Efficient electronic structure calculation for molecular ionization dynamics at high x-ray intensity
Yajiang Hao, Ludger Inhester, Kota Hanasaki, Sang-Kil Son, and Robin Santra

Citation: Structural Dynamics 2, 041707 (2015); doi: 10.1063/1.4919794
View online: http://dx.doi.org/10.1063/1.4919794
View Table of Contents: http://scitation.aip.org/content/aca/journal/sdy/2/4?ver=pdfcov
Published by the American Crystallographic Association, Inc.

Articles you may be interested in
Reactive scattering of H2 from Cu(100): Comparison of dynamics calculations based on the specific reaction parameter approach to density functional theory with experiment

Intense X-ray FEL-molecule physics: Highly charged ions

Calculation of electronic structures and magnetic moments of Nd 2 Fe 14 B and Dy 2 Fe 14 B by using linear-combination-of-pseudo-atomic-orbital method

Multiphoton ionization of inner valence electrons and fragmentation of polyatomic molecules in an intense Ti:sapphire laser pulse

Excess-photon ionization spectra and atomic structure in intense laser fields
Efficient electronic structure calculation for molecular ionization dynamics at high x-ray intensity

Yajiang Hao (郝亚江),1,2,3,a) Ludger Inhester,1,3,b) Kota Hanasaki,1,3,c) Sang-Kil Son (손상길),1,3,d) and Robin Santra1,3,e) 1Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany 2Department of Physics, University of Science and Technology Beijing, Beijing 100083, People’s Republic of China 3The Hamburg Centre for Ultrafast Imaging, Luruper Chaussee 149, 22761 Hamburg, Germany 4Department of Physics, University of Hamburg, Jungiusstrasse 9, 20355 Hamburg, Germany 1,23, 041707-1

(Received 7 March 2015; accepted 24 April 2015; published online 7 May 2015)

We present the implementation of an electronic-structure approach dedicated to ionization dynamics of molecules interacting with x-ray free-electron laser (XFEL) pulses. In our scheme, molecular orbitals for molecular core-hole states are represented by linear combination of numerical atomic orbitals that are solutions of corresponding atomic core-hole states. We demonstrate that our scheme efficiently calculates all possible multiple-hole configurations of molecules formed during XFEL pulses. The present method is suitable to investigate x-ray multiphoton multiple ionization dynamics and accompanying nuclear dynamics, providing essential information on the chemical dynamics relevant for high-intensity x-ray imaging. © 2015 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License. [http://dx.doi.org/10.1063/1.4919794]

I. INTRODUCTION

The advent of x-ray free-electron lasers (XFELs)1,2 opens up a new era in science and technology.3–5 Unprecedentedly ultraintense and ultrafast hard x-ray pulses generated from XFELs enable us to measure molecular structures on the atomic scale and to explore the structural dynamics on the femtosecond scale. One of the most prominent XFEL applications is imaging of biological macromolecules. X-ray crystallography with XFELs, after demonstration of the proof-of-principle,6 has started to reveal previously unknown protein structure,7 promising a breakthrough in structural biology (see reviews in Refs. 8–11). Recent advances in time-resolved serial femtosecond crystallography12–16 enable us to take a step forward towards molecular movies. To investigate molecular structure and structural dynamics with XFELs, one needs to understand radiation damage dynamics—x-ray multiphoton ionization dynamics and accompanying nuclear dynamics. Furthermore, the phase problem17 is the bottleneck to reconstruct molecular structures in serial femtosecond crystallography as much as in conventional x-ray crystallography. To overcome the phase problem for x-ray crystallography with XFELs, one uses conventional phasing technique at intermediate x-ray intensity,18 or one takes an advantage of the large degree of ionization at high x-ray intensity. The latter brings in high-intensity phasing (HIP) methods,19 including high-intensity multwavelength anomalous

[a]haoyj@ustb.edu.cn
[b]ludger.inhester@cfel.de
[c]kota.hanasaki@cfel.de
[d]sangkil.son@cfel.de
[e]robin.santra@cfel.de
diffraction\textsuperscript{20,21} and high-intensity radiation damage induced phasing.\textsuperscript{22} The HIP techniques require detailed description of deep-inner-shell decay dynamics of heavy atoms influenced by the molecular environment. Therefore, understanding of radiation damage dynamics is the key for successful molecular imaging.

Modeling of biological macromolecules exposed to XFEL radiation was pioneered by the seminal work of Neutze \textit{et al.}\textsuperscript{23} Since then, there have been several computational tools to address molecular imaging problems, for example, classical molecular dynamics model,\textsuperscript{24,25} particle-in-cell approach,\textsuperscript{26,27} transport model,\textsuperscript{28,29} and Coulomb complex model.\textsuperscript{30,31} Some of these methods have been recently applied to ion fragment spectra\textsuperscript{32} and electron spectra\textsuperscript{33} from C\textsubscript{60} molecules interacting with intense x-ray pulses. So far, most computational methods have been based on the independent-atom model or the plasma model. The description of the molecular environment is omitted by assumption or incorporated in an \textit{ad hoc} manner. When a molecule absorbs x-ray photons, inner-shell multiple ionization induces fragmentation dynamics.\textsuperscript{34,35} Chemical bonds are weakened, and electrons and holes rearrange before the molecule breaks apart.\textsuperscript{36–38} Detailed electronic structure calculations for chemical bonding and charge rearrangement are thus crucial to describe molecular effects in modeling of the XFEL–matter interaction.

The electronic response of atoms and molecules to XFEL pulses is in essence characterized by multiphoton multiple ionization dynamics.\textsuperscript{39–41} The \textit{xatom} toolkit\textsuperscript{42,43} has been developed to simulate the XFEL–atom dynamical interaction, and the ionization dynamics model has been tested with a series of experiments.\textsuperscript{44–48} The unprecedentedly large number of x-ray photons delivered by an XFEL pulse induces sequential x-ray absorptions, creating a variety of different \(q\)-hole configurations for each charge state \(+q\). To simulate ionization dynamics, one needs to calculate photoionization cross section, Auger rate, and fluorescence rate for each configuration and solve a set of coupled rate equations for the time-dependent populations of the configurations.\textsuperscript{39,49,50} The \(q\)-hole electronic configurations are energetically highly excited, and theoretical treatment of such highly-excited states is challenging. Another challenge is the complexity of the ionization dynamics. Even for the atomic case, one must solve more than \(2 \times 10^7\) coupled rate equations for Xe \(L\)-shell-initiated ionization dynamics.\textsuperscript{46} To address this formidable problem, a Monte-Carlo approach has been proposed for solving the rate equations\textsuperscript{43,45} and sampling the most probable configurations.\textsuperscript{46} In this scheme, the electronic structure is calculated for every single configuration selected on the fly as part of the Monte Carlo sampling. For the molecular case, the complexity increases even further because of the degrees of freedom associated with atomic motions, so the Monte Carlo approach seems to be the only way to simulate molecular response to an intense XFEL pulse. However, the most important question still remains: How to describe the electronic structure of molecules on the fly for the Monte Carlo method?

Here, we present a new efficient method for electronic structure calculation of polyatomic molecules and implement a dedicated toolkit, \textit{xmolecule}. The proposed method is well suited for calculations of molecular multiple-hole configurations that are formed during x-ray multiphoton ionization dynamics. To efficiently describe molecular orbitals of core-hole configurations, the method employs atomic orbitals as basis functions that are numerical solutions of atomic core-hole states, calculated by \textit{xatom}.\textsuperscript{42} For any given molecular electronic configuration and any given molecular geometry, \textit{xmolecule} calculates molecular orbitals and orbital energies, which are essential components for dynamical simulations of x-ray multiphoton multiple ionization. We demonstrate that \textit{xmolecule} is capable to calculate the whole spectrum of multiple-hole configurations at a given molecular geometry and potential energy surfaces for given multiple-hole configurations of molecules. Also, performance scalability with the system size is discussed. In this paper, we focus on the implementation of a molecular electronic-structure approach. Calculating cross sections and rates and solving coupled rate equations to simulate ionization dynamics will be described elsewhere. Having achieved these results, \textit{xmolecule} aims to play a key role in molecular imaging at high x-ray intensity.

The paper is organized as follows. Section II formulates our scheme to calculate molecular multiple-hole configurations. It includes theoretical and computational schemes for basis function generation with numerical atomic orbitals, multicenter integration on a molecular grid, and
direct Coulomb integral evaluation. In Sec. III, we show benchmark calculations for XMOL-EUTE, and then numerical results for the potential energy curves of various electronic configurations of carbon monoxide, and single- and double-core ionization potentials of several polyatomic molecules. We discuss the scalability of our scheme to a molecular size of hundreds of atoms. This is followed by the conclusion in Sec. IV.

II. COMPUTATIONAL METHODS

A. The Hartree–Fock–Slater method

We consider a molecular system composed of $N_{\text{atom}}$ atoms with $N_{\text{elec}}$ electrons. The $A$th nuclear charge and coordinates are denoted by $Z_A$ and $R_A$, respectively. The molecular charge state $+q$ is given by $q = \sum_A Z_A - N_{\text{elec}}$. We use the Hartree-Fock-Slater (HFS) method in which molecular orbitals (MO), $\psi_i(\mathbf{r})$, and orbital energies, $e_i$, are obtained by solving the effective single-electron Schrödinger equation (atomic units are used unless specified otherwise),

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{ext}}(\mathbf{r}) + V_H(\mathbf{r}) + V_X(\mathbf{r})\right] \psi_i(\mathbf{r}) = e_i \psi_i(\mathbf{r}),$$

(1)

where $V_{\text{ext}}(\mathbf{r})$ is the external potential due to the nuclei,

$$V_{\text{ext}}(\mathbf{r}) = -\sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|},$$

(2)

and the Hartree potential $V_H(\mathbf{r})$ represents the classical Coulomb interaction among the electrons,

$$V_H(\mathbf{r}) = \int d^3 r' \left| \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \right|,$$

(3)

and the last term $V_X(\mathbf{r})$ represents the exchange interaction, which is approximated by the Slater exchange potential,

$$V_X(\mathbf{r}) = -\frac{3}{2} \left[ \frac{3}{\pi} \rho(\mathbf{r}) \right]^{\frac{1}{3}}.$$

(4)

The electronic density $\rho(\mathbf{r})$ is obtained by the sum of squared MO’s weighted by the occupation numbers $\{n_i\}$ as

$$\rho(\mathbf{r}) = \sum_i n_i |\psi_i(\mathbf{r})|^2,$$

(5)

where $n_i \in \{0, 1, 2\}$. In contrast to conventional ground-state electronic structure calculations, in which the $N_{\text{elec}}$ spin-orbitals with the lowest energies are filled, we consider all possible $\{n_i\}$ subject to $\sum_i n_i = N_{\text{elec}}$, in order to take account of electronic excited states representing $q$-hole configurations.

The total energy within the HFS method is given by the sum of the nucleus–nucleus repulsion energy and the electronic energy,

$$E_{\text{total}} = \sum_{A<B} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} + \sum_i n_i e_i - \frac{1}{2} \int d^3 r \int d^3 r' \frac{\rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{3}{8} \left( \frac{3}{\pi} \right)^{\frac{3}{4}} \int d^3 r \rho(\mathbf{r})^\frac{3}{2}.$$

(6)
B. Linear combination of numerical atomic orbitals

For atomic systems, the orbital is represented with spherical harmonics as
\[ \phi_{nlm}(r) = \frac{\mu_{nl}(r)}{r} Y_{lm}(\theta, \phi), \]
where \( n, l, \) and \( m \) are the principal quantum number, the orbital angular momentum quantum number, and the associated projection quantum number, respectively. The radial wavefunction \( \mu_{nl}(r) \) can be solved by a numerical grid-based method. The \texttt{xatom} toolkit\textsuperscript{42} has been developed to solve the atomic HFS equation. By employing the generalized pseudospectral (GPS) method\textsuperscript{52,53} and imposing a spherically symmetric potential, \texttt{xatom} accurately calculates \( \mu_{nl}(r) \) for a given \((n, l)\)-subshell, and accordingly \( \phi_{nl}(r) \) for a given \( \mu \equiv (n, l, m) \). This numerical atomic orbital has been used to successfully calculate multiple-hole configurations formed during x-ray multiphoton ionization dynamics in the atomic case.\textsuperscript{41,42}

For molecular systems, we employ the linear combination of atomic orbitals (LCAO) scheme to construct molecular orbitals,
\[ \psi_i(r) = \sum_{\mu} C_{\mu i} \phi_{\mu}(r), \]
where \( \phi_{\mu}(r) \) is the \( \mu \)th atomic orbital (AO) and \( C_{\mu i} \) is the coefficient of the \( \mu \)th AO for the \( i \)th MO. Using Eq. (8) transforms the self consistent field (SCF) Eq. (1) into the corresponding Roothaan-Hall equation,\textsuperscript{54}
\[ HC = SCE, \]
where \( E \) is a diagonal matrix of MO energies and \( C \) is the MO coefficient matrix. The elements of the Hamiltonian matrix \( H \) and the overlap matrix \( S \) are given as
\[ H_{\mu\nu} = \int d^3r \phi_\mu(r) \left[ -\frac{1}{2} \nabla^2 + V_{\text{eff}}(r) \right] \phi_\nu(r), \]
\[ S_{\mu\nu} = \int d^3r \phi_\mu(r) \phi_\nu(r), \]
where the effective potential \( V_{\text{eff}}(r) \equiv V_{\text{ext}}(r) + V_H(r) + V_X(r) \). Equation (9) is solved in a self-consistent manner. To accelerate convergency, we employ the direct inversion in the iterative subspace (DIIS) method.\textsuperscript{55,56} When we encounter convergence problems at large bond distances, where the energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) is very small, we apply level shifts\textsuperscript{57} in the SCF iterations.

Here, our choice of basis set for the LCAO scheme is the numerical atomic orbitals (NAOs) obtained by \texttt{xatom} described above. In Fig. 1, we plot the squared radial function

FIG. 1. Numerical atomic orbitals for different core-hole states of the nitrogen atom.
\[ |\mu_\alpha(r)|^2 \] for the 1s, 2s, and 2p orbitals of the ground state of the neutral nitrogen (N) atom, the single-core-hole (SCH) state of \( N^+ \), and the double-core-hole (DCH) state of \( N^{2+} \), respectively. Comparison among different core-hole states shows significant deformation of valence orbitals in states with core holes. To cover these effects efficiently in the molecular calculation, we use NAOs that are numerical solutions of the corresponding atomic core-hole states. For instance, \( N^+ \) with one core hole at each atomic site (a DCH state) is calculated with basis functions optimized for \( N^+ (1s^{-1}) \) on both N atoms, whereas \( N^{2+} \) with a single-site DCH state is calculated with basis functions optimized for \( N^{2+} (1s^{-2}) \) on which the core hole is located and basis functions optimized for neutral N on the other side. In this way, we expect core-hole MOs are well described by core-hole-adapted NAOs.

To achieve utmost efficiency towards complex ionization dynamics, we employ the minimal basis set. Each AO with \((n, l, m)\) in Eq. (7) corresponds to a single basis function. Fully or partially occupied \((n, l)-subshells contribute to a set of basis functions and each \(l\) gives \((2l + 1)\) basis functions \(|m| \leq l\). For example, the N atom has 1s, 2s, and 2p (partially) occupied subshells, which constitute 5 basis functions \((\phi_{1s}, \phi_{2s}, \phi_{2p_1}, \phi_{2p_2}, \phi_{2p_3})\) in total. This basis set is denoted as \([2s1p]\). According to the minimal-basis-set scheme, the chemical elements from B to Ne have the same number of basis functions \((N_{\text{basis}} = 5)\). In Sec. III A, we will discuss limitations and extensions of the minimal-basis-set scheme.

C. Molecular grid and multicenter integration

Equations (10) and (11) require evaluation of the corresponding integrals in three dimensions. In our case, the \( \phi_\mu(r) \) and \( \phi_\nu(r) \) are represented with a radial grid and spherical harmonics. To perform 3D integrals involving many atomic centers, we employ the multicenter integration proposed by Becke. 58 Molecular grid points are constructed as a combination of sets of atomic grid points. Each set of atomic grid points, centered at one of the nuclei, consists of \( N_r \) radial grid points and \( N_{\text{ang}} \) angular grid points. The radial grid points are exactly the same as those used for NAO calculations with the GPS method. 52, 53 The angular grid points are obtained by the Lebedev grid scheme 59 with an angular momentum cutoff at \( l_{\text{max}} \). The number of angular grid points is approximately given by \( N_{\text{ang}} \approx 4(l_{\text{max}} + 1)^2 / 3 \). A detailed description of constructing multicenter molecular grid points is found in Refs. 60 and 61. We use an atomic radial grid size \( \ell_{\text{max}} \) large enough \((\sim 10 \text{ Å})\) so that the atomic grids of many neighboring atoms overlap with each other. In principle, different atomic grid parameters can be used for individual atoms in a molecule. For convenience, however, we use the same grid parameters for all atoms. Then, the total number of molecular grid points is given by \( N_{\text{grid}} = N_{\text{atom}} \times N_r \times N_{\text{ang}} \).

Becke’s multicenter integration scheme 58 introduces a set of smooth nuclear weight functions \( \{w_A(r)\} \), subject to the constraint \( \sum_A w_A(r) = 1 \). The nuclear weight functions are generated by the third-order polynomial cutoff profile in the fuzzy cell scheme. 58 Then, any integral of a given function \( f \) can be evaluated by the sum of individual atomic integrals,

\[
I = \int d^3 r f(r) = \sum_A \int d^3 r f(r) w_A(r) = \sum_A \int d^3 r_A f(r_A) w_A(r_A), \tag{12}
\]

where \( r_A = r - R_A \). Each atomic integral can be readily performed using the spherical coordinate system of \( r_A \), centered at the \( A \)th atom,

\[
I_A = \int d^3 r_A f(r_A) w_A(r_A) \approx \sum_{k \in A} f(r_k) w_A(r_k) w_k, \tag{13}
\]

where \( k \) is the index of the grid points of the \( A \)th atom and \( w_k \) is defined as a product of the radial Legendre-Gauss-Lobatto quadrature weights 62, 63 and the angular Lebedev quadrature weights. 59
D. Implementation of direct Coulomb integrals

In electronic structure calculations, one of the most time-consuming parts is the evaluation of electron repulsion integrals. In order to achieve fast calculation within a desired accuracy, we have developed a multipole expansion scheme with an adaptive cut off. First, the integral involved in the Hartree potential in Eq. (3) can be decomposed into individual atomic integrals,

\[ V_H(\mathbf{r}) = \int d^3r' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \approx \sum_A \int d^3r_A' \frac{\rho_A(\mathbf{r}'_A)}{|\mathbf{r}_A - \mathbf{r}'_A|}, \]  

(14)

where \( \rho_A(\mathbf{r}) \equiv \rho(\mathbf{r})w_A(\mathbf{r}) \). Each single-center density \( \rho_A(\mathbf{r}) \) can then be regarded as the atomic contribution to the total electronic density. To implement the integral, we expand the single-center density with real spherical harmonics \( Y_{lm}(\theta, \varphi) \) as

\[ \rho_A(\mathbf{r}_A) = \sum_{l=0}^{\text{max}} \sum_{m=-l}^{l} \rho_A^{lm}(r_A) Y_{lm}(\theta_A, \varphi_A), \]  

(15)

where \( \rho_A^{lm}(r) \) is the \((l, m)\)-component of the spherical expansion,

\[ \rho_A^{lm}(r_A) = \frac{2\pi}{2l+1} \int_0^\pi d\theta_A \int_0^{2\pi} d\varphi_A \rho_A(\mathbf{r}_A) Y_{lm}(\theta_A, \varphi_A). \]  

(16)

With this single-center decomposition and spherical harmonic expansion of the electronic density, \( \rho_A^{lm}(r) \), the Hartree potential in Eq. (3) is obtained as

\[ V_H(\mathbf{r}) = \sum_A \sum_{l,m} V_A^{lm}(r_A) Y_{lm}(\theta_A, \varphi_A), \]  

(17)

where \( V_A^{lm} \) is given by

\[ V_A^{lm}(r_A) = \frac{4\pi}{2l+1} \int_0^{r_{\text{max}}(r_A)} dr_A' l_A^2 r_{\text{<}}^{2l+1} \rho_A^{lm}(r_A'), \]  

(18)

where \( r_{\text{<}} = \min(r'_A, r_A) \) and \( r_{\text{>}} = \max(r'_A, r_A) \). This radial integral is numerically evaluated in combination with various truncation methods (see the Appendix).

E. Molecular electronic configuration

Keeping the energetically lowest orbitals doubly occupied, the SCF procedure obtains the HFS solution for the electronic ground state. In order to obtain a solution for an excited electronic state of a \( q \)-hole configuration, each molecular orbital has to be assigned a specific occupation number. This can be done, as in the ground state calculation, by identifying the orbitals by their HFS energy eigenvalue. However, during the SCF iterations, the energetic order of MOs may change. Thus, identifying the orbitals by ordering them according to their HFS energy eigenvalue may lead to failure of the above SCF procedure or yield a solution for a different electronic state than required. This is called variational collapse.64–66

To prevent this situation, we employ a variant of the maximum overlap method.66 In the maximum overlap method, the desired excited electronic state is specified by a set of initial guess orbitals \( \{\psi_j^{\text{guess}}\} \) in combination with a set of occupation numbers \( \{n_j\} \). In each SCF iteration, the occupation number \( n_i \) of the calculated orbital \( \psi_i \) is chosen according to its projection onto the subspace spanned by the guess orbitals with respective occupation number. Specifically, we calculate the overlap of the \( i \)th current MO with the \( j \)th guess MO,
\[ O_{ij} = \langle \psi_i | \psi_j^{\text{guess}} \rangle = \sum_{\mu, \nu} C_{\mu j} S_{\mu \nu} C_{\nu i}^{\text{guess}}, \]

where \( S_{\mu \nu} = \int d^3r \phi_\mu(r) \phi_\nu^*(r) \). Note that the basis set for the initial guess orbitals is not necessarily the same as the one used for the expansion of the actual molecular orbitals, because different NAOs can be used for different \( g \)-hole configurations. Therefore, \( S_{\mu \nu} \) can be different from the overlap matrix \( S_{\mu \nu} \) defined in Eq. (11). Then, the projections of the \( j \)th orbital onto the span of the guess orbitals for the unoccupied \((n = 0)\), singly occupied \((n = 1)\), and doubly occupied \((n = 2)\) cases are given by

\[ P_j^{(n)} = \sum_j |O_{ij}|^2, \]

where \( j \) runs over all initial guess orbitals whose occupation number \( n_j \) equals \( n \). To preserve the character of the required electronic configuration during the SCF procedure, we choose the set of the occupation numbers of the current orbitals, \( \{n_j\} \), such that \( \sum P_j^{(n)} \) is maximized, while the total number of doubly and singly occupied orbitals is maintained.

This procedure to determine the orbital occupation critically depends on the initial guess MOs. Thus, it is essential that the provided guess MOs \( \{\psi_i\} \) together with the provided occupation numbers \( \{n_j\} \) describe a wavefunction that is close to the required solution. For the calculations performed here, we choose initial guess MOs obtained from a previous calculation for a lower ionized electronic state or for the same electronic state with an altered molecule geometry. For the single-core-hole state in \( N_2 \), we obtain a localized core hole on a specific nucleus by performing a Boys-orbital-localization procedure\(^{67} \) of the two guess core orbitals. Having obtained a converged solution, we verify that the obtained set of MOs is indeed close to the initial guess, by inspecting the individual overlap \( O_{ij} \).

### III. RESULTS AND DISCUSSION

#### A. Benchmark calculations

We first estimate the accuracy of our calculations using the numerical multicenter integration in comparison with conventional calculations using the analytic Gaussian integration by GAMESS.\(^{68} \) Here, we employ the 6–31G Gaussian basis set\(^{69} \) to calculate the SCF-level ground-state energy of a water molecule. The internuclear distance of \( R(\text{OH}) = 0.957 \) \( \text{Å} \) and the bond angle of \( \angle(\text{HOH}) = 104.48^\circ \) are used. Only in this test, we employ the restricted Hartree-Fock (RHF) method instead of the HFS method, in order to directly compare with the GAMESS results. Figure 2 shows that our numerical calculations converge to the GAMESS results as the number of radial grid points per atom \( (N_r) \) and the number of angular grid points per atom (determined by \( l_{\text{max}} \)) are increased. The total number of molecular grid points for \( N_r = 50 \) and \( l_{\text{max}} = 8 \) is \( 3 \times 50 \times 110 = 16500 \). The maximum radius \( r_{\text{max}} \) is 20 a.u., and the GPS mapping parameter\(^{52,53} \) \( L = 1 \) a.u. is used. Note that all grid parameters utilized here provide a numerical accuracy \( |\Delta E| < 1.5 \) eV. If chemical accuracy is required (typically 1 kcal/mol \( \approx 0.04 \) eV), our study for the water molecule shows that it is achievable with \( N_r \geq 200 \) and \( l_{\text{max}} \geq 11 \), keeping the same \( L \) and \( r_{\text{max}} \). As to be shown in Sec. III B, the energy scale of x-ray-induced dynamics of highly-charged molecules will extend into the keV regime. Therefore, the worst grid parameters (for example, \( N_r = 30 \) and \( l_{\text{max}} = 4 \)) shown in Fig. 2 would be sufficient to describe the molecular ionization dynamics at high x-ray intensity.

We next examine the performance of our NAO basis set scheme. In Fig. 3, we show the calculated HFS energy of (a) the ground state of neutral \( N_2 \) molecule with NAOs optimized for neutral N atom and (b) the quadruple-core-hole (QCH) state of \( N_2^{4+} \) ion with NAOs optimized for the DCH state of \( N_2^2 \). The internuclear distance \( R = 1.096 \) \( \text{Å} \) is fixed. \( N_r = 200 \), \( L = 1 \) a.u., \( r_{\text{max}} = 20 \) a.u., and \( l_{\text{max}} = 11 \) are used. The results are shown together with those obtained by the equivalent calculations using conventional Gaussian-type-orbital (GTO) basis sets of different sizes (STO-3G\(^{70} \)) and a series of Dunning’s correlation-consistent basis sets.\(^{71} \) All GTO
basis sets are obtained from the EMSL Basis Set Library. $\Delta E$ is the energy difference from the total energy calculated with the uncontracted version of cc-pV6Z, [16s10p5d4f3g2h1i] with 161 basis functions, which is considered here the complete basis set limit. Thus, $\Delta E$ indicates the numerical error due to lack of basis functions.

In both Figs. 3(a) and 3(b), one can see that the minimal NAO basis set is superior to the conventional minimal basis set of STO-3G, illustrating that fully optimized NAOs are a practical choice for the basis set in the LCAO scheme. Also, Fig. 3 shows convergence of GTOs with respect to the number of basis functions. Interestingly, the conventional GTOs for QCH $N_2^{+}$ perform almost one order of magnitude less accurate than GTOs for neutral $N_2$. The reason is that GTOs are optimized to be used for neutral ground-state calculations. In contrast, NAOs optimized for corresponding atomic $q$-hole configuration provide similar accuracy for both neutral $N_2$ and QCH $N_2^{+}$. Thus, NAO functions provide an ideal basis set for our minimal-basis-set HFS scheme.

To improve accuracy, we try to increase the number of NAOs in a systematic manner by including unoccupied atomic orbitals with higher $(n, l)$ such as 3s, 3p, and so on. As shown in Fig. 3, the NAOs are somewhat inefficient to achieve higher accuracy by simply extending to higher $(n, l)$, as previously reported in Ref. 73. This is attributed to the fact that additional series of higher $(n, l)$-orbitals, whose mean square radius is far from the atomic center, are inefficient for representing bonding molecular orbitals. Instead, we propose a scheme for adding compact $p$-type and $d$-type functions to the minimal NAO basis set in order to improve the

![Graph showing convergence of total HF energy with respect to the number of grid points.](image)

FIG. 2. Convergence of the total HF energy with respect to the number of grid points. $N_r$ is the number of radial grid points, and $l_{\text{max}}$ controls the number of angular grid points per atom. The ground-state energy calculation of $H_2O$ with RHF/6-31G is performed using the numerical multicenter integration, and $\Delta E$ is the difference from the result obtained using the analytic Gaussian integrals.

![Graph comparing convergence in total energy with respect to the number of basis functions.](image)

FIG. 3. Comparison of convergency in total energy with respect to the number of basis functions, using the GTO scheme and the NAO scheme: (a) neutral $N_2$ and (b) QCH $N_2^{+}$. $\Delta E$ is defined by the total energy difference from the complete basis set limit (see the text).
description of chemical bonding. Additional functions are constructed by use of radial wavefunctions of occupied subshells multiplied by $r$, where $r$ is the radial coordinate in the atomic system. For the chemical elements from B to Ne, the $p$-type functions are $u_{2p}(r)Y_{1m}(\theta, \phi)$, where $m = 0, \pm 1$, and the $d$-type functions are $u_{2d}(r)Y_{2m}(\theta, \phi)$, where $m = 0, \pm 1, \pm 2$. By adding these functions, as denoted by extended NAO (NAO[e]) and as marked with the black rectangle in Fig. 3, the accuracy is much improved; the total energy of neutral N$_2$ is close to the cc-pVDZ level, and the total energy of QCH N$_2^{2+}$ is close to the cc-pVQZ level. The number of basis functions for NAO[e] is only 13 per atom, whereas cc-pVQZ has 55 basis functions. There have been several approaches for extension of the minimal NAO basis set, where additional basis functions are constructed in a systematic way.

B. Potential energy curves for various hole configurations

Figure 4 shows the HFS total energies in Eq. (6) using core-hole-adapted NAO basis functions for all possible $q$-hole configurations that can be accessed by x-ray multiphoton ionization of the neutral carbon monoxide molecule. The internuclear distance $R = 1.128$ Å is fixed, and the grid parameters of $N_r = 50$, $L = 1$ a.u., $r_{\text{max}} = 20$ a.u., and $l_{\text{max}} = 8$ are used. For convenience, the figure shows these configurations grouped into charge states. The lowest horizontal line for each charge state indicates the ground-state energy for a given charge $+q$. This figure then illustrates how much energetically excited the $q$-hole configurations are. For example, the energy of DCH CO$^{2+}$ (O1$s^{-2}$) is about 1 keV higher than the ground-state energy of CO$^{2+}$. Ionization dynamics induced by intense x-ray pulses may occur step by step, visiting lots of these electronic states. Therefore, it is crucial to efficiently calculate this set of electronic states of $q$-hole configurations.

We further investigate the behavior of the potential energy curves (PEC) obtained using the NAO basis set. In Fig. 5, we show the calculated HFS total energies for three different types of CO$^{2+}$ DCH states: (a) C1$s^{-2}$, (b) C1$s^{-1}$O1$s^{-1}$, and (c) O1$s^{-2}$. The dashed red line indicates PECs calculated with core-hole-adapted NAO[e]. The solid red line indicates PECs with core-hole-adapted NAO without additional functions. Both results are compared with the solid blue line calculated with the conventional cc-pVTZ basis set. Previous theoretical studies of core-hole states suggested that calculations of the cc-pVTZ level are reasonably
Our NAO\[e\] scheme reproduces well the cc-pVTZ results, even though the size of NAO\[e\] (N\text{basis} = 13) is much smaller than that of cc-pVTZ (N\text{basis} = 30). For comparison, we also plot PECs with NAOs optimized for neutral ground-state atoms, denoted by NAO\[n\], which shows a trend similar to what a conventional STO-3G minimal basis set would be. The NAO\[n\] results represent a poor estimate of PECs due to missing the core-hole effect on orbitals. On the other hand, PECs from NAOs, which are optimized for atomic core-hole states, show dramatic improvement over NAO\[n\], even though NAO and NAO\[n\] have the same number of basis functions (N\text{basis} = 5).

C. Single- and double-core ionization potentials of molecules

To further test the accuracy of our calculation scheme, we compare core ionization potentials for a series of small molecules obtained from the HFS calculation. The molecular geometries are taken from Ref. 77 and the grid parameters are the same as those used in Sec. III B. We derive the single-core ionization potential from the HFS orbital energy of a neutral ground-state calculation using NAOs with and without additional basis functions. The double-core ionization potential is calculated as the sum of the first and the second core ionization potential, where the second core ionization potential is taken from the orbital energy of the SCH state calculation. For the DCH states with core holes on different nuclear sites, thus, two values are obtained for the two different ionization sequences.

Table I lists the ionization potentials compared with the values obtained from complete-active-space SCF (CASSCF) calculations\(^\text{75}\) and experimental values.\(^\text{78–83}\) For our calculations of two-site DCH states, a mean value and a deviation are listed for two values from the different ionization sequences. For the CASSCF results of two-site DCH states, only triplet spin states are listed and the difference between singlet and triplet states is smaller than 0.7 eV. Note that N\text{2}O is a linear molecule N\text{t}—N\text{c}—O, where N\text{t} indicates the terminal N atom and N\text{c} means N at the center. As can be seen, the CASSCF results\(^\text{75}\) show agreement within less than 4 eV with the available experimental values. The single ionization potentials we extract from the much simpler HFS calculation using the minimal NAO basis set show for all molecules a similar agreement within 5.1 eV, except F1s\textsuperscript{-1} in LiF (28 eV). For the DCH states, where the core holes are located on different nuclear sites, with the minimal NAO basis set, we also see a similar agreement within 7 eV to the CASSCF values and, where available, the experimental values. Again, LiF is an exception showing a much larger discrepancy of \(\approx 30\) eV. For the DCH state with core holes on the same nucleus, we find a systematically larger disagreement of about 20–30 eV (for F1s\textsuperscript{-2} in LiF 78 eV).

The inclusion of the \(p\)-type and \(d\)-type functions in the basis set leads in most cases to a larger deviation to the literature values than the results obtained with the minimal basis set. For these calculations, we get ionization potentials that tend to be lower than the literature values (from 3.4 eV for Li1s\textsuperscript{-1} to 60.5 eV for F1s\textsuperscript{-2} in LiF). Clearly, the extended NAO basis set should improve the quality of the electronic structure model, as the electronic wave function...
has more flexibility. Thus, we conclude that the good agreement with the minimal basis set might be an artifact due to cancellation of errors.

For the results obtained with the larger basis set, we attribute the remaining deviations to the literature values mainly to relaxation energy contributions associated with the core hole electron removal. The applied scheme of taking orbital energies as ionization potentials cannot account for these effects. For core holes on the same nuclear site, where the core hole relaxation contributions are particularly strong, we see the strongest deviations (18.2–60.5 eV). Also, the extreme deviations for LiF may be explained from these contributions: The core hole on the F atom in LiF shows a particular large core hole relaxation effect, whereas for the core hole on Li it is very small.75

### D. Performance scaling

Our implementation of xMOLECULE aims for large-scale molecular calculations, especially for a large number of repeated calculations where time and resources available for each calculation are severely limited. At the same time, it requires the capability of calculating moderate-size systems in order to describe molecular-environment effects. Here, we demonstrate the
performance scalability of our scheme toward molecular calculations with a few hundred atoms. Our grid-based method has the potential to achieve linear scaling in the number of atoms.\textsuperscript{85–88}

In the HFS method, the two-body interaction is divided into the exchange interaction and the direct Coulomb interaction. The former is replaced with the local density approximation, and the latter is treated with the Hartree potential as described in Sec. II D. The computational complexity of the Hartree potential is $O(N_{\text{grid}}^2)$, where $N_{\text{grid}}$ is linearly proportional to $N_{\text{atom}}$, because the potential $V_{\mu\nu}(r)$ in Eq. (3) contains the integral over molecular grid points and has to be evaluated at every single molecular grid point. By introducing the truncation methods described in the Appendix, this complexity can be reduced to $O(N_{\text{grid}}N_{\text{atom}})$. These truncation schemes do not change the quadratic scaling behavior with respect to $N_{\text{atom}}$, but reduce the actual computational time by several times (for example, a factor of two in our following calculations).

Another truncation can be made in the evaluation of one-body matrix elements in Eqs. (10) and (11). Both $H_{\mu\nu}$ and $S_{\mu\nu}$ are decomposed into atomic contributions by the multicenter integration: $H_{\mu\nu} \approx \sum_{A} H_{\mu\nu}^{A}$ and $S_{\mu\nu} \approx \sum_{A} S_{\mu\nu}^{A}$. We define an AO pair $\phi_{\mu}(r)\phi_{\nu}(r)$ and its contribution to each atomic grid,

$$Q_{\mu\nu}^{A} = \int A(r_{A}) \phi_{\mu}(r_{A}) \phi_{\nu}(r_{A}) |w_{A}(r_{A})|. \quad (21)$$

Then we set $H_{\mu\nu}^{A}$ and $S_{\mu\nu}^{A}$ to zero if $Q_{\mu\nu}^{A} < \epsilon$, where $\epsilon$ is a control parameter. The complexity of the integrals in Eqs. (10) and (11) is $O(N_{\text{basis}}^2N_{\text{grid}})$, where both $N_{\text{basis}}$ and $N_{\text{grid}}$ are linearly proportional to $N_{\text{atom}}$. By using our truncation scheme described above, we can reduce it to a quadratic behavior with respect to $N_{\text{atom}}$.

Figure 6 shows the size dependence of the computation time of XMOLECULE with the current truncation schemes. We calculate the HFS ground state of C$_{24}$H$_{12}$ molecule (coronene) in its equilibrium molecular geometry taken from Ref. 77. Then, we perform calculations for $n$ such molecules ($n = 1, \ldots, 7$) stacked in the vertical direction with an interlayer separation of 3.3 Å. The minimal NAO basis set is used with $N_{f} = 20$, $L = 1$ a.u., $r_{\text{max}} = 10$ a.u., and $l_{\text{max}} = 4$. The $y$ axis is the CPU time per SCF iteration in seconds on a lab workstation (Intel Xeon X5660 2.80 GHz), and the $x$ axis indicates the number of atoms in the stacked (C$_{24}$H$_{12}$)$_{n}$ molecule. When all truncations are off (blue curve), the computational performance shows close to a cubic dependence. On the other hand, when the truncation method of Eq. (21) is applied with $\epsilon = 10^{-3}$ (red curve), the scaling shows a quadratic dependence on the system size. Note that when the truncation of Eq. (21) is used, the complexity of the matrix element calculations is $O(N_{\text{atom}}^2)$.
reduced to a quadratic relation, while the Hartree potential calculation becomes the most time-consuming step, which is also governed by a quadratic scaling. The difference in the total energy between the calculations with and without this truncation is less than 0.14 eV/atom, whereas the truncated calculation is about 7.5 times faster than the calculation with no truncation. The calculation with 216 atoms \((n=6)\) takes 40 second per single SCF iteration on the lab workstation. The whole computation time takes about 14 min including the overhead costs for numerical grid construction and 12 SCF iterations. When additional truncation schemes for the Hartree potential (see the Appendix) are applied with \(\epsilon_0 = 0.1\) and \(\epsilon_1 = 0.01\) (green curve), the complexity is a bit reduced towards a linear relation and the errors in the total energy are less than 0.93 eV/atom. The actual computational time per iteration is improved by a factor of two for the 216-atom case.

**IV. CONCLUSION**

In summary, we present a new method to calculate various multiple-hole electronic states for polyatomic molecules that may be formed by x-ray multiphoton ionization dynamics at high x-ray intensity. The method is based on the Hartree-Fock-Slater method, employing the LCAO scheme, where NAOs are used as a minimal basis set for molecular orbital calculations. Usage of NAOs has two advantages over conventional Gaussian-type basis functions. First, NAOs are obtained from numerical solutions for atomic core-hole states at the same computational level. Second, accuracy and efficiency of numerical integration with NAOs are controllable by grid parameters and truncation schemes. The NAOs presented here are accurately solved by using the numerical grid-based method that is implemented in the xatom toolkit.

Using core-hole-adapted NAOs, molecular orbitals for core-hole states are efficiently calculated. We present benchmark calculations for multiple-core-hole states of \(\text{N}_2\). The NAO results show consistent accuracy for different charge states, which is not the case for conventional basis sets that are optimized for neutral systems. We demonstrate that our scheme is able to calculate all possible configurations that may be formed by removing zero, one or more electrons from the ground-state configuration of neutral \(\text{CO}\) molecule. The electronic state during x-ray multiphoton ionization dynamics may visit several of these multiple-hole configurations, which are energetically excited by about 4 keV with respect to the ground-state configuration of neutral \(\text{CO}\). For molecular and ionization dynamics during XFEL pulses, we need not only all different multiple-hole states but also potential energy surfaces for individual electronic states. For double-core-hole states of \(\text{CO}^{2+}\), we calculate potential energy curves with core-hole-adapted NAOs, in good agreement with converged results with respect to the basis-set size. Also, we present single- and double-core-hole ionization potentials for several molecules in comparison with available theoretical and experimental data.

Efficient electronic structure calculations for molecules are essential for dynamical modeling of molecules at high x-ray intensity. We have implemented xMOLECULE to make a step toward dynamical simulation of molecular imaging with XFELs. Calculations of photoionization cross sections, fluorescence rates, and Auger rates for all possible configurations formed during molecular ionization dynamics are in progress.

**ACKNOWLEDGMENTS**

We thank Oriol Vendrell for helpful discussions. This work has been supported by the excellence cluster ‘The Hamburg Centre for Ultrafast Imaging–Structure, Dynamics and Control of Matter at the Atomic Scale’ of the Deutsche Forschungsgemeinschaft. Yajiang Hao was supported by the National Natural Science Foundation of China (Grant No. 11004007) and the Fundamental Research Funds for the Central Universities of China.

**APPENDIX: TRUNCATION SCHEMES FOR EVALUATING THE HARTREE POTENTIAL**

Here, we introduce truncation schemes on \(V_{\text{AO}}^\alpha(r_\lambda)\). The upper bound of \(|V_{\text{AO}}^\alpha(r_\lambda)|\) is given by
\[ |V^A_{lm}(r_A)| = \frac{4\pi}{2l+1} \left[ \int_0^{r_A} dr'_A r'^2_A \left( \frac{r'_A}{r_A} \right)^l \rho^A_{lm}(r'_A) \right] \]

\[ \leq \frac{4\pi}{2l+1} \left[ \int_0^{r_A} dr'_A r'^2_A \left( \frac{r'_A}{r_A} \right)^l \rho^A_{lm}(r'_A) \right] + \frac{1}{r_A} \int_{r_A}^{r_{\text{max}}} dr'_A r'^2_A \rho^A_{lm}(r'_A) \]

\[ \leq \frac{4\pi}{2l+1} \int_0^{r_A} dr'_A r'^2_A \rho^A_{lm}(r'_A) \]

\[ = \frac{4\pi}{2l+1} \int_0^{r_{\text{max}}} dr'_A r'^2_A \rho^A_{lm}(r'_A) \]  

(A1)

Then, we define

\[ d^A_{lm} = \int_0^{r_{\text{max}}} dr'_A r'^2_A |\rho^A_{lm}(r'_A)| \]  

(A2)

to be used as a truncation indicator. Note that the number of electrons in the A-th atomic electronic density is given by \( Q_A = \int d^3 r \rho_A(r) = \sqrt{4\pi d^A_{00}} \). Within the atom, we consider higher multipole moments of the density to be less relevant. Thus, if \( d^A_{lm} \) is small enough in comparison with \( d^A_{00} \), then the contribution of \( l \) and \( m \) is truncated, i.e.,

\[ V^A_{lm}(r_A) \rightarrow 0 \quad \text{when} \quad \frac{d^A_{lm}}{d^A_{00}} < \varepsilon_1, \]  

(A3)

where \( \varepsilon_1 \) is a truncation control parameter.

Another truncation is that if the distance from the origin of the A-th atom is large enough, the Hartree potential contributed from A is approximately evaluated by the monopole only, and all \( l > 0 \) contributions are truncated, i.e.,

\[ V^A_{lm}(r_A) \rightarrow 0 \quad \text{when} \quad r_A > r_c, \]  

(A4)

where \( r_c \) is a cut-off radius given by \( r_c = Q_A/\varepsilon_0 = \sqrt{4\pi d^A_{00}}/\varepsilon_0 \). Here, \( \varepsilon_0 \) is another truncation control parameter.


