communications physics

A Nature Portfolio journal

Review article

https://doi.org/10.1038/s42005-026-02671-y

Tutorial: theoretical methods for attosecond molecular ionization and dynamics

Fernando Martín [1,2] ✉, Jakub Benda [3], Jimena D. Gorfinkiel [4], Zdeněk Mašín [3], Laura Rego [5] & Armin Scrinzi[6]

Ionization of molecules by ultrashort extreme ultraviolet and X-ray pulses triggers complex ultrafast electron-nuclear dynamics central to attochemistry. Unlike atoms, molecules feature multicentric potentials, reduced symmetry, and intrinsic vibrational motion, all of which complicate the theoretical description of continuum states and call for approaches that extend beyond standard atomic physics and quantum chemistry. Although atomic ionization is well understood, existing methods for molecular ionization are often specialized, fragmented across disciplines, and suited only to limited aspects of the problem. In this tutorial, we highlight the key physical features that set molecular ionization apart from its atomic counterpart, examine the challenges involved in modelling these processes, and present selected computational strategies applicable across weak and strong field regimes.

Attosecond pulses currently produced in the laboratory from high-harmonic generation or free-electron lasers lie in the extreme ultraviolet and X-ray spectral regions. Therefore, absorption of a single photon contained in such pulses leads to ionization of any atom or molecule. Intense ultrashort infrared (IR) pulses can also produce a substantial amount of ionization through multi-photon absorption. In molecules, any of these laser sources can lead to very rich electron and nuclear dynamics, generating unusual charge migration, molecular breakup or reactivity, which are the subject of intense investigations in attochemistry[1,2]. Such dynamics are usually resolved in time by using a pump-probe scheme in which the ionizing pulse is combined with a delay-controlled (probe) pulse that further interacts with the molecule. Hence, for the theoretical modeling of these processes, one needs first to describe the eigenstates of the unbound system, which are not L^2 integrable, i.e., do not vanish asymptotically, and often show a much richer node structure than bound states. These features imply that specific computational methods, different from those developed for bound states, are needed[3]. Such methods were introduced for atomic systems many decades ago and are well established. However, for molecules, the number of available methods is much scarcer and developments are still in progress[4].

The additional features not present in atoms that complicate the theoretical description of electronic continuum states in molecules are diverse. An obvious one is the multicentric nature of the potential seen by the escaping electron: in atoms, this potential has spherical symmetry, while in molecules it has not. For small molecules, the potential can still exhibit point-group symmetry, but for larger ones, it may be completely non-symmetric. In all cases, multicentric effects manifest in the electronic continuum states especially at low and medium photoelectron energies, so that single-center methods developed for atoms may not be of practical use except for very small molecules, mainly diatomics. As a consequence of this, the total angular momentum is not a good quantum number in molecules, so that a large, formally infinite number of final-state angular momentum eigenstates is accessible for the electrons in the continuum. Thus, expansions of the electronic continuum wave functions in terms of partial waves may require inclusion of very large values of the photoelectron angular momentum, as, for example, when an electron initially localized in a given molecular region is ionized and is subsequently diffracted by the neighboring atomic nuclei. This is at variance with atomic systems, where angular momentum selection rules limit the number of accessible angular momenta.

Another complication is that the lower symmetry of the molecular system causes coupling of more electronic continua (associated to different ionic states) compared to the atomic case. Hence, molecular systems generally require a higher level of correlation description than atomic systems. Also, at variance with atoms, molecules can have a permanent dipole

[1]Departamento de Química, Universidad Autónoma de Madrid, Madrid, Spain. [2]Instituto Madrileño de Estudios Avanzados en Nanociencia, Campus de Cantoblanco, Madrid, Spain. [3]Institute of Theoretical Physics, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic. [4]School of Physical Sciences, The Open University, Milton Keynes, United Kingdom. [5]Instituto de Ciencia de Materiales de Madrid, Consejo Superior de Investigaciones Científicas, Madrid, Spain. [6]Ludwig Maximilian University, Munich, Germany. ✉e-mail: fernando.martin@uam.es

moment. In strong laser fields, this can lead to molecular alignment. Furthermore, as in atoms, a strong field can polarize the molecule's electronic charge density in the initial and/or the final states.

Finally, but not less important, in a molecule, atomic nuclei move significantly around their equilibrium positions or, equivalently, with respect to its center of mass. As a consequence:

- Ionization does not proceed through a single molecular geometry. All geometries associated with the molecule's zero-point vibrational motion, i.e., belonging to the so-called Franck-Condon region of the system, contribute.
- Ionization does not always precede nuclear motion generated after light absorption, as one would in principle expect due to the much larger mass of the nuclei. This is the case when ionization is "slow", e.g., in the vicinity of autoionizing states with lifetimes of the order of tens or hundreds of femtoseconds, or even picoseconds. In these cases, electronic and nuclear motions must be treated on an equal footing, i.e., ionization must be described taking into account that the nuclei move at the same time (coupled electron-nuclear dynamics). This requires going beyond the usual Born-Oppenheimer approximation. Therefore, electron correlation can be severely affected by nuclear motion, Feshbach resonances being an extreme case of strongly coupled electron-nuclear dynamics due to their slow autoionization decay.

The theoretical description of this complex phenomenology requires the development of approximate methods that integrate concepts from strong-field physics and quantum chemistry. However, the foundational principles of these approaches are often dispersed across specialized literature, typically without direct reference to the molecular ionization problem. Moreover, each method tends to be best suited for specific aspects of the ionization process, complicating the selection of the appropriate framework for a given scenario. This tutorial provides a comprehensive overview of the distinctive features that differentiate molecular ionization from atomic ionization, and outlines strategies to model these processes under both weak and strong field conditions. Also, a selection of computational tools that have successfully addressed molecular ionization is briefly presented. Our aim is to provide a concrete guide to all the theoretical and technical aspects for treating molecular photoionization induced by ultrashort and attosecond pulses, which we hope will be useful both for scientists active in the field and entering it.

Hartree atomic units ($e = \hbar = m_e = 4\pi\varepsilon_0 = 1$) are assumed throughout this article unless stated otherwise.

Key concepts

Scattering states

The simplest model for an unbound electron is a plane wave $\exp[i\mathbf{k}\cdot\mathbf{r}]$ with constant momentum $\mathbf{k}$. This is a reasonable zeroth-order approximation when the electron is scattered by a neutral molecule, as in electron-molecule collisions. However, the unbound electron feels the potential field of the molecule, which is reflected in the phase it accumulates on its way to the detector (see Fig. 1). This is the concept of scattering phase-shift. When the electron wave function is decomposed into its angular momentum components (partial-waves), $\psi_l(r;k)$, one has

$$\psi_l(r;k) \sim \sin(kr - \frac{l\pi}{2} + \delta_l(k)),\ r \to \infty, \tag{1}$$

where l is the angular momentum of partial wave $\psi_l(r;k)$ and $\delta_l(k)$ is the partial-wave phase-shift. The phase-shift becomes negligible for high-energy electrons, which do not have enough time to notice the molecular field. The phase-shift is a key quantity in studies of continuum spectroscopy. In collision problems (e.g. electron-molecule scattering) it is all that is needed to compute all observables (cross-sections, time-delays, etc.) associated with the electron[5].

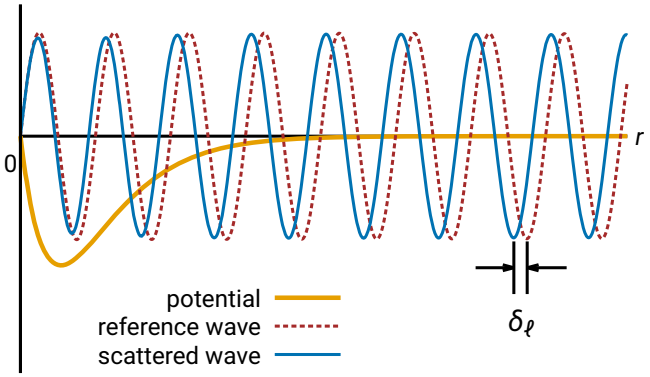

Fig. 1 | Photoionization phase-shift δ_ℓ. When, on its way out of the molecule, the photoelectron experiences an additional attractive short-range potential in addition to the central Coulomb field, its wave function will appear "compressed", implying a positive phase-shift compared to a reference wave function that is not affected by this potential.

In the case of photoionization of neutral molecules, the residual molecule is an ion and the photoelectron partial waves must asymptotically be described by a Coulomb wave instead of a plane wave[5], so that

$$\psi_l(r;k) \sim \sin(kr - \eta\ln(2kr) - \frac{l\pi}{2} + \sigma_l(\eta) + \delta_l(k)),\ r \to \infty, \tag{2}$$

where $\eta = -\frac{Z}{k}$, Z is the residual ion charge, and $\sigma_l(\eta) = \arg\Gamma(l+1+i\eta)$ is the Coulomb phase, but the underlying principles remain the same.

As mentioned in the introduction and described in more detail in the "Boundary Conditions" subsection in the Key Concepts, due to the non-spherical symmetry of the molecular potential, the number of partial waves that effectively contribute to the ionization process, hence the number of phase shifts, can be very large. So, it is more convenient to consider the so-called eigenphase sum[6], obtained from diagonalization of the scattering S-matrix.

Both in scattering and photoionization problems, the scattering phase-shift contains information about the molecular potential. Therefore, monitoring photoelectrons allows one to study molecular properties, which is the key concept behind time-resolved photoelectron spectroscopy[7]. To model photoionization, however, we require not only the phase-shift but crucially also the dipole matrix element between a bound (initial) and a continuum (final) wave function to describe the transition induced by an external electromagnetic field between these states.

Boundary conditions

In photochemistry, IR or UV photons are used to excite the molecule vibrationally or electronically to one of its discrete states, but the electrons remain bound to the molecule or its fragments, and the electronic molecular wave function decays exponentially with the distance from the atomic nuclei. As a consequence, it is associated with a discrete value of the energy, which is typically the object of many bound-state calculations. This bound state character of the wave function is easily embedded in most quantum chemistry calculations by the use of Gaussian-type orbitals, see below. In a stationary (time-independent) picture, when the electron is unbound, the wave function no longer decays exponentially (see Eqs. (1) and (2)) and is composed of infinitely many continuum states for each of the different asymptotic angular momenta[8]. If one is not interested in a particular electron emission direction, the N-electron continuum wave function satisfies the boundary conditions:

$$\begin{aligned} r_{N_e} \to \infty,\ & \Psi^-_{\alpha E}(\mathbf{r}_1,\ldots,\mathbf{r}_{N_e}) \\ &= \frac{1}{N_e}\sum_\beta N_{\beta E}\Upsilon_\beta(\mathbf{r}_1,\ldots,\mathbf{r}_{N_e-1};\hat{r}_{N_e},\zeta_{N_e})\frac{u^-_{\beta,\alpha E}(r_{N_e})}{r_{N_e}}, \end{aligned} \tag{3}$$

where $\mathbf{r}_i$ denotes the position and spin coordinates of electron i, $\hat{r}_{N_e}$ are the angular coordinates of electron N_e, ζ_{N_e} is the spin component of electron N_e, Υ_β is the electronic wave function of the remaining $(N-1)$ electrons coupled to the spin and angular momentum of the continuum electron, $N_{\beta E}$ is a normalization factor that ensures the correct asymptotic behavior, and $u^-_{\beta,\alpha E}(r)$ is the radial function that describes the continuum electron, which is asymptotically given by

$$u^-_{\beta,\alpha E}(r) = \delta_{\alpha\beta}\sqrt{\frac{2}{\pi k_\alpha}}e^{i\Theta_\alpha(r)} - \sqrt{\frac{2}{\pi k_\beta}}e^{-i\Theta_\beta(r)}S^*_{\beta\alpha}, \tag{4}$$

with

$$\Theta_\alpha(r) = k_\alpha r + \frac{Z}{k_\alpha}\ln(2k_\alpha r) - \ell_\alpha\pi/2 + \sigma_{\ell_\alpha}(k_\alpha), \tag{5}$$

where $S_{\alpha\beta}$ is the *on-shell* scattering matrix[5], Z is again the parent-ion charge, k_α the absolute value of the momentum of the continuum electron in the α channel, and σ_{ℓ_α} its Coulomb phase. Thus, the scattering wave function $\Psi^-_{\alpha E}$ behaves asymptotically as a combination of incoming spherical waves for all open channels (α, β, and γ in Fig. 2) and an outgoing spherical wave in channel α corresponding to an outgoing flux of 1.

At variance with calculations of bound states, scattering and photoionization problems require solving the Schrödinger equation for the unknown scattering phases (or amplitudes) at a fixed kinetic energy of the unbound electron. A number of specialized theoretical methods have been developed to solve the time-independent Schrödinger equation by imposing the proper boundary conditions. These include the R-matrix, Schwinger and complex Kohn variational methods, the Exterior Complex Scaling (ECS) method, and many others[9]. They obtain the complex-valued solutions obeying the above incoming-wave boundary conditions appropriate for photoionization problems, outgoing-wave boundary conditions for scattering problems or real-valued standing wave boundary conditions. Obtaining one of the three solutions is sufficient to obtain the other ones by a simple basis transformation. When the continuum state is obtained, it can be used to evaluate, for example, amplitudes for weak-field 1-photon ionization by computing the dipole matrix elements $d_{i,\alpha}(E) = E_0\langle\Psi_i|\mathbf{u}\cdot\widehat{\mathbf{D}}|\Psi^-_{\alpha,E}\rangle$, where $\widehat{\mathbf{D}}$ is the dipole operator, the monochromatic field with frequency ω is represented by the electric field amplitude $E_0\mathbf{u}$ and $E = E_i + \hbar\omega$, see the "General methodologies" Section.

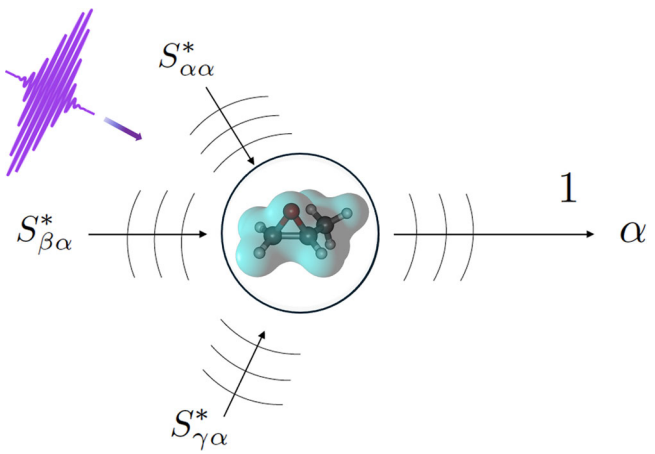

Fig. 2 | Scattering boundary conditions. The S-matrix element $S^*_{\beta\alpha}$ describes the probability amplitude for transition of the system from the channel β into the channel α induced by the non-spherical field of the molecule in the absence of an external field. The equipotential surface of a highly non-spherical molecule is shown in the middle. Probabilities associated with the channel α must sum up to 1. In weak fields, the external field (represented as a wave packet) is accounted for perturbatively by a dipole matrix element between the field-free scattering state and the initial bound state.

As we will see in "Time-dependent Schrödinger equation (TDSE) methods" subsection in "General methodologies", to describe the interaction of a short pulse with a molecule, very often it is more convenient to solve the time-dependent Schrödinger equation. In this case, the photoelectron is represented by a wave packet containing components that will eventually travel away from the molecule to infinity. Representing such an escaping wave packet usually requires a very large simulation domain (represented by a finite difference grid or numerical functions such as B-splines) and/or the use of absorbing boundary conditions, exterior complex scaling[9] or other approaches such as tSurff[10].

Spherical vs. point group symmetry

In atomic, spherically symmetric, systems the electronic Hamiltonian commutes with arbitrary rotation operators. Hence, the eigenstates of the Hamiltonian can be partitioned into subspaces classified by angular momentum quantum numbers. In other words, the total electronic angular momentum quantum number, L, and the total magnetic angular momentum quantum number, M_L, are good quantum numbers. In molecules, there are generally only a finite number of symmetry operators, comprising discrete rotations and reflections as well as inversion, that commute with the molecular Hamiltonian. These operators form a point group. The finite number of combinations of quantum numbers associated with them (characters of the point group) then leads to a finite number of subspaces (irreducible representations)[11]. The total electronic angular momentum is not a good quantum number for molecules (although the projection of M_L along the internuclear axis is a good quantum number for linear molecules). However, every molecular state will belong to an irreducible representation.

The discrete molecular symmetries, similarly to spherical symmetry, may lead to selection rules[12,13]. In one-photon ionization, the photoionization amplitude $A_{fi} \sim \langle\Psi_i|V|\Psi^-_{\alpha,E}\rangle$ from the initial state Ψ_i to the final state $\Psi^-_{\alpha,E}$—being a scalar—is non-zero only when the combined symmetry of Ψ_i, $\Psi^-_{\alpha,E}$ and the potential V ($V = \mathbf{u}\cdot\widehat{\mathbf{D}}$ in the dipole approximation) belongs to the totally symmetric irreducible representation, i.e., the representation in which application of any of the symmetry operators of the point group has no effect. In the case of atoms this translates into clear selection rules for L: only states with specific values of L will be populated. However, in molecular systems, the selection rules do not impose very strict constraints on angular momentum: they affect the irreducible representation the final state belongs to, but this can be associated to many angular momenta of the photoelectron. As a result, even simple absorption of a single polarized photon by a molecule will populate many angular momentum states of the photoelectron, reflecting the complicated shape of the molecular orbital from where the electron was ejected[14].

Experiments are often performed on randomly oriented molecules in the gas phase, which requires averaging of the calculated quantities over orientations. This can be done analytically from transition amplitudes calculated for three different orientations, in both time-independent and time-dependent cases when the laser-molecule interaction is perturbative (see, e.g., ref. 15 for a particular application), but elsewhere, typically in time-dependent calculations for molecules in strong fields, it requires several separate calculations at selected orientations because the distortion of the molecular potential by the external field also depends on orientation.

Short vs. long wavelengths and low vs. high intensities

At wavelengths in the XUV or shorter and low field intensities, the dominant ionization process is one-photon absorption, which can easily be described within first-order perturbation theory. As the intensity of the field rises, the molecule can absorb several photons. This is strictly necessary for ionization when IR wavelengths are used (multi-photon ionization). In fact, for high enough intensity, the electron can absorb more photons than those required to overcome the ionization potential of the molecule, I_p, in a process called above-threshold ionization (ATI), which has played a key role

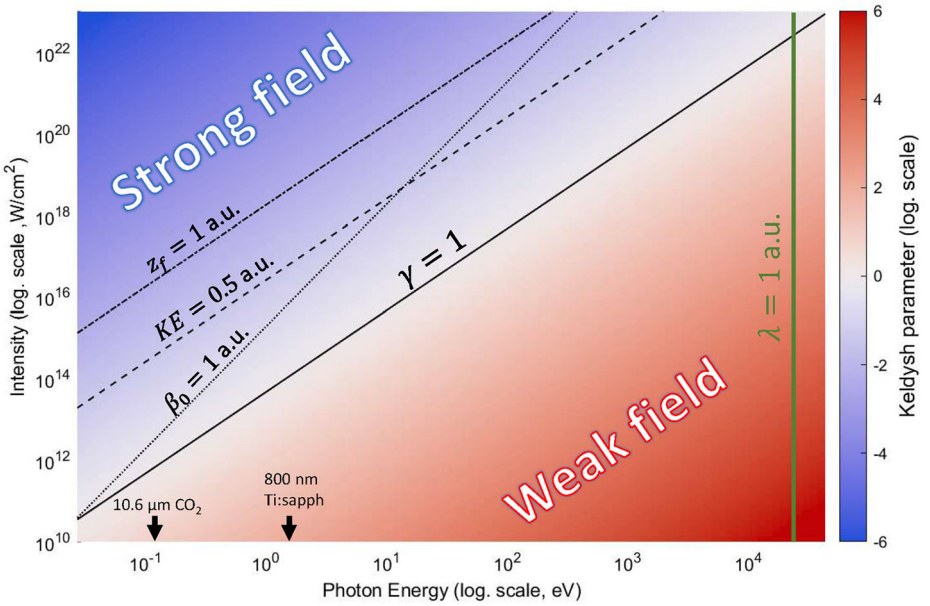

Fig. 3 | Partition of the laser parameter space into the weak- and strong-field domains. The background color indicates the value of the Keldysh parameter[17], $\gamma = \sqrt{\frac{I_P}{2U_P}}$, for a Hydrogen atom (ionization potential of $I_P = 0.5$ a.u.). The tunneling regime is achieved for $\gamma < 1$, while for $\gamma \gg 1$ the regime is perturbative (multi-photon), and over-barrier ionization (i.e. barrier suppression) occurs for $\gamma \ll 1$. The vertical green line indicates an estimated upper limit of the dipole approximation. The dashed and dotted curves indicate alternative lower limits of the dipole approximation. The $z_f = 1$ a.u. indicates the relativistic limit ($U_P = \frac{1}{2}mc^2$). The $KE = 0.5$ a.u. indicates the limit imposed by the radiation pressure. The $\beta_0 = 1$ a.u. line indicates the limit imposed by the displacement in the field propagation direction caused by the magnetic field. The black arrows point out the commonly used Ti:sapph and CO_2 lasers photon energy. Note that the distinction between strong and weak fields is system-dependent and it is sensitive to the description of the electric field as plane wave vs. as a quasistatic field[73]. Figure modified from ref. 73.

in the development of strong-field physics[16]. More complex processes can occur in which the photoelectron not only absorbs but also emits photons. As long as the number of absorbed and emitted photons is reasonably small, the picture of individual absorptions and emissions is meaningful and perturbation theory, treating the field as a perturbation, is a practical computational tool.

An electron moving under the influence of an even stronger oscillating laser field follows a drifting oscillatory trajectory (quiver motion). The amplitude of that motion grows linearly with the field strength and *quadratically* with the wavelength. As a result, the wave function can grow to very large sizes during a pulse that lasts just a few field oscillations. For example, at a moderate laser intensity of 10^{14} W/cm^2 and wavelength 800 nm, the wave function expands to radii greater than 50 Bohr during a single oscillation of the field, and expands to hundreds of Bohr during subsequent cycles. The energy acquired during those oscillations is given by the so called ponderomotive potential U_p.

Apart from intensity, the laser photon energy $\hbar\omega$ crucially determines whether an interaction must be considered strong-field or not. For example, at photon energies of ~100 eV, typically found at a free-electron laser, even intensities of 10^{16} W/cm^2 can be treated in first-order perturbation theory, while the same intensity at near-infrared wavelengths around 800 nm (~1.5 eV) causes complete obliteration of the valence electron structure. The frontier between the weak and strong-field regimes is approximately described by the Keldysh parameter[17], $\gamma = \sqrt{\frac{I_P}{2U_P}}$, which combines the molecule's ionization potential I_p with the field wavelength and intensity through U_P (see Fig. 3). When $\gamma \gg 1$, the laser-molecule interaction is reasonably described by perturbation theory (single- and multi-photon ionization), while when $\gamma \sim 1$ or $\ll 1$ one enters the strong field regime and a completely different picture must be adopted.

We can reach an extreme case when the photoelectron dynamics is completely dominated by the field. Then we can ignore the effect of the molecular field on the photoelectron and instead replace the photoelectron wave function with a modified plane wave called Volkov function[5] describing the motion of a free electron in the external field. In length gauge and in dipole approximation (see the "General methodologies" section)[16], it is given by the expression

$$\Phi(\mathbf{r}, t, t') = \frac{1}{2\pi^{3/2}} \exp[iS_V(\mathbf{p}, t, t')] \exp[i(\mathbf{p} + \mathbf{A}(t)) \cdot \mathbf{r}], \quad (6)$$

$$S_V(\mathbf{p}, t, t') = \frac{1}{2}\int_{t'}^{t} d\tau[\mathbf{p} + \mathbf{A}(\tau)]^2, \quad (7)$$

where S_V is the Volkov action and t' and t are the starting and current times of the laser pulse, respectively. The instantaneous electron momentum is now determined by the vector potential $\mathbf{A}(t) = -\int^t \mathbf{E}(t')\mathrm{d}t'$, where $\mathbf{E}(t)$ is the electric field.

Perturbation theory can still be used in this extreme case, but in a sense reversed from the weak field case: now the molecular field is a perturbation to the strong field of the laser. This is the essence of the Strong Field Approximation (SFA), which is frequently employed in semi-classical models of strong-field phenomena including High Harmonic Generation (HHG)[18]. The division between the weak (perturbative) and the SFA regimes is not sharp and in many practical situations both types of processes contribute, which is why it is often necessary to use ab initio methods to solve the actual time-dependent dynamics of the multi-electron molecule in the external field.

Fixed vs. moving nuclei

The wave function describing a molecular state Ψ of a non-rotating molecule in the center of mass frame depends both on the $3n$ electronic ($\mathbf{r}$) and $3N_A$ nuclear vibrational coordinates ($\mathbf{R}$), where n is the number of electrons and $N_A = 3N - 6$ ($3N - 5$ for linear molecules) with N being the number of atomic nuclei. In a non-relativistic approach, it is the eigenfunction of a Hamiltonian $\mathcal{H}(\mathbf{r}, \mathbf{R})$ that is usually decomposed as a sum of an electronic Hamiltonian $\mathcal{H}_{el}$, which contains the electron kinetic energy operators, the electron-nuclei, electron-electron and the nucleus-nucleus Coulomb interactions, and a nuclear Hamiltonian $\mathcal{H}_N$, which contains the nuclear kinetic energy operators[19]:

$$\left[\mathcal{H}_{el}(\mathbf{r}, \mathbf{R}) + \mathcal{H}_N(\mathbf{R})\right]\Psi(\mathbf{r}, \mathbf{R}) = E\Psi(\mathbf{r}, \mathbf{R}). \quad (8)$$

Except for the simplest H_2^+ molecule, this equation cannot be solved exactly in full dimensionality. Since in many cases electronic motion is much faster than nuclear motion, one usually relies on the Born–Oppenheimer (BO) approximation in which one assumes that the molecular wave function can be written as the product of an electronic wave function $\psi_{el}(\mathbf{r}; \mathbf{R})$, which parametrically depends on the nuclear coordinates and satisfies the

eigenvalue equation

$$\mathcal{H}_{el}(\mathbf{r};\mathbf{R})\psi_{el}(\mathbf{r};\mathbf{R}) = E_{el}(\mathbf{R})\psi_{el}(\mathbf{r};\mathbf{R}), \quad (9)$$

and a nuclear wave function that satisfies

$$\mathcal{H}_N(\mathbf{R})\chi(\mathbf{R}) = E\chi(\mathbf{R}), \quad (10)$$

so that

$$\Psi(\mathbf{r},\mathbf{R}) = \psi_{el}(\mathbf{r};\mathbf{R})\chi(\mathbf{R}). \quad (11)$$

Thus, in the BO approximation one has to solve the electronic Schrödinger equation (9) in a dense grid of $\mathbf{R}$ coordinates and, from the resulting $E_{el}(\mathbf{R})$ eigenvalues, which depend parametrically on $\mathbf{R}$ and represent the potentials felt by the nuclei (and are included in $\mathcal{H}_N$), obtain the nuclear wave functions and the eigenvalues E associated with the energy levels of the molecule. Therefore, in this scheme, the nuclei move in a sort of mean field created by the electrons in their rapid motion. In the case of ionization, this can be a reasonable approximation when the kinetic energy of the ejected electron is not too low, i.e., when ionization occurs far enough from the ionization threshold or does not proceed from the slow decay of autoionizing resonances. Still, obtaining the electronic continuum states in a dense grid of $\mathbf{R}$ coordinates from Eq. (9) can be a major computational challenge for most molecules, except for diatomics. So, a usual additional approach consists in disregarding the parametric dependence on $\mathbf{R}$ and solving Eq. (9) for the equilibrium geometry $\mathbf{R}_{eq}$ only, leading to the so-called fixed-nuclei approximation (FNA):

$$\mathcal{H}_{el}(\mathbf{r};\mathbf{R}_{eq})\psi_{el}(\mathbf{r};\mathbf{R}_{eq}) = E_{el}(\mathbf{R}_{eq})\psi_{el}(\mathbf{r};\mathbf{R}_{eq}). \quad (12)$$

This is by far the most widely used approximation in describing molecular ionization. In general, it provides reasonable descriptions of total photoionization spectra as a function of photon energy in those regions where the BO approximation is expected to work. However, it has important limitations. In high energy-resolution experiments, measured photoelectron spectra at a given photon energy exhibit vibrational progressions that cannot be accounted for by the FNA. Nevertheless, when the relevant ionization dynamics occur in a narrow region around the equilibrium geometry, one can assume that ψ_{el} will not be too dependent on $\mathbf{R}$ and then use this wave function for all $\mathbf{R}$. For electronic continuum states, $E_{el}(\mathbf{R}) = E_{el}^{+}(\mathbf{R}) + \epsilon$, where E_{el}^{+} is the potential energy surface of the remaining cation (see Fig. 4) and ϵ is the kinetic energy of the ionized electron, which does not depend on $\mathbf{R}$. Hence, solving Eq. (10) to obtain E and $\chi(\mathbf{R})$ only requires knowledge of the potential energy surface of the bound electronic state of the cation, which can be readily obtained by using standard quantum chemistry approaches. The resulting wave function can then be written:

$$\Psi(\mathbf{r},\mathbf{R}) = \psi_{el}(\mathbf{r};\mathbf{R}_{eq})\chi(\mathbf{R}), \quad (13)$$

which is the Franck-Condon (FC) approximation. This has been widely used to evaluate vibrationally resolved photoelectron spectra in those energy regions where the BO approximation works. When the conditions for the validity of the latter are not fulfilled, one has to look for methods that are able to handle the original eigenvalue equation (8). This will be discussed in the "Coupled electronic and nuclear motion" section.

Electronic continuum in molecules

Basis sets or numerical approaches

The computational description of electronic bound states of molecules is usually based on the use of Gaussian-type orbitals (GTOs):

$$G_{\alpha,l,m}(\mathbf{r}) = N_{\alpha,l}^{GTO} r^{l} X_{l,m}(\hat{\mathbf{r}}) \exp[-\alpha r^2], \quad (14)$$

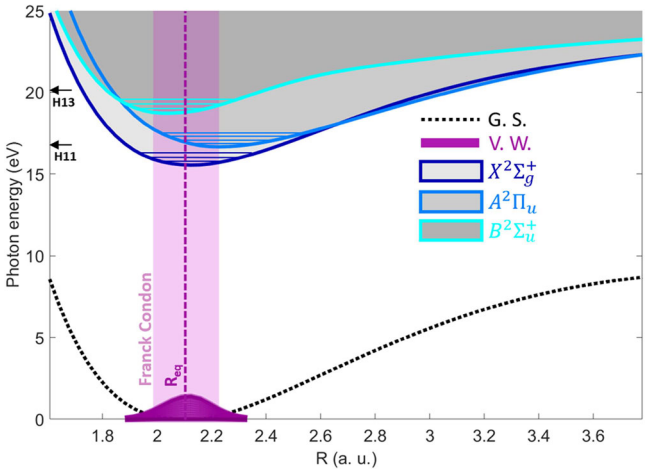

Fig. 4 | Potential energy curves of the N_2 and N_2^+ molecules as a function of the internuclear distance. The black dotted curve corresponds to the ground state of N_2, while the dark blue, light blue and cyan curves correspond to the first ($X^2\Sigma_g^+$), second ($A^2\Pi_u$) and third ($B^2\Sigma_u^+$) states of N_2^+, respectively. The gray areas indicate the electronic continua associated to each ionic state and the horizontal lines show the lowest vibrational levels of each parent ion. The filled magenta area shows the $v = 0$ vibrational wavepacket of the ground state, which determines the Franck-Condon region (pink area) around the equilibrium internuclear distance, R_{eq}, indicated by the magenta dashed line. The horizontal arrows point out the photon energy of the harmonics 11 and 13 generated from a typical 800 nm driving wavelength during HHG, as a reference.

where $X_{l,m}(\hat{\mathbf{r}})$ is a real spherical harmonic, as the basis functions for the expansion of the single-particle orbitals. The use of GTOs has clear numerical advantages: most of the integrals that are needed can be evaluated quickly and accurately in closed form. For this reason, a wide range of GTO basis sets with different characteristics are readily available[20,21]. Although GTOs do not have the correct behavior near the nuclei, they can provide an excellent description of bound electrons in molecules.

The situation is more complicated when describing an unbound electron. The functions describing an electron in the continuum are required to cover a much larger spatial domain and have, in general, significantly more nodes than the functions describing bound electrons; cf. Fig. 5. Molecules introduce an additional complication. When dealing with atoms, the basis functions, both bound and continuum, are centered on the nucleus of the atom. In the case of molecules, the bound basis sets are usually multicentric as this ensures a better description of the orbitals (single-center basis can still be used and can provide a reasonable description for diatomic molecules[22]). When combined with the continuum basis, this can lead to issues of over-completeness and linear dependence and usually requires the implementation of orthogonalization procedures in the software employed.

To avoid this problem, a number of approaches have been implemented and used to describe the electronic continuum. These include use of numerical Bessel and Coulomb functions, particularly for linear targets, specially designed GTO basis sets[23–25], complex GTOs[26], B-splines[22,27], combinations of (real) GTOs and B-splines[28–30], discrete variable representations[31], etc. B-spline orbitals (BTOs)

$$\mathcal{B}_{i,l,m}(\mathbf{r}) = N_i \frac{B_i(r)}{r} X_{l,m}(\hat{\mathbf{r}}) \quad (15)$$

are built as the product of a B-spline function of the radial coordinate (piecewise polynomial with compact support) and a (real) spherical harmonic. When GTOs are used, alone or in combination with other functions, the number and exponents of the Gaussians for each angular momentum need to be selected. The UKRmol+ suite, for example, makes use of GTOs fitted to Bessel or Coulomb functions[24,32]: the fit is performed for a specific radial range and a maximum kinetic energy of the continuum

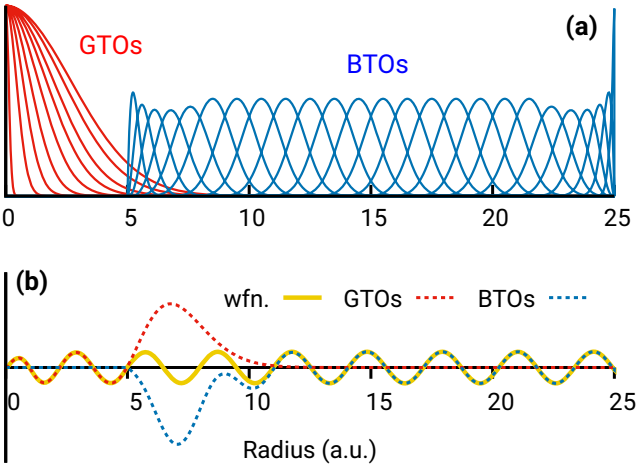

Fig. 5 | Radial basis functions. The representation of continuum wave functions (yellow line in panel **b**) requires an extensive radial basis, for example, (panel **a**) a combination of GTOs (red lines in both panels) and B-splines (BTOs, blue lines). The GTO part on its own cannot reproduce distant nodes of the continuum wave function (yellow line) accurately as shown in panel (**b**). Only the radial part of the wave function is shown.

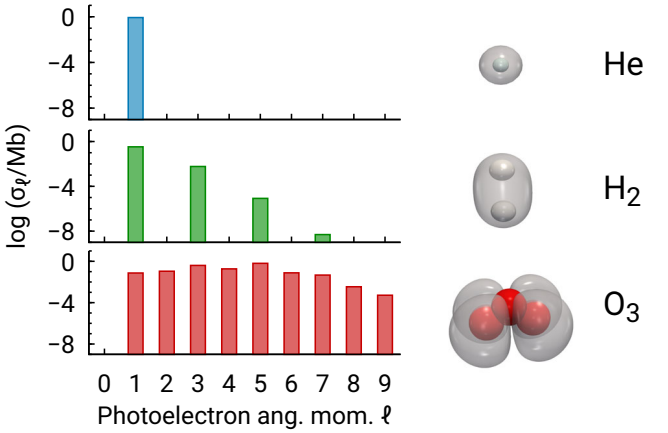

Fig. 6 | Partial-wave cross sections. Angular momentum contributions σ_ℓ to the total photoionization cross section σ for several targets with decreasing symmetry. Calculations performed using UKRmol+ for single-channel photoionization from the highest occupied molecular orbital (HOMO, pictured) at the Hartree–Fock level for a photoelectron energy E_k = 50 eV.

electron. In the XChem package[33], the exponents follow a geometrical sequence. Other approaches have been proposed (see ref. 34 for an evaluation of how some of these perform when modeling high-harmonic generation spectra). Building a B-spline basis set is more straightforward and is done by choosing the order and number of the polynomials, and the radial and angular momentum ranges to be covered. Following a similar spirit, the continuum can also be described using a finite element discrete variable representation[31] (FE-DVR, e.g. in the tRecX package[35]) for the radial coordinate and spherical harmonics for the angular coordinate or other grid-based finite difference approaches (e.g. the RMT package[36]). The construction of basis sets to model molecules in strong fields remains an area of active research[37,38].

The angular momentum range to cover depends on the (non-)symmetry of the molecule, as discussed above, as well as on the range of kinetic energies of the ionized electron and also on the method used. In the time-independent picture, the number of angular momenta accessible by photoionization, though infinite in principle, is in practice limited to a finite range, as the ionization probability becomes very small for very large values of the angular momentum (see Fig. 6). In time-dependent simulations, though, the need to model the instantaneous effect of the field on the photoelectron wave packet generally requires significantly larger angular momentum expansions, in particular, when strong laser pulses are involved. This becomes even more important with increasing intensity of the field.

In contrast to the one-particle continuum, description of the double continuum—simultaneous photoemission of two electrons—is much less developed. Most successful approaches rely on the exterior complex scaling method[39], which has been shown to provide very accurate results in diatomic molecules, even for properties fully differential in the momentum of each of the ejected electrons[40]. At the same time, for highly correlated systems, inclusion of coupling to the double electron continuum may be important and may affect ionization observables in some channels. In state-of-the-art methods, the double continuum is either included directly for small, two-electron systems, or simulated by inclusion of many positive-energy pseudostates for more complex targets.

Integral evaluation problem

To determine observables one needs to calculate the electronic wave function of the system. This is done by solving the time-independent (Eq. (9)) or time-dependent Schrödinger equations. The interactions of electrons with nuclei, with the external field and with each other are described by matrix elements of the molecular Hamiltonian between configuration state functions (CSFs), i.e., multi-electronic functions built from molecular orbitals. These matrix elements are then reduced, by means of Slater rules[41], to sums of integrals over the underlying 1-electron GTOs and/or numerical functions. The most difficult ones to evaluate are the Coulombic electron-electron repulsion (also called two-electron) integrals. They have the form of a 6-dimensional integral over all space

$$(ab|cd) = \int d^3\mathbf{r}_1 \int d^3\mathbf{r}_2 \frac{a(\mathbf{r}_1)b(\mathbf{r}_1)c(\mathbf{r}_2)d(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}, \tag{16}$$

where each of a, b, c, d stands for either a GTO or a numerical orbital. As mentioned above, GTOs are usually located in different atomic centers. When all orbitals are GTOs, the integrals can be evaluated analytically using fast and numerically stable algorithms developed since the 1950s[42]. However, if one of the functions is an orbital based on a numerical function (e.g., a B-spline) either analytic methods cannot be used at all (when the numerical functions and the GTOs are in different centers) or only in a very limited form (when all of them are in the same center). While each individual integral can be evaluated in a reasonable time purely numerically, the total number of required integrals scales as n_o^4, where n_o is the number of orbitals in the basis. In practice, this leads to the challenge of evaluating hundreds of millions to billions of such integrals, depending on the size of the system and the energy of the photoelectron (the higher the energy the more basis functions are needed). Therefore, highly sophisticated and optimized integral evaluation algorithms had to be developed. In atomic physics, numerical functions have been used since 1970s, since single-center atomic integrals are much simpler and reduce to a finite sum of 2-dimensional radial quadratures.

The multi-center nature of the molecular integrals had for a long time prevented a detailed study of multi-electron processes in molecular ultrafast physics. Instead, simpler 1-electron models were used, which do not require evaluation of 2-electron integrals. Over time a number of different approaches have been developed to tackle the two-electron integral problem. This includes the Legendre expansion approach[28,33], solving the equivalent Poisson equation[43] or grid-based methods[3,35]. All these approaches optimize the set-up of the numerical basis to minimize the number of difficult 2-electron integrals involving numerical functions, and require significant computational resources.

General methodologies

In most common applications, the treatment of the interaction with the external field involves several approximations. First, the external field is

treated classically (while the molecule is described in a quantum manner). Second, the magnetic component of the field is neglected (as its interaction with the molecule is much weaker than that of the electric field). Third, the electric field is assumed not to vary spatially, $\mathbf{E}(\mathbf{r}, t) \approx \mathbf{E}(t)$, which is known as the dipole approximation (valid as long as its wavelength is much larger than the size of the molecule). Under these approximations, the interaction of the system with the external laser field is described as $V(\mathbf{r}, t) = e\mathbf{E}(t) \cdot \mathbf{r}$ (length gauge) or $V(\mathbf{r}, t) = \frac{e}{m}\mathbf{A}(t) \cdot \mathbf{p}$ (velocity gauge). To specify the field pulse, one can choose the vector potential $\mathbf{A}(t) = -\mathbf{u}E_0F(t)\sin(\omega t + \phi)/\omega$ and obtain $\mathbf{E}(t)$ by time-derivation, see Section "Short vs. long wavelengths and low vs. high intensities". This excludes any unphysical zero-frequency component from $\mathbf{E}(t)$. E_0, ω and ϕ are the field's amplitude, frequency and phase, respectively, and the unit vector $\mathbf{u}$ is the polarization direction. For the pulse envelope $F(t)$ one typically chooses Gaussian, trapezoidal, or sine-like functions.

Perturbation methods

If the intensity of the applied field is low enough, $V(\mathbf{r}, t)$ can be treated as a small perturbation of the field-free Hamiltonian $\mathcal{H}_0(\mathbf{r}, \mathbf{R})$, allowing a fairly accurate description of the problem by means of perturbation theory, either time-dependent or time-independent[44]. Photoionization (photodetachment in the case of an anion) is often modeled in the first order of perturbation theory, as a one-photon process, giving time-independent ionization amplitudes as transition matrix elements of $V(\mathbf{r}, t)$ between the initial bound state and the final continuum state with the correct boundary condition[1,4]. Multi-color interference experiments, as for instance the reconstruction of attosecond beating by interference of two-photon transitions, have been successfully modeled using second-order perturbation theory[15] and also generalized to higher orders. At weak fields, the yield from ionization by absorption of a higher number of photons decreases exponentially. Enhancement of the yield requires increasing the intensity, which then naturally leads out of the perturbation regime and into the strong-field domain discussed below.

Time-dependent Schrödinger equation (TDSE) methods

The TDSE provides the quantum mechanical description of the temporal evolution of a quantum system. When the external electric field is weak and the number of absorbed photons is more than one, the most straightforward procedure to solve this equation is to expand the wave function in a large enough basis of molecular states (see the "Coupled electronic and nuclear motion" section). In contrast, for stronger fields, since the interaction is governed by the instantaneous value of the electric field, expansion on other basis function may be more convenient. The general expression of the TDSE is:

$$i\frac{\partial\Psi(\mathbf{r}, \mathbf{R}, t)}{\partial t} = \mathcal{H}(\mathbf{r}, \mathbf{R}, t)\Psi(\mathbf{r}, \mathbf{R}, t), \tag{17}$$

where $\Psi(\mathbf{r}, \mathbf{R}, t)$ is the time-dependent wave function and $\mathcal{H}(\mathbf{r}, \mathbf{R}, t) = \mathcal{H}_0(\mathbf{r}, \mathbf{R}) + V(\mathbf{r}, t)$ is the total Hamiltonian operator. To solve the TDSE, the partial differential Eq. (17) is transformed into a set of ordinary differential equations by representing the wave function by a set of coefficients in the chosen basis, FE-DVR, Gaussians, B-splines, molecular states represented in a basis of those functions (spectral method, see the "Coupled electronic and nuclear motion" section), etc. Then, the solution of the TDSE can be obtained through different numerical methods, such as the Split-Operator or other more general integration methods for differential equations (Finite-Difference Time-Domain Method, Crank-Nicholson, Runge-Kutta, etc), which discretize the time variable and compute the wave function of the system at each time step. In the following, we will briefly explain the Split-Operator procedure. The temporal evolution of the system's wave function at each time step can be described as

$$\Psi(\mathbf{r}, \mathbf{R}, t + \mathrm{dt}) = U(t)\Psi(\mathbf{r}, \mathbf{R}, t), \tag{18}$$

where $U(t) = e^{-i\mathcal{H}\mathrm{dt}}$ is the time evolution propagator. The Split-Operator method considers this propagator as separable into a potential and a kinetic term $U(t) = e^{-i(\widehat{T}+\widehat{V}(t))\mathrm{dt}}$. This exponential is not separable into the product of two exponentials (since $\widehat{T}$ and $\widehat{V}$ do not commute), but a reasonable approximation is to split it symmetrically as $U(t) = e^{-i\widehat{V}(t)\frac{\mathrm{dt}}{2}}e^{-i\widehat{T}\mathrm{dt}}e^{-i\widehat{V}(t)\frac{\mathrm{dt}}{2}}$ (second-order Trotter–Suzuki decomposition). As $\widehat{V}$ is diagonal in position space and $\widehat{T}$ is diagonal in momentum space, using Fast-Fourier Transforms to switch between spaces allows one to solve Eq. (18) as pointwise multiplications, in its simplest form in Cartesian coordinates (or "1D"). After solving the TDSE, the computed wave function evolution can be used to obtain the different desired observables by projecting onto appropriate eigenstates of the field-free Hamiltonian.

Methods for strong laser fields

Beyond perturbation theory. When the laser field is strong, it can significantly disturb the electronic structure of the molecule, rendering perturbation theory ineffective or invalid[45]. Typical effects in strong fields are strong polarization, substantial ionization, and a highly non-linear dipole response leading to the generation of high harmonic radiation. In the strong field regime, the energy structure of the valence-shell electrons is increasingly dominated by the field's electric forces, with a large number of excited bound states and continuum states participating in the dynamics, up to the point where the original field-free electronic structure becomes nearly irrelevant, as for SFA (see Section "Short vs. long wavelengths and low vs. high intensities"). In molecules, where, in general, excitation energies are lower and the structure is more complex, the threshold for strong-field phenomena is typically lower than in atoms.

In single- and multi-photon ionization, where perturbation theory applies, the ionization rates, i.e., the ionization probabilities per unit time, are proportional to powers of the intensity, I^n, where n is the number of absorbed photons. When one reaches the regime of tunnel-ionization, rates increase with field strength, $E = \sqrt{2I}$, as

$$\Gamma_K \sim \exp\left(-2\frac{(2I_p)^{3/2}}{3E}\right), \tag{19}$$

where the ionization potential I_p sets the typical energy scale of the molecule. A signature of strong fields is polarization leading to significant distortion of the wave function. Even when distortion is adiabatic and the system largely returns to the initial state after the pulse, a great number of intermediate virtual states, including continuum states, can be involved, and the binding potential felt by electrons may be temporarily lowered. In this situation, an expansion into bound electronic states alone is insufficient, and one needs to include intermediate electronic continuum states.

With strong fields, the distinction between length and velocity gauge becomes important: while both descriptions are mathematically equivalent *in exact calculations* and are closely related in lowest-order perturbation theory, they have very different properties in the general case. One reason is technical: unbound electrons have much smoother wave functions in velocity gauge, which can be more easily approximated by numerical functions. The other reason is that gauge affects the meaning of approximate models: the transition from length to velocity gauge amounts to a boost in momentum space. In strong fields and velocity gauge, the original field-free wave function does *not* represent the actual field-free state. Therefore, if bound states carry the essence of a model, one needs to work in length gauge. If the process is dominated by the motion of continuum electrons, velocity gauge or a mixed-gauge approach is preferable[46].

Computing rates and spectra from the TDSE. With strong fields, semi-classical or even classical models can provide important insight. However, when specific properties of a molecule are important or when

quantitative results are needed, solving the TDSE is the only resort. One needs to represent the system in a basis that accounts for the essential molecular structure and that can accommodate the strong-field aspects of electronic motion. Once an accurate solution for the time-dependent wave function $\Psi(t)$ (we have omitted the dependence on the coordinates ($\mathbf{r}$, $\mathbf{R}$) for simplicity) is obtained, one needs to extract experimentally relevant observables. These are, in sequence of increasing complexity, ionization yields, the harmonic yield, single-electron emission and multiple-electron emission.

The total ionization yield is obtained from the time-dependent electron density $\rho(\mathbf{r}, t)$ as

$$\lim_{r_{max}\to\infty}\lim_{t\to\infty}\int_{r<r_{max}} d^3\mathbf{r}\rho(\mathbf{r},t). \tag{20}$$

In practice, r_{max} can be limited to the size of the most diffuse molecular orbitals. In the absence of long-lived states, t can be constrained to the scale of the pulse duration, allowing some time for the low-energy electrons to leave the volume $|\mathbf{r}| < r_{max}$. The approach will miss contributions from the decay of long-lived resonances, as one needs $t \gg 1/\Gamma$ for resonance width Γ to account for their contribution. In that case, alternative methods will be needed that are described below.

Harmonic radiation occurs when the acceleration of the system's dipole $\mathbf{d}(t)$

$$\frac{d^2}{dt^2}\mathbf{d}(t) = \frac{d^2\langle\Psi(t)|\mathbf{r}|\Psi(t)\rangle}{dt^2} = \frac{d\langle\Psi(t)|[i\mathcal{H}(t),\mathbf{r}]|\Psi(t)\rangle}{dt} \tag{21}$$

becomes non-linear. The first expression ("length form" of the dipole) emphasizes contributions from large r and therefore usually performs worse than the second, so-called "velocity form". It is important to be aware that, while the harmonic response is qualitatively a single-electron effect, for quantitatively correct high-harmonic spectra, multi-electron effects, and exchange effects, must be included.

Photoelectron spectra, just as ionization yields, are obtained from the asymptotic behavior of $\Psi(T)$ at times T after the pulse has passed. If, in addition to $\Psi(T)$, one knows the scattering solutions in momentum space, e.g. the Volkov solutions $\Phi_{\mathbf{p}}$, Eq. (6), the spectral amplitude for emission with momentum $\mathbf{p}$ is given by $b_{\mathbf{p}} = \langle\Phi_{\mathbf{p}}|\Psi(T)\rangle$[33]. If $\Psi(T)$ becomes too extended, one can resort to recording outgoing flux at some distance r_{max} from the molecule (typically $r_{max} \lesssim 50$ Bohr) and discard $\Psi(t)$ beyond r_{max}. Care has to be taken for the fact that the dipole interaction of the laser has infinite range, and all momenta will be modified as long as a field is present. This is achieved by the method described in ref. 10, where the flux is picked up at some radius outside the reach of the molecular potential.

Finally, a few comments pertinent to solving the TDSE in cases where long-lived resonances and electron correlation effects play an important role. In the case of long-lived resonances, one can replace the time integration after the end of the pulse with a more practical approach. For any T after the end of the pulse, one can use a closed-form expression for the remaining contribution to the spectrum that involves applying the Green's function to the wave function at the end of the pulse: $[H - p^2/2 - i\epsilon]^{-1}\Psi(T)$[47]. Electron correlation manifests itself in the high sensitivity of ionization to the ionization threshold I_p, as can be seen, for example, in the tunnel formula (Eq. (19)). Both initial (neutral) and ionic state energies enter I_p. In the presence of a non-perturbative field, the energy shifts due to polarization further change I_p. For the ions, the combined polarization by the laser field and Coulomb force of the emitted electron creates a significant multi-electron effect[48]. It is also important to mention that, in the present context, exchange can also have an impact on the emission patterns: without including exchange, even seemingly simple observables such as the dependence of photoemission yields on laser polarization are incorrect[48].

Coupled electronic and nuclear motion

In this section, we will explore the methods going beyond the Born-Oppenheimer approximation, which are imperative for interactions in which the electronic and nuclear motion are strongly coupled (such as in the presence of autoionizing Feshbach resonances). The general expression of the TDSE, introduced in the "Time-dependent Schrödinger equation (TDSE) methods" subsection in "General methodologies", couples electronic and nuclear motions. Solving the TDSE in a grid is computationally very demanding, and becomes unfeasible for medium-to-large molecules. To overcome this obstacle, one can resort to the spectral method[4]. For this, one must first obtain the solutions of the time-independent Schrödinger equation, see Eq. (8), and then write the time-dependent molecular wave function as a superposition of the time-independent eigenstates:

$$\Psi(\mathbf{r},\mathbf{R},t) = \sum_b c_b(t)e^{-iE_bt}\Psi_b(\mathbf{r},\mathbf{R}) + \sum_c c_c(t)e^{-iE_ct}\Psi_c(\mathbf{r},\mathbf{R}), \tag{22}$$

where $\Psi_b(\mathbf{r}, \mathbf{R})$ and $\Psi_c(\mathbf{r}, \mathbf{R})$ are the bound and continuum states of the molecule, respectively, and E_b and E_c are their corresponding energies. To find the unknown time-dependent coefficients $c_i(t)$, one must substitute Eq. (22) into (17), which is reduced to the following system of coupled differential equations:

$$i\frac{\partial}{\partial t}\begin{pmatrix}\mathbf{c}_b\\ \mathbf{c}_c\end{pmatrix} = \begin{pmatrix}\mathbf{V}_b^{b'}(t)\mathbf{V}_b^{c'}(t)\\ \mathbf{V}_c^{b'}(t)\mathbf{V}_c^{c'}(t)\end{pmatrix}\begin{pmatrix}\mathbf{c}_{b'}\\ \mathbf{c}_{c'}\end{pmatrix}, \tag{23}$$

where every element in the coupling matrix is a matrix block involving bound-bound or continuum-continuum couplings (diagonal blocks) or bound-continuum couplings (off-diagonal blocks). As mentioned above, this system of equations can be solved through standard numerical methods (Runge-Kutta, Crank-Nicholson, etc.).

The molecular eigenstates $\Psi_i(\mathbf{R}, \mathbf{r})$ in Eq. (22) are usually written as products of electronic and nuclear wave functions as given by Eq. (11). This means that the time-dependent wave function is expanded in a basis of Born-Oppenheimer states, which results in the appearance of off-diagonal non-adiabatic couplings (i.e., couplings between electronic and nuclear motion) in addition to the off-diagonal couplings induced by the external field in Eq. (23). Alternatively, non-adiabatic couplings can be ignored by transforming the basis of adiabatic states into a set of diabatic states[49], which requires evaluating and introducing off-diagonal Hamiltonian couplings in Eq. (23). Another option is to resort to the Born-Huang approach[50], by expanding the time-dependent wave function in a basis of time-independent electronic states multiplied by time-dependent nuclear wave functions:

$$\Psi(\mathbf{r},\mathbf{R},t) = \sum_b \psi_b^{el}(\mathbf{r};\mathbf{R})\chi_b(\mathbf{R},t) + \sum_c \psi_c^{el}(\mathbf{r};\mathbf{R})\chi_c(\mathbf{R},t). \tag{24}$$

In this case, the unknowns are the time-dependent nuclear wave functions $\chi_i(\mathbf{R}, t)$, which are the solutions of a system of coupled differential equations containing, as in the previous case, diagonal and off-diagonal coupling matrix elements. The main difference with the spectral method is that integration of the system of differential equations must now be performed over both nuclear coordinates $\mathbf{R}$ and time t instead of only time. In either approach, one accounts for the non-adiabatic coupling terms arising from the derivatives of the electronic molecular eigenstates with respect to the position of the nuclei, due to the presence of the $-\frac{\hbar^2}{2M}\nabla_R^2$ nuclear kinetic energy operator.

For polyatomic molecules, a full quantum treatment for both electrons and nuclei considering the ionizing electromagnetic field during the time-propagation becomes prohibitively expensive. Still such a treatment is possible in medium-size molecules if one knows the coherent superposition of electronic states induced by the ionizing pulse and the TDSE is integrated starting just at the end of the ionizing pulse. In this more restricted scenario,

Box 1 | Some computational codes for the electronic continuum of molecules

XChem

XChem[29,33] is a solution for an all-electron ab initio calculation of the electronic continuum of molecular systems. XChem combines the tools of quantum chemistry (as implemented in OpenMolcas[60]) and scattering theory to accurately account for electron correlation in the single-ionization continuum of atoms, small and medium-size molecules. XChem combines standard polycentric Gaussian basis functions with B-spline functions to describe the short and long-range parts of the continuum wave functions, respectively. These are connected by a large even-tempered basis of monocentric Gaussian functions that avoids numerical evaluation of expensive integrals involving both polycentric Gaussians and B-splines. As a consequence, it keeps computational cost at the same level as that required to evaluate bound states at the same level of theory. The method is specially designed to accurately represent molecular ionization where electron correlation plays an important role, as e.g. in the vicinity of autoionizing resonances.

UKRmol+

The UKRmol+[28] suite can calculate stationary photoionization states, which in turn may be used to evaluate photoionization amplitudes in the first (or higher) order of perturbation theory. Similarly to XChem, the single-center continuum basis may comprise Gaussians as well as B-splines, both potentially overlapping with the polycentric molecular basis. The solution of the continuum states is performed in the R-matrix paradigm by splitting the problem into an "inner" region, where the many-electron problem is solved using quantum-chemical methods, and the "outer" (or single-ionization) region, where one of the electrons is assigned the role of the photoelectron and moves in the central Coulomb field of the residual ion. The R-matrix method is well suited for evaluation of ionization amplitudes for densely spaced photoelectron kinetic energy grids, where a significant part of the computation may be reused. UKRmol+ also produces the time-independent molecular data required for R-matrix with time[65] calculations using RMT[36].

Tiresia

Tiresia[43] provides solutions of the electronic continuum wave functions of complex polyatomic molecules within the static exchange approximation by solving an effective one-particle Schrödinger equation for the ejected electron while allowing for a full-electron description of the remaining cation when necessary. The continuum wave function is expanded in a local multicentric basis set, with primitive functions built as products of a radial B-spline and a real spherical harmonic. The wave function of the remaining cation can be described at either density functional theory (DFT) or configuration interaction (CI) levels. This provides a computationally efficient description of photoionization properties of complex systems, both in the weak and strong field regimes.

tRecX-haCC

The tRecX-haCC[10] code is a general TDSE solver. It has mostly been used for strong-field problems like two-electron emission from Helium and single-electron electron emission from molecules at high intensities and wavelengths up to 800 nm. For molecules, it uses "hybrid anti-symmetrized Coupled Channels" (haCC) basis. This combines CI states for the neutral molecule and the ions, which account for electron correlation on the field-free quantum chemistry level. The CI basis is augmented by all molecular orbitals and a numerical FE-DVR basis for describing polarization and broad continua that are particularly important in strong fields. The code implements irECS (infinite range exterior complex scaling) and the tSurff method described in Ref. 10, methods that were specifically developed for strong field interactions.

the multi-configuration time-dependent Hartree (MCTDH) method has been shown to be a very efficient computational approach[51,52]. For large molecules, one has to rely on semiclassical approaches in which the electrons are treated quantum mechanically and the nuclei follow classical trajectories. In practice, this implies replacing the nuclear wave functions in Eqs. (22) and (24) by a set of classical trajectories $\mathbf{R}(t)$ starting from a distribution of initial molecular geometries that mimic the initial nuclear wave function at $t = 0$:

$$\Psi(\mathbf{r}, \mathbf{R}(t), t) = \sum_b c_b(t)\psi_b^{el}(\mathbf{r}; \mathbf{R}(t)) + \sum_c c_c(t)\psi_c^{el}(\mathbf{r}; \mathbf{R}(t)). \quad (25)$$

In most applications of semiclassical approaches, transitions between different electronic states due to non-adiabatic couplings are accounted for through a stochastic algorithm, as for example in the trajectory surface hopping (TSH) method[53] and further extensions of such methods[54].

A sample of recent molecular ionization codes

Box 1 briefly outlines some of the available codes to describe the electronic continuum of molecules and photoionization, namely XChem, UKRmol+, Tiresia and tRecX-haCC. They constitute the core of those taught at the "COST/ZCAM School on New Computational Methods for Attosecond Molecular Processes". There are several other codes available, some of them following a philosophy similar to that of the codes described below, as for example, ePolyScat[55], ASTRA[56], those based on the Schwinger variational principle[57], etc. Readers interested in the latter codes are referred to the given references.

In brief, XChem and UKRmol+ have the capability of using both GTOs and B-spline basis sets to provide an accurate description of both the inner and asymptotic regions of the electronic continuum wave functions, as well as of electron correlation by explicitly including couplings between all open ionization channels. They basically differ in the way they impose the appropriate scattering boundary conditions. Tiresia is a full B-spline code that aims at describing large molecular systems. It basically follows the same philosophy, but makes use of the static-exchange approximation to simplify the computations, so that part of electron correlation is not accounted for. Finally, tRecX-haCC has been especially designed to describe strong-field ionization. It makes use of a numerical FE-DVR basis and of infinite range exterior complex scaling to account for the appropriate boundary conditions.

XChem, UKRmol+, and tRecX-haCC are installed on the AMOS Gateway[58], a user-friendly interface for academic use. All four codes interface with well-established quantum chemistry packages to describe bound electronic states, either via Molden format files[59], as UKRmol+ and XChem, or directly, as XChem (OpenMolcas[60] and MOLPRO[61]), Tiresia (ADF[62]), and tRecX-haCC (Columbus[63]). Although all of them provide a description of the electronic continuum of molecules, some of them may be more suitable for specific applications than others. Table 1 provides a ranking of the four codes according to the number of applications reported so far to describe ionization of molecules of different sizes with electromagnetic fields

Table 1 | Capabilities of the XChem, UKRmol+, Tiresia and tRecX-haCC codes ranked from one to three stars

		XChem	UKRmol+	TIRESIA	tRecX
Molecular size	Small	★★★	★★★	★	★★★
	Medium	★	★	★★	—
	Large	—	—	★★★	—
Field intensity	Weak	★★★	★★★	★★★	★★★
	Strong	—	★★★†	—	★★★
Resonances	Shape	★★★	★★★	★★★	★★★
	Feshbach	★★★	★★★	★	★★★
Wavelength	IR	★	★★★	★	★★★
	XUV/X-ray	★★★	★★★	★★★	★★★
Observables	X-sections	★★★	★★★	★★★	★★★
	Phases	★★★	★★★	★★★	★
	MFPADs	★★★	★★★	★★★	★★★
	HHG	—	—	—	★★★
	Time-resolved	★★★∘	★★★†	★★★∘	★★★
Nuclear motion	Fixed nuclei	★★★	★★★	★★★	★★★
	Moving (Q)	★★★*	★★★×	★*	—
	Moving (C)	★□	—	★★★□	—

In the row labels, "medium" molecular size is considered 10–20 nuclei, "IR" wavelength refers to perturbative treatment of multi-photon processes at infrared wavelengths, MFPAD refers to Molecular Frame Photoelectron Angular Distribution, (Q) and (C) in Nuclear motion refer to quantum and classical dynamics. Combinations marked with a superscript require an additional auxiliary code: (∘) see e.g. refs. 66,67; (*) see e.g. refs. 49,68; (□) see e.g. refs. 69,70; (†) RMT[36]; (×) non-local nuclear dynamics codes[71,72].

of different intensities and wavelengths, to accurately account for electron correlation in the ionization process (in particular, Feshbach resonances), to evaluate observables and spectra directly accessible in experiments, and to perform time-dependent treatments without and with inclusion of nuclear motion (eventually accounting for the coupling with electronic motion). That a particular code has not been widely used in a particular category does not mean that it is not appropriate for this task, but that it might be less obvious to use or computationally more demanding than others. Also, as pointed out in the caption of Table 1, for some specific tasks, some of these codes require the use of accessory codes, which are usually accessible in local servers. It is also important to highlight that among the four codes only tRecX-haCC can provide a description of the double ionization continuum.

Outlook

In this tutorial, we have given a broad overview of the main concepts and obstacles that make the description of photoionization in molecules much more challenging than in atoms, and explained numerical and computational solutions that successfully overcome the difficulties inherent to any molecule, both in the weak and strong field regimes. Some of the available computational codes, namely those employed in the "COST/ZCAM School on New Computational Methods for Attosecond Molecular Processes"[64] have been briefly described. These tools allow for a reasonably accurate description of molecular ionization induced by current extreme ultraviolet and X-ray attosecond and few-femtosecond pulses, as well as much longer pulses, in different scenarios, and cover a wide range of molecular sizes, from diatomics to medium-size and large polyatomics, at different levels of approximation. They have played a crucial role in the development of attosecond science, in particular attochemistry. For small molecules, existing methods are able to take into account the interplay between electronic and nuclear motions during and after the ionization process in a full quantum mechanical way. Ongoing efforts are focusing on the implementation of semiclassical approaches in which nuclear motion is treated classically while preserving a full quantum mechanical description of electronic motion. Such methods are the only ones that are affordable for

large molecular systems and therefore can have a significant impact in problems of chemical interest.

Received: 31 October 2025; Accepted: 24 April 2026;
Published online: 20 May 2026

References

1. Nisoli, M., Decleva, P., Calegari, F., Palacios, A. & Martín, F. Attosecond electron dynamics in molecules. *Chem. Rev.* **117**, 10760–10825 (2017).
2. Calegari, F. & Martin, F. Open questions in attochemistry. *Commun. Chem.* **6**, 184 (2023).
3. Schneider, B. I. & Gharibnejad, H. Numerical methods every atomic and molecular theorist should know. *Nat. Rev. Phys.* **2**, 89–102 (2020).
4. Palacios, A. & Martín, F. The quantum chemistry of attosecond molecular science. *WIREs Comput. Mol. Sci.* **10**, e1430 (2020).
5. Friedrich, H. *Scattering theory* Lecture notes in physics; (Springer-Verlag, Berlin, 2013).
6. Morrison, M. The physics of low-energy electron-molecule collisions: a guide for the perplexed and the uninitiated. *Aust. J. Phys.* **36**, 239 (1983).
7. Wu, G., Hockett, P. & Stolow, A. Time-resolved photoelectron spectroscopy: from wavepackets to observables. *Phys. Chem. Chem. Phys.* **13**, 18447–18467 (2011).
8. Child, M. S. *Theory of Molecular Rydberg States*. Cambridge Molecular Science (Cambridge University Press, 2011).
9. Huo, W. M. & Gianturco, F. A. (eds.) *Computational Methods for Electron Molecule Collisions* (Plenum Press, New York, 1995).
10. Scrinzi, A. t-surff: fully differential two-electron photo-emission spectra. *N. J. Phys.* **14**, 085008 (2012).
11. Herzberg, G. *Molecular Spectra and Molecular Structure. II. Infrared and Raman Spectra of Polyatomic Molecules* (D. van Nostrand Coompany, Inc., 1945).
12. Xie, J. & Zare, R. N. Selection rules for the photoionization of diatomic molecules. *J. Chem. Phys.* **93**, 3033–3038 (1990).
13. Signorell, R. & Merkt, F. General symmetry selection rules for the photoionization of polyatomic molecules. *Mol. Phys.* **92**, 793–804 (1997).
14. Demekhin, P. V., Ehresmann, A. & Sukhorukov, V. L. Single center method: a computational tool for ionization and electronic excitation studies of molecules. *J. Chem. Phys.* **134**, 024113 (2011).
15. Ertel, D. et al. Anisotropy parameters for two-color photoionization phases in randomly oriented molecules: theory and experiment in methane and deuteromethane. *J. Chem. Phys. A* **128**, 1685–1697 (2024).
16. Reiss, H. R. *Foundations of Strong-Field Physics*, 41–84 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011). https://doi.org/10.1007/978-3-540-95944-1_2.
17. Keldysh, L. V. Ionization in the field of a strong electromagnetic wave. *J. Exp. Theor. Phys.* **20**, 1307 (1964).
18. Smirnova, O. & Ivanov, M. Multielectron high harmonic generation: simple man on a complex plane. https://arxiv.org/abs/1304.2413 (2013).
19. Levine, I. N. *Molecular Spectroscopy* (Wiley, New York, 1975).
20. Basis set exchange. See https://www.basissetexchange.org/.
21. Pritchard, B. P., Altarawy, D., Didier, B., Gibson, T. D. & Windus, T. L. New basis set exchange: an open, up-to-date resource for the molecular sciences community. *J. Chem. Inf. Model.* **59**, 4814–4820 (2019).
22. Martín, F. Ionization and dissociation using B-splines: photoionization of the hydrogen molecule. *J. Phys. B* **32**, R197–R231 (1999).
23. Kaufmann, K., Baumeister, W. & Jungen, M. Universal Gaussian basis sets for an optimum representation of rydberg and continuum wavefunctions. *J. Phys. B* **22**, 2223 (1989).
24. Faure, A., Gorfinkiel, J. D., Morgan, L. A. & Tennyson, J. GTOBAS: fitting continuum functions with Gaussian-type orbitals. *Comput. Phys. Commun.* **144**, 224–241 (2002).
25. Coccia, E. et al. Gaussian continuum basis functions for calculating high-harmonic generation spectra. *Int. J. Quantum Chem.* **116**, 1120–1131 (2016).
26. Ammar, A., Ancarani, L. U. & Leclerc, A. A complex Gaussian approach to molecular photoionization. *J. Comput. Chem.* **42**, 2294–2305 (2021).
27. Bachau, H., Cormier, E., Decleva, P., Hansen, J. E. & Martin, F. Applications of B-splines in atomic and molecular physics. *Rep. Prog. Phys.* **64**, 1815–1943 (2001).
28. Mašín, Z., Benda, J., Gorfinkiel, J. D., Harvey, A. G. & Tennyson, J. UKRmol+: a suite for modelling electronic processes in molecules interacting with electrons, positrons and photons using the R-matrix method. *Comput. Phys. Commun.* **249**, 107092 (2020).
29. Marante, C., Argenti, L. & Martín, F. Hybrid Gaussian–B-spline basis for the electronic continuum: photoionization of atomic hydrogen. *Phys. Rev. A* **90**, 012506 (2014).
30. Marante, C. et al. Hybrid-basis close-coupling interface to quantum chemistry packages for the treatment of ionization problems. *J. Chem. Theory Comput.* **13**, 499–514 (2017).
31. Rescigno, T. N. & McCurdy, C. W. Numerical grid methods for quantum-mechanical scattering problems. *Phys. Rev. A* **62**, 032706 (2000).
32. Nestmann, B. M. & Peyerimhoff, S. D. Optimized Gaussian basis sets for representation of continuum wavefunctions. *J. Phys. B* **23**, L773 (1990).
33. Borràs, V. J., Fernández-Milán, P., Argenti, L., González-Vázquez, J. & Martín, F. Photoionization cross sections and photoelectron angular distributions of molecules with XChem-2.0. *Comput. Phys. Commun.* **296**, 109033 (2024).
34. Morassut, C., Coccia, E. & Luppi, E. Quantitative performance analysis and comparison of optimal-continuum Gaussian basis sets for high-harmonic generation spectra. *J. Chem. Phys.* **159**, 124108 (2023).
35. Scrinzi, A. tRecX – an environment for solving time-dependent schrödinger-like problems. *Comput. Phys. Commun.* **270**, 108146 (2022).
36. Brown, A. C. et al. RMT: R-matrix with time-dependence. Solving the semi-relativistic, time-dependent Schrödinger equation for general, multielectron atoms and molecules in intense, ultrashort, arbitrarily polarized laser pulses. *Comput. Phys. Commun.* **250**, 107062 (2020).
37. Woźniak, A. P. et al. A systematic construction of Gaussian basis sets for the description of laser field ionization and high-harmonic generation. *J. Chem. Phys*. **154**, https://doi.org/10.1063/5.0040879 (2021).
38. Durden, A. S. & Schlegel, H. B. Evaluation of diffuse basis sets for simulations of strong field ionization using time-dependent configuration interaction with a complex absorbing potential. *J. Phys. Chem. A* https://www.scopus.com/inward/record.uri?eid=2-s2.0-105001654697&doi=10.1021%2facs.jpca.5c00195&partnerID=40&md5=555d510bbac0c809683c36e70ed8224d (2025).
39. McCurdy, C. W. & Martín, F. Implementation of exterior complex scaling in B-splines to solve atomic and molecular collision problems. *J. Phys. B* **37**, 917 (2004).
40. Vanroose, W., Martín, F., Rescigno, T. N. & McCurdy, C. W. Complete photo-induced breakup of the H_2 molecule as a probe of molecular electron correlation. *Science* **310**, 1787–1789 (2005).
41. Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Courier Corporation, 2012).
42. Helgaker, T., Jörgensen, P. & Olsen, J. *Molecular Electronic Structure Theory* (John Wiley & Sons, LTD, Chichester, 2000).

43. Toffoli, D., Coriani, S., Stener, M. & Decleva, P. Tiresia: a code for molecular electronic continuum states and photoionization. *Comput. Phys. Commun.* **297**, 109038 (2024).
44. Faisal, F. H. M. *Theory of Multiphoton Processes* (Springer Science +Bussiness Media, 1987).
45. Faisal, F. H. M. Introduction to atomic dynamics in intense light fields. In Yamanouchi, K. (ed.) *Lectures on Ultrafast Intense Laser Science*, Vol. 1, 1–40 (Springer, 1987).
46. Majety, V. P., Zielinski, A. & Scrinzi, A. Mixed gauge in strong laser-matter interaction. *J. Phys. B* **48**, 025601 (2014).
47. Morales, F., Bredtmann, T. & Patchkovskii, S. isurf: a family of infinite-time surface flux methods. *J. Phys. B* **49**, 025601 (2016).
48. Majety, V. P. & Scrinzi, A. Dynamic exchange in the strong field ionization of molecules. *Phys. Rev. Lett.* **115**, 103002 (2015).
49. Borràs, V. J., González-Vázquez, J., Argenti, L. & Martín, F. Attosecond photoionization delays in the vicinity of molecular feshbach resonances. *Sci. Adv.* **9**, eade3855 (2023).
50. Born, M. & Huang, K. *Dynamical Theory of Crystal Lattices* (Oxford University Press, 1954).
51. Meyer, H.-D., Manthe, U. & Cederbaum, L. The multi-configurational time-dependent hartree approach. *Chem. Phys. Lett.* **165**, 73–78 (1990).
52. Vendrell, O. & Meyer, H.-D. Multilayer multiconfiguration time-dependent hartree method: implementation and applications to a henon-heiles hamiltonian and to pyrazine. *J. Chem. Phys.* **134**, 044135 (2011).
53. Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **93**, 1061–1071 (1990).
54. Agostini, F. & Curchod, B. F. E. Different flavors of nonadiabatic molecular dynamics. *WIREs Comput. Mol. Sci.* **9**, e1417 (2019).
55. Gianturco, F. A., Lucchese, R. R. & Sanna, N. Calculation of low-energy elastic cross sections for electron-cf4 scattering. *J. Chem. Phys.* **100**, 6464–6471 (1994).
56. Randazzo, J. M. et al. ASTRA: transition-density-matrix approach to molecular ionization. *Phys. Rev. Res.* **5**, 043115 (2023).
57. Lucchese, R. R. & McKoy, V. Application of the Schwinger variational principle to electron-ion scattering in the static-exchange approximation. *Phys. Rev. A* **21**, 112–123 (1980).
58. AMOS gateway. https://sciencegateways.org/resources/amosgateway.
59. Schaftenaar, G. & Noordik, J. H. Molden: a pre- and post-processing program for molecular and electronic structures. *J. Comput. Aided Mol. Des.* **14**, 123–134 (2000).
60. Li Manni, G. et al. The OpenMolcas web: a community-driven approach to advancing computational chemistry. *J. Chem. Theory Comput.* **19**, 6933–6991 (2023).
61. Werner, H.-J., Knowles, P. J., Knizia, G., Manby, F. R. & Schütz, M. MOLPRO: a general-purpose quantum chemistry program package. *WIREs Comput. Mol. Sci.* **2**, 242–253 (2012).
62. Baerends, E. J. et al. The Amsterdam modeling suite. *J. Chem. Phys.* **162**, 162501 (2025).
63. Plasser, F. et al. Columbus: an efficient and general program package for ground and excited state computations including spin–orbit couplings and dynamics. *J. Phys. Chem. A* **129**, 6482–6517 (2025).
64. COST/ZCAM school on new computational methods for attosecond molecular processes. https://www.cecam.org/workshop-details/costzcam-school-on-new-computational-methods-for-attosecond-molecular-processes-1411 (2026).
65. Moore, L. R. et al. The RMT method for many-electron atomic systems in intense short-pulse laser light. *J. Mod. Opt.* **58**, 1132 (2011).
66. Plésiat, E., Lara-Astiaso, M., Decleva, P., Palacios, A. & Martín, F. Real-time imaging of ultrafast charge dynamics in tetrafluoromethane from attosecond pump-probe photoelectron spectroscopy. *Chem. Eur. J.* **24**, 12061–12070 (2018).
67. Barreau, L. et al. Disentangling spectral phases of interfering autoionizing states from attosecond interferometric measurements. *Phys. Rev. Lett.* **122**, 253203 (2019).
68. Pranjal, P., González-Vázquez, J., Bello, R. Y. & Martín, F. Vibrationally resolved photoionization delays in the water molecule. *Phys. Rev. Lett.* **135**, 223202 (2025).
69. Mai, S., Marquetand, P. & González, L. Nonadiabatic dynamics: the sharc approach. *WIREs Comput. Mol. Sci.* **8**, e1370 (2018).
70. Grell, G., González-Vázquez, J., Fernández-Villoria, F., Palacios, A. & Martín, F. Modeling the evolution of laser-induced electronic coherences with trajectory surface hopping. *J. Chem. Theory Comput.* **21**, 10645–10668 (2025).
71. Čížek, M. Photodetachment dynamics using nonlocal dicrete-state-in-continuum model. https://doi.org/10.48550/arXiv.2309.05830 [physics.chem-ph] (2023).
72. Zawadzki, M. et al. Resonances and Dissociative Electron Attachment in HNCO121. *Phys. Rev. Lett.* **121**, 143402 (2018).
73. Reiss, H. R. The tunnelling model of laser-induced ionization and its failure at low frequencies. *J. Phys. B* **47**, 204006 (2014).

Acknowledgements

The authors are grateful to the Center Européen de Calcul Atomique et Moléculaire (CECAM) and the Zaragoza Scientific Center for Advanced Modeling (ZCAM) for their continuous support to the School on New Computational Methods for Attosecond Molecular Processes on which this tutorial is based. F.M., J.B., J.G., Z.M. L.R. and A.S. have been supported by COST (European Cooperation in Science and Technology) through the grants CA18222 (AttoChem) and CA22148 (NEXT). F.M. has been partly funded by the Ministerio de Ciencia, Innovación y Universidades (project no. PID2022-138288NB-C31, MCIN/AEI/10.13039/501100011033/FEDER UE) and the Severo Ochoa Centers of Excellence program (Grant CEX2024-001445-S). L. R. acknowledges funding from "la Caixa" Foundation (ID 100010434), under the agreement "LCF/BQ/PR24/12050018", the Spanish Ministry of Science, Innovation and Universities & the State Research Agency through the project ref. PID2024-163024NA-I00 (MICIU/AEI/10.13039/501100011033/FEDER, UE) and support from the Severo Ochoa Centers of Excellence program through Grant CEX2024-001445-S.

Author contributions

F.M., J.B., J.G., Z.M. L.R. and A.S. contributed to design and wrote the manuscript. F.M. coordinated the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-026-02671-y.

Correspondence and requests for materials should be addressed to Fernando Martín.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.