Computer Physics Communications 292 (2023) 108882

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

Attosecond Chemistry

# PyStructureFactor: A Python code for the molecular structure factor in tunneling ionization rates ☆,☆☆

Shanshan Song [a,1], Mingyu Zhu [a,b,1], Hongcheng Ni [a,c,d,*], Jian Wu [a,c,d,e,*]

[a] State Key Laboratory of Precision Spectroscopy, East China Normal University, Shanghai 200241, China
[b] School of Physics and Electronic Science, East China Normal University, Shanghai 200241, China
[c] NYU-ECNU Joint Institute of Physics, New York University at Shanghai, Shanghai 200062, China
[d] Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China
[e] CAS Center for Excellence in Ultra-intense Laser Science, Shanghai 201800, China

ARTICLE INFO

*Article history:*
Received 28 February 2023
Received in revised form 19 July 2023
Accepted 29 July 2023
Available online 8 August 2023

*Keywords:*
Molecular structure factor
Tunneling ionization
Weak-field asymptotic theory
PySCF

ABSTRACT

Tunneling ionization is at the core of strong-field and attosecond science. In this paper, we present `PyStructureFactor` — a general `Python` code towards the calculation of the structure factor in the tunneling ionization rate of common molecules under intense laser fields. The numerical implementation is based on the well-developed weak-field asymptotic theory in the integral representation. The information of the electronic structure of the molecules is obtained via the `PySCF` quantum chemistry package. `PyStructureFactor` is a general computational framework that can be utilized to compute the molecular structure factor of various types of molecules, including polar and nonpolar diatomic molecules, degenerate molecules, and open-shell molecules. Examples are given that are benchmarked against known results with good agreements. The present `PyStructureFactor` is implemented in an efficient manner and is easily applicable towards larger molecules.

**Program summary**
*Program title:* PyStructureFactor
*CPC Library link to program files:* https://doi.org/10.17632/td88mhw2sg.1
*Developer's repository link:* https://github.com/TheStarAlight/PyStructureFactor
*Licensing provisions:* Apache-2.0
*Programming language:* Python 3
*Nature of problem:* The structure factor of a molecule in intense laser fields determines its orientation-dependent tunneling ionization rate, which is crucial for the studies of ultrafast molecular dynamics and its control. However, the strong-field community lacks an open-source code to calculate the molecular structure factor, and can only resort to known results of a few molecules.
*Solution method:* We developed the `PyStructureFactor` program with the structure factor of arbitrary molecules calculated using the weak-field asymptotic theory in the integral representation. The underlying electronic structure necessary for the calculation is obtained using the `PySCF` quantum chemistry package.
*Additional comments including restrictions and unusual features:* The accuracy of the molecular structure factor calculated by `PyStructureFactor` is restricted by the level of precision of the electronic structure information extracted from the `PySCF` package.
The running time depends on the size of the molecule, the basis set of the calculation, the level of precision of the electronic structure calculations, and other parameters passed to the program. The example in Fig. 2 took 1.2 seconds to finish on an AMD Ryzen 9 7950X CPU on the WSL Ubuntu 22.04 LTS.

© 2023 Elsevier B.V. All rights reserved.

☆ The review of this paper was arranged by Prof. Jimena Gorfinkiel.
☆☆ This paper and its associated computer program are available via the Computer Physics Communications homepage on ScienceDirect (http://www.sciencedirect.com/science/journal/00104655).

* Corresponding authors at: State Key Laboratory of Precision Spectroscopy, East China Normal University, Shanghai 200241, China.
*E-mail addresses:* hcni@lps.ecnu.edu.cn (H. Ni), jwu@phy.ecnu.edu.cn (J. Wu).
[1] These authors contributed equally.

https://doi.org/10.1016/j.cpc.2023.108882
0010-4655/© 2023 Elsevier B.V. All rights reserved.

## 1. Introduction

Advances in the laser technology have made intense laser fields widely accessible in typical table-top lab settings. When matter interacts with intense laser fields, a plethora of novel strong-field phenomena emerge that are of fundamental interest, such as high-order harmonic generation (HHG) [1–3], above-threshold ionization [4–6], nonsequential double ionization [7], and laser-induced electron diffraction [8,9]. HHG is the key to produce extreme ultraviolet light pulses, which has facilitated the improvement of time resolution of ultraprecise measurements down to the attosecond level [10–13]. Common to many of these strong-field phenomena is their first step, known as tunneling ionization [14,15], which is thus a cornerstone of strong-field and attosecond science. It has attracted widespread interest not only in physics but also in chemistry and related fields [16–22]. Given the recent development in optochemistry [23,24], a complementary emerging field to photochemistry, the importance of the tunneling ionization of molecules has been elevated to new heights.

In contrast to photoionization where the photonic feature of light plays a dominant role, tunneling ionization is better represented by the highly nonlinear ionization induced by intense optical fields in the time domain $F(t)$, with a tunneling ionization rate [25,26]

$$W[F(t)] \sim \mathrm{e}^{-2\kappa^3/3F(t)} \tag{1}$$

depending exponentially on the field strength $F(t)$, where $\kappa = \sqrt{-2E_0}$ with $E_0$ the orbital energy of the target. For typical intense infrared laser fields, the photon energy is substantially lower than the ionization potential of the bound electron. Moreover, the timescale of the field oscillation is much longer than the intrinsic attosecond timescale of the bound electron. Therefore, the strong-field ionization process can be approximated by tunneling in a static electric field represented by the instantaneous laser electric field $F(t)$ [27,28].

In addition to the field factor [Eq. (1)], the tunneling ionization rate of molecules depends on a factor related to the electronic structure of the ionizing orbital, known as the molecular structure factor. It determines the orientation-dependent rate of tunneling ionization that is crucial to the analysis of subsequent strong-field molecular dynamics and photoelectron spectrum [29]. It is the aim of the current work to present the numerical solution to the molecular structure factor.

The molecular structure factor can be obtained either by the molecular Ammosov–Delone–Krainov theory (MO-ADK) [29–33] or via the weak-field asymptotic theory (WFAT) [34–49]. WFAT is applicable in laser fields below the critical field strength that leads to over-barrier ionization. Based on the parabolic adiabatic expansion approach [50], WFAT generalizes isotropic atomic potentials to arbitrary molecular potentials and naturally accounts for the effect of molecular permanent dipole moment (if any), and has been adopted as the theoretical method in some experimental research on molecular strong-field ionization [51–55]. In this work, we employ WFAT towards a numerical implementation of the computation of the molecular structure factor. Under its framework, the obtained tunneling ionization rate, in its leading order, factorizes into the field factor [Eq. (1)] and the absolute square of the structure factor, the latter of which is an inherent property of the ionizing molecular orbital independent of the external field [37].

In an essence of the perturbation theory, the structure factor of WFAT is found through the behavior of the wave function in the asymptotic region in early stages, known as the tail representation [36,38,42]. Hence, its reliable evaluation requires an accurate description of the asymptotic tail of the ionizing orbital. To extract the orbital wave function of polyatomic molecules by the Hartree-Fock (HF) method or density functional theory (DFT) with quantum chemistry software packages [38,41], one normally resorts to the basis-based approaches. However, compared to the correct exponential behavior obtained by the grid-based approach, the standard Gaussian-type basis functions facilitating multicenter integral calculations decay too rapidly in the asymptotic region and thus could result in erratic oscillations. To reproduce the correct local property of the asymptotic tail of the orbital, large basis sets with quadruple or pentuple-zeta quality and variationally optimized exponents are required [37]. This approach significantly increases the computational cost and becomes infeasible in many cases. The efficiency of such numerical computation needs to be improved urgently.

As a remedy, an alternative integral approach of WFAT was developed with the spirit of scattering theory to overcome the above issue [46–48]. In this approach, the structure factor is expressed as an integral form involving the ionizing orbital and a known analytical function. The integral accumulates at the localized regions where the orbital has large amplitudes, and it is insensitive to the representation of the asymptotic tail far from the nuclei. Hence, the integral approach can be implemented using standard quantum chemistry software packages with Gaussian-type orbitals (GTOs) for arbitrary molecules.

In this work, an open-source program implementing WFAT in the integral representation is developed that can easily and quickly calculate the molecular structure factor of common molecules. In our implementation, named the `PyStructureFactor`, the ionizing molecular orbital is expanded as a linear combination of GTOs at the HF or complete-active-space self-consistent field (CASSCF) level of theory calculated by the `PySCF` package [56,57]. The integrals are evaluated numerically using a three-dimensional grid-based method, which exhibits much better numerical stability and is readily applicable to large molecules, compared to the standard analytical integration procedure. Our present implementation of `PyStructureFactor` features flexible control of accuracy and efficient computation towards the calculation of the molecular structure factor.

This work is organized as follows. In Sec. 2, the definition and derivation of the tunneling ionization rate, the molecular structure factor, and other parameters within the WFAT framework are given. In Sec. 3, the details of numerical implementation are presented. In Sec. 4, we detail the usage of the `PyStructureFactor` program to calculate the molecular structure factor. In Sec. 5, we illustrate the application of the `PyStructureFactor` program with examples of various types of common molecules. Conclusions are given in Sec. 6. Atomic units are used throughout unless stated otherwise.

## 2. Theoretical formulation

Tunneling ionization of molecules can be usually modeled within the Born-Oppenheimer and the single-active-electron approximation. Under such a framework, the wave function of the ionizing orbital $\psi_0$ is the solution to an effective one-electron Schrödinger equation

$$\left[-\frac{1}{2}\nabla^2 + V(\mathbf{r})\right]\psi_0(\mathbf{r}) = E_0\psi_0(\mathbf{r}), \tag{2}$$

where the HF potential $V(\mathbf{r})$ describes the interaction between the ionizing electron and the parent ion.

Within WFAT, tunneling ionization is investigated in parabolic coordinates. We assume the electric field vector $\mathbf{F} = F\hat{\mathbf{z}}$, and define the orientation of the molecule using a set of Euler angles $(\alpha, \beta, \gamma)$ (within the $z$-$y'$-$z''$ convention [46], which is shown schematically in Fig. 1) representing a rotation $\hat{\mathbf{R}}$ from the laboratory frame (LF) to the molecular frame (MF). The ionization rate does not depend

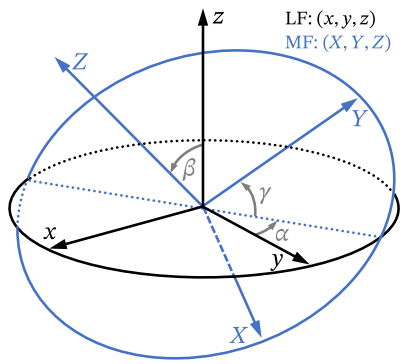


**Fig. 1.** Illustration of the rotation from the LF to the MF using Euler angles of $z$-$y'$-$z''$ convention. The rotation consists of three steps: (1) The coordinate system $(x, y, z)$ rotates about the $z$ axis by angle $\alpha$, then (2) rotates about the new $y$ axis by angle $\beta$, finally (3) rotates about the rotated $z$ axis by $\gamma$, and becomes $(X, Y, Z)$. The rotations mentioned above are all counter-clockwise.

on $\alpha$ and so we set $\alpha = 0$ throughout. The total ionization rate is split into different parabolic channels:

$$\Gamma(\beta, \gamma) = \sum_{\nu} \Gamma_{\nu}(\beta, \gamma), \tag{3}$$

where $\Gamma_{\nu}(\beta, \gamma)$ are partial rates of parabolic quantum number indices

$$\nu = (n_{\xi}, m), \quad n_{\xi} = 0, 1, 2, \cdots, \quad m = 0, \pm 1, \pm 2, \cdots \tag{4}$$

In the leading-order approximation, the asymptotic expansion of the partial rates for $F \to 0$ gives

$$\Gamma_{\nu}(\beta, \gamma) = |G_{\nu}(\beta, \gamma)|^2 W_{\nu}(F), \tag{5}$$

which can be separated into two factors: The structural part $|G_{\nu}(\beta, \gamma)|^2$ and the field part $W_{\nu}(F)$. The field factor, $W_{\nu}(F)$, has the form [25,26]

$$W_{\nu}(F) = \frac{\kappa}{2} \left( \frac{4\kappa^2}{F} \right)^{2Z/\kappa - 2n_{\xi} - |m| - 1} \mathrm{e}^{-2\kappa^3/3F}. \tag{6}$$

The structure factor, namely $G_{\nu}(\beta, \gamma)$, in the integral representation of WFAT, is given as an integral:

$$G_{\nu}(\beta, \gamma) = \mathrm{e}^{-\kappa \mu_z} \int \Omega_{\nu}^{*} \left( \hat{\mathbf{R}}^{-1} \mathbf{r} \right) V_{\mathrm{c}}(\mathbf{r}) \psi_0(\mathbf{r}) \mathrm{d}\mathbf{r}, \tag{7}$$

where

$$\boldsymbol{\mu} = \int \psi_0^{*}(\mathbf{r}) \mathbf{r} \psi_0(\mathbf{r}) \mathrm{d}\mathbf{r} \tag{8}$$

denotes the orbital dipole moment in the LF with $\mu_z$ its component along the field direction; $\Omega_{\nu}$ is a reference function which can be expanded into spherical harmonics:

$$\Omega_{\nu}(\mathbf{r}) = \sum_{l=|m|}^{\infty} \Omega_{lm}^{\nu}(\mathbf{r}), \tag{9}$$

where

$$\Omega_{lm}^{\nu}(\mathbf{r}) = \sum_{l=|m|}^{\infty} R_l^{\nu}(r) Y_{lm}(\theta, \phi); \tag{10}$$

$V_{\mathrm{c}}(\mathbf{r}) = V(\mathbf{r}) + Z/r$ is the core potential without the Coulomb tail (i.e., $V_{\mathrm{c}} \to 0$ as $r \to \infty$) with $Z$ the asymptotic charge of the parent molecular ion. The radial part $R_l^{\nu}$ of the reference function $\Omega_{lm}^{\nu}$ in Eq. (10) reads

$$R_l^{\nu}(r) = \omega_l^{\nu} \, (\kappa r)^l \, \mathrm{e}^{-\kappa r} \, \mathrm{M}(l + 1 - Z/\kappa, 2l + 2, 2\kappa r), \tag{11}$$

where $\mathrm{M}(a, b, x)$ is the confluent hyper-geometric function [58] and $\omega_l^{\nu}$ is the normalization coefficient:

$$\begin{aligned} \omega_l^{\nu} = {} & (-1)^{l + (|m| - m)/2 + 1} \, 2^{l + 3/2} \, \kappa^{Z/\kappa - (|m| + 1)/2 - n_{\xi}} \\ & \times \sqrt{(2l + 1)(l + m)!(l - m)!(|m| + n_{\xi})! n_{\xi}!} \, \frac{l!}{(2l + 1)!} \\ & \times \sum_{k=0}^{\min(n_{\xi}, l - |m|)} \frac{\Gamma(l + 1 - Z/\kappa + n_{\xi} - k)}{k!(l - k)!(|m| + k)!(l - |m| - k)!(n_{\xi} - k)!}. \end{aligned} \tag{12}$$

Under the framework of the HF method, the HF potential $V(\mathbf{r})$ consists of three parts, namely the nuclear Coulomb potential ($V_{\mathrm{nuc}}$), the direct ($V_{\mathrm{d}}$) and exchange ($V_{\mathrm{ex}}$) parts of inter-electron interactions:

$$V(\mathbf{r}) = V_{\mathrm{nuc}}(\mathbf{r}) + V_{\mathrm{d}}(\mathbf{r}) + V_{\mathrm{ex}}(\mathbf{r}), \tag{13}$$

where

$$V_{\mathrm{nuc}}(\mathbf{r}) = -\sum_{A=1}^{N_{\mathrm{atm}}} \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|}, \tag{14}$$

$$V_{\mathrm{d}}(\mathbf{r}) = \sum_{i=1}^{N} \int \frac{\psi_i^{*}(\mathbf{r}') \psi_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}', \tag{15}$$

$$V_{\mathrm{ex}}(\mathbf{r}) \psi_0(\mathbf{r}) = -\sum_{i=1}^{N} \psi_i(\mathbf{r}) \int \frac{\psi_i^{*}(\mathbf{r}') \psi_0(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \langle \sigma_i | \sigma_0 \rangle \, \mathrm{d}\mathbf{r}', \tag{16}$$

where $N$ is the number of electrons, $N_{\mathrm{atm}}$ is the number of nuclei, $\psi_i(\mathbf{r})$ and $\sigma_i$ denote the molecular orbital and the spin state of the electron of index $i$ ($\langle \sigma_i | \sigma_j \rangle = 1$ if electrons $i$ and $j$ have the same spin, and $\langle \sigma_i | \sigma_j \rangle = 0$ otherwise), while $Z_A$ and $\mathbf{R}_A$ are the nuclear charge and position of atom of index $A$.

## 3. Numerical implementation

For the practical numerical implementation of the computation of the molecular structure factor within the framework of WFAT, we first refer to the evaluation of the integral in Eq. (7). Directly evaluating the integrals for each molecular orientation $(\beta, \gamma)$ is inefficient. Nevertheless, we may utilize the spherical harmonic expansion of the reference function $\Omega_{\nu}$, and express the rotated spherical harmonic functions as a linear combination of spherical harmonic functions using the Wigner-$d$ matrix. Substituting Eqs. (9) and (10) into Eq. (7) yields [46]

$$G_{\nu}(\beta, \gamma) = \mathrm{e}^{-\kappa \mu_z} \sum_{l=|m|}^{\infty} \sum_{m'=-l}^{l} I_{lm'}^{\nu} d_{mm'}^{l}(\beta) e^{-\mathrm{i}m'\gamma}, \tag{17}$$

where $d_{mm'}^{l}(\beta)$ is the Wigner-$d$ matrix, and the integral $I_{lm'}^{\nu}$ is expressed as

$$I_{lm'}^{\nu} = \int \Omega_{lm'}^{\nu *}(\mathbf{r}) V_{\mathrm{c}}(\mathbf{r}) \psi_0(\mathbf{r}) \mathrm{d}\mathbf{r}. \tag{18}$$

In this way the computational overhead is significantly reduced — with the values of the integrals $I_{lm'}^{\nu}$, we may obtain the structure factor of arbitrary molecular orientation $G_{\nu}(\beta, \gamma)$ with little computational cost.

The next topic is concerned with the calculation of the HF potential operator acting on the ionizing orbital, i.e., $V(\mathbf{r})\psi_0(\mathbf{r})$. In the practical application, the molecular orbitals are usually obtained by

**Table 1**
Essential input parameters of the `pyscf.M` method in the `PyStructureFactor` program.

| Parameter | Description | Default |
|---|---|---|
| `atom` | Information of the molecular structure. Refer to Fig. 2 for an example. | — |
| `unit` | The unit used for the position coordinates of the atoms in the molecule. `'B'` or `'AU'` indicates Bohr and otherwise Angström. | `'Angström'` |
| `basis` | Basis set used for calculation. Refer to `pyscf.gto.basis` for more information. An accurate basis set is suggested for optimal convergence. | `'sto-3g'` |
| `charge` | Charge of the molecule or molecular ion. | `0` |
| `spin` | Number of spin (i.e., $2S$) of the molecule. | `0` |

**Table 2**
Input parameters and return value of the `PyStructureFactor.get_structure_factor` method.

| Parameter | Description | Default |
|---|---|---|
| `mol` | The `PySCF` molecule object. Initialized by invoking `pyscf.M` or `pyscf.gto.M`. | — |
| `orbital_index` | Index of the ionizing orbital relative to HOMO, e.g., HOMO $\to 0$, LUMO $\to +1$, HOMO $-1 \to -1$, … | `0` |
| `channel` | Parabolic channel $\nu = (n_\xi, m)$. The program would calculate the structure factor of channel $(n_\xi, m)$, i.e., $G_{n_\xi,m}$. | `(0,0)` |
| `lmax` | The cut-off limit of the angular quantum number (larger $l$ would be neglected) used in the summation in Eq. (17). | `10` |
| `hf_method` | Indicates whether `'RHF'` or `'UHF'` should be used in the molecular HF calculation. Note: `'UHF'` must be used for open-shell molecules. | `'RHF'` |
| `casscf_conf` | Configuration of CASSCF calculation consisting of `(n_active_orb, n_active_elec)`. Specifying `None` (by default) indicates employing the primitive HF method instead of the more accurate CASSCF method. | `None` |
| `atom_grid_level` | Level of fineness of the grid used in integration (see also `pyscf.dft.Grid`), which controls the number of radial and angular grids around each atom in the evaluation of the integration, ranging from 0 to 9. | `3` |
| `orient_grid_size` | Size of the output $(\beta, \gamma)$ grid, which defines the orientation of the molecule with respect to the polarization direction of the laser field. The grid is uniform, with $\beta$ ranging from 0 to $\pi$ and $\gamma$ ranging from 0 to $2\pi$. Setting the $\gamma$ grid count to 1 indicates that $\gamma$ would be zero throughout the calculation. | `(90,1)` |
| `move_dip_zero` | Indicates whether to shift the molecular coordinates such that the dipole moment of the parent ion [Eq. (23)] vanishes (i.e., $\mathbf{D} = \mathbf{0}$). | `True` |
| `rmax` | Indicates the cut-off limit of the radial grid points, points of radii larger than `rmax` would be neglected in the calculation. | `40` |
| Returns | A `NumPy` array containing the structure factor $G_{n_\xi,m}$ of the given channel $(n_\xi, m)$ on the $(\beta, \gamma)$ orientation grid, whose shape is given by `orient_grid_size`. | |

the self-consistent-field (SCF) approach, which is commonly implemented in standard quantum chemistry packages. The molecular orbitals are usually expanded into linear combinations of Gaussian-type orbitals (GTOs):

$$\psi_i(\mathbf{r}) = \sum_\alpha C_{i\alpha}\chi_\alpha(\mathbf{r} - \mathbf{R}_\alpha), \tag{19}$$

where $\chi_\alpha(\mathbf{r} - \mathbf{R}_\alpha)$ denotes a GTO of index $\alpha$ centering at $\mathbf{R}_\alpha$, and the coefficients $C_{i\alpha}$ are obtained through the SCF approach. In this way we obtain $V_\mathrm{d}(\mathbf{r})$ and $V_\mathrm{ex}(\mathbf{r})\psi_0(\mathbf{r})$ in Eqs. (15) and (16) expressed in coefficients $C_{i\alpha}$ for practical calculations. Defining the integral

$$J_{\alpha\beta}(\mathbf{r}) = \int \frac{\chi_\alpha(\mathbf{r}' - \mathbf{R}_\alpha)\chi_\beta(\mathbf{r}' - \mathbf{R}_\beta)}{|\mathbf{r} - \mathbf{r}'|}\mathrm{d}\mathbf{r}' \tag{20}$$

which can be calculated analytically for GTOs, Eqs. (15) and (16) can be written as

$$V_\mathrm{d}(\mathbf{r}) = \sum_{i=1}^{N}\sum_{\alpha,\beta} C_{i\alpha}C_{i\beta}J_{\alpha\beta}(\mathbf{r}), \tag{21}$$

$$V_\mathrm{ex}(\mathbf{r})\psi_0(\mathbf{r}) = -\sum_{i=1}^{N}\langle\sigma_i|\sigma_0\rangle \sum_{\alpha,\beta,\gamma} C_{i\alpha}C_{i\beta}C_{0\gamma}J_{\beta\gamma}(\mathbf{r})\chi_\alpha(\mathbf{r} - \mathbf{R}_\alpha). \tag{22}$$

The last topic is related to the convergence of the algorithm. The integral in Eq. (18) is numerically evaluated on a grid using the Becke fuzzy cell integration scheme [59] implemented in the `PySCF` package, which uses higher density of grid points near the nuclei to accelerate convergence of the calculation. Apart from adopting the Becke grids in the numerical integration, we also choose the origin such that the dipole of the parent ion

$$\mathbf{D} = \left[\sum_{A=1}^{N_\mathrm{atm}} Z_A\mathbf{R}_A - \sum_{i=1}^{N}\int \psi_i^*(\mathbf{r})\mathbf{r}\psi_i(\mathbf{r})\mathrm{d}\mathbf{r}\right] - \boldsymbol{\mu} \tag{23}$$

vanishes, which reduces the grid size required for the integral Eq. (18) to converge.

## 4. Program usage

The usage of the program is simple and straightforward. Firstly, import necessary packages, including `PyStructureFactor`, `PySCF` and `NumPy` in Python. Next, initialize the molecule object by calling the method `pyscf.M`, and provide necessary information of the molecule. Finally, get the structure factor of the molecule by invoking the method `PyStructureFactor.get_structure_factor`, whose return value is a `NumPy` array containing the structure factor at different molecular orientations $(\beta, \gamma)$.

To better illustrate the usage of the program, we present a description of the input parameters of the two main methods, including `pyscf.M`, which defines the molecular and orbital information, and `PyStructureFactor.get_structure_factor`, which outputs the corresponding structure factor, in Tables 1 and 2, respectively. Shown in Fig. 2 is a minimal example of the program which calculates the molecular structure factor $G_{00}(\beta)$ corresponding to the highest occupied molecular orbital (HOMO) of the hydrogen molecule. For the diatomic molecules, the $\gamma$ Euler angle does not come into play, and $\beta$ denotes the angle between the electric field vector and the molecular axis in the LF. More examples are available in Sec. 5.

Now we present convergence and reliability tests of the program in Figs. 3 and 4 with $N_2$ and CO molecules, respectively. To this end, we obtain their structure factors as a function of $\beta$ and compare the results to reference data extracted from the literature [41] which is obtained using the tail representation of WFAT. In both figures, the pc-4 basis is adopted in the calculations, and the cut-off limit of the index $l$ in the summation of Eq. (17), i.e.,

```
from PyStructureFactor import get_structure_factor
import numpy as np
import pyscf
n_beta  = 90
n_gamma = 1
molH2 = pyscf.M(atom="H␣0,0,0.37;␣H␣0,0,-0.37", basis="pc-1",
      spin=0)
beta_grid = np.linspace(0, np.pi, n_beta)
G_grid = get_structure_factor(
            mol = molH2, orbital_index = 0, channel = (0,0),
            lmax = 10, hf_method = "RHF",
            atom_grid_level = 3,
            orient_grid_size = (n_beta, n_gamma))
```

**Fig. 2.** A minimal example of the program calculating the structure factor $G_{00}(\beta)$ of the HOMO of the hydrogen molecule ($H_2$).

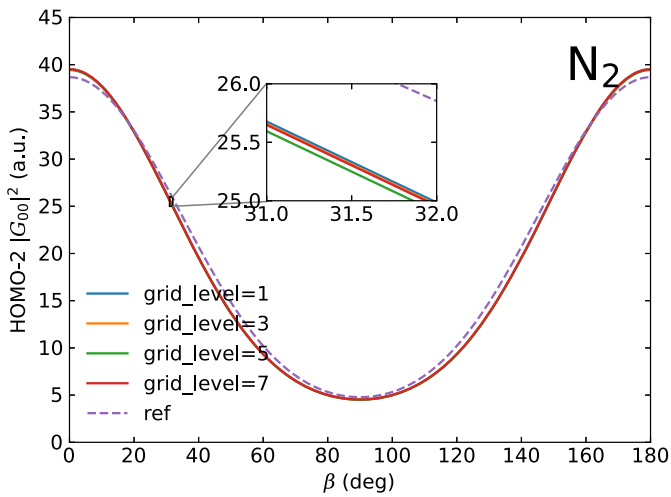


**Fig. 3.** Squared structure factor $|G_{00}|^2$ of the HOMO-2 of the nitrogen molecule ($N_2$). The reference data are extracted from Ref. [41].

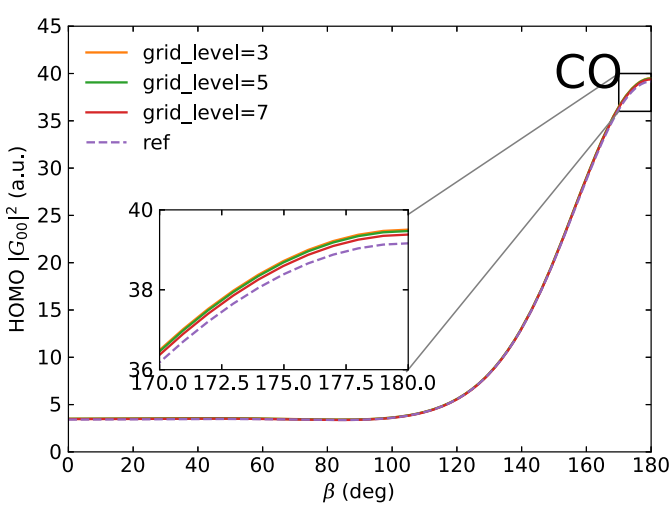


**Fig. 4.** Squared structure factor $|G_{00}|^2$ of the HOMO of the carbon monoxide molecule (CO). The reference data are extracted from Ref. [41].

$l_{max}$, is set to the default value of 10. The "grid_level" labels denote the input parameter "`atom_grid_level`" of the program. Here, grid levels 1, 3, 5, and 7 correspond to radial and angular sizes (40, 194), (75, 302), (105, 770), and (135, 1202) for each spherical grid around the nuclei, respectively. It is clear from the figures that for "`atom_grid_level`" larger than 3, the obtained results are hardly distinguishable from each other even after enlargement,

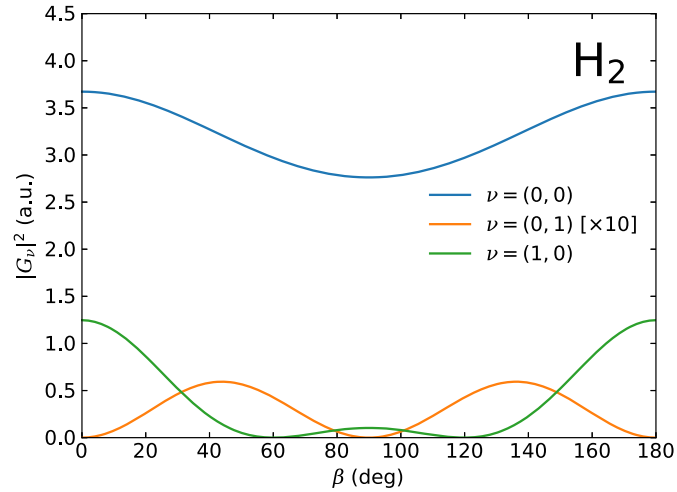


**Fig. 5.** Squared structure factor $|G_\nu|^2$ of the HOMO of the hydrogen molecule ($H_2$). The molecular axis is aligned along the $z$-axis in the MF and the internuclear distance is 0.74 Å. The basis used for calculation is pc-4.

which reveals a stable and fast convergence of the present program. In addition, the converged result is close to the reference data. This illustrates that the calculation of the molecular structure factor is able to achieve accurate results under default parameters. Minor discrepancy from the reference data results from the difference in the tail and integral presentations of WFAT and the choice of basis.

## 5. Illustrative examples

In this section we take a number of common molecules as examples, and present their structure factors obtained from `PyStructureFactor`. The examples cover molecules with various properties, as shown in Figs. 5 to 9: diatomic molecules ($H_2$, CO, and $O_2$), linear polyatomic molecules ($C_2H_2$) and planar molecules ($C_6H_6$); open-shell molecules ($O_2$); molecules with degenerate HOMOs (double HOMO degeneracy for $O_2$, $C_2H_2$ and $C_6H_6$).

For all examples presented, we set the input parameter $l_{max}$ to the default value of 10 and set the "`atom_grid_level`" parameter to the default value of 3. For linear molecules, the $\gamma$ dependence of the structure factor drops out and it only depends on $\beta$. For nonlinear molecules, on the other hand, the structure factor is presented on the $(\beta, \gamma)$ grid, see Fig. 9 for an example of the $C_6H_6$ molecule.

It is worthwhile noting that the molecular structure factor varies not only from molecule to molecule, but also changes from orbital to orbital for a particular molecule. Novel laser technologies have enabled the exploration of ionization from different molecular orbitals, and studies show that subsequent molecular dynamics critically depends on the ionizing orbital. Not surprisingly, the structure factor also depends on the orbital from which ionization occurs, as shown in Figs. 5 and 6.

Furthermore, for molecules with degenerate orbitals, the structure factor depends on the orientation of the orbital. The HOMOs of $O_2$, for example, are $\pi$ orbitals that are two degenerate orbitals perpendicular to each other. Each of them has a nodal plane, and we use the nodal plane to identify the two degenerate orbitals. "HOMO-$yz$" corresponds to the orbital that has a nodal plane on the $y$-$z$ plane while "HOMO-$xz$" represents that with a nodal plane on the $x$-$z$ plane. Clearly, the structure factor depends on the orientation of the orbital and thus varies for different degenerate HOMO orbitals, as shown in Figs. 7 and 8. We note that tunneling ionization could occur simultaneously from multiple orbitals,

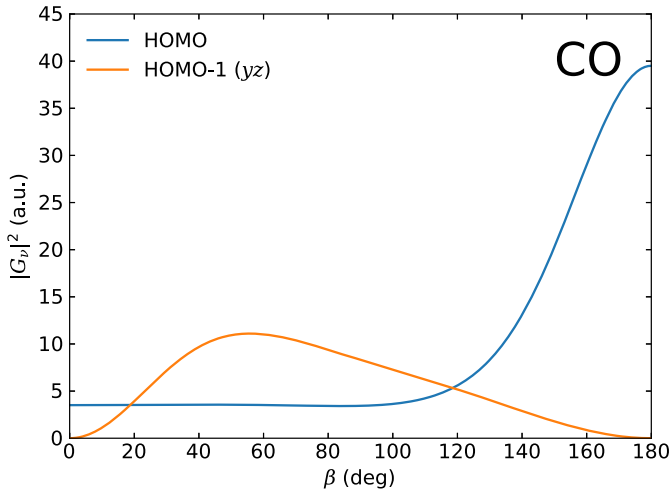


**Fig. 6.** Squared structure factor $|G_{00}|^2$ of the HOMO and HOMO-1 of the carbon monoxide molecule (CO, with the nodal plane on the $y$-$z$ plane). The molecular axis is aligned along the $z$-axis in the MF with the $+z$ axis pointing from the C to the O nucleus, and the internuclear distance is 1.13 Å. The basis used for calculation is pc-4.

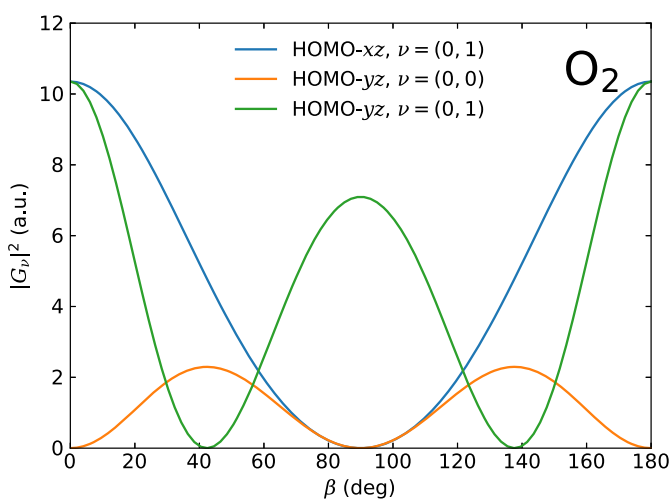


**Fig. 7.** Squared structure factor $|G_\nu|^2$ of the two degenerate HOMOs of the oxygen molecule ($O_2$, denoted by their nodal planes $xz$ and $yz$). The molecular axis is aligned along the $z$-axis in the MF and the internuclear distance is 1.21 Å. The basis used for calculation is pc-4. The oxygen molecule at the ground state has two unpaired electrons and thus has a total spin $S = 1$, which requires parameters `spin=2` and `hf_method="UHF"` to be passed to the program.

and the final tunneling ionization rate might be a summation over various contributions.

## 6. Conclusions

The tunneling ionization rate of molecules depends on the molecular orientation. Such dependence is determined by the molecular structure factor. In this work, we have presented the `PyStructureFactor` program, which computes the molecular structure factor within the WFAT theoretical framework based on the `PySCF` quantum chemistry package. `PyStructureFactor` is implemented in a manner that is both computationally efficient and easily accessible to the end users. It provides a straightforward approach to obtain the molecular structure factor of virtually all kinds of molecules, and therefore lays a solid foundation for the study of strong-field molecular dynamics.

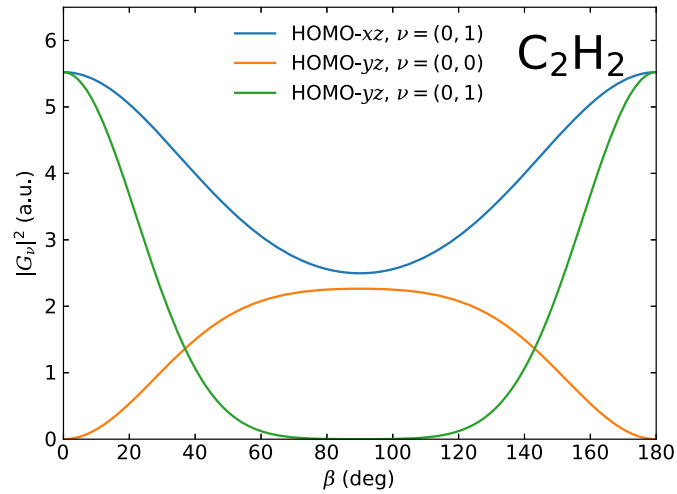


**Fig. 8.** Squared structure factor $|G_\nu|^2$ of the two degenerate HOMOs of the acetylene molecule ($C_2H_2$). The molecular axis is aligned along the $z$-axis in the MF, the internuclear distance between the C nuclei is 1.20 Å, and it is 1.06 Å between the C nucleus and the neighboring H nucleus. The basis used for calculation is pc-2.

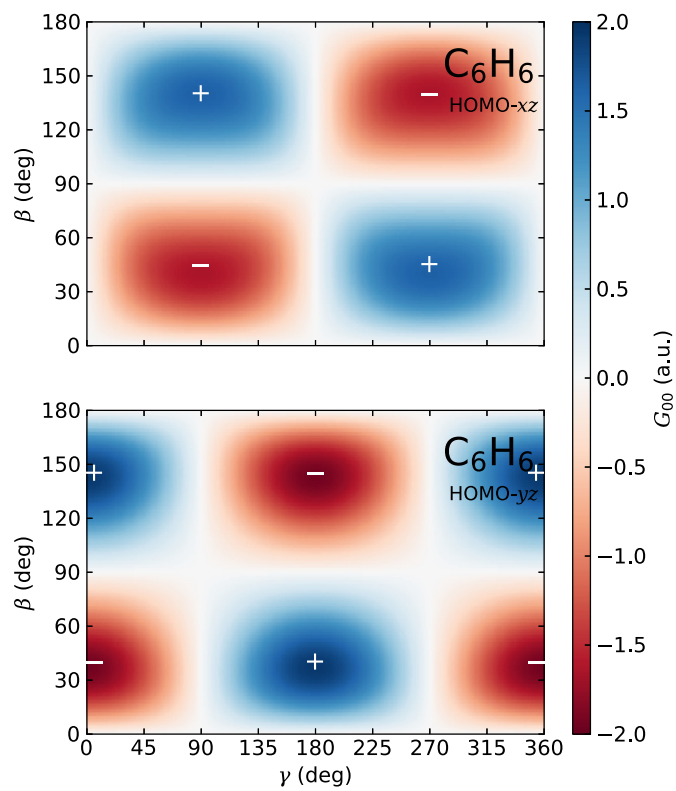


**Fig. 9.** Structure factor $G_{00}$ of the two degenerate HOMOs of the benzene molecule ($C_6H_6$). The molecule is placed in the $x$-$z$ plane in the MF, with the $z$-axis pointing from one C nucleus to its neighboring H nucleus. The basis used for calculation is pc-1.

## CRediT authorship contribution statement

Shanshan Song: Methodology (equal); Software (equal); Visualization (equal); Writing (equal). Mingyu Zhu: Methodology (equal); Software (equal); Visualization (equal); Writing (equal). Hongcheng Ni: Conceptualization (lead); Methodology (equal); Writing (equal); Supervision (equal). Jian Wu: Conceptualization (equal); Writing (equal); Supervision (equal).

## Declaration of competing interest

The authors declare no conflicting interests.

## Data availability

The code is made open-source in github, and our manuscript has included the link to the source code. In addition, all figures used in the manuscript can be generated from the examples folder in code.

## Acknowledgements

We would like to thank Qiming Sun, Hao Huang, and Shengzhe Pan for helpful discussions. This work is supported by the National Natural Science Foundation of China (Grant Nos. 92150105, 11834004, 12227807, and 12241407) and the Science and Technology Commission of Shanghai Municipality (Grant No. 21ZR1420100). Numerical computations were in part performed on the East China Normal University Multifunctional Platform for Innovation (001).

## References

[1] J.L. Krause, K.J. Schafer, K.C. Kulander, Phys. Rev. Lett. 68 (1992) 3535–3538, https://doi.org/10.1103/PhysRevLett.68.3535.
[2] P.B. Corkum, Phys. Rev. Lett. 71 (1993) 1994–1997, https://doi.org/10.1103/PhysRevLett.71.1994.
[3] T. Popmintchev, M.-C. Chen, P. Arpin, M.M. Murnane, H.C. Kapteyn, Nat. Photonics 4 (2010) 822–832, https://doi.org/10.1038/nphoton.2010.256.
[4] P. Agostini, F. Fabre, G. Mainfray, G. Petite, N.K. Rahman, Phys. Rev. Lett. 42 (1979) 1127–1130, https://doi.org/10.1103/PhysRevLett.42.1127.
[5] W. Becker, F. Grasbon, R. Kopold, D. Milošević, G. Paulus, H. Walther, Adv. At. Mol. Opt. Phys. 48 (2002) 35–98, https://doi.org/10.1016/S1049-250X(02)80006-4.
[6] D.B. Milošević, G.G. Paulus, D. Bauer, W. Becker, J. Phys. B 39 (2006) R203, https://doi.org/10.1088/0953-4075/39/14/R01.
[7] B. Walker, B. Sheehy, L.F. DiMauro, P. Agostini, K.J. Schafer, K.C. Kulander, Phys. Rev. Lett. 73 (1994) 1227–1230, https://doi.org/10.1103/PhysRevLett.73.1227.
[8] C.I. Blaga, J. Xu, A.D. DiChiara, E. Sistrunk, K. Zhang, P. Agostini, T.A. Miller, L.F. DiMauro, C.D. Lin, Nature 483 (2012) 194–197, https://doi.org/10.1038/nature10820.
[9] B. Wolter, M.G. Pullen, A.-T. Le, M. Baudisch, K. Doblhoff-Dier, A. Senftleben, M. Hemmer, C.D. Schröter, J. Ullrich, T. Pfeifer, R. Moshammer, S. Gräfe, O. Vendrell, C.D. Lin, J. Biegert, Science 354 (2016) 308–312, https://doi.org/10.1126/science.aah3429.
[10] F. Krausz, M. Ivanov, Rev. Mod. Phys. 81 (2009) 163–234, https://doi.org/10.1103/RevModPhys.81.163.
[11] J.M. Dahlström, A. L'Huillier, A. Maquet, J. Phys. B 45 (2012) 183001, https://doi.org/10.1088/0953-4075/45/18/183001.
[12] R. Pazourek, S. Nagele, J. Burgdörfer, Rev. Mod. Phys. 87 (2015) 765–802, https://doi.org/10.1103/RevModPhys.87.765.
[13] A.S. Kheifets, J. Phys. B 56 (2023) 022001, https://doi.org/10.1088/1361-6455/acb188.
[14] L.V. Keldysh, Sov. Phys. JETP 20 (1965) 1307.
[15] S.L. Chin, From Multiphoton to Tunnel Ionization, Advances in Multi-Photon Processes and Spectroscopy, vol. 16, World Scientific, 2004.
[16] O. Smirnova, S. Patchkovskii, Y. Mairesse, N. Dudovich, D. Villeneuve, P. Corkum, M.Y. Ivanov, Phys. Rev. Lett. 102 (2009) 063601, https://doi.org/10.1103/PhysRevLett.102.063601.
[17] S. Sukiasyan, S. Patchkovskii, O. Smirnova, T. Brabec, M.Y. Ivanov, Phys. Rev. A 82 (2010) 043414, https://doi.org/10.1103/PhysRevA.82.043414.
[18] X. Xie, K. Doblhoff-Dier, S. Roither, M.S. Schöffler, D. Kartashov, H. Xu, T. Rathje, G.G. Paulus, A. Baltuška, S. Gräfe, M. Kitzler, Phys. Rev. Lett. 109 (2012) 243001, https://doi.org/10.1103/PhysRevLett.109.243001.
[19] M. Spanner, J. Mikosch, A.E. Boguslavskiy, M.M. Murnane, A. Stolow, S. Patchkovskii, Phys. Rev. A 85 (2012) 033426, https://doi.org/10.1103/PhysRevA.85.033426.
[20] K. Doblhoff-Dier, K.I. Dimitriou, A. Staudte, S. Gräfe, Phys. Rev. A 88 (2013) 033411, https://doi.org/10.1103/PhysRevA.88.033411.
[21] K. Doblhoff-Dier, M. Kitzler, S. Gräfe, Phys. Rev. A 94 (2016) 013405, https://doi.org/10.1103/PhysRevA.94.013405.
[22] S. Patchkovskii, M.S. Schuurman, Phys. Rev. A 96 (2017) 053405, https://doi.org/10.1103/PhysRevA.96.053405.
[23] H. Li, X. Gong, H. Ni, P. Lu, X. Luo, J. Wen, Y. Yang, X. Qian, Z. Sun, J. Wu, J. Phys. Chem. Lett. 13 (2022) 5881–5893, https://doi.org/10.1021/acs.jpclett.2c01119.
[24] J. Ma, L. Xu, H. Ni, C. Lu, W. Zhang, P. Lu, J. Wen, F. He, O. Faucher, J. Wu, Phys. Rev. Lett. 127 (2021) 183201, https://doi.org/10.1103/PhysRevLett.127.183201.
[25] M.V. Ammosov, N.B. Delone, V.P. Krainov, Sov. Phys. JETP 64 (1986) 1191.
[26] N.B. Delone, V.P. Krainov, Phys. Usp. 41 (1998) 469, https://doi.org/10.1070/PU1998v041n05ABEH000393.
[27] O.I. Tolstikhin, T. Morishita, S. Watanabe, Phys. Rev. A 81 (2010) 033415, https://doi.org/10.1103/PhysRevA.81.033415.
[28] O.I. Tolstikhin, T. Morishita, Phys. Rev. A 86 (2012) 043417, https://doi.org/10.1103/PhysRevA.86.043417.
[29] D. Pavičić, K.F. Lee, D.M. Rayner, P.B. Corkum, D.M. Villeneuve, Phys. Rev. Lett. 98 (2007) 243001, https://doi.org/10.1103/PhysRevLett.98.243001.
[30] X.M. Tong, Z.X. Zhao, C.D. Lin, Phys. Rev. A 66 (2002) 033402, https://doi.org/10.1103/PhysRevA.66.033402.
[31] L. Holmegaard, J.L. Hansen, L. Kalhøj, S. Louise Kragh, H. Stapelfeldt, F. Filsinger, J. Küpper, G. Meijer, D. Dimitrovski, M. Abu-samha, C.P.J. Martiny, L. Bojer Madsen, Nat. Phys. 6 (2010) 428–432, https://doi.org/10.1038/nphys1666.
[32] R. Johansen, K.G. Bay, L. Christensen, J. Thøgersen, D. Dimitrovski, L.B. Madsen, H. Stapelfeldt, J. Phys. B, At. Mol. Opt. Phys. 49 (2016) 205601, https://doi.org/10.1088/0953-4075/49/20/205601.
[33] S.-F. Zhao, J.-K. Li, G.-L. Wang, P.-C. Li, X.-X. Zhou, Commun. Theor. Phys. 67 (2017) 289, https://doi.org/10.1088/0253-6102/67/3/289.
[34] O.I. Tolstikhin, T. Morishita, L.B. Madsen, Phys. Rev. A 84 (2011) 053423, https://doi.org/10.1103/PhysRevA.84.053423.
[35] L.B. Madsen, O.I. Tolstikhin, T. Morishita, Phys. Rev. A 85 (2012) 053404, https://doi.org/10.1103/PhysRevA.85.053404.
[36] L.B. Madsen, F. Jensen, O.I. Tolstikhin, T. Morishita, Phys. Rev. A 87 (2013) 013406, https://doi.org/10.1103/PhysRevA.87.013406.
[37] V.H. Trinh, O.I. Tolstikhin, L.B. Madsen, T. Morishita, Phys. Rev. A 87 (2013) 043426, https://doi.org/10.1103/PhysRevA.87.043426.
[38] L.B. Madsen, F. Jensen, O.I. Tolstikhin, T. Morishita, Phys. Rev. A 89 (2014) 033412, https://doi.org/10.1103/PhysRevA.89.033412.
[39] O.I. Tolstikhin, L.B. Madsen, T. Morishita, Phys. Rev. A 89 (2014) 013421, https://doi.org/10.1103/PhysRevA.89.013421.
[40] I.Y. Tolstikhina, T. Morishita, O.I. Tolstikhin, Phys. Rev. A 90 (2014) 053413, https://doi.org/10.1103/PhysRevA.90.053413.
[41] R. Saito, O.I. Tolstikhin, L.B. Madsen, T. Morishita, At. Data Nucl. Data Tables 103–104 (2015) 4–49, https://doi.org/10.1016/j.adt.2015.02.001.
[42] V.H. Trinh, V.N.T. Pham, O.I. Tolstikhin, T. Morishita, Phys. Rev. A 91 (2015) 063410, https://doi.org/10.1103/PhysRevA.91.063410.
[43] A.I. Dnestryan, O.I. Tolstikhin, Phys. Rev. A 93 (2016) 033412, https://doi.org/10.1103/PhysRevA.93.033412.
[44] J. Svensmark, O.I. Tolstikhin, L.B. Madsen, Phys. Rev. A 93 (2016) 053426, https://doi.org/10.1103/PhysRevA.93.053426.
[45] V.H. Trinh, O.I. Tolstikhin, T. Morishita, J. Phys. B, At. Mol. Opt. Phys. 49 (2016) 195603, https://doi.org/10.1088/0953-4075/49/19/195603.
[46] L.B. Madsen, F. Jensen, A.I. Dnestryan, O.I. Tolstikhin, Phys. Rev. A 96 (2017) 013423, https://doi.org/10.1103/PhysRevA.96.013423.
[47] A.I. Dnestryan, O.I. Tolstikhin, L.B. Madsen, F. Jensen, J. Chem. Phys. 149 (2018) 164107, https://doi.org/10.1063/1.5046902.
[48] P.K. Samygin, T. Morishita, O.I. Tolstikhin, Phys. Rev. A 98 (2018) 033401, https://doi.org/10.1103/PhysRevA.98.033401.
[49] H. Matsui, O.I. Tolstikhin, T. Morishita, Phys. Rev. A 103 (2021) 033102, https://doi.org/10.1103/PhysRevA.103.033102.
[50] P.A. Batishchev, O.I. Tolstikhin, T. Morishita, Phys. Rev. A 82 (2010) 023416, https://doi.org/10.1103/PhysRevA.82.023416.
[51] H. Ohmura, N. Saito, T. Morishita, Phys. Rev. A 89 (2014) 013405, https://doi.org/10.1103/PhysRevA.89.013405.
[52] P.M. Kraus, O.I. Tolstikhin, D. Baykusheva, A. Rupenyan, J. Schneider, C.Z. Bisgaard, T. Morishita, F. Jensen, L.B. Madsen, H.J. Wörner, Nat. Commun. 6 (2015) 7039, https://doi.org/10.1038/ncomms8039.
[53] S.G. Walt, N. Bhargava Ram, A. von Conta, O.I. Tolstikhin, L.B. Madsen, F. Jensen, H.J. Wörner, J. Phys. Chem. A 119 (2015) 11772–11782, https://doi.org/10.1021/acs.jpca.5b07331.
[54] T. Endo, A. Matsuda, M. Fushitani, T. Yasuike, O.I. Tolstikhin, T. Morishita, A. Hishikawa, Phys. Rev. Lett. 116 (2016) 163002, https://doi.org/10.1103/PhysRevLett.116.163002.
[55] H. Fujise, M. Uemura, H. Hasegawa, D. Ikeya, A. Matsuda, T. Morishita, L.B. Madsen, F. Jensen, O.I. Tolstikhin, A. Hishikawa, Phys. Chem. Chem. Phys. 24 (2022) 8962–8969, https://doi.org/10.1039/D1CP05858D.
[56] Q. Sun, T.C. Berkelbach, N.S. Blunt, G.H. Booth, S. Guo, Z. Li, J. Liu, J.D. McClain, E.R. Sayfutyarova, S. Sharma, S. Wouters, G.K. Chan, WIREs Comput. Mol. Sci. 8 (2018) e1340, https://doi.org/10.1002/wcms.1340.
[57] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N.S. Blunt, N.A. Bogdanov, G.H. Booth, J. Chen, Z.-H. Cui, J.J. Eriksen, Y. Gao, S. Guo, J. Hermann, M.R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J.D. McClain, M. Motta, B. Mussard, H.Q. Pham, A. Pulkin, W. Purwanto, P.J. Robinson, E. Ronca, E.R. Sayfutyarova, M. Scheurer, H.F. Schurkus, J.E.T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L.K. Wagner, X. Wang, A. White, J.D. Whitfield, M.J. Williamson, S.

Wouters, J. Yang, J.M. Yu, T. Zhu, T.C. Berkelbach, S. Sharma, A.Y. Sokolov, G.K.-L. Chan, J. Chem. Phys. 153 (2020) 024109, https://doi.org/10.1063/5.0006074.
[58] F. Olver, D. Lozier, R. Boisvert, C. Clark, The NIST Handbook of Mathematical Functions, Cambridge University Press, 2010.
[59] A.D. Becke, J. Chem. Phys. 88 (1988) 2547–2553, https://doi.org/10.1063/1.454033.