coupled with high-performance computer architectures. The hybrid MC/MD method developed in this work was implemented in the LAMMPS [93] molecular simulation code. LAMMPS is a well-written modular (C/C++) code that supports parallelization with MPI and OpenMP on CPUs (and GPUs) and can be easily extended with new algorithms without significantly degrading parallel performance of the base code. This enables the MC/MD simulations to take advantage of new methods and algorithms added to the base LAMMPS code (e.g. load-balancing particles across processes) with little to no modification of the MC/MD algorithm. The bulk of the LAMMPS fix written contains all of the essential physics associated with particle interactions with the x-ray pulse and executing the MC portion of the algorithm sampling all electron processes in figure 1. The current implementation is equally suited for calculations of small systems on personal laptops to larger systems on large-scale computing resources. In fact, many of the results discussed for \( \text{Ar}_{1415} \) in the previous sections were performed on laptops and workstations. For the Argon clusters examined here, the MC portion of the algorithm amounted to roughly 5%–10% of the total runtime, a nominal overhead considering the complex physics introduced. So far, we have assumed that all the atomic data needed during each MC step can be predetermined and stored. For calculations that involve a large number of ECs that would exceed the available per-node memory to store, like resonant excitations and high Z systems, the required atomic data will be recomputed and updated on the fly populating look-up tables. To achieve good computational performance with the MD portion of the algorithm, especially for increasingly large system sizes, there are a number of considerations to be mindful of. The simulation timestep must be set small enough in order to accurately integrate the classical equations of motion for both fast moving electrons and relatively slow ions. Additionally, unlike typical molecular simulations, clusters irradiated with intense x-ray pulses in MC/MD simulations result in creation of many new particles (excited electrons), rapid expansion of the region containing particles (thousand times larger than volume of original system), and highly nonuniform particle distributions due to escape of fast electrons and cluster disintegration. This requires a simulation domain that is sufficiently large to minimize possible finite size effects and an optimal distribution of particles across processes for efficient parallel computation.

Performance measurements of the MC/MD implementation in LAMMPS is reported in figure 8 for Argon clusters ranging from 0.24 to 8.16 million particles (ions + electrons) on Mira, a 10-petaflops IBM Blue Gene/Q supercomputer at ALCF. Simulation productivity is reported in units of \( 10^6 \) timesteps per day. For simulations using an attosecond timestep, one is able to generate picosecond long trajectories for the range of cluster sizes considered with a reasonable time-to-solution. For each data point in figure 8, there is the optimal combination of MPI ranks and OpenMP threads was determined. In all cases, a total of 64 processes per node was found to be optimal and usually in the combination of 8 MPI ranks and 8 OpenMP threads per rank. For the largest node-count considered here, this corresponds to a maximum of 262 144 processes. For the smallest Argon cluster (0.24 M), the strong scaling parallel efficiency on 64 nodes is 58.3% relative to 16 nodes with improved productivity observed up to 256 nodes at more than 2 million timesteps per day. For the largest Argon cluster considered, a parallel efficiency of 63.2% is observed on 2 048 nodes relative to 256 nodes with a productivity of nearly 1 million timesteps per day generated on 4 096 nodes.

One of the primary sources for loss of parallel efficiency in these cluster systems is due to the fact the spatial decomposition scheme [93] adopted in LAMMPS does not efficiently account for the rapid changes in large spatial inhomogeneities in particle distribution throughout the simulation cell created by ultrafast ionization. During a simulation, this gives rise to imbalanced workloads of pairwise force computation across MPI processes. The optimal ratio of MPI and OpenMP processes are used on each node count, such that a total of 64 processes per node (i.e. there are 262,144 processes running on 4,096 BG/Q nodes).

![Figure 8](image_url)

**Figure 8.** (a) Productivity and (b) parallel efficiency of MC/MD calculation as a function of compute nodes for Argon clusters containing 0.24–8.16 million particles (ions and electrons). The optimal ratio of MPI and OpenMP processes are used on each node count, such that a total of 64 processes per node (i.e. there are 262,144 processes running on 4,096 BG/Q nodes).
outperform the pairwise calculation from the spatial decomposition scheme once the Argon cluster increases beyond one million particles.

4. Summary and outlook

Theoretical investigation of the intense x-ray interaction of nano-sized systems in the x-ray regime is challenging. It requires the description of multi-photon physics, ultrafast transient electron dynamics involving a large number of EC, and nonequilibrium nanoplasma dynamics from formation to disintegration that spans multiple time scales (sub-femtoseconds to nanoseconds) and length scales. This paper outlines the hybrid MC/MD theoretical method suitable for investigating the complex interaction of intense x-ray pulses with nanosized clusters, both homogeneous and heterogeneous in composition. The power of the MC/MD method lies in its ability to provide atomistic tracking of the time-dependent x-ray induced electronic excitation/relaxation dynamics in atoms and the subsequent highly excited transient cluster dynamics with attosecond time resolution. Currently, the combined effects of eight different electronic processes can be accounted. We show the importance of including the electron–ion recombination process in calculations of ion charge state distributions and scattering signals, even for a 2 fs pulse, where it was observed to be responsible for the formation of a nonuniform spatial charge density profile, a lowly ionized core and highly charged shell layers, in the x-ray excited cluster over femtosecond timescales. Also, the Ar cluster fluorescence spectrum is very different from the Ar atom spectrum, in which recombination processes give enhanced fluorescence signal near the $K_{\alpha}$ region by producing additional pathways to reach the required EC for fluorescence transitions in an intense 2 fs XFEL pulses.

The multiscale atomistic tracking enabled by the MC/MD calculations is computationally intensive, especially for systems involving heavy elements and large number of particles (ions and electrons). Several algorithms, which include spatial decomposition and MSM method for electrostatics, have been employed within the LAMMPS molecular simulation code to take advantage of large-scale computing resources. We show that picosecond calculations with attosecond time resolution for 8-million particle systems can be completed within a single day using upwards of a quarter million processes on Mira (IBM BG/Q) at ALCF. Also, the MSM method was observed to outperform the pairwise potential with long cutoff scheme for systems containing millions of particles. Efforts to continue developing and improving advanced methodologies and algorithms (e.g. electronic structure, parallelization schemes, evaluation of Coulomb interactions, multiscale dynamics), and multi-threaded performance will further allow the MC/MD method to harness the computational power from next generation HPC resources to systematically investigate the induced dynamics in finite-sized systems with increasing complexity at the ultrafast timescale and in the non-linear x-ray interaction regime.

Acknowledgments

The authors thank Christoph Bostedt, Wei Jiang, Sang-Kil Son and Linda Young, for fruitful discussions. This work was supported in part by the US Department of Energy, Office of Science, Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division under Contract DE-AC02-06CH11357. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357, through ALCF DD projects and a 2016 ALCC award.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (‘Argonne’). Argonne, a US Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The US Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

Appendix A. HFS model

Consider a multi-electron atom with an EC given by a set of occupation numbers $N^{(i)} = \{N^{(1)}, N^{(2)}, \ldots, N^{(M)}\}$ in $M$ sub-shells. Here, $i$ is the label of this EC and $N^{(j)}$ is the number of electrons occupying the $j$th subshell. It is clear that the total number of electrons in this atom is related to $N^{(j)}$ via

$$N_{\psi}^{(i)} = \sum_{j=1}^{M} N^{(j,i)}.$$  \hfill (A.1)

We describe the electronic structure of this atom using the HFS model [95], which is a simplified version of the Hartree–Fock model with the exchange potential being approximated in the framework of a free-electron gas. This model is effectively a one-electron theory with central field approximation, in which a set of individual one-electron wave functions in the $i$th orbital with quantum numbers $n_{\ell m}$ can be written as

$$\psi_{n_{\ell m}}^{(i)}(r) = \frac{u_{n_{\ell m}}^{(i)}(r)}{r} Y_{\ell m}(\Omega),$$ \hfill (A.2)

where $u_{n_{\ell m}}^{(i)}(r)$ is the radial wave function and the $Y_{\ell m}(\Omega)$ are the spherical harmonics. The radial wave functions of these orbitals
can be obtained by solving the radial Schrödinger equation
\[
\left[ -\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l + 1)}{2r^2} + V(r) \right] u_{nl}(r) = E_{nl}(r)u_{nl}(r),
\]
(A.3)
where \(E_{nl}(r)\) is the orbital energy and
\[
V(r) = \begin{cases} V_0(r), & r < r_0, \\ (-Z + N_i - 1)/r, & r \geq r_0, \end{cases}
\]
is the central potential and it includes the Latter correction [96] for \(r \geq r_0\) in order to account for the self-Coulomb potential and to give a correct asymptotic description for the one-electron wave function at large \(r\). To ensure continuity of the potential, \(r_0\) is the distance \(r\) at which \(V_0(r) = (-Z + N_i - 1)/r\) for the \(r < r_0\), the potential is given by
\[
V_0(r) = -\frac{Z}{r} + \int_{r}^{\infty} dr' \rho(r') \frac{1}{|r - r'|} + V_{ex}(r).
\]
The first term is the Coulomb potential from the nucleus with charge \(Z\). The second term is the total electronic Coulomb potential and the third term is the exchange potential. In our HFS model, the form of the exchange potential proposed by Slater [97] is used where
\[
\rho(r) = \frac{1}{4\pi r^2} \sum_i N_i |u_{nl}(r)|^2,
\]
(A.7)
where \(N_i\) is the occupation number of the \(i\)th subshell with quantum number \(n, l_i\).

We model the photo-ionization process in a multi-electron system as a dipole transition. Using the first-order perturbation and following the \(LSJM\) scheme given in [98], the cross section of an electron in \(i\)th orbital being ionized by an x-ray photon with energy \(\omega_X\) is given by [95, 98–103]
\[
\sigma_p^{(i)}(\omega_X; i) = \frac{4\pi^2 \alpha N_i |\omega_X|}{3(2l_i + 1)} \left[ l_i |\tilde{R}_{l_i-1}(\epsilon)|^2 + (l_i + 1)|\tilde{R}_{l_i+1}(\epsilon)|^2 \right],
\]
(A.8)
where \(\alpha\) is the fine structure constant,
\[
\tilde{R}_{l_i-1}(\epsilon) = \int_0^\infty u_{nl}^{(i)}(r)\tilde{u}_{n-l}^{(i)}(r) r dr
\]
(A.9)
and \(\epsilon\) is the energy of the photoelectron with a radial continuum wave function \(\tilde{u}_{nl}(r)\) and angular momentum \(l_i\). From conservation of energy, \(\epsilon = \omega_X + E^{(i)}_{nl}\). For a given x-ray flux \(\dot{j}_{X}\), the rate of the photoionization that leads to change of EC from \(I\) to \(J\) can be computed as
\[
\Gamma_{I,J} = \sigma_p^{(i)}(\omega_X; i)\dot{j}_{X}.
\]
(A.10)
In our treatment of resonant excitation, we include only electric dipole transitions. An electron is resonantly excited when the difference between the incident photon energy and the transition energy is within the bandwidth of the pulse. The resonant excitation cross sections convolved with a Gaussian XFEL pulse with a bandwidth of \(\delta \omega\) can be expressed as
\[
\sigma_{RE}^{(i)}(i, j) = \frac{2\pi^2 \alpha}{\delta \omega} f_{abs}^{(i)}(i, j),
\]
(A.11)
where \(\alpha\) is the fine structure constant and \(f_{abs}^{(i)}(i, j)\) is the absorption oscillator strength for making a transition from a lower state \(j\) to an upper state \(i\), and it can be calculated as [104]
\[
f_{abs}^{(i)}(i, j) = \frac{2}{3} \frac{(I_{i} - E_{j})\tilde{N}_{j}}{2(l_i + 1)(2l_i + 1)} |R_{ij}|^2.
\]
(A.12)
Here, \(E_{i}\) and \(E_{j}\) are the eigenenergies of \(i\)th and \(j\)th subshells respectively, \(N_j\) is the number of electrons in state \(j\), \(N_i = 4l_i + 2 - N^{(i)}_{j}\) is the number of available vacancies in state \(i\), \(l_i\) and \(l_j\) are the \(l\) quantum number of state \(i\) and \(j\), respectively, \(l_s = \max(l_i, l_j)\) and
\[
R_{ij} = \int_0^\infty u_{nl}^{(i)}(r)\tilde{u}_{n-l}^{(i)}(r) r dr.
\]
(A.13)
Since Rydberg atoms/ions usually have long lifetimes, we account for the possibility that they can undergo further sequential, multiple resonant excitation in addition to photo-ionization and inner-shell decay. Similar to the photoionization rate, the rate of the resonant absorption can be computed as
\[
\Gamma_{I,J}^{RE} = \sigma_{RE}^{(i)}(\omega_X; i)\dot{j}_{X}.
\]
(A.14)
An excited atom has a finite lifetime. The atom can decay through a fluorescence process with a photon being released to carry away the excess energy. In this paper, the fluorescence process is treated as a dipole transition, in which a hole vacancy in the \(n,l\) shell is filled by an electron from the \(n,l+1\) shell. Using first order perturbation theory the fluorescence rate can be calculated as [105]
\[
\gamma_{F}^{(i)}(i, j) = \frac{2\alpha^3}{3} f_{em}^{(i)}(i, j),
\]
(A.15)
where \(f_{em}^{(i)}(i, j)\) is the emission oscillator strength
\[
f_{em}^{(i)}(i, j) = -f_{abs}^{(i)}(i, j).
\]
(A.16)
In addition to photon emission, the excess energy can be partially or fully released through a radiationless process that ejects an electron into the continuum. Depending on the EC, this radiationless process can be classified as one of three types: the Auger [106], Coster–Kronig [107] or auto-ionization [108]. This radiationless process is due to the electrostatic interaction between two electrons that leads to rearrangement of EC in the excited atom. From first-order perturbation theory the Auger decay rate can be expressed as
\[
\gamma_{A}^{(i)}(i, j') = 2\pi \left| \sum_{l, j} \left( \Phi_j(\{r\}) \right) \sum_{l, j'} \left( \Phi_j(\{r\}) \right) \right|^2.
\]
(A.17)
Using the LSJM scheme [109], the total Auger (decay) rate of an \( n_l f \) electron filling a vacancy in \( n_l f \) subshell leading to ejection of an electron from \( n_l f \) subshell to a continuum with energy \( E \) without disturbing the angular momentum state of the Auger electron, \( l_f \), is given by

\[
\gamma_A^{(j)} (i, j'; i') = \sum_{K} \sum_{l_f} \sum_{l_i} \sum_{l_j} \frac{N_{ij} \times (2S + 1)(M_{ij} (j, j', i, i'))^2}{2(l_i + 1) L_{l_f} (l_f - l_i)}.
\]  

(A.18)

where \( N_{ij} \) is the weighting factor,

\[
N_{ij} = \begin{cases} \frac{N_{ij}}{4l_f + 2(4l_f + 2)}, & j = j', \\ \frac{N_{ij}}{4l_f + 2(4l_f + 2)}, & j = j', \end{cases}
\]  

(A.19)

and

\[
M_{ij} (j, j', i, i') = (-1)^{I_{l_f}} \sum_{K} R_K (j, j', i, i') A_K (j, j', i, i')
\]  

(A.20)

with

\[
R_K (j, j', i, i') = \int_{0}^{\infty} \int_{0}^{\infty} \frac{r^K}{r^{2l_f + 1}} u_{n_l f} (r_1) u_{n_l f} (r_2) \times u_{n_f} (r_1) u_{n_f} (r_2) dr_1 dr_2,
\]  

(A.21)

and

\[
A_K (j, j', i, i') = \langle l_i || C_{K} || l_f \rangle \langle l_f || C_{K} || l_j \rangle \times \begin{pmatrix} l_i & l_f & K \\ l_j & l_f & L \end{pmatrix}.
\]

(A.22)

Here, \( \langle l_i || C_{K} || l_f \rangle \) is the Clebsh Gordon coefficient and can be written as

\[
\langle l_i || C_{K} || l_f \rangle = \frac{(-1)^{l_f}}{\sqrt{(2l_i + 1)(2l_f + 1)}} \begin{pmatrix} l_j & K \\ l_f & L \end{pmatrix}.
\]

(A.23)

Note that the Auger rate is a total rate of various final configurations (\( K \) sum in equation (A.20)) and it is averaged over various initial configuration (\( L \) and \( S \) sum in equation (A.18)).

### A.1. Numerical grid

To calculate the radial wave functions and the energies of the bound state orbitals, we diagonalize equation (A.3) on a Gauss–Lobatto grid [110–114], in which its grid points are the roots of the first derivative of the \( N \)th order Legendre polynomial \( (P_N (x) = 0) \). We map these roots, which lie on the interval \( x \in [-1, 1] \), onto the radial space of the atom [111] from \( r = 0 \) to \( r_{\text{max}} \) using

\[
r (x) = \frac{r_{\text{max}}}{2} \left( 1 + \frac{1 + x + \zeta}{2} \right).
\]

(A.24)

where \( \zeta \) is a parameter that controls the density of the mapped points near the origin. Typically, depending on the atom and its charge state, \( \zeta \) in the range of 0.001–0.5 is used to give a higher density of grid points near the origin since the potential drops off rapidly.

With the mapping of equation (A.24), the radial Schrödinger equation in equation (A.3) is transformed to

\[
\left[ -\frac{1}{2} \frac{d^2}{dr^2} + \frac{1}{r^2} + W_i (x) \right] \phi_{nl}^{(i)} (x) = E_{nl}^{(i)} \phi_{nl}^{(i)} (x),
\]

(A.25)

where \( r' (x) \) the derivative of \( r (x) \) with respect to \( x \), and

\[
W_i (x) = \frac{l(l + 1)}{2r^2} + V (r (x))
\]

(A.26)

and

\[
\phi_{nl}^{(i)} (x) = \sqrt{r'} (x) u_{nl}^{(i)} (r (x)).
\]

(A.27)

Using the Gauss–Lobatto grid, we can arrive at a discretized radial Schrödinger equation

\[
\sum_{j=0}^{N} \left[ -\frac{1}{2} \frac{d^2}{dx^2} + W_j (x) \right] \phi_{nl}^{(i)} (x) = E_{nl}^{(i)} \phi_{nl}^{(i)} (x),
\]

(A.28)

where

\[
\phi_{nl}^{(i)} (x) = \frac{1}{P_N (x)} \phi_{nl}^{(i)} (x),
\]

(A.29)

and

\[
d_{nl}^{(i)} (-) = \frac{N (N + 1)}{3(1 - x_i^2)},
\]

(A.30)

For our bound state calculation, 800 grid points that reach \( r_{\text{max}} = 50 \) a.u. are used to obtain a set of radial wave functions with quantum number \( n \) going from 1 to 10 and \( l \) from 0 to 4. So, a total of 40 subshells. The highest Rydberg level is 10 g. These maximum \( n \) and \( l \) values are found to be sufficient to account for the ionization dynamics of Ar in the range of pulse parameters investigated in this paper and give a converged value for each of the observables of interest. Increasing these values, which leads to longer computational time, does not change the values of the observables significantly. By diagonalizing the time-independent Schrödinger equation in equation (A.28), we obtain the wave functions and orbital energies of the 40 subshells. These orbitals are then arranged in energy ascending order.

On the other hand, we get the continuum wave function \( \tilde{u}_{nl}^{(i)} (r) \) for a photoelectron with energy \( \epsilon \) and angular momentum \( l \) by numerically integrating the radial Schrödinger equation

\[
\left[ -\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l + 1)}{2r^2} + V (r) \right] \tilde{u}_{nl}^{(i)} (r) = \epsilon \tilde{u}_{nl}^{(i)} (r),
\]

(A.31)

via the Numerov method on a dense grid with 50 000 points reaching \( r_{\text{max}} = 50 \) a.u., as in the bound states. The Numerov method is more efficient than the Runge–Kutta algorithm as it requires fewer functional evaluations to achieve the same accuracy. To energy normalize this continuum wave function, we follow the algorithm of Cooper and Manson [101, 102].
Using the computed bound and continuum wave functions, various excitation and relaxation matrix elements can be obtained. In particular, the absorption cross sections and fluorescence rates are computed on the Gauss–Lobatto grid since only bound orbitals are involved. As for the computation of ionization cross sections and Auger rates, which need both the bound and continuum state orbitals, an interpolation is employed since two different numerical grids are used.

Appendix B. Parallel efficiency of MC/MD simulations

One of the primary sources for loss of parallel efficiency in these cluster systems is due to a load-imbalance of particles across MPI ranks. In parallel computations, LAMMPS makes use of a spatial domain decomposition scheme [93] to efficiently distribute particles and compute pairwise interactions. In this scheme, the 3-d spatial volume occupied by the Argon cluster in these MC/MD simulations is partitioned into non-overlapping sub-domains, such that each processor is responsible for calculating the forces of individual particles within its sub-domain. For cluster calculations, this means that processors responsible for sub-domains near the edges of the simulation domain will contain very few, if any, particles and thus largely sit idle during most of the calculation. Furthermore, the ultrafast ionization resulting from interactions with the x-ray pulse create large spatial inhomogeneities in particle distribution throughout the simulation cell during the pulse. During a simulation, this gives rise to imbalanced workloads across MPI processes and is the primary reason for performance loss in the strong scaling limit. One way to reduce the bottleneck is to incorporate dynamic load-balancing schemes, such as those implemented in LAMMPS, that update the boundaries of individual sub-domains to reduce the variation in the number of particles per processor. For the 0.24 M cluster on 256 nodes (2048 MPI ranks), 44% of the MPI ranks do not have particles within their respective sub-domains. With load-balancing, this is improved to only 6% of MPI ranks having empty sub-domains, but the issue persists and increases (e.g. 10% of 16 384 MPI ranks for 2.17 M cluster). An additional complication for load-balancing MC/MD simulations is the rapid change in the number of active electrons within rather small regions of space that would require load-balancing to be performed at every single simulation step, particularly while the x-ray pulse is incident on the cluster, which would further degrade performance. LAMMPS and the MC/MD fix can both take advantage of OpenMP parallelism to use threads to compute interparticle interactions and updating EC via the MC procedure (e.g. selection collisional ionization and recombination events), which partially alleviate the load-balancing issues encountered. It should be stressed that these load-balancing performance concerns are largely relevant in the strong-scaling limit and are less of an issue for smaller scale calculations. In fact, the results presented for the smaller cluster Ar$_{4145}$ in the main text were performed on laptops and workstations with upwards of ten cores using the same MC/MD code in LAMMPS. For these smaller calculations, which do not require MPI parallelization, load imbalance is not an issue and, for similar reasons, these smaller calculations benefit from the multi-threaded OpenMP parallelization.

For the simulation productivity reported in figure 8, the Coulombic interactions were evaluated using real-space pairwise interactions with a long cutoff. However, in the current spatial decomposition implementation, such a long cutoff for the pairwise potential can significantly increase the communication costs between neighboring processors due to the larger number of particle neighbors that must be tracked. Due to the open boundary conditions of the simulation domain, typical periodic methods for long-ranged electrostatics based on Ewald summation [115, 116] are not appropriate. A recently developed MSM [94] is, however, able to support open boundary conditions and was found to have a parallel performance competitive to PPPM on large numbers of computer nodes. The main features of the MSM method are that it is a linear scaling algorithm and separates the electrostatic interaction potential into short-range part plus a long-range, slowly varying part. The potential advantage of the MSM method over the truncated Coulomb potential is the use of considerably shorter pairwise cutoffs and that all processors would contribute to the calculation of the long-range term improving overall utilization of all processes.

We have added appropriate extensions for the MSM method in LAMMPS to support the MC/MD simulations in order to investigate its computational performance for the Argon clusters. Figure 9 shows the performance of the MSM solver relative to the pairwise calculations for three smaller Argon atom clusters containing 0.24, 1.05, 2.17 and 5.21 million particles. For the MSM method, 16 MPI ranks and 4 OpenMP threads was found to be the optimal ratio for each data point. The main finding from this initial investigation is that the MSM method begins to outperform the pairwise calculation once the Argon cluster increases beyond one million particles. To further investigate the appropriateness of the truncated potential and MSM evaluations of the Coulombic potential in MC/MD calculations, a parallel force decomposition scheme was implemented in LAMMPS to efficiently evaluate the full Coulomb pairwise potential including all pairs of particles (i.e. no neighbor lists or pairwise cutoffs used). With this method for computing the exact Coulomb interaction energy and forces, errors resulting from parameters and approximations of the computationally more efficient methods can be determined exactly to guide development of robust and efficient calculations on large-scale computing resources.

Appendix C. Comparison of two three-body recombination models

In section 2, we presented two methods for computing the recombination processes. Method 1 is based on equation (8) and method 2 is based on the energy and distance criteria. Figure C1 shows that the recombination dynamics obtained from these two models agree rather well with each other.
Appendix D. Charge state distributions at high intensity

Figure D1 shows the charge state distributions obtained for a 2 fs pulse with 5 keV, $4.1 \times 10^{14}$ photons $\mu m^{-2}$ pulse and a 2 fs pulse with 8 keV, $2.6 \times 10^{14}$-photons $\mu m^{-2}$. Both distributions are peaked at high charge states, rather than neutral states as illustrated in figures 5(g) and (i). This is due to the fact that at these high fluence levels fast ablation of the cluster shells reduces the number of recombination events.

References

[1] Ackermann W et al 2007 Nat. Photon. 1 336
[2] Emma P et al 2010 Nat. Photon. 4 641
[3] Ishikawa T et al 2012 Nat. Photon. 6 540
[4] Gaffney K J and Chapman H N 2007 Science 316 1444
[5] Young L et al 2010 Nature 466 56
[6] Bostedt C et al 2016 Rev. Mod. Phys. 88 015007
[7] Fergusson K R et al 2016 Sci. Adv. 2 e1500837
[8] Neutze R, Wouts R, van der Spoel D, Weckert E and Hajdu J 2000 Nature 406 752
[9] Rohringer N et al 2012 Nature 481 488
[10] Yoneda H et al 2015 Nature 524 446
[11] Shwartz S et al 2014 Phys. Rev. Lett. 112 163901
[12] Aquila A et al 2015 Struct. Dyn. 2 041701
[13] Chapman H N et al 2011 Nature 470 73
[14] Caleman C, Huldt G, Maia F R N C, Ortiz C, Parak F G, Hajdu J, van der Spoel D, Chapman H N and Timneanu N J A N 2011 ACS Nano 5 139
[15] Barty A et al 2012 Nat. Photon. 6 35
[16] Boutet S et al 2012 Science 337 362
[17] Seibert M M et al 2011 Nature 470 78 02
Manson S T and Cooper J W 1968 Phys. Rev. 165 126
Yeh J J and Lindau I 1985 At. Data Nucl. Data Tables 32 1
Rozsnyai B F 1975 J. Quant. Spectrosc. Radiat. Transfer 15 695
Hilborn R C 1982 Am. J. Phys. 50 982
Auger P 1925 J. Phys. Radium 6 205
Coster D and Kronig R 1935 Physica 2 13
Kiang A T, Ma S T and Wu T Y 1936 Phys. Rev. 50 673
McGuire E 1975 Atomic Inner-Shell Processes ed B Crasemann (New York: Academic) pp 293–330

Boyd J 2001 Chebyshev and Fourier Spectral Methods: Second Revised Edition (Dover Books on Mathematics) (New York: Dover)
Wang J Y, Chu S I and Laughlin C 1994 Phys. Rev. A 50 3208
Yao G H and Chu S I 1993 Chem. Phys. Lett. 204 381
Szalay V 1995 J. Chem. Phys. 99 1978
Greenman L, Ho P J, Pabst S, Kamarchik E, Mazziotti D A and Santra R 2010 Phys. Rev. A 82 023406
Ewald P 1921 Ann. Phys. 64 253
Hockney R and Eastwood J 1988 Computer Simulation Using Particles (London: Taylor and Francis)