Chapter 1

First-principles many-electron dynamics with B-spline algebraic diagrammatic construction approach

¹Department of Physics, Imperial College London, Prince Consort Road,
London SW7 2AZ, United Kingdom
v.averbukh@imperial.ac.uk

M. Ruberti¹ and V. Averbukh^{1*}

December 12, 2017

In this chapter we present a first-principles theoretical and numerical method based on the many-electron algebraic diagrammatic construction [ADC(n)] schemes for electronic excitations, able to describe the correlated multi-electron ionisation dynamics induced in atomic and molecular systems by laser pulses both in the perturbative and non-perturbative regime. Within the ADC(n) framework electron correlation is described at different levels of approximation depending on the specific ADC method n used within the ADC hierarchy. An accurate representation of the electronic ionisation continuum is achieved by the implementation and computational optimisation of the first- [ADC(1)] and second-order [ADC(2)] schemes in the monocentric B-spline basis set which makes it possible to describe highly oscillatory discretised continuum wave-functions. The implementation of the time-dependent (TD) version of the B-spline ADC method is made by solving the many-electron TDSE via the Arnoldi Lanczos algorithm. As illustrative examples we present applications of this method to the calculations of both static quantities (photoionisation cross-sections of noble gas atoms) and dynamical quantities such as the high harmonic generation (HHG) spectra of Ar and CO₂ and the attosecond transient absorption spectrum in laser dressed He atoms.

1.1 Theoretical Method: TD B-spline ADC

The ab initio many-body Greens function ADC methods have been first introduced for calculation of excitation and ionization energies of closed shell species [1, 2] and since then generalized for the description of double [3, 4] and triple [5] ionization energies. The ADC(n) schemes of various orders(n) are size consistent and compact relative to the corresponding truncated CI expansions [6, 7]. While initially developed within the many-body Greens function approach, the ADC can be reformulated as wave-function method using the intermediate state representation (ISR) [8, 9]. This has led to applications of the ADC-ISR technique to calculation of properties of and transition moments between excited, [9] singly ionized, [10] and doubly ionized [11] bound states. Recently, it has been shown that the ADC schemes in conjunction with iterative block-Lanczos (BL) diagonalization [12, 13, 14] and the Stieltjes-Chebyshev moment theory [15, 16] can be used also for accurate and efficient characterization of bound-continuum transitions, i.e., for calculations of the decay widths of resonance states [17, 18] and photoionization cross-sections [19, 20, 21]. In the bulk of the ADC work cited above, Gaussian type orbitals (GTOs) have been used as a single-electron

basis as is indeed customary in the ab initio quantum chemistry. However, a very recent benchmark study [20] indicates that it is the use of the GTOs in the excitation ADC schemes that lead to the onset of major inaccuracies in molecular photoionization cross-section calculations at about 70 eV above threshold. This trend has been shown to be general for a series of molecular species and independent of the order of the employed ADC scheme [20]. Moreover, GTO-based ADC-Lanczos-Stieltjes method fails to reproduce not only the narrow features of the cross-section due to excitation resonances (as is fully expected of a moment theory technique), but even the much broader features, such as Cooper minima, e.g., in argon photoionization. Very similar behaviour has been observed in a recent implementation of the GTO/Stieltjes approach within linear response coupled cluster models for electronic excitations [22]. Even very careful GTO selection cannot afford converged high order moments, i.e., high energy features and high resolution, without running into linear dependence problems [22]. The inadequacy of the GTO bases for the characterization of the molecular photoionization calls for an introduction of new basis sets, better suited for the description of the oscillatory continuum wave functions across the interaction region. Several such basis sets have been already introduced in the 1990s within the framework of the many-body methods that do not fully include double electronic excitations and were successfully used for the solution of the time-independent many-electron problems, e.g., for photoionisation cross-section calculations, see, e.g., Ref. [65]. The present work addresses the single-electron basis set issue by construction of ADC schemes explicitly treating not only single [ADC(1)], but also double [ADC(2)] excitations and applying them to the solution of both time-independent and time-dependent many-electron problems [24]. The employed single-electron basis is composed of the spherical harmonics for the angular part and a B-spline expansion for the radial coordinate. B-splines [25] are one-variable piecewise polynomial functions designed to generalize polynomials for the purpose of approximating arbitrary functions on some interval $[0, R_{max}]$. They have the property of approaching completeness as much as desired by refining the corresponding breaking points sequence, i.e., the set of points that divide the full radial interval into sub-intervals [25]. B-spline basis is known to lead to negligible linear dependences even for large bases (i.e., dense breaking points sequences), and are computationally competitive with the local finite-difference methods [26]. Moreover, the use of such a pre-determined quasi-complete set of one-particle orbitals eliminates the need of any apriori procedure in selecting the parameters for the Gaussian-type or Slater-type functions required in the standard L^2 quantum chemistry calculations which often depend upon

non-trivial treatment to minimize the linear dependence embedded in the choice of basis functions. The theory presented in this work is formulated within the fixed-nuclei approximation. This means that the nuclear degrees of freedom are frozen throughout the calculation and the nuclear geometry is assumed not to change in the course of the electronic transition. The neutral equilibrium geometry is assumed for both the initial neutral system and the final ionic system and the electronic states are calculated making use of the Born-Oppenheimer approximation [28]. Therefore, our calculations provide a frozen-nuclei benchmark to which the experimental results can be compared in order to identify the effects of nuclear motion.

1.1.1 Intermediate State Representation Approach to ADC *Ab Initio* Schemes

Within our approach the many-electron correlated states are described by the ADC(n) *ab initio* schemes [2]. Although the ADC(n) schemes for excited states of closed-shell systems were originally derived as approximations to the polarisation propagator [27], based on an algebraic reformulation of its diagrammatic perturbation theory expansion [2], they were later recognised [8] as being interpretable as wave-function methods as well. In fact, ADC establishes a connection between propagator and wave-function methods. The latter interpretation comes from the explicit identification and construction of the intermediate states representation (ISR) that gives rise to the ADC form of the propagator, providing an alternative and direct approach to the hierarchy of the ADC schemes [8, 9]. The starting point of the ISR approach to the hierarchy of the ADC schemes [8, 9] is the construction of the so-called correlated excited states (CES), defined as

$$| \Psi_I^0 \rangle = \hat{C}_I^\dagger | \Psi_0 \rangle , \quad (1.1)$$

where the operators $\hat{C}_I^\dagger$ denote the physical excitation operators corresponding respectively to one hole - one particle (1h1p), 2h2p etc... excitations,

$$\hat{C}_I^\dagger = \left\{ \hat{a}_a^\dagger \hat{a}_i ; \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_k \quad (a < b, j < k) \dots\dots \right\} , \quad (1.2)$$

and $| \Psi_0 \rangle$ is the exact correlated ground state of the system, with energy E_0 . In Eq. (1.1.1) $\hat{a}_i$ and $\hat{a}_a^\dagger$ represent annihilation and creation operators in second quantization formalism [28], that respectively annihilate and create an electron in occupied (denoted by index i) and unoccupied (denoted by index a) single-particle orbitals. These formulas are written adopting the notation in which the subscripts $a, b, c, \dots$ refer to unoccupied virtual orbitals

(with respect to the mean-field single Slater determinant ground state), and $i, j, k, \dots$ to the occupied orbitals [28].

This non-orthogonal CES basis set is complete in the space of the excited states of the N -electron system [29] and has the advantage that ground state correlation is already built into every basis vector. Different orthogonalization procedures for the CES basis set lead to different ISR schemes. The ISR of the ADC schemes is the excitation class orthogonalized (ECO) representation resulting from a Gram-Schmidt orthogonalization procedure for the CES. The CES are orthonormalized in a two-step procedure. First, one performs Gram-Schmidt orthogonalisation of each excitation class with respect to all the lower excitation classes. The states $|\Psi_y^{m\#}\rangle$ formed in this first step are referred to as precursor states. The second step is symmetric (Löwdin) orthonormalisation [28] of the resulting precursor states within each excitation class. As an example the procedure for the first (1h1p) excitation class gives the following precursor states:

$$|\Psi_{ai}^{1\#}\rangle = \hat{a}_a^\dagger \hat{a}_i |\Psi_0\rangle - |\Psi_0\rangle \langle \Psi_0 | \hat{a}_a^\dagger \hat{a}_i | \Psi_0\rangle. \quad (1.3)$$

The second step gives

$$|\tilde{\Psi}_{ai}^1\rangle = \sum_{bj} |\Psi_{bj}^{1\#}\rangle \left(S^{-\frac{1}{2}}\right)_{bj,ai}, \quad (1.4)$$

where $\mathbf{S}$ is the overlap matrix of the first excitation class precursor states, i.e.

$$S_{bj,ai} = \langle \Psi_{bj}^{1\#} | \Psi_{ai}^{1\#} \rangle \quad (1.5)$$

In a compact notation the excitation class orthogonalized (ECO) states can be written as

$$|\tilde{\Psi}_x^m\rangle = \hat{Q}^{m-1} \sum_y |\Psi_y^m\rangle \left(S_{yx}^m\right)^{-\frac{1}{2}}, \quad (1.6)$$

where S_{yx}^m is defined as

$$S_{yx}^m = \langle \Psi_y^m | \hat{Q}^{m-1} | \Psi_x^m \rangle, \quad \hat{Q}^m = \hat{1} - \sum_{l=0}^m \hat{P}^l \quad (1.7)$$

and $\hat{Q}^m$ is the projector operator onto the space orthogonal to the first m excitation classes. Finally, every intermediate state can be expressed as

$$|\tilde{\Psi}_I\rangle = \tilde{C}_I^\dagger |\Psi_0\rangle, \quad (1.8)$$

where all the effects of the consecutive orthonormalisations are encoded in the new creation operators $\tilde{C}_I^\dagger$. The orthonormalisation procedure is bijective in the sense that there is a one-to-one correspondence between the initial states and the final orthonormal states. Hence, since both the projective and the symmetric orthonormalisation steps preserve the association of the final states with the indices of the original overlapping correlated excited states, each of the orthonormal ECO-CES states is largely characterised, as the initial states, by specific hole and particle indices and by a specific excitation class index.

The ADC secular matrix is the representation of the shifted electronic Hamiltonian operator $\hat{H} - E_0$ in the ECO-CES space:

$$\mathcal{H}_{IJ} = \langle \tilde{\Psi}_I | \hat{H} - E_0 | \tilde{\Psi}_J \rangle = \langle \Psi_0 | \tilde{C}_I [\hat{H}, \tilde{C}_J^\dagger] | \Psi_0 \rangle. \quad (1.9)$$

Inserting the expansions of Eq. (1.8) in the respective sub-blocks of the ADC secular matrix, one obtains closed form expressions for $\mathcal{H}_{IJ}$ depending on the exact ground state wave-function $|\Psi_0\rangle$ and energy E_0 . At this point Møller-Plesset (MP) perturbation theory is introduced to describe the ground state correlation, i.e. $|\Psi_0\rangle$ and E_0 :

$$|\Psi'_0\rangle = |\Phi_0^{HF}\rangle + |\Psi_0^{[1]'}\rangle + |\Psi_0^{[2]'}\rangle + |\Psi_0^{[3]'}\rangle + \dots, \quad (1.10)$$

where the first order correction $|\Psi_0^{[1]'}\rangle$ contains only double excitations (2h2p), while $|\Psi_0^{[2]'}\rangle$ contains single, double, triple and quadruple excitations with respect to the Hartree-Fock (HF) ground state $|\Phi_0^{HF}\rangle$.

The vertical excitation energies are obtained by solving the eigenvalue problem $\mathbf{H}\mathbf{V} = \omega\mathbf{V}$, where the matrix $\mathbf{V}$ is the eigenvectors matrix, and the excited eigenstates of the system $\mathbf{V}_n$ are therefore given, in the basis of the intermediate states, as:

$$|\Psi_n\rangle = \sum_I V_{I,n} |\tilde{\Psi}_I\rangle. \quad (1.11)$$

Using this explicit expression for the excited states of the system, one obtains the corresponding transition moments as

$$\langle \Psi_m | \hat{D} | \Psi_0 \rangle = \mathbf{V}_m^\dagger \cdot \mathbf{F} = \sum_{rs} d_{rs} \mathbf{V}_m^\dagger \cdot \mathbf{f}_{rs}, \quad (1.12)$$

where d_{rs} are the matrix elements of the dipole operator on the one particle orbitals chosen as basis set functions. The matrix of effective transition amplitudes $\mathbf{f}$ and the associated vector $\mathbf{F}$ are defined by:

$$f_{I,rs} = \langle \tilde{\Psi}_I | \hat{a}_r^\dagger \hat{a}_s | \Psi_0 \rangle, \quad F_I = \langle \tilde{\Psi}_I | \hat{D} | \Psi_0 \rangle. \quad (1.13)$$

The hierarchy of ADC(n) approximations is obtained for each order n by truncating the intermediate state manifold at some limiting excitation class and, also, by truncating the resulting perturbation expansion for the matrix elements of the ADC secular matrix between the included classes in the way consistent with the polarisation propagator approach to ADC [2]. In other words, the perturbation expansion of each excitation block of the Hamiltonian is truncated, at different orders depending on the excitation classes involved, in such a way to reproduce the results of the polarisation propagator based derivation of the ADC schemes. For example, at the ADC(2) level, the Hamiltonian matrix of the system can be represented as

$$\mathbf{H}^{\text{ADC}[2]} = \mathbf{H}^{[0]} + \mathbf{H}^{1,1[1]} + \mathbf{H}^{1,1[2]} + \mathbf{H}^{1,2[1]} + \mathbf{H}^{2,1[1]} \quad (1.14)$$

where $\mathbf{H}^{i,j[n]}$ denotes the ipih-jpjh excitations block of the Hamiltonian computed at the order n of perturbation theory.

Thus, in ADC(2) the perturbation expansion of the secular matrix elements extends through second, first and zero-th order in the 1h1p block, the 1h1p-2h2p coupling block and the diagonal 2h2p block, respectively. In a similar way the 1h1p and 1h2p parts of the effective transition amplitudes are given by perturbation expansions through second and first order respectively. An extension of the ADC scheme, not strictly consistent with the polarisation propagator and referred to as ADC(2) extended [ADC(2)x], is obtained by using the first order expansion for the 2h2p block that accounts for the couplings between the 2h2p intermediate states. The dipole matrix between the excited states of the system is constructed in the same way and can be represented as

$$\mathbf{D}^{\text{ADC}[2]} = \mathbf{D}^{[0]} + \mathbf{D}^{1,1[1]} + \mathbf{D}^{1,1[2]} + \mathbf{D}^{1,2[1]} + \mathbf{D}^{2,1[1]}. \quad (1.15)$$

In the ADC(1) scheme the Hamiltonian reduces to the configuration interaction singles (CIS) one, while the transition moment with respect to the ground state are expressed as

$$\langle \tilde{\Psi}_i^a | \hat{D} | \Psi_0 \rangle = d_{ai} + \sum_v \sum_o \frac{\langle va || oi \rangle}{\epsilon_o + \epsilon_i - \epsilon_v - \epsilon_a} d_{ov} \quad (1.16)$$

where $\langle va || oi \rangle$ is the antisymmetrized two-particle Coulomb integral in physicists' notation and the two indices v and o run over the virtual and the occupied canonical (Hartree-Fock) orbitals respectively.

One of the peculiar features of ADC is that the ADC matrices have the property that ground state correlation is built into each matrix element via

perturbation theory, and the extent to which this is done is described by the ADC order n . In the ADC ground state correlation is estimated in a size consistent manner, via perturbation theory. Excitations on top of the ground state feel a dressed interaction, due to the local ground state correlation, but the ground and excited states have the exact same correlation character away from this region. Furthermore, much of the intrinsic error of ground state MP perturbation theory is cancelled out since only the energy differences between the ground and the excited states of the system are directly computed. While correlation is treated in a balanced way between ground and excited states, excitations of different classes are handled differently. The scheme discussed above, i.e. ADC(2), is most accurate for excitations of mainly p-h character (the qualitative definition of such a state can be made independent of methodology by the use of reduced density matrices [30])). These should be the most experimentally relevant excitations of a system, although excitations of 2h2p character are an important manifestation of electron correlation, and this effective truncation is closely related to the efficiency of the scheme.

One of the best features of ADC(n) methods is that, while at second order the number of degrees of freedom is the same as in either the configuration interaction with single and double excitations (CISD) [28] and the equation of motion coupled cluster with single and double excitations (EOM-CC-SD) methods [31], the equation to be solved for excited states energies is a Hermitian eigenvalue problem and, at the same time, the method is size-consistent, therefore obtaining two of the primary advantages of the CISD and EOM-CCSD methods respectively. The ADC equations are indeed separable in the sense that local and nonlocal excitations are strictly decoupled, and the ADC results for both energies and for transition moments are size-consistent (size-intensivity property) [6, 7]. Moreover the ADC(n) schemes are compact relative to the corresponding truncated CI expansions [6, 7], i.e. the dimension of explicit configurations space required in the ADC(n) secular problem, in order for the energies of the resulting principal (1h1p character) eigenstates to satisfy a certain level of perturbation theoretical accuracy, is systematically smaller than the dimension needed in CI expansions of comparable accuracy.

1.1.2 B-spline Basis Set

Monocentric B-spline basis [25], which is composed of a spherical harmonics for the angular part and B-splines for the radial coordinate, is used to

represent single particle orbitals which are therefore expressed as

$$\psi_{ilm}(r, \theta, \phi) = \frac{1}{r} B_i(r) Y_{lm}(\theta, \phi). \quad (1.17)$$

B-splines [25] are one-variable *piecewise polynomial functions* (pp-functions), characterised by their polynomial order k (maximum degree $k - 1$), designed to generalise polynomials for the purpose of approximating arbitrary functions on some finite interval $[0, R_{max}]$. Since they are defined in a restricted space, usually referred to as a box, one has to define a certain points sequence with which discretising the radial coordinate, depending on the target system. The radial interval $[0, R_{max}]$ is therefore divided into l adjacent subintervals $I_j = [\xi_j, \xi_{j+1}]$ by a sequence of $l + 1$ points ξ_j in strict ascending order $0 = \xi_1 < \xi_2 < \dots < \xi_{l+1} = R_{max}$; the ξ_j are usually called breakpoints [25].

The reason for employing B-splines as radial basis functions is that, with respect to our purpose, the description of the continuum states requires basis sets which could reproduce the oscillating behaviour of the electronic wave-function up to big distances from the molecular region. Traditional $\mathcal{L}^2$ basis functions, such as Slater Type Orbitals (STOs) or Gaussian Type Orbitals (GTOs), are not adequate to represent the rapidly oscillating continuum states of an ionised electron far away from the molecular region, since numerical linear dependencies rapidly occur as the basis set increases. B-splines, conversely, are very flexible functions and are able to provide a very accurate representation of both bound and continuum states without running into numerical dependencies even for large bases. The use of such a pre-determined set of one-particle orbitals eliminates the need of any a priori procedure in selecting the parameters for the Gaussian-type or Slater-type functions required in the standard $\mathcal{L}^2$ quantum chemistry calculations which often depend upon non-trivial treatment to minimise the linear dependence embedded in the choice of basis functions. This feature makes them superior to the other more conventional $\mathcal{L}^2$ basis sets and computationally competitive with the “local” finite-difference methods [26]. B-splines produce smooth curves with continuous derivatives at every point.

Care has to be taken with the choice of the breakpoints sequence defining the B-spline basis set. For simple equidistant breakpoints each $B_i(r)$ is just a translation by one interval of the previous one. If the breakpoints are not equidistant the shape of $B(r)$ changes smoothly with the change of the breakpoint. In the short-range region near the nuclei, where the wave-function oscillates quickly and correlation effects are stronger, a more densely populated set of B splines is needed in order to accommodate the fast

rising inner s orbitals at small radius r . On the other hand, a more evenly populated set of B-splines at larger r is required if the oscillating behaviour of the positive-energy orbitals at large distance has to be properly represented. Unlike bound states, continuum states are ever-oscillating functions and thus are non-zero up to infinite r . Therefore, for large distances from the origin, there are no reasons for not choosing a linear breakpoint sequence which divides the radial coordinates into equally spaced grid points. The B-spline breakpoint sequence we use is the parabolic-linear sequence [24], which is a mix of these two types of sequences: a short-range dense one which concentrates points near the nucleus with quadratically increasing breakpoint spacing, which is better adapted to bound states, and a linear breakpoint spacing sequence for larger values of r necessary for continuum states description. The connection between these two regions is made at an intermediate radial distance r_0 , imposing continuity of the breaking point function ξ_i and of its first derivative. The explicit parabolic-linear breakpoint sequence formula is:

$$\xi_i = \alpha(i - i_0)^2 \text{ for } 1 \leq i < i_0, \xi_i = h(i - i_0) \text{ for } i_0 \leq i \leq n \quad (1.18)$$

The intermediate radius is defined by the index i_0 in the following way $r_0 = \xi_{i_0} = \frac{r_{max}(i_0-1)}{2n-i_0-1}$ and the quadratic and linear coefficients by $\alpha = \frac{r_0}{(i_0-1)^2}$ and $h = \frac{r_{max}}{(n-i_0)}$ respectively.

The range of kinetic energies accurately described by the B-spline basis set can be tuned changing the breakpoint spacing in the linear region. The long range breakpoint spacing is related mainly to the minimum wavelength we want to describe for the photoelectron, considered as a free particle, which cannot be much smaller than the grid step otherwise the B-spline basis hardly describes the oscillating behaviour of the wave-function.

The maximum energy E_{max} of the discretised continuum states is set by both the B-spline order k and the breakpoint spacing h . At large distance, a continuum wave-function oscillates with a wavelength given by $\frac{2\pi}{\sqrt{2E}}$ and the number of B-splines contained in one wavelength has to be large enough to fully reproduce the two sign switches. For B-splines of order $k=10$, such as the ones used in the present work, it is sufficient to satisfy the following condition: $h \approx \frac{\lambda_{min}}{\pi}$, where λ_{min} represents the minimum value of the continuum wave-function wavelength to be described.

Finally it must be mentioned that one of the most useful advantages of making use of B-spline functions is that they have the property of approaching completeness as much as desired by refining the corresponding breakpoint sequence [25]. For example, the density of discretised continuum

states represented by the basis can be increased arbitrarily by simply increasing the extension of the grid, i.e. R_{max} , while maintaining at the same time fixed the spatial density of breakpoints.

1.1.3 Computational Procedure

While this choice of the single-particle basis functions leads to higher computational effort than in the GTO case, it at the same time simplifies tremendously the numerical algorithm mainly because of the absence of linear dependencies to take care of; moreover, the minimal localised support of the individual primitive B-spline basis functions allows to reduce the number of radial B-spline one-electron and two-electron integrals to be calculated and dynamically stored in the RAM at the beginning of the calculation. Computer time for the evaluation of the radial two-electron integrals over the primitive B-spline functions is a minor part of the total computational cost of the present calculations.

The main numerical procedure required in a B-spline-ADC calculation consist of four key steps, namely:

1. Solution of the B-spline HF equations.
2. One- and two-electron integral transformations with simultaneous calculation of the first order Hamiltonian and dipole matrix-elements on the fly.
3. Calculation of the second-order matrix elements in the ADC(2) method.
4. Full/iterative diagonalisation of the ADC Hamiltonian and calculation of the static physical quantity of interest (e.g. a cross-section) via the SI technique [15, 16] or time-propagation of an initial state in order to obtain the dynamic quantity of interest (e.g. HHG spectrum).

The first step in the B-spline ADC calculations consists of solving a set of one-particle integro-differential equations with non-local terms, i.e., the Hartree-Fock equations,

$$\hat{h}^{HF}\psi_n = \epsilon_n\psi_n, \quad (1.19)$$

where $\hat{h}^{HF}$ is the spin-free ground-state Fock operator of the neutral system [28].

In the molecular case Eq. (1.19) is solved for every irreducible representation of the molecular point symmetry group, and the expansion of the wave-function ψ_n contains all the spherical harmonics angular functions $Y_{lm}(\theta, \phi)$

belonging to the particular irreducible representation. For example, in the case of linear molecules with $D_{\infty h}$ as symmetry point group, such as N_2 or CO_2 , the axial angular momentum m is an exact quantum number for the system. This means that the number of spherical harmonics belonging to the irreducible representations, and therefore entering the expansion of the corresponding symmetric orbitals, is equal to $L_{max} + 1$ for σ_g and to L_{max} for all the other irreducible representations (σ_u, π_g, π_u , etc.), where L_{max} is the maximum angular momentum used in the calculation. In the atomic case, the HF equations can be projected upon the l-m spherical harmonic subspace and simplify depending upon the radial coordinate only.

Solving Eq. (1.19) self-consistently, we obtain a set of discretised one-particle functions $\psi_{\lambda,n}$ corresponding to electronic orbitals with defined symmetry λ (in the atomic case λ stands for the pair of orbital angular momentum quantum numbers l, m) and variable energy (both negative and positive) defined by the HF effective one-particle Hamiltonian. Therefore, both the set of occupied HF orbitals and the full orthogonal complement of virtual orbitals, expressed in terms of B-splines basis functions, are obtained. These canonical orbitals are then used to construct the ADC many-body Hamiltonian. Therefore the *ab initio* calculation is not performed working directly with primitive B-splines and as a consequence the one- and two-electron integral transformation from the original B-spline basis to the HF basis set must be computed.

The second step makes the integrals transformation from the primitive B-spline basis set to the basis of the Hartree-Fock (HF) canonical orbitals. Two-electron integral transformation from primitive orbital sets to atomic/molecular working orbital sets is the most time consuming process for large-scale B-spline ADC calculations.

The two-electron integrals over the primitive monocentric basis set functions of Eq. (1.17), indexed with the integers 1,2,3,4, are defined as

$$(12 | 34) = \int \int \Phi_1(\mathbf{r})^* \Phi_2(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \Phi_3(\mathbf{r}')^* \Phi_4(\mathbf{r}') d^3\mathbf{r} d^3\mathbf{r}', \quad (1.20)$$

where

$$\Phi_\alpha(\mathbf{r}) = \frac{1}{r} B_{i_\alpha}(r) Y_{l_\alpha, m_\alpha}(\theta, \phi), \quad (\alpha = 1, \dots, 4). \quad (1.21)$$

Two-electron integrals over the primitive monocentric basis set functions can be factorised, using Eq. (1.21), into products of radial and angular

contributions and consequently expressed as

$$(12 | 34) = \sum_{k=k_{min},2}^{k_{max}} ANG(k, l_1, m_1, l_2, m_2, l_3, m_3, l_4, m_4) R^k(i_1 i_2, i_3 i_4), \quad (1.22)$$

Here

$$R^k(i_1 i_2, i_3 i_4) = \int B_{i_1}(r_1) B_{i_2}(r_1) \frac{r_{\leq}^k}{r_{>}^{k+1}} B_{i_3}(r_2) B_{i_4}(r_2) dr_1 dr_2, \quad (1.23)$$

is the radial two-electron integral over the primitive B-spline functions, evaluated by direct two-dimensional numerical integration as reported in [32], and the two bounds of the summation over the k index are given by

$$k_{min} = \max(|l_1 - l_2|, |l_3 - l_4|), \quad k_{max} = \min(l_1 + l_2, l_3 + l_4). \quad (1.24)$$

The function $ANG(k, l_1, m_1, l_2, m_2, l_3, m_3, l_4, m_4)$, which depends only on the angular momentum quantum numbers of the orbitals and is the result of the two-dimensional integration onto the angular variables, can be expressed as

$$ANG(k, l_1, m_1, l_2, m_2, l_3, m_3, l_4, m_4) = \sum_{q=-k}^{+k} YYY(l_2, k, l_1, m_2, q, m_1) YYY(l_3, k, l_4, m_3, q, m_4) \frac{4\pi}{2k+1}, \quad (1.25)$$

where the quantity YYY stands for the integral of the product of three spherical harmonics

$$YYY(l_1, l_2, l_3, m_1, m_2, m_3) = \int_0^{2\pi} \int_0^\pi Y_{l_1, m_1}^*(\theta, \phi) Y_{l_2, m_2}(\theta, \phi) Y_{l_3, m_3}(\theta, \phi) \sin \theta d\theta d\phi. \quad (1.26)$$

Using the relation between the canonical HF orbitals and the primitive monocentric basis functions, which reads explicitly as

$$\psi_n = \sum_j \Phi_j C_j^n = \sum_{l,m} \sum_i C_{l,m,i}^n Y_{l,m}(\theta, \phi) \frac{1}{r} B_i(r), \quad (1.27)$$

the two-electron integrals over the primitive monocentric basis set are transformed to the HF canonical basis set as

$$(pq | rs) = \sum_{i_1=1}^N \sum_{i_2=1}^N \sum_{i_3=1}^N \sum_{i_4=1}^N C_{i_1}^p C_{i_2}^q (12 | 34) C_{i_3}^r C_{i_4}^s, \quad (1.28)$$

where N is the total number of HF orbitals.

However, a straightforward application of Eq. (1.28), simply looping over all p, q, r, s , and then looping over all the primitive indices 1,2,3,4, gives rise to a total number of numerical operations which scales as N^8 , where N is the number of basis states. Indeed, this elementary approach would give eight nested loops which, with a basis size required for a realistic calculation, is beyond the power of present computers. However, a careful analysis shows that the size of the problem can be reduced to N^5 . Therefore a different sequential algorithm, which allows the $O(N^8)$ process to be converted into four $O(N^5)$ consecutive steps, has been implemented in major *ab initio* quantum chemistry MO codes [28].

The trick is to convert one index at a time from the primitive orbitals to the final orbitals basis. That is, starting with the matrix (tensor) $(12 | 34)$, convert one of the indices (a "quarter transform"), then convert the next one, until a total of four quarter-transforms are accomplished. Therefore the total 4-indices transformation can be thought of in the following way

$$(pq | rs) = \left(\sum_{i_1=1}^N C_{i_1}^p \left(\sum_{i_2=1}^N C_{i_2}^q \left(\sum_{i_3=1}^N C_{i_3}^r \left(\sum_{i_4=1}^N C_{i_4}^s (12 | 34) \right) \right) \right) \right). \quad (1.29)$$

This is called factorizing the expression, and it helps reducing the cost immensely. Therefore the transformation is performed with this common four-step algorithm, giving a scaling of $4N^5$ with respect to the number of orbitals to be transformed. In the B-spline basis set case an N^5 scaling is still large. Further reductions are possible depending on the type of basis set, as discussed in the following for monocentric B-splines, and on the level of ADC theory used.

Indeed, the first transformation step reads:

$$(12 | 3s) = \sum_{(l_1, m_1)} \sum_{i_1} \sum_{(l_2, m_2)} \sum_{i_2} \sum_{(l_3, m_3)} \sum_{i_3} \sum_{d=1}^N \sum_{(l_4, m_4)} \sum_{i_4} C_{l_4, m_4, i_4}^s \langle 12 | 34 \rangle \quad (1.30)$$

and, with the monocentric B-splines as primitive basis set, scales therefore as $O(N_{ang}^4 N_{B-splines}^2 (2j_{ord} - 1)^2 2l_{max} N)$. Here the factor $2l_{max}$ comes from

the k loop in Eq. (1.22), l_{max} denotes the maximum angular momentum used in the monocentric expansion and j_{ord} stands for the order of the B-spline basis set used. $N_{B-splines}$ and N_{ang} denote the total number of radial B-splines and the number of spherical harmonics angular functions belonging to the irreducible representations the transformed orbitals belong to respectively. Similar procedures are repeated three more times until all the four indices have been transformed.

Because of the number of basis functions involved, both the memory requirements and the time of the transformation can quickly become unaffordable with the increase of the number of spherical harmonics and/or of the radial B-spline functions used in the monocentric expansion (Eq.(17)). We approach this problem by calculating only the two-electron integral types required by the ADC Hamiltonian and dipole matrix elements: in the ADC(1) case only the integrals involving two virtual canonical orbitals indices are needed, namely, integrals of the type (vv—oo) and (vo—vo). In the ADC(2) calculations two other types of integrals are required, namely, the (vv—vv), (vv—vo), and the less demanding (vo—oo) types [2].

Furthermore, use has been made of the Hamiltonian symmetry group to divide every type of integral calculation with respect to the quadruplet of orbital irreducible representations involved. While in the atomic and linear molecules calculations the full symmetry of the system is exploited, i.e. the full rotational symmetry and the full $D_{\infty h}$ or $C_{\infty v}$ point group symmetry respectively, in the general, non-linear, molecular case the highest Abelian symmetry subgroup of the molecule has been used. Following this strategy it is possible to end up with an increased number of less demanding single calculations to be performed.

In the cases where one of the final orbital indices is restricted to the occupied HF orbitals, that is (vv—oo), (vo—vo), (vo—oo) and (vv—vo), this index is the first one to be transformed. Doing so allows the scaling of the first transformation to be considerably improved, as the loop over the final transformed index reduces to a smaller loop over the occupied orbitals in a certain symmetry space of the system, and to minimise the memory allocation requirements after the first step. Moreover, the scaling can be further improved noticing that the occupied orbitals are localised in space within the molecular region; therefore, the loop over the radial B-spline functions with the same radial argument as the transformed orbitals (i.e. i_3) can be truncated at an index value for which their overlap with the occupied orbital becomes smaller than a certain threshold value. Looping over higher indices would be useless as the resulting two-electron integrals would have a negligible value; the same argument allows us to reduce the

Eq. (1.29) step number	(vo—vo)	(vv—oo)
1	$O\left(N_{ang}^4 \left(N_{Bspl}^{R_{occ}}\right)^2 (2j_{ord}-1)^2 2l_{max} N_{occ}\right)$	$O\left(N_{ang}^4 N_{Bspl} N_{Bspl}^{R_{occ}} (2j_{ord}-1)^2 2l_{max} N_{occ}\right)$
2	$O\left(N_{ang}^3 \left(N_{Bspl}^{R_{occ}}\right)^2 (2j_{ord}-1) N_{occ} N_{virt}\right)$	$O\left(N_{ang}^3 N_{Bspl} N_{Bspl}^{R_{occ}} (2j_{ord}-1) N_{occ}^2\right)$
3	$O\left(N_{ang}^2 N_{Bspl}^{R_{occ}} (2j_{ord}-1) N_{occ}^2 N_{virt}\right)$	$O\left(N_{ang}^2 N_{Bspl} (2j_{ord}-1) N_{occ}^2 N_{virt}\right)$
4	$O\left(N_{ang} N_{Bspl}^{R_{occ}} N_{occ}^2 N_{virt}^2\right)$	$O\left(N_{ang} N_{Bspl} N_{occ}^2 N_{virt}^2\right)$

Table 1.1: Scaling of the number of numerical operations required at any step of the two-electron integral transformation for the cases of the (vo—vo) and (vv—oo) types of integrals. Here N_{virt} stands for the number of virtual HF orbitals. When the parallel algorithm described in the text and schematically represented in Fig. 1.1 is used, the scalings for each transformation step reduce by a factor approximately equal to the number of cpu's used.

memory allocation requirements for the intermediate two-electron integral vectors as a considerable number of integrals is not required anymore in the subsequent steps.

The scaling becomes then $O\left(N_{ang}^4 N_{Bspl} N_{Bspl}^{R_{occ}} (2j_{ord}-1)^2 2l_{max} N_{occ}\right)$, where N_{occ} stands for the number of occupied HF orbitals and $N_{Bspl}^{R_{occ}}$ denotes the number of B-splines that overlap with the occupied orbitals. If extended radial grids are used, this can give a considerable gain. In Table 1.1 the scalings of the number of numerical operations required at any step of the two-electron integral transformation for the cases of the (vo—vo) and (vv—oo) types of integrals are reported.

However, even though these procedures considerably reduce the computational effort required by the transformation, the number of basis functions used in B-spline ADC is well above the standards of a typical *ab initio* calculation and as a result, the numerical algorithm has to be further highly

Figure 1.1: Schematic representation of the way the two-electron integral transformation is parallelised. The indices I,J,K,L refer to the primitive basis functions of Eq. (1.21); the indices P,Q,R,S refer to the HF canonical orbitals. "Reproduced with permission from ["B-spline algebraic diagrammatic construction: Application to photoionization cross-sections and high-order harmonic generation", M. Ruberti, V. Averbukh and P. Decleva, J. Chem. Phys. 141, 164126 (2014). <http://dx.doi.org/10.1063/1.4900444>]. Copyright [2014], AIP Publishing LLC.

optimised. Consequently, it becomes necessary to take advantage of the parallel computing environment [33, 34], which has become increasingly easy to manage, in order to minimise the computer time for the transformation. Each specific two-electron integrals transformation Eq. (1.29) has been massively parallelized, both inter-nodes, with the standard MPI (message passing interface) technology in parallel computing environment, and intra-node with the OpenMP (shared memory) technology. A schematic representation of the way every specific quadruple of two-electron integrals undergoes the 4-indices transformation in parallel is given in Fig. 1.1. In details, since the first transformation step, the two-electron integral vector memory allocation is divided between the computational MPI nodes available. This division is performed onto the first index of the four-indices two-electron integral tensor V_{IJKL} . Each of these pieces of the two-electron integral vector undergoes the four indices transformations from V_{IJKL} to V_{PQRS} locally on every node in parallel, and only after the last index transformation is performed the resulting fully transformed partial vectors from every node are sent to a specific node where they are summed together. The advantage of this algorithm is that it turns out to never be necessary, at any intermediate stage of the transformation, to have the full integral vector dynamically allocated on a single node. Moreover, even if the full final transformed integrals vector can be bigger than the full partially transformed intermediate ones, its complete allocation onto a single node is not required because it can be computed pieces by pieces sequentially. The contributions to the first order matrix elements of the ADC Hamiltonian and of the ADC transition moments are calculated on the fly at the last step of the integral transformation. The integrals involved in the second-order ADC matrix elements are instead stored on disk. The OpenMP technology is used to parallelize every intermediate transformation performed on every node, reducing therefore considerably the time of the computation.

With the use of the parallel algorithm described in the text and schematically represented in Fig. 1.1, the scalings for each transformation step of Table 1.1, must be divided by a factor approximately equal to the number of cpu's used. The parallel transformation algorithm presented here makes it possible to carry out large-scale calculations within the B-spline based post HF ADC(n) theories on parallel computers. However it must be mentioned that in general, although the MPI technology allows to substantially reduce the memory requirement per single-node, substantial global amount of RAM (spread across a high number of nodes) is still needed for B-spline ADC calculations requiring a big dimension of the one-electron basis set expansion and that, for such basis sets, assembling the two-electron integrals is still a

time-consuming task.

The third step, in the case of the ADC(2) level of theory, evaluates the remaining terms of the ADC Hamiltonian matrix and transition moments using the relevant two-electron integrals stored on disk. The matrix we deal with in ADC(2) are of a size often up to 1,000,000 or greater. In ADC(1), the linear scaling with respect to the number of basis functions is however much more favourable. The value of the individual matrix element, which at the first order ADC(1) level represents a two-variable integral over a two-body Coulomb interaction is in general small but non-zero and, as a result, the matrix is non-sparse. At the second order level, the 1h1p-2h2p coupling block is quite sparse, around 10/15%, while the 2h2p block is diagonal; this simplifies enormously the computation and actually makes such types of calculation possible. The calculation is performed by dividing the matrix into a large number of sub-matrices of a size storable on a single node. Every sub-matrix is calculated in parallel, with every node computing a partial contribution to it. The same process is repeated for every sub-matrices, the number of times this is done depending on the memory storage capabilities available and on the matrix dimension. Therefore for memory reasons, the different sub-matrices are calculated sequentially one after the other, even if they are independent one of each other. Nevertheless the extension to the parallel computing environment speeds up substantially this step as well, because every node just calculates its own specific contribution to the given submatrix and moreover this calculation is OpenMP parallelized as well.

The fourth step involves the extraction of the relevant physical information from the ADC Hamiltonian. In the particular case of a total cross-section calculation, performed using the SI technique, we are interested in the pseudospectrum of the Hamiltonian and in the transition moments of the pseudo-eigenstates. A newly-implemented, parallelised, routine based on the block-Lanczos diagonalisation algorithm [12, 14] allows us to obtain a limited number of energy eigenvalues and their corresponding eigenvectors by diagonalizing the ADC(2) sparse matrices generated in the third step. A direct diagonalization is instead performed in the ADC(1) calculations presented here. This is possible because with the B-spline basis sets used the ADC(1) matrices have typically small enough dimension (around a few thousands). The range of the B-spline and spherical harmonics expansions over which such a full diagonalization procedure is still applicable, is quite large for the ADC(1) Hamiltonian, because of the linear scaling of its dimension with respect to the number of one-electron basis functions. Finally, given the pseudospectrum, it is possible to calculate the transition amplitudes and ultimately the cross section. This latter calculation can be

performed on any single-node CPU. In the case of the dynamical calculation, the time-dependent Schrödinger equation (TDSE) in the basis of the ADC intermediate states is solved via the Arnoldi-Lanczos algorithm. More detail on the time-dependent method are given in the following subsection.

1.1.4 Time Propagation

In order to apply the correlated ADC schemes to the solution of time-dependent many-electron problems involving ionisation, we have implemented a time-dependent version of the *ab initio* B-spline ADC method. The time-dependent problem is solved within TD-ADC making the following ansatz for the time-dependent electronic wave-function:

$$|\Psi(t)\rangle = C_0(t) |\Psi_0(t)\rangle + \sum_{n \neq 0} C_n(t) |\Psi_n(t)\rangle \quad (1.31)$$

where the coefficients $C_0(t)$ and $C_n(t)$ refer to the ground-state and to the ECO-CES ADC configuration basis states respectively.

Propagation of the B-spline ADC many-electron wave-function allows one to achieve a realistic *ab initio* description of a series of many-electron phenomena in the time-resolved fashion on the atto- and femto-second scale. A complex absorbing potential (CAP) [35] has been used in order to eliminate wave-packet reflection effects from the radial grid boundaries. The form of the CAP used is the following:

$$\hat{W} = \eta(r - r_{CAP})^2, \quad r \geq r_{CAP} \quad (1.32)$$

and with the addition of the CAP term the form of the total time-dependent Hamiltonian of the system reads

$$\hat{H} = \hat{H}_0 + \hat{z}E(t) - i\hat{W} \quad (1.33)$$

where $\hat{H}_0$ is the field-free Hamiltonian and $\hat{z}E(t)$ is the laser-atom interaction in length form and within the dipole approximation [36]. With the addition of a complex-absorbing potential, the Hamiltonian becomes complex symmetric. The time propagation of the unknown coefficients C_0, C_n of the B-spline ADC many-electron wave-function is performed by means of the general complex Lanczos, or Arnoldi-Lanczos, algorithm [37, 38, 39]. The Lanczos algorithm has been proved to be very efficient as a time propagation scheme for the non-relativistic Schrödinger equation [40] and is widely used in treating nuclear-dynamics problems [41, 42, 43]. The more general

Arnoldi-Lanczos algorithm is suited to describe time-propagation with non-Hermitian Hamiltonians as well as the usual Hermitian ones and it can be used efficiently for complex symmetric Hamiltonians.

It is known that the TDSE has a formal solution

$$| \Psi(t) \rangle = \hat{U}(t, 0) | \Psi(0) \rangle = T \exp \left(-\frac{i}{\hbar} \int_0^t \hat{H}(t') dt' \right) | \Psi(0) \rangle. \quad (1.34)$$

where $T \exp$ is the so-called chronological exponent which represents the evolution operator $\hat{U}(t, 0)$.

For a sufficiently small time increment Δt , over which the variation of the Hamiltonian is negligible, the evolution operator describing the evolution of the system from time t to time $t + \Delta t$ can be written as

$$\hat{U}(\Delta t) = \hat{U}(t + \Delta t, t) = \exp \left(-\frac{i}{\hbar} \hat{H} \Delta t \right). \quad (1.35)$$

The Arnoldi-Lanczos algorithm leads to the following polynomial expansion of the exponential in the evolution operator:

$$\hat{U}(\Delta t) \approx \sum_{j=1}^K c_j P_j \left(-\frac{i}{\hbar} \hat{H} \Delta t \right) \quad (1.36)$$

where $P_j \left(-\frac{i}{\hbar} \hat{H} \Delta t \right)$ is a polynomial of degree j , and c_j are the expansion coefficients.

This polynomial expansion is obtained by projecting the Hamiltonian of the system at time t onto the Krylov subspace constructed starting from the normalised state vector of the system at the same time t , i.e. $\mathbf{V}_0 = \psi(t)$, and spanned by the vectors $\mathbf{V}_j = \mathbf{H}^j \psi(t)$ $j = 0, 1, 2, \dots, K-1$. Here $\mathbf{H}$ and $\psi(t)$ are respectively the matrix representations of $\hat{H}$ and $| \Psi(t) \rangle$ in the ECO-CES ADC basis.

The recursive procedure for building the reduced Hamiltonian matrix $\mathbf{H}_K$ within the K -dimensional Krylov subspace reads as

$$\tilde{\mathbf{V}}_{j+1}^{(1)} = \mathbf{H} \mathbf{V}_j$$

for $i = 1, \dots, j$

$$\beta_{ij} = \mathbf{V}_i^\dagger \tilde{\mathbf{V}}_{j+1}^{(i)} \quad \tilde{\mathbf{V}}_{j+1}^{(i+1)} = \tilde{\mathbf{V}}_{j+1}^{(i)} - \beta_{ij} \mathbf{V}_i$$

$$\mathbf{V}_{j+1} = \frac{\tilde{\mathbf{V}}_{j+1}^{(j+1)}}{\beta_{j+1,j}} \quad \beta_{j+1,j} = || \tilde{\mathbf{V}}_{j+1}^{(j+1)} ||$$

(1.37)

In this way, starting with a normalised state $\mathbf{V}_0 = \psi(0)$ one constructs an orthonormal basis $\mathbf{V}_0, \dots, \mathbf{V}_{K-1}$ for the K-dimensional Krylov subspace, which represents a very good approximation to the Hilbert subspace to which the state of the system at time $t + \Delta t$ belongs and in which the reduced Hamiltonian $\mathbf{H}^K$ is represented as a complex upper Hessenberg matrix, i.e. a matrix which has zero entries only below the first subdiagonal,

$$H_{ij}^K = \mathbf{V}_i^\dagger \mathbf{H} \mathbf{V}_j =$$

$$= \begin{pmatrix} \beta_{0,0} & \beta_{0,1} & \cdots & \cdots & \beta_{0,K-1} \\ \beta_{1,0} & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \beta_{K-2,K-1} \\ 0 & 0 & 0 & \beta_{K-1,K-2} & \beta_{K-1,K-1} \end{pmatrix}$$

(1.38)

In order to assure numerical orthogonality of the Krylov space basis set, at each step, each state vector is orthogonalised twice with respect to the previous ones. This re-orthogonalisation procedure is also known as the Kahan-Parlett "twice is enough" algorithm [12] and provides a very good level of accuracy for the orthogonality of the resulting Krylov basis states.

The propagation operator $\hat{U}(\Delta t)$ can be approximated by its representation in the Krylov subspace:

$$\mathbf{U}_K(\Delta t) = \exp\left(-\frac{i}{\hbar} \mathbf{H}_K \Delta t\right). \quad (1.39)$$

The propagation of the initial wave-function is hence performed in the projected subspace by the upper Hessenberg Hamiltonian $\mathbf{H}_K$, which is much smaller than the full Hamiltonian and therefore much easier to diagonalise.

After the diagonalisation the propagator can be expressed in the following form:

$$\mathbf{U}_K(\Delta t) = \mathbf{Z}^\dagger \exp\left(-\frac{i}{\hbar} \mathbf{D}_K \Delta t\right) \mathbf{Z}, \quad (1.40)$$

where $\mathbf{Z}$ is the eigenvector matrix of $\mathbf{H}_K$, and $\mathbf{D}_K$ is its diagonal matrix of eigenvalues. The propagated vector is constructed from the eigenvalues and eigenvectors of the matrix $\mathbf{H}_K$ and reads:

$$\Psi(t) = \exp\left(-\frac{i}{\hbar} \mathbf{H}_K \Delta t\right) \Psi(0) = \sum_{j=1}^{K-1} a_j \mathbf{V}_j, \quad (1.41)$$

where its coefficients onto the Krylov space basis vectors read

$$a_j = \sum_{n=0}^{K-1} Z_{jn} \exp\left(-\frac{i}{\hbar} \lambda_n \Delta t\right) (Z^{-1})_{n0}. \quad (1.42)$$

In Eq. (1.42) the λ_n denote the eigenvalues of the matrix H_K . The propagated wave-function is then used as the initial vector for the next iteration, Eq. (1.37).

In the computational implementation of this scheme used in TD B-spline ADC, it is possible to fix the maximum number of iterations requested, i.e. the maximum allowed dimension of the Krylov space. Moreover, the algorithm is stopped when the norm of the $(j+1)$ -th Krylov vector candidate becomes smaller than a certain fixed threshold value, which, in this work, has been taken as the numerical precision of the calculator. The error, i.e. the difference between the propagated and exact wave-functions, when using the Lanczos integrator is sensitive to both the time-step size Δt or the maximum order K allowed.

Using this algorithm, the time propagation reduces to repetitive diagonalisation, at each time step, of Hessenberg matrices whose dimension is significantly smaller with respect to the full ADC Hamiltonian. This allows the treatment of dynamical problems in which the size of the resulting system Hamiltonian is relatively large and time-propagation by full diagonalisation of the secular matrix at each time step becomes practically inaccessible. Another very important feature of the Arnoldi-Lanczos algorithm is that, when computing the tridiagonal matrix $\mathbf{H}_K$, the full Hamiltonian matrix $\mathbf{H}$ is needed only for a simple matrix vector product and remains unchanged during the whole procedure. Each of these reduced single calculations has been massively parallelized, both inter-nodes, with standard MPI technology in parallel computing environment [34], and intra-node with OpenMP (shared memory) technology [33].

The time propagation can be performed representing the operators matrices both in the ECO-CES configuration basis set or in the Hamiltonian eigenvectors basis set. In the second scheme, after the Hamiltonian has been diagonalised, the dipole and CAP matrices are transformed to their representations in the basis of the Hamiltonian eigenvectors. The main advantages of this approach are that the dense Hamiltonian matrices do not have to be stored anymore, reducing the dynamical memory requirements of the calculation. However the CAP matrices will generally become dense and therefore will need to be fully stored. The second advantage of this scheme is that the basis states immediately have a clearer physical interpretation. Indeed, it becomes easier to calculate the rate of ionisation caused by the laser pulse, just by looking at the population of the bound eigenstates of the system at the end of the pulse. Therefore, it is not necessary to wait for the whole outgoing ionised wavepacket to be absorbed by the CAP and to look at the resulting loss of norm of the wave-function, as is the case for the first scheme. However, the full diagonalisation can be performed with reasonable expense only for relatively small basis sets.

In order to guarantee the convergence of the Arnoldi-Lanczos algorithm, which depends crucially on the spectral radius of the Hamiltonian, in the first propagation scheme we removed from the numerical simulation the ECO-CES configurations with a zeroth-order energy higher than a certain threshold value. In an analogous way, in the second propagation scheme the unphysical high energy eigenstates have been removed from the propagation. The threshold value for the cutoff energy which has been used in the calculations performed in this work is 40 a.u.; with this choice the Arnoldi-Lanczos algorithm converges well at each time step and, moreover, the removal of these high energy electron orbitals has no effect on the physical observables we are interested in.

1.2 Results and Discussion

1.2.1 Atomic Photoionisation Cross-Sections by Combination of B-spline-ADC and Stieltjes Methods

In this section we present the application of the B-spline versions of ADC schemes explicitly treating not only single [ADC(1)], but also double [ADC(2)] excitations to the *ab initio* calculation of static (photoionisation cross-sections) quantities. In order to calculate the total photoionisation cross-sections of noble gas atoms and to perform a direct comparison of the performance of the B-spline basis against the Gaussian type one, the Stieltjes imaging

Figure 1.2: Total photoionisation cross-section of the Ar atom. Crosses – experimental result of Ref. [44], pluses – experimental result of Ref. [45], circles – B-spline-ADC(1) result; squares – B-spline-ADC(2)-Lanczos-Stieltjes cross-section converged after 39600 Lanczos algorithm iterations. ”Reproduced with permission from [”B-spline algebraic diagrammatic construction: Application to photoionization cross-sections and high-order harmonic generation”, M. Ruberti, V. Averbukh and P. Decleva, J. Chem. Phys. 141, 164126 (2014). <http://dx.doi.org/10.1063/1.4900444>]. Copyright [2014], AIP Publishing LLC.

(SI) [15, 16, 19, 20, 21] technique has been used. This technique allows one to obtain the correct continuum oscillator-strength density, without going through the direct calculation of the correctly energy-normalised wave-functions.

The argon total photoionisation cross-section is of particular interest for testing the accuracy of the method due to the presence of the Cooper minimum that is known to present a challenge for the GTO calculations. Fig. 1.2 shows the calculated total photoionisation cross section of Ar in the region of the Cooper minimum [20], together with the experimental result of Samson *et al.* [44] and Chang *et al.* [45]. It is worth noting that the Cooper minimum is due to a change of sign of the radial dipole matrix elements, which pass through zero at a certain energy [46], and therefore it is already treatable within one-electron models such as LDA [47], although its

precise energy position may be sensitive to correlation effects. Moreover the possibility of describing properly the Cooper minimum feature is strongly related to the ability of the basis set to properly represent the oscillating continuum single-electron orbitals.

The calculation is performed using a parabolic-linear B-spline basis set with radial grid radius $R_{max} = 120$ and 200 radial grid points. The spherical-harmonics expansion extends up to $L_{max} = 3$. In the ADC(2) Ar calculations, the core $1s$, $2s$ and $2p$ orbitals were frozen in both the singly and doubly excited intermediate configurations. Therefore, only the valence holes in the $1h1p$ and $2h2p$ state manifold are allowed. Both ADC(1) and ADC(2) models correctly reproduce the shape of the minimum, with the ADC(2) providing a better quantitative agreement with the experiment in terms of the Cooper minimum position. In fact, ADC(2) gives an almost perfect prediction of the position of the Cooper minimum at 49 eV, while the ADC(1) predicted value is at about 53 eV. The position of the following maximum at 79 eV is overestimated by almost 4 eV from the ADC(1) calculation, while it is underestimated by the nearly the same amount from the ADC(2) one.

The superior accuracy of the newly-implemented basis set is manifested also by its ability to correctly reproduce the high-energy tails of the cross-sections that have been shown to present a notorious difficulty for the GTO-based schemes [20].

In Fig. 1.3 we show the relative deviations from experimental results of the three *ab initio* methods as function of the photon energy for a series of molecular photo-ionisation cross-section calculations performed using a GTO basis set [20]. One observes that below 60 eV both ADC(2) and ADC(2)x methods lead to impressive agreement with experiment with the relative deviations below 10%, while at higher photon energies inaccuracy of all the ADC schemes grows reaching 20% level around 80 eV. Since this behaviour does not depend on the level of *ab initio* theory, it is concluded that it has to do with the limitations of the Gaussian single-electron basis sets.

In Fig. 1.4 the same analysis is shown for the case of a series of atomic photo-ionisation cross-section calculations performed using B-spline basis sets instead [24]. As it is possible to observe, the precision deterioration at higher energies typical of the GTO calculations [20] is not present any more. The high-energy tails of the cross-sections are indeed reproduced by the B-spline ADC in excellent agreement with the experiment, the average error stabilizing around 5% for both ADC(1) and ADC(2) methods.

Figure 1.3: Relative deviations of the ADC-Stieltjes photoionisation cross-sections from the experimental results averaged on eight medium-sized molecules (CO_2 , H_2O , CH_4 , NH_3 , HF , C_2H_2 , C_2H_4 , CH_2O) as a function of the energy in the energy range of ionisation threshold to 100 eV. Dashed-dotted line – ADC(1)-Stieltjes result, dashed line – ADC(2)-Lanczos-Stieltjes cross-section, full line – ADC(2)x-Lanczos-Stieltjes cross-section. "Reproduced with permission from ["Total molecular photoionization cross-sections by algebraic diagrammatic construction-Stieltjes-Lanczos method: Benchmark calculations", M. Ruberti, R. Yun, K. Gokhberg, S. Kopelke, L. S. Cederbaum, F. Tarantelli and V. Averbukh, J. Chem. Phys. 139, 144107 (2013). <http://dx.doi.org/10.1063/1.4824431>]. Copyright [2013], AIP Publishing LLC.

Figure 1.4: Relative deviations of the B-spline-ADC-Stieltjes photoionisation cross-sections from the experimental results averaged on He, Ne, Ar and Kr, as a function of the energy in the energy range of He ionisation threshold to 260 eV. Green line – ADC(1)-Stieltjes result, red line – ADC(2)-Lanczos-Stieltjes result. "Reproduced with permission from ["B-spline algebraic diagrammatic construction: Application to photoionization cross-sections and high-order harmonic generation", M. Ruberti, V. Averbukh and P. Decleva, J. Chem. Phys. 141, 164126 (2014). <http://dx.doi.org/10.1063/1.4900444>]. Copyright [2014], AIP Publishing LLC.

1.2.2 ADC(1) High Harmonic Generation Spectra of Ar Atom and Aligned CO₂ Molecules

In this section a first application of the B-spline TD-ADC technique, by calculating the high-order harmonic generation spectrum (HHG) spectra of the Ar atom and CO₂ molecule, is presented.

High-order harmonic generation (HHG) spectroscopy has become a powerful tool to resolve ultrafast electron dynamics in atoms and molecules [48, 49] and also to investigate the geometric structure of molecules on an ultrafast time scale [50, 51, 52, 53, 54, 55, 56]. HHG occurs when an electron is removed from the parent molecule under the action of an intense infra-red (IR) laser field and is forced to recollide with the parent ion less than one period of the field oscillations later.

In order to compute the dipole acceleration which is required for describing high-harmonic generation, we first calculate the expectation value of the electric dipole moment $z(t)$ [57] and then we calculate its second time derivative. The HHG spectrum then reads

$$S_{\text{HHG}}(\omega) = \frac{1}{20} \frac{1}{3\pi c^3} \left| \int_{-\infty}^{\infty} \left[\frac{d^2}{dt^2} \langle z \rangle(t) \right] e^{-i\omega t} dt \right|^2. \quad (1.43)$$

In the high-intensity and low-frequency limit, this quantity takes into account the contributions to the HHG spectrum from both the short and long trajectories, including the interference between the two [58]. Differently, in a molecular gas macroscopic propagation of harmonic radiation associated with short and long trajectories is different in terms of both the phase-matching (which allows the harmonic field emitted from different molecules in the gas to sum coherently) and the divergence. In typical experiments on high harmonic generation, the geometry of the experimental setup suppresses the so-called 'long' trajectories and favours the short ones.

The expectation value of the electric dipole moment $z(t)$, as well as the one of any single-particle operator $\hat{O}$, onto the time-dependent wave-function of the system is calculated using the expansion of Eq. (1.31) and reads

$$\langle \hat{O} \rangle(t) = \sum_{n=0}^N \sum_{m=0}^N C_n(t) C_m^*(t) O_{m,n}. \quad (1.44)$$

The coefficients C_n are known from the time propagation calculation. The matrices $O_{m,n}$ in the ECO-CES representation, as well as the transition moments from the ground state of the system to the ECO-CES states themselves, are evaluated consistently at the same level of the ADC(n) hierarchy

as the one used to time propagate the state of the system. In the following calculations we have used the first-order method of the ADC-hierarchy, namely ADC(1).

In the case of Ar, we again concentrate on the effect of the Cooper minimum, since besides its basic importance for photoionisation, it also has recently drawn much attention in connection to the HHG by rare gases [59, 60, 61, 62]. Calculation of the HHG spectrum of the Ar atom interacting with an intense and short infrared (IR) laser pulse allows for the illustration of the effect of the Cooper minimum on the HHG spectrum by a fully *ab initio* single-atom simulation.

The results presented here were obtained using two laser pulses with a carrier frequency $\omega = 0.057$ a.u. (800 nm), a full width at half maximum (FWHM) pulse duration of $\tau = 413$ a.u. (10 fs) and with a peak field strength of $E_{max} = 0.075$ a.u. and $E_{max} = 0.095$ a.u. respectively. These two values correspond to peak intensities of $I = 1.5 \times 10^{14}$ W/cm² and $I = 3.2 \times 10^{14}$ W/cm² respectively. The time-step used in the calculation is equal to 0.2 a.u. (1/500 of the IR period) at which value the convergence of our final results is obtained.

In the Ar HHG spectrum calculation, the size of the computational box, R_{max} , is dictated by the semiclassical picture of the process [48], i.e. it must be large enough to contain the longest recolliding electronic trajectories. Therefore the CAP starts at a radius $r_{CAP} = 100a.u.$ which is larger than the classical quiver amplitude of the electron for the higher intensity pulse used here ($r_{HHG} = \frac{E_{max}}{\omega^2} = 30$ a.u.). The calculation is performed using a radial grid radius $R_{max} = 120$ and 200 radial grid points. The CAP strength used was $\eta = 0.0005$ and the maximum angular momentum employed was $l_{max} = 70$.

The HHG spectral intensity profile is shown in Fig. 1.5, for both the two different IR intensities used. The smaller intensity used gives a cutoff energy for the HHG emission at about 50 eV. The photon energy range of 30-50 eV corresponds to a recollision electron energy range of 15-35 eV and therefore no Cooper minimum is observed in the corresponding HHG spectrum. Conversely, the higher intensity provides higher kinetic energy for the returning electrons, and therefore makes it possible to see the Cooper minimum shape in the HHG spectrum.

Our simulation shows that the Cooper minimum in the HHG spectra can be reproduced by the B-spline TDADC method and its position indeed lies slightly above 50 eV as found in the recent experiments [60, 61, 62]. The results are also in good agreement with those of Ref. [63], obtained using the time-dependent CIS technique. The position of the HHG Cooper

Figure 1.5: ADC(1) high harmonics generation emission spectrum of the the Ar atom interacting with an IR field. The red and black curves refer to the $1.5 \times 10^{14} \text{ W/cm}^2$ and $3.2 \times 10^{14} \text{ W/cm}^2$ values for the IR peak intensity respectively. The time duration of the IR laser pulse used is about 21 fs and the wavelength is 800 nm. "Reproduced with permission from ["B-spline algebraic diagrammatic construction: Application to photoionization cross-sections and high-order harmonic generation", M. Ruberti, V. Averbukh and P. Decleva, J. Chem. Phys. 141, 164126 (2014). <http://dx.doi.org/10.1063/1.4900444>]. Copyright [2014], AIP Publishing LLC.

minimum is very much consistent with the photoionisation cross-section one. This has to do with the final step of the HHG process [48], where the laser-driven continuum electron recombines into the ground state emitting a single photon in a process that is directly reversed relative to the photoionisation.

In the following we present a fully 3D *ab initio* ADC(1) time-dependent calculation for the electron dynamic of the CO₂ molecule interacting with an high-intensity ultra-short infrared (IR) laser pulse, calculating the resulting HHG spectra and quantitatively investigating the effect of its multi-channel nature on the dynamical minimum [64].

Despite the fact that strong-field ionisation is exponentially sensitive to the ionisation potential I_p , in molecular ions electronic excitations often lie within a couple of electron volts from the ground state leading to sizeable population amplitudes even for exponential scaling of ionisation rates. The importance of different orbitals in high harmonic generation has been pointed out in various experiments [65, 66]. Moreover, clear evidence for strong-field ionisation from lower orbitals has demonstrated the importance of including multiple orbitals in the description of HHG [56, 65, 67].

Different states of the molecular ion between ionisation and recombination provide different channels for the HHG process. For each channel, specific electronic structure is encoded onto the amplitude and the phase of the harmonic emission signal. These channels are different pathways that take the system from the same initial to the same final state via different states of the ion. Their interference records the relative phases between the channels by mapping them onto the amplitude modulation in the harmonic spectra. The phases between the channels, accumulated between ionisation and recombination, encode the underlying electronic dynamics with attosecond time resolution. The coherent sum of the multiple channels can thus reveal both structural minima due to the recombination matrix elements or dynamical minima that occur from destructive interferences between channels.

If the minimum is structural as in argon, its position is largely independent of the laser parameters [60, 61, 62, 68, 69, 70]. On the contrary, if the minimum is dynamical, its position can be controlled directly by the intensity and the wavelength of the IR driving laser field, its mobility reflecting attosecond multielectron dynamics within the ion, including electron rearrangement upon ionisation. This is indeed the case of the CO₂ molecule where minima have been found at vastly different positions in different experiments [52, 71, 72, 73, 74].

The calculations have been performed using a linear B-spline sequence with a radial box radius $R_{max} = 120$ a.u. and 180 radial grid points. In

order to smoothly remove the long trajectories contribution from the time-dependent calculated dipole $D(t)$, we use a CAP with starting radius $r_{CAP} = 30$ a.u. and strength $\eta = 0.0005$. The laser pulses used in the calculation have a carrier frequency of $\omega = 0.057$ a.u. (800 nm) and a full width at half maximum (FWHM) pulse duration of $\tau = 2900$ a.u. (70 fs). The time-step used in the calculation is equal to 0.2 a.u. (1/500 of the IR period) at which value the convergence of our final results has been obtained. Moreover, since the laser field used is linearly polarised along the molecular axis, it is enough to consider only the two multi-electron spaces corresponding to the irreducible representations Σ_g^+ and Σ_u^+ .

The list of active HF occupied orbitals of the CO₂ molecule used in the following calculations are, in energetically increasing order, $4\sigma_g$, $3\sigma_u$, $1\pi_u$ and $1\pi_g$. Therefore, in this work, we have included a total of four ionisation channels that contribute to harmonic emission of CO₂ molecules; they respectively correspond to creating the CO₂ ion in the ground state $^1\Pi_g$ (channel X), in the first excited state $^1\Pi_u$ (channel A), in the second excited state $^1\Sigma_u^+$ (channel B) and in the third excited state $^1\Sigma_g^+$ (channel C) respectively. Within ADC(1) they respectively consist of, in the molecular orbital picture, the removal of a bound electron from the *HOMO*, *HOMO* − 1, *HOMO* − 2 and *HOMO* − 3 HF orbitals.

The ADC(1) HHG spectra of CO₂ molecule, calculated for three different laser peak-intensities ($I_1 = 0.7 \times 10^{14}$ W/cm², $I_2 = 1.1 \times 10^{14}$ W/cm², $I_3 = 1.75 \times 10^{14}$ W/cm²), are shown in Fig. 1.6. As can be seen, the *ab initio* calculations performed here present a clear minimum in the high harmonic spectrum of CO₂, which shifts to higher orders with increasing intensity, from H_{19} at $I_1 = 0.7 \times 10^{14}$ W/cm² to H_{27} at $I_3 = 1.75 \times 10^{14}$ W/cm². Varying the intensity of the fundamental field allows one to vary the relative phase of the high-harmonic emission channels, thereby controlling the photon energy at which the destructive interference occurs. Therefore, the main features of the CO₂ molecule HHG spectra, i.e. the position of the dynamical minimum and its field intensity dependence, are found to be reproduced by the B-spline ADC(1) calculations in very good agreement with the various experiments which have been performed during the past few years [52, 71, 72, 73, 74]. Indeed, experiment [52] confirms that the harmonic minimum shifts approximately linearly with intensity.

A quantitative study of the role of the different ionisation channels in the high-harmonic emission and of the effect of their interaction (inter-channel couplings) on the dynamical minimum can be found in [64].

Figure 1.6: ADC(1) high harmonic generation emission spectrum of the CO₂ molecule interacting with an IR field. The pulse duration is 70 fs and the central wavelength is 800 nm. The laser field is linearly polarised along the molecular axis. Top panel – The peak intensity of the IR laser pulse used is $0.7 \times 10^{14} \text{ W/cm}^2$; central panel – The peak intensity of the IR laser pulse used is $1.1 \times 10^{14} \text{ W/cm}^2$; bottom panel – The peak intensity of the IR laser pulse used is $1.75 \times 10^{14} \text{ W/cm}^2$.

1.2.3 Attosecond Transient Absorption Spectroscopy Calculation in He Atom

In the following we present a calculation about the vectorial response of the transient absorption of singly-excited helium dressed by an 800 nm IR strong laser field with polarisation oriented either parallel or perpendicular to the polarisation of the probing attosecond pulse train (APT), and intensity approaching the strong-field ionisation threshold of the ground state (1.2×10^{14} W/cm²) [75]. In the experiment [75] the XUV APT was produced using HHG in a krypton effusive gas jet and its spectrum contained harmonics of the 800 nm field from order 11 to order 27.

The frequency dependent absorption is calculated from the expectation value of the electric dipole moment of the atom $z(t)$ and the incident XUV field $E_X(t)$ as

$$S(\omega; \tau_d) = -\text{Im} \left[\tilde{E}_X^*(\omega) \langle \tilde{z}(\omega; \tau_d) \rangle \right], \quad (1.45)$$

where tilde denotes the Fourier transform from time to XUV frequency ω , star denotes complex conjugation, and τ_d is the IR-XUV delay. The generalised cross-section, $\sigma(\omega; \tau_d)$ was calculated using the following equation

$$\sigma(\omega; \tau_d) = \frac{4\pi\alpha\omega S(\omega; \tau_d)}{|\tilde{E}_X|^2}, \quad (1.46)$$

where α is the fine structure constant.

Varying the time delay between the IR dressing field and the APT in the transient absorption measurement results in modulations to the recorded absorption amplitude for the harmonic orders around the respective ionisation thresholds. Experimental results are presented as well as the ones obtained by means of a TD-ADC(1) calculation. Delay-dependent absorption with half and quarter-cycle periodicity with a strong dependence on the relative polarisations of the dressing and probing fields was observed in both experiment and in our calculations [75].

In Fig. 1.7. we show the measured and calculated delay-dependent absorption for parallel and perpendicularly polarised dressing and probe fields, for both above-threshold harmonics (17 and 19) and below-threshold harmonics (13 and 15). As it is possible to see, several aspects of the experimental results are reproduced by the single-atom TD-ADC(1) calculations. We have studied the dependence of field induced absorption modulation on the relative polarisation of dressing and probe fields. The experimental measurement shows a strong modulation for the case of parallel polarisation, but the modulation is greatly suppressed in the case of perpendicular

Figure 1.7: Delay-dependent absorption for parallel (left column) and perpendicularly (right column) polarised dressing and probe fields, for (a)(d) above-threshold harmonics 17 (red) and 19 (cyan), and (e)(h) below-threshold harmonics 13 (blue) and 15 (green). Representative error bars are shown. The dashed lines indicate the experimental field-free (FF) absorption [30], and the theoretical field-free absorption from the calculation. Dashed lines are absent from the below threshold harmonics since the field-free absorption for harmonics 13 and 15 should be zero.

polarisations for the above threshold harmonics. As it is possible to see, this is also reproduced by the single-atom TD-ADC(1) calculations. In the experiment, little difference in the modulation amplitude was observed between the parallel and perpendicular polarisations for the below threshold harmonics. This behaviour is reproduced by the theory for harmonic 13, but not so well captured for the 15th harmonic. We understand this to be due to small offsets in the calculated field-free bound energy states in helium arising from the use of a truncated ADC hierarchy, i.e. ADC(1) [75].

1.3 Conclusions

In this chapter we presented an efficient first-principles theoretical and numerical method based on the many-electron algebraic diagrammatic construction [ADC(n)] schemes, in order to describe the correlated multi-electron ionisation dynamics induced in atomic and molecular systems by laser pulses both in the perturbative and non-perturbative regime. Within the ADC(n) framework it is possible to describe the electron correlation at different levels of approximation depending on the specific ADC method n used within the ADC hierarchy. A monocentric B-spline basis set expansion is used to represent single-electron orbitals. The main advantage is that the B-splines basis functions give an accurate representation of the true electronic continuum, making it possible to describe highly oscillating discretised continuum wave-functions and therefore overcoming the limitations inherent to the use of Gaussians. The implementation of the time-dependent version of the B-spline method has been based on the solution of the TDSE via the Arnoldi Lanczos algorithm.

In view of the large (compared to GTO) size of the B-spline bases, optimised implementation of the B-spline method, both the time-independent and -dependent versions, within the MPI and OpenMP protocols has been realised, making the future calculations of molecular systems entirely possible.

By performing a series of atomic photoionisation cross-section calculations, we have demonstrated the advantages of the B-spline ADC showing that the cross-section features that pose a challenge for the Gaussian basis calculations, such as the argon Cooper minimum shape and position, are found to be reproduced by the B-spline ADC in excellent agreement with the experiment [24]. The superior accuracy of the newly implemented basis set is also manifested by the ability to correctly reproduce the high-energy tails of the cross-sections that we have shown to present a notorious difficulty for the GTO-based schemes [20].

By calculating the HHG spectrum of the Ar atom we presented the first strong-field time-dependent application of the new *ab initio* B-spline ADC technique [24], showing that the effect of the Cooper minimum on the high-order harmonic generation spectrum of Ar is correctly predicted by the time-dependent ADC calculation in the B-spline basis.

We also presented a fully *ab initio* molecular calculation, focusing on the HHG spectrum of CO_2 molecule and showing that the position of the dynamical minimum and its field intensity dependence, are found to be reproduced by the B-spline ADC(1) calculations in a very good agreement

with the various experiments which have been performed during the past few years [52, 71, 72, 73, 74].

Finally we presented an application of the method to the problem of attosecond transient absorption in laser dressed helium, showing how many features of the experimental results can be reproduced by the single-atom TD-ADC(1) calculations [75].

Future applications of the TD B-spline ADC method involve full simulation of attosecond molecular pump-probe experiments. We plan to apply our method to describe sequential molecular double ionization by the pump and probe laser pulses with controlled delay from first principles with full account of the continuum dynamics of the photoelectrons. Our calculations allows us to give a quantitative prediction about the creation of a coherent superposition of molecular ionic states in the photo-ionization process, by accurately extracting the coherent ionic wave function produced after the first ionization event, and simulate the probe of the ensuing attosecond dynamics by a second ionising pulse within a single first-principles many-electron framework. Full characterization of the initial ionic state will allows us to simulate the electron-hole dynamics of molecular systems, such as glycine, where we expect the results to uncover effects, such as the molecular-orbital (MO) picture breakdown [76], which would never appear in a simplified theoretical model. The strong electron correlation in an ion with energy near its double ionization potential causes the excited state configurations to overwhelm the ground state in the wave-function of the ion. The electrons no longer have well-defined orbitals, and the MO picture of the ionisation breaks down. The time dependence of the wave-function becomes significant due to the abundance of the excited state configurations. Theoretical studies of such molecules with TD B-spline ADC may also help us understand attosecond hole migration in large molecules such as peptides, short chains of aminoacids, as suggested in [77].

The monocentric B-spline basis set expansion used in this work allows us to accurately treat compact symmetric molecules up to the size of Benzene. However, in order to be able to apply the TD B-spline ADC method to molecules as large as glycine, we plan to extend the actual implementation based on a One Centre Expansion (OCE), where the functions are centred on a single origin, to a multicentre basis set approach. The multicentre approach consists of adding off-centre functions, located at non-equivalent nuclei, to the B-spline expansion and permits to improves dramatically the convergence of the calculation for most molecules [78]. Two different types of approach are possible: a mixed B-spline/GTO, where the added off-centre radial functions are GTO's or a full multicentre B-spline basis approach,

where all the radial functions located at the different centres are represented as B-splines. Traditionally, a multicentre basis set is constituted by GTO or STO functions that yield fast convergence for the lowest bound states with a limited number of basis functions. Increasing the basis represents a difficult task because numerical linear dependencies rapidly come up. This is due to the large overlap between functions on different centres. One of the advantages of using a complete spherical B-spline functions approach lies on the fact that their local nature permits to control the overlap between functions and avoid numerical linear dependence problems [25]. In this approach the radial and angular parts of the basis set are expanded over several suitable centres to correctly describe bound and continuum states. These centres are:

- The centre of mass of the molecule, associated with a large spherical box of radius R_{max}^O to correctly describe the long range behaviour of the continuum wave-functions.
- The position occupied by each nucleus i , associated with smaller spherical boxes of radius R_{max}^i ; this boxes are generally quite small (1 a.u.) in order to avoid significant overlap with the expansions performed on neighbouring centres.

The addition of these off-centre basis functions will solve the convergence problem that the spherical harmonics monocentric expansion presents when applied to the description of orbitals localised in distant off-centre atomic nuclei. Our goal is to efficiently implement the calculation of the two-electron integrals in this multicentre B-spline basis set, and consequently apply the TD B-spline ADC correlated method to large molecules, such as glycine and amino-acids.

Acknowledgements

The authors acknowledge the EU support through the Marie Curie ITN CORINF. V.A. acknowledges the financial support of the Engineering and Physical Sciences Research Council (EPSRC, UK) through the Career Acceleration Fellowship (Award No. EP/H003657/1) and the Programme Grant on Attosecond Dynamics (Award No. EP/I032517). M.R. acknowledges the financial support by the EPSRC/DSTL MURI grant EP/N018680/1.

Bibliography

- [1] L. S. Cederbaum, in The Encyclopaedia of Computational Chemistry, edited by P. v. R. Schleyer, N. L. Allinger, T. Clask, J. Gasteiger, P. A. Kollmann, H. F. Schaefer III, and P. R. Schreiner (Wiley, Chichester, 1998).
- [2] J. Schirmer, Phys. Rev. A, 1982, **26**, 2395.
- [3] J. Schirmer and A. Barth, Z. Phys. A 317, 267 (1984).
- [4] A. Tarantelli and L.S. Cederbaum, Phys.Rev.A 39, 1656 (1989).
- [5] A. Tarantelli and L. S. Cederbaum, Phys.Rev.A 46, 81 (1992).
- [6] F. Mertins, J. Schirmer and A. Tarantelli, Phys. Rev. A, 1996, **53**, 2153.
- [7] F. Mertins and J. Schirmer, Phys. Rev. A, 1996, **53**, 2140.
- [8] J. Schirmer, Phys. Rev. A, 1991, **43**, 4647.
- [9] J. Schirmer and A. B. Trofimov, J. Chem. Phys., 2004, **120**, 11449.
- [10] A. B. Trofimov and J. Schirmer, J. Chem. Phys., 2005, **123**, 144115.
- [11] Y. Velkov, T. Miteva, N. Sisourat, and J. Schirmer, J. Chem. Phys., 2011, **135**, 154113.
- [12] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, NJ, 1980.
- [13] H.-D. Meyer and S. Pal, J. Chem. Phys. **91**, 6195 (1989).
- [14] H.-G. Weikert, H.-D. Meyer, L. S. Cederbaum, and F. Tarantelli, J. Chem. Phys. **104**, 7122 (1996).
- [15] P. Langhoff, Chem. Phys. Lett., 1973, **22**, 60.

- [16] P. W. Langhoff and C. T. Corcoran, *J. Chem. Phys.*, 1974, **61**, 146.
- [17] S. Kopelke, K. Gokhberg, L. S. Cederbaum, F. Tarantelli, and V. Averbukh, *J. Chem. Phys.*, 2011, **134**, 024106.
- [18] S. Kopelke, K. Gokhberg, V. Averbukh, F. Tarantelli, and L. S. Cederbaum, *J. Chem. Phys.*, 2011, **134**, 094107.
- [19] K. Gokhberg, V. Vysotskiy, L. S. Cederbaum, L. Storchi, F. Tarantelli, and V. Averbukh, *J. Chem. Phys.*, 2009, **130**, 064104.
- [20] M. Ruberti, R. Yun, K. Gokhberg, S. Kopelke, L. S. Cederbaum, F. Tarantelli and V. Averbukh, *J. Chem. Phys.*, 2013, **139**, 144107. "Reproduced with permission from ["Total molecular photoionization cross-sections by algebraic diagrammatic construction-Stieltjes-Lanczos method: Benchmark calculations", M. Ruberti, R. Yun, K. Gokhberg, S. Kopelke, L. S. Cederbaum, F. Tarantelli and V. Averbukh, *J. Chem. Phys.* 139, 144107 (2013). <http://dx.doi.org/10.1063/1.4824431>]. Copyright [2013], AIP Publishing LLC.
- [21] M. Ruberti, R. Yun, K. Gokhberg, S. Kopelke, L. S. Cederbaum, F. Tarantelli and V. Averbukh, *J. Chem. Phys.*, 2014, **140**, 184107. "Reproduced with permission from ["Total photoionization cross-sections of excited electronic states by the algebraic diagrammatic construction-Stieltjes-Lanczos method", M. Ruberti, R. Yun, K. Gokhberg, S. Kopelke, L. S. Cederbaum, F. Tarantelli and V. Averbukh, *J. Chem. Phys.* 140, 184107 (2014). <http://dx.doi.org/10.1063/1.4874269>]. Copyright [2014], AIP Publishing LLC.
- [22] J. W. Cukras, P. Decleva, and S. Coriani, *J. Chem. Phys.*, 2013, **139**, 094103.
- [23] I. Cacelli, V. Carravetta, A. Rizzo, and R. Moccia, *Phys. Rep.*, 1991, **205**, 283.
- [24] M. Ruberti, V. Averbukh and P. Decleva, *J. Chem. Phys.*, 2014, **141**, 164126. "Reproduced with permission from ["B-spline algebraic diagrammatic construction: Application to photoionization cross-sections and high-order harmonic generation", M. Ruberti, V. Averbukh and P. Decleva, *J. Chem. Phys.* 141, 164126 (2014). <http://dx.doi.org/10.1063/1.4900444>]. Copyright [2014], AIP Publishing LLC.

- [25] H. Bachau, E. Cormier, P. Decleva, J. E. Hansen and F. Martin, Rep. Prog. Phys., 2001, **64**, 1815.
- [26] C. Bottcher, Phys. Rev. Lett., 1982, **48**, 85.
- [27] R. D. Mattuck, *A Guide to Feynman Diagrams in the Many-Body Problem*, Dover, New York, 1992.
- [28] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover, New York, 1996.
- [29] E. Dalgaard, Int. J. Quantum Chem., 1979, **15**, 169.
- [30] P.-O. Löwdin, Phys. Rev., 1955, **97**, 1474.
- [31] J. F. Stanton and R. J. Bartlett, J. Chem. Phys., 1993, **98**, 7029.
- [32] P. Decleva, A. Lisini and M. Venuti, J. Phys. B: At. Mol. Opt. Phys., 1994, **27**, 4867.
- [33] B. Chapman, G. Jost and R. van der Pas, *Using OpenMP: Portable Shared Memory Parallel Programming*, The MIT Press Cambridge, Massachusetts, 2007.
- [34] M. Snir, S. Otto, S. Huss-Lederman, D. Walker, and J. Dongarra, *MPI: The Complete Reference*, The MIT Press Cambridge, Massachusetts, 1996.
- [35] U. V. Riss and H.-D. Meyer, J. Phys. B, 1993, **26**, 4503.
- [36] Farhad H.M. Faisal, *Theory of Multiphoton Processes*, Springer US, New York, 1987.
- [37] W. E. Arnoldi, Q. Appl. Math., 1951, **9**, 17.
- [38] R. A. Friesner, L. S. Tuckerman, B. C. Dornblaser, and T. V. Russo, J. Sci. Comput., 1989, **4**, 327.
- [39] H. Hennig, J. Breidbach, and L. S. Cederbaum J. Chem. Phys., 2005, **122**, 134104.
- [40] C. Leforestier, R. H. Bisseling, C. Cerjan, M.D. Feit, R. Friesner, A. Guldborg, A. Hammerich, G. Jolicard, W. Karrlein, H.-D Meyer, N. Lipkin, O. Roncero and R. Kosloff, J. Comput. Phys., 1991, **94**, 59.

- [41] T. J. Park and J. C. Light, J. Chem. Phys., 1986, **85**, 5870.
- [42] G. Jolicard and E. Austin, J. Chem. Phys., 1991, **95**, 5056.
- [43] M. H. Beck, A. Jnckle, J. A. Worth, and H.-D. Meyer, Phys. Rep., 2001, **324**, 1.
- [44] J. A. R. Samson and W. C. Stolte, J. Electron Spectrosc. Relat. Phenom., 2002, **123**, 265.
- [45] Chan W. F., Cooper G., Guo X., Burton G. R. and Brion C. E., Phys. Rev. A, 1992, **46**, 149.
- [46] A.F. Starace, in *Theory of atomic photoionization Handbuch der Physik*, ed. S Flügge, Springer, Berlin, 1982, Vol. **31**.
- [47] M. Stener, P. Decleva and A. Lisini, J. Phys. B: At. Mol. Opt. Phys., 1995, **28**, 4973.
- [48] P. B. Corkum, Phys. Rev. Lett., 1993, **71**, 1994.
- [49] Y. Mairesse, J. Higuët, N. Dudovich, D. Shafir, B. Fabre, E. Mvel, E. Constant, S. Patchkovskii, Z. Walters, M. Yu. Ivanov, and O. Smirnova, Phys. Rev. Lett., 2010, **104**, 213601.
- [50] J. Itatani, J. Levesque, D. Zeidler, H. Niikura, H. Pepin, J. Kieffer, P. Corkum, and D. Villeneuve, Nature (London), 2004, **432**, 867.
- [51] C. Vozzi, M. Negro, F. Calegari, G. Sansone, M. Nisoli, S. De Silvestri and S. Stagira, Nature Physics, 2011, **7**, 822.
- [52] T. Kanai, S. Minemoto, and H. Sakai, Nature (London), 2005, **435**, 470.
- [53] S. Baker, J. S. Robinson, C. A. Haworth, H. Teng, R. A. Smith, C. C. Chirila, M. Lein, J. W. G. Tisch and J. P. Marangos, Science, 2006, **312**, 424.
- [54] J. P. Marangos, S. Baker, N. Kajumba, J. S. Robinson, J. W. G. Tisch and R. Torres, Phys. Chem. Chem. Phys., 2008, **10**, 35.
- [55] C. Vozzi, R. Torres, M. Negro, L. Brugnera, T. Siegel, C. Altucci, R. Velotta, F. Frassetto, L. Poletto, P. Villoresi, S. De Silvestri, S. Stagira, and J. P. Marangos, Appl. Phys. Lett., 2010, **97**, 241103.

- [56] O. Smirnova, Y. Mairesse, S. Patchkovskii, N. Dudovich, D. Villeneuve, P. Corkum, and M. Y. Ivanov, *Nature (London)*, 2009, **460**, 972.
- [57] A. Gordon and F. Kärtner, *Opt. Express*, 2005, **13**, 2941.
- [58] M. Lewenstein, Ph. Balcou, M. Yu. Ivanov, A. LHuillier and P. B. Corkum, *Phys. Rev. A*, 1994, **49**, 2117.
- [59] J. Higuier, H. Ruf, N. Thiré, R. Cireasa, E. Constant, E. Cormier, D. Descamps, E. Mvel, S. Petit, B. Pons, Y. Mairesse, and B. Fabre, *Phys. Rev. A*, 2011, **83**, 053401.
- [60] A. D. Shiner, B. E. Schmidt, C. Trallero-Herrero, P. B. Corkum, J.-C. Kieffer, F. Legare and D. M. Villeneuve, *J. Phys. B: At. Mol. Opt. Phys.*, 2012, **45**, 074010.
- [61] H. J. Wörner, H. Niikura, J. B. Bertrand, P. B. Corkum, and D. M. Villeneuve, *Phys. Rev. Lett.*, 2009, **102**, 103901.
- [62] J. P. Farrell, L. S. Spector, B. K. McFarland, P. H. Bucksbaum, M. Guhr, M. B. Gaarde, and K. J. Schafer, *Phys. Rev. A*, 2011, **83**, 023420.
- [63] S. Pabst, L. Greenman, D. A. Mazziotti and R. Santra, *Phys. Rev. A*, 2012, **85**, 023411.
- [64] M. Ruberti, P. Decleva and V. Averbukh, *New J. Phys.*, submitted.
- [65] B. K. McFarland, J. P. Farrell, P. H. Bucksbaum, and M. Guhr, *Science*, 2008, **322**, 1232.
- [66] W. Li, X. Zhou, R. Lock, S. Patchkovskii, A. Stolow, H. C. Kapteyn, and M. M. Murnane, *Science*, 2008, **322**, 12071211.
- [67] A.-T. Le, R. R. Lucchese, and C. D. Lin, *J. Phys. B*, 2009, **42**, 211001.
- [68] T. Morishita, A.-T. Le, Z. Chen, and C. D. Lin, *Phys. Rev. Lett.*, 2008, **100**, 013903.
- [69] A.-T. Le, R. R. Lucchese, S. Tonzani, T. Morishita, and C. D. Lin, *Phys. Rev. A*, 2009, **80**, 013401.
- [70] S. Minemoto, T. Umegaki, Y. Oguchi, T. Morishita, A.-T. Le, S. Watanabe, and H. Sakai, *Phys. Rev. A*, 2008, **78**, 061402(R).

- [71] C. Vozzi, F. Calegari, E. Benedetti, J.-P. Caumes, G. Sansone, S. Stagira, M. Nisoli, R. Torres, E. Heesel, N. Kajumba, J. P. Marangos, C. Altucci, and R. Velotta, *Phys. Rev. Lett.*, 2005, **95**, 153902.
- [72] W. Boutu, S. Haessler, H. Merdji, P. Breger, G. Waters, M. Stankiewicz, L. J. Frasinski, R. Taieb, J. Caillat, A. Maquet, P. Monchicourt, B. Carre and P. Salieres, *Nature Phys.*, 2008, **4**, 545549.
- [73] X. Zhou, R. Lock, W. Li, N. Wagner, M. M. Murnane, and H. C. Kapteyn, *Phys. Rev. Lett.*, 2008, **100**, 073902.
- [74] H. J. Wörner, J. B. Bertrand, P. Hockett, P. B. Corkum, and D. M. Villeneuve, *Phys. Rev. Lett.*, 2010, **104**, 233904.
- [75] E. R. Simpson, A. Sanchez-Gonzalez, D. R. Austin, Z. Diveki, S. E. E. Hutchinson, T. Siegel, M. Ruberti, V. Averbukh, L. Miseikis, C. S. Struber, L. Chipperfield and J. P. Marangos, *New J. Phys.*, 2016, **18**, 083032.
- [76] L. S. Cederbaum, W. Domcke, J. Schirmer and W. Von Niessen, *Adv. Chem. Phys.*, 1986, **65**, 115.
- [77] F. Remacle and R. D. Levine, *Proceedings of the National Academy of Science* **103**, 6793 (2006).
- [78] D. Toffoli, M. Stener, G. Fronzoni, and P. Decleva, *Chem. Phys.*, 2002, **276**, 25.