ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

iSPECTRON: A simulation interface for linear and nonlinear spectra with ab-initio quantum chemistry software

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*
Segatta F., Nenov A., Nascimento D.R., Govind N., Mukamel S., Garavelli M. (2021). iSPECTRON: A simulation interface for linear and nonlinear spectra with ab-initio quantum chemistry software. JOURNAL OF COMPUTATIONAL CHEMISTRY, 42(9), 644-659 [10.1002/jcc.26485].

*Availability:*
This version is available at: https://hdl.handle.net/11585/845939 since: 2023-05-12

*Published:*
DOI: http://doi.org/10.1002/jcc.26485

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

12 October 2024

This is the final peer-reviewed accepted manuscript of:

**iSPECTRON: A simulation interface for linear and nonlinear spectra with ab-initio quantum chemistry software Segatta F.; Nenov A.; Nascimento D. L.; Govind N.; Mukamel S.; Garavelli M.**

**J. Comp. Chem. 2021, 42, 644–659**

The final published version is available online at: **https://dx.doi.org/10.1002/jcc.26485**

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

*This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)*

***When citing, please refer to the published version.***

# iSPECTRON: a simulation interface for linear and nonlinear spectra with ab-initio quantum chemistry software

Francesco Segatta[1], Artur Nenov[1], Daniel R. Nascimento[2], Niranjan Govind[2,*], Shaul Mukamel[3,*], and Marco Garavelli[1,*]

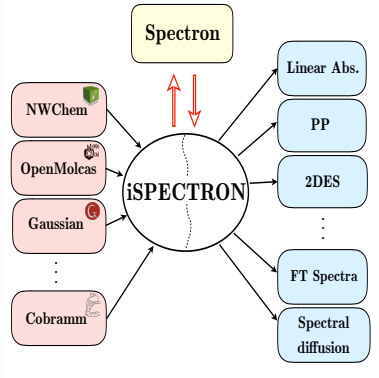
[1]Dipartimento di Chimica Industriale "Toso Montanari", Università degli Studi di Bologna, Viale del Risorgimento 4, I-40136 Bologna, Italy
[2]Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington 99352, United States
[3]Department of Chemistry and Department of Physics and Astronomy, University of California, Irvine, 92697, USA

December 22, 2020

**Abstract**

We introduce iSPECTRON, a program that parses data from common quantum chemistry software (NWChem, OpenMolcas, Gaussian, Cobramm, etc.), produces the input files for the simulation of linear and nonlinear spectroscopy of molecules with the Spectron code, and analyzes the spectra with a broad range of tools. Vibronic spectra are expressed in term of the electronic eigenstates, obtained from quantum chemistry computations, and vibrational/bath effects are incorporated in the framework of the displaced harmonic oscillator model, where all required quantities are computed at the Franck-Condon point. The program capabilities are illustrated by simulating linear absorption, transient absorption and two dimensional electronic spectra of the pyrene molecule. Calculations at two levels of electronic structure theory, TDDFT (with NWChem) and RASSCF/RASPT2 (with OpenMolcas) are presented and compared where possible. The iSPECTRON program is available online at https://github.com/ispectrongit/iSPECTRON/ and distributed open source under the terms of the Educational Community License version 2.0 (ECL 2.0).

Keywords: linear and nonlinear optical spectroscopy simulation, CASSCF/CASPT2, TDDFT, displaced harmonic oscillator, line-shape functions; ■




(iSPECTRON is a general purpose interface between popular QM packages (Gaussian, NWChem, OpenMolcas) and Spectron, a platform for simulation of coherent multidimensional optical spectroscopy of molecules. It allows the simulation, visualization and post-processing of an ample body of linear and nonlinear spectra from first principles. The interface is aimed at quantum chemists who seek to connect their QM calculations with spectroscopic observables, as well as expert spectroscopists who seek easy access to QM data and a tool for spectra manipulation.)

# INTRODUCTION

Optical spectroscopy is a valuable experimental tool for investigating electronic and vibrational properties and dynamics of molecular systems. The past few decades have witnessed the birth of multi-dimensional multi-pulse techniques, which allow one to study elementary molecular events such as energy and charge transfer processes, formation and evolution of vibrational and electronic coherences, conformational and solvent dynamics and even to study the evolution of a system simultaneously in time and space (4D spectroscopy)[1-8]. In combination with ultra-short laser pulses with high phase stability, these novel techniques have equipped researchers with the necessary tools to unravel gas- and condensed-phase dynamics in the sub-femtosecond (fs, $10^{-15}$ s) and even in the attosecond (as, i.e. $10^{-18}$ s)[9,10] regimes with an unprecedented level of detail. Nonetheless, connecting the optical response of the system to the underlying quantum-chemical structure and vibronic dynamics remains an open computational challenge. Theoretical simulations constitute an invaluable tool for unravelling the complex underlying dynamics. The simulation of molecular spectroscopy is based on two ingredients: i) an accurate description of the electronic and vibrational structure and dynamics of the molecular system, and ii) an effective approach to account for the light-matter interaction and response.

Step i) is usually tackled by employing the tools of quantum chemistry, i.e., by solving the Schrödinger equation at some level of approximation, which yields the manifold of electronic states of the system and the transition dipole between them. Broadly, quantum mechanical calculations fall into two families - wavefunction (WF) based and density functional theory (DFT) based - both offering advantages and limitations due to the underlying approximations and computational cost. The description of molecular vibrations, environment fluctuations, as well as the mechanism and time scale of energy dissipation (e.g. internal conversion, energy or charge transfer) are key additional ingredients to properly access the physical phenomena that govern the recorded line-shapes. This can be achieved at various levels of sophistication ranging from numerically exact schemes solving the time-dependent Schrödinger equation (quantum dynamics) to model effective Hamiltonians which parametrize the couplings between the electronic states, as well as between the electronic

and vibrational degrees of freedom.

Step ii) is usually treated by quantum master equations which employ the non-perturbative or perturbative (i.e. weak-field) treatment of the light-matter interactions. A comprehensive perturbative theoretical framework for linear and nonlinear optical spectroscopies was first introduced a few decades ago.[11] The key quantity, the response function, incorporates all the relevant molecular information for the response to external light-induced perturbations. The response function requires, as a minimal input, the electronic levels of the system under study and the transition dipoles between them, but it also incorporates the vibronic dynamics at various degrees of sophistication.

Over the years, specialized programs that address steps i) and ii) have been developed by different groups and made available to the user community. For the former, various proprietary and open source packages for QM calculations exist - Gaussian[12], Molcas[13,14], NWChem[15,16], GAMESS[17] to name a few - but none of them encompass the entire array of existing and emerging electronic structure methods. General purpose codes for simulating spectroscopy are less common, reflecting the complexity associated with system specific input parameters required to simulate the nonlinear response. Among the publicly available codes we note FCclasses[18] and EXAT[19] for simulating the linear response in the eigenstate and excitonic picture, respectively, as well as Spectron, a platform developed by the Mukamel group specifically to describe coherent multidimensional optical spectroscopy of molecules, chromophore aggregates and semiconductor nanostructures from the IR to the X-ray.[20,21][1]. There is a clear gap between Steps (i) and (ii) that needs to be bridged for the general user. In the present paper we introduce iSPECTRON, an open source program and distributed under the Educational Community License version 2.0 (ECL 2.0), conceived as a stepping stone towards a general purpose interface that can fill this gap. As schematically shown in Figure 1), it has been designed with a twofold goal:

(a) Automatize the parsing of the output data from a range of popular quantum chemistry programs (NWChem, OpenMolcas, Gaussian, Cobramm, etc.) to produce the necessary input files to simulate linear and nonlinear spectroscopy with the Spectron code;

[1]Spectron can be obtained for free from https://mukamel.ps.uci.edu/software.html

(b) Read the nonlinear response data produced by the Spectron code and produce, display and analyze a broad arsenal of possible spectra (transient absorption, two-dimensional electronic spectroscopy (2DES), power spectra, etc.).

The interface is aimed at quantum chemists who seek to connect their QM calculations with spectroscopic observables, as well as at expert spectroscopists who seek easy access to QM data and a tool for spectra manipulation. In this work we present the capabilities of version 1.0 of the interface, which facilitates the first principles simulation of linear and nonlinear spectra within the electronic eigenstate expansion of the response function, relying on model Hamiltonians to describe the coupling to a bath[2]. In the **METHODOLOGY** Section we provide a synopsis of the basic equations at the heart of the Spectron code, with the focus on the link to QM calculations. The workflow and capabilities of the interface are presented in a nutshell in the **Interface Workflow(s)** Section and employed in the **EXAMPLES** Section to track the photo-induced vibrational dynamics and ultrafast internal conversion of pyrene, a polycyclic aromatic hydrocarbon (PAH) consisting of four fused benzene rings. We compare linear spectra at the time-dependent (TD) DFT (DFT-based) and RASSCF/RASPT2 (WF-based) levels of theory, employing data obtained from the NWChem and OpenMolcas software packages, respectively. Finally we present nonlinear spectra at the RASSCF/RASPT2 level of theory and analyze them with the post-processing tools of the interface. A summary and future developments are reported in the last section.

# METHODOLOGY

## Theory

In this section we summarize the theoretical background used by the interface and by the Spectron code to simulate a broad variety of spectroscopic signals.

[2]The bath comprises the intra-molecular and the environment (e.g. a protein cage, a solvent, a nearby molecule) degrees of freedom (DOF)

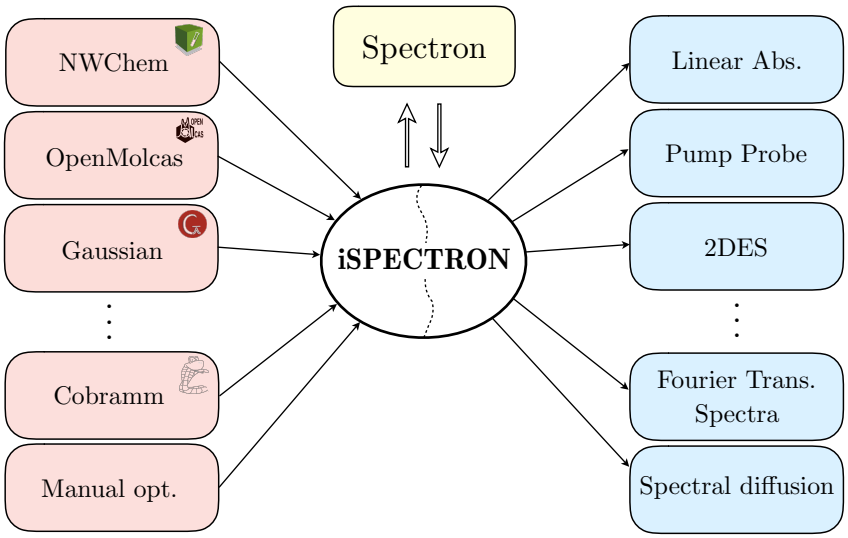



Figure 1: Schematic representation of the iSPECTRON links with quantum chemistry software, its double-sided flow of data with the Spectron software, and the range of possible spectroscopic techniques and properties that can be produced with it.

## Electronic Structure Methods

The essential quantities for the simulation of spectroscopy are the manifold of energies of each state and the transition dipole moments between these states. Additionally, calculations of normal modes and frequencies at the ground state (GS) minimum, and evaluation of vertical gradients for the states of interest may be required. All of these properties may be obtained at the QM level by employing either WF-based or DFT-based approaches. Here, our focus is on multi-configurational WF techniques labelled as multi-configurational self-consistent field (MCSCF) methods, and the techniques emerging through the linear-response time-dependent (TD) extension of DFT categorized as TDDFT methods. As part of this work, we demonstrate the iSPECTRON interface with the widely used OpenMolcas[13,14] and NWChem[15,16] computational chemistry programs, respectively.

- **MCSCF techniques**, such as the complete active space self consistent field (CASSCF, and its variants, as, e.g., the restricted active space (RAS) SCF)[22-25], employed in tandem with second order perturbative corrections (CASPT2, or RASPT2)[26], represent the state-of-the-art method for the computation of excited state properties in generic

molecular systems. The CASSCF/CASPT2 protocol has been extensively used over the past 30 years and is widely considered the gold standard for a number of applications in photochemistry[27]. The major bottlenecks of the protocol lies in its computational cost, which rapidly increases with the system size. The OpenMolcas program contains the most advanced suite of MCSCF techniques.

With the wavefunction at hand, it is possible to obtain (virtually) all properties necessary for the spectroscopy simulation: state energies, permanent and transition dipole moments between them, frequencies, gradients, etc. Particularly important is the ability of the CASSCF/CASPT2 protocol to describe higher lying states which may be accessed in nonlinear spectroscopy techniques. Therefore, it has been utilized in the past decade to simulate transient absorption, two-dimensional IR, Raman and electronic spectra of organic compounds (e.g. pyrene, uracil and thio-uracil, azobenzene) from first principles.

- **DFT/TDDFT techniques**, are the most widely used theories for electronic structure calculations. They offer a good compromise between accuracy and computation cost, and are therefore broadly applicable over the chemical and materials space across the periodic table. Because of their broad applicability and computational efficacy, DFT and TDDFT have been implemented in all established quantum chemistry packages, such as Gaussian[12] and NWChem[15,16] to name a few. However, double excitations and charge transfer states have been found to be problematic with TDDFT. In addition, transition dipole moments between excited states within TDDFT require solving the costly quadratic-response equations, although recent theoretical advances have been reported to address this issue.[28,29] This has limited their use in the simulation of nonlinear spectroscopies. Here, we utilize the DFT/TDDFT implementation in NWChem, where excited-state transition dipole moments have been implemented recently using the TDDFT-based pseudowavefunction approach.[30]

We refer interested readers to several recent reviews on the state-of-the-art of WF-based and DFT-based QC methods[28,31-34]. A brief summary of the two techniques can be also found in the Supporting Information.

### Accounting for environment effects

The study of molecular systems often requires considering their complex native condensed-phase environments. This can either be described implicitly with an analytical approach known as the polarizable continuum model (PCM)[35] or explicitly, where the entire system is treated within a common level of theory or via hybrid embedding schemes like ONIOM[36] and quantum mechanics/molecular mechanics (QM/MM). iSPECTRON has also been interfaced with the COBRAMM[37] program, which provides a flexible QM/MM framework between different electronic structure programs and the Amber molecular mechanics suite.[38]

## Spectroscopy Simulations

The term "spectroscopy" comprises the conspicuous body of techniques that use radiation to obtain information on the structure and dynamical properties of matter. In the dipole (long wavelength) approximation, where the field does not vary across the molecule, the system is described by the Hamiltonian

$$\hat{H} = \hat{H}_{\text{mol}} - \hat{\boldsymbol{\mu}} \cdot \boldsymbol{E}(t) \tag{1}$$

where the first term represent the molecular Hamiltonian, which can be handled via quantum chemistry approaches; the second term represents the (semi-classical) interaction between the external (classical) electromagnetic field and the molecular system, through the dipole operator $\hat{\boldsymbol{\mu}}$.

A density matrix (rather than a wavefunction) representation of the system is required when dealing with open quantum systems and time-dependent perturbation theory is applied with respect to the field-matter interaction term: different orders of the perturbative expansion give rise to specific signals. We focus here on first- (linear) and third-order (nonlinear) signals.

The interaction between the external electromagnetic field and the electrons produces a time-dependent polarization. This is the source of a new field, which eventually creates the measured signal of spectroscopic experiments.

### The linear and nonlinear response functions

The linear (first-order) polarization is connected to the incident radiation field through the linear relation[20]

$$\boldsymbol{P}^{(1)}(\boldsymbol{r},t) = \int_0^\infty dt_1 R^{(1)}(t_1) E(\boldsymbol{r}, t-t_1) \tag{2}$$

$R^{(1)}(t_1)$ is the (first-order) system *response function*, and contains the information of the first-order perturbed system. It may be recast as:

$$R^{(1)}(t_1) = \left(\frac{i}{\hbar}\right) Tr\left[\hat{\boldsymbol{\mu}}\mathbb{G}(t_1)\mu^\times\hat{\rho}_0\right] \tag{3}$$

where $\hat{\rho}_0$ is the equilibrium density matrix (i.e. prior to any field interaction), $\mathbb{G}(t_1)$ is a superoperator that describes the density matrix evolution in absence of the external field (defined as $\mathbb{G}(t)\bullet = e^{-\frac{i}{\hbar}\hat{H}_{\rm mol}t}\bullet e^{\frac{i}{\hbar}\hat{H}_{\rm mol}t}$), and $\mu^\times$ is the dipole superoperator (defined as $\mu^\times\bullet = [\hat{\boldsymbol{\mu}},\bullet]$) that accounts for the interaction between the system and the external field. Similarly, the third-order polarization reads[20]

$$\begin{aligned}\boldsymbol{P}^{(3)}(\boldsymbol{r},t) \;=\; & \int_0^\infty dt_3 \int_0^\infty dt_2 \int_0^\infty dt_1 R^{(3)}(t_3,t_2,t_1) \times \\ & E(\boldsymbol{r},t-t_3)E(\boldsymbol{r},t-t_3-t_2)E(\boldsymbol{r},t-t_3-t_2-t_1)\end{aligned} \tag{4}$$

where the corresponding response function, $R^{(3)}(t_3,t_2,t_1)$, is given by

$$R^{(3)}(t_3,t_2,t_1) = \left(\frac{i}{\hbar}\right)^3 Tr\left[\hat{\boldsymbol{\mu}}\mathbb{G}(t_3)\mu^\times\mathbb{G}(t_2)\mu^\times\mathbb{G}(t_1)\mu^\times\hat{\rho}_0\right] \tag{5}$$

Eqs. (3) and (5), can be interpreted as follows:

1. Start from the the equilibrium density matrix $\rho_0$;
2. Interact once with the laser via the dipole superoperator $\mu^\times$;
3. Evolve the perturbed density matrix by the field-free molecular Hamiltonian $H_{\rm mol}$ for a time $t_1$;
4. Repeat steps 2. and 3. for times $t_2$ and $t_3$;
5. Compute, with the perturbed and time-dependent density matrix $\rho^1(t_1)$ ($\rho^3(t_1,t_2,t_3)$), the first (third) order response function $P^{(1)}(t_1) = Tr\left[\boldsymbol{\mu}\rho^{(1)}\right]$ ($P^{(3)}(t_1,t_2,t_3) = Tr\left[\boldsymbol{\mu}\rho^{(3)}\right]$);

The polarization must be averaged over all possible system orientations to represent a randomly oriented sample.[3]

A general solution to eqs. (3) and (5) can be derived[11] which applies to arbitrary level schemes (see Supporting Information). In the case of third-order polarization, the number of terms that appear in the general expression is very large due to the various Liouville space pathways[4] and the permutation over the incident fields. Therefore, the response is often recast for approximate models of two or three distinct manifolds (i.e. bands, see section "Sum-over-states expansion" for definition of the manifolds) in combination with temporally well separated pulses, resonant with the energy gaps between the manifolds. Fortunately, in resonant conditions, only a few of the terms survive when a specific (called phase-matched) combination of the incoming fields wave vectors $\boldsymbol{k}_i, i = 1, 2, 3$ is chosen, as highly oscillatory terms (where molecular and field frequencies) add up rather than subtract can be safely neglected, an approximation known as the rotating wave approximation (RWA). Experimentally, different types of linear and nonlinear signals that radiate from the sample can be distinguished by placing the detector in different directions, identified by a specific combination of the incoming fields wave vectors $\pm\boldsymbol{k}_1 \pm \boldsymbol{k}_2 \pm \boldsymbol{k}_3$.

Signals in a given direction can be further decomposed into ground state bleaching (GSB), stimulated emission (SE) and excited state absorption (ESA)[5]. The nomenclature reflects the physical processes induced by the light-matter interactions.[20] The response function of eq. (5) in a given phase-matching direction can be then expressed as the sum of the individual (GSB, SE, ESA) contributions. Notably, ESA contributions have an opposite sign to GSB and SE contributions, thus their spectral overlap could potentially lead to complex interference patterns and occasionally to signal cancellation. For this reason, the calculation of the individual components offered by Spectron is essential.

[3]Note that in some specific cases the systems in the sample may be globally oriented in a specific direction: in that case such orientational average should not be performed.

[4]The Liouville pathways are the terms that appear when the action of the super-operator $\mu^\times$ is written out explicitly: being a commutator, it can act either on the left or on the right, producing multiple contributions known as Liouville space pathways.

[5]Note that, in a real experiment, GSB, SE and ESA associated with the same phase-matching condition cannot be detected separately.

In linear spectroscopy the time-dependent polarization (eq. (2)) is Fourier transformed with respect to $t$

$$\boldsymbol{P}^{(1)}(\Omega) \;=\; \int_0^\infty dt \boldsymbol{P}^{(1)}(t) \exp\{i\Omega t\} \tag{6}$$

to provide an intensity vs. $\Omega$ spectrum - the LA spectrum. In third-order spectroscopy (rephasing, non-rephasing and quasi absorptive signals) any pair of time variables can be used to generate a 2D plot. The time-dependent polarization (eq. (4)) is usually Fourier transformed with respect to variables $t_1$ and $t_3$

$$\boldsymbol{P}^{(3)}(\Omega_1, t_2, \Omega_3) \;=\; \int_0^\infty dt_3 \int_0^\infty dt_1 \boldsymbol{P}^{(3)}(t_3, t_2, t_1) \exp\{i\Omega_3 t_3 + i\Omega_1 t_1\} \tag{7}$$

to provide a two-dimensional map (intensity vs. $\Omega_1$ vs. $\Omega_3$) - the 2D spectrum - for each (waiting) time $t_2$. The double quantum coherence signal is Fourier transformed with respect either to intervals $t_1$ and $t_2$ or $t_2$ and $t_3$. Finally, the transient absorption (pump probe) spectra (intensity vs. $\Omega_3$ at different $t_2$ times) can be readily obtained from the two-dimensional maps by integrating over $\Omega_1$.

$$\boldsymbol{P}^{(3)}(t_2, \Omega_3) \;=\; \int_0^\infty d\Omega_1 \boldsymbol{P}^{(3)}(\Omega_1, t_2, \Omega_3) \tag{8}$$

### Sum-over-states (SOS) expansion: connecting the electronic structure of the system to a spectroscopic signal

In the SOS approach, the response function is expanded in the eigenstate basis. These states are usually divided in three sets: on the bottom of the energy ladder there is the unique ground state (denoted by $g$). Above it, there is the single-excitation manifold (with states denoted by the index $e$). Finally, one finds the double-excitation manifold (index $f$). The manifolds are connected by the dipole operator $\boldsymbol{\mu}$, and we assume that only transitions between $g \to e$ (characterized by energy $\omega_{eg}$ and transition dipole moment $\boldsymbol{\mu}_{eg}$) and $e \to f$ ($\omega_{fe}$ and $\boldsymbol{\mu}_{fe}$) can occur (Figure 2(d)). Linear techniques only interrogate states of the first-excitation manifold, while in third-order techniques the second excitation manifold should be considered as well.

We note that the terms "single-excitation" and "double-excitation" are historically related to the study of excitonic systems and referred to exciton states having a single site or two sites simultaneously excited.[20,21] In the framework of the current level scheme (Figure 2(d)), the $f$-manifold can include states generated by single-excitation from the ground state which are coupled to states from the $e$-manifold by a non-vanishing transition dipole moment. Furthermore, as typical in long conjugated systems, doubly excited states can lie among the lowest excited states of the system and, while not coupled to the ground state (i.e they are dark) they participate in internal conversion. A more convenient way to think of the $e$ and $f$ manifolds is in terms of being resonant with the energy of the pump ($e$ manifold, $\omega_{eg}$ energies) and probe ($f$ manifold, $\omega_{fe}$ energies) pulses.[39,40] Thus, the assignment of states to either manifold is connected to the experimental conditions.

The $f$ states are not directly involved in the photo-induced dynamics (during $t_2$). However, as each $e$ state has its own characteristic manifold of higher lying $f$ states the ESA associated with their population upon interaction with the probe pulse can serve as indirect signatures of the underlying dynamics in the photo-active states. This is particularly valuable for resolving states intermediately populated during the non-adiabatic decay which are dark with respect to the ground state (i.e. dark or phantom states). Thus, states from the $f$ manifold could be considered "spectator states" in the terminology of transient IR spectroscopy.[14,39]

Three non vanishing (under the RWA) independent third-order signals can be obtained for the three-level scheme (Figure 2(d)) from different combinations of the wave vectors of the incident pulses $\boldsymbol{k}_1$, $\boldsymbol{k}_2$ and $\boldsymbol{k}_3$. The signals are referred to as rephasing ($\boldsymbol{K}_I = -\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3$), non-rephasing ($\boldsymbol{K}_{II} = +\boldsymbol{k}_1 - \boldsymbol{k}_2 + \boldsymbol{k}_3$) and double-quantum coherence ($\boldsymbol{K}_{III} = +\boldsymbol{k}_1 + \boldsymbol{k}_2 - \boldsymbol{k}_3$)(see Liouville space ladder diagrams in Fig. 2). The various signals encompass a different complementary information about the system.

By adding the simulated/recorded rephasing and non-rephasing contributions one obtains (quasi-)absorptive spectra which provide the highest spectral resolution thanks to favorable cancellation of dispersive spectral components. The simulation of the various third-order signals is an integral part of the Spectron code.

In the SOS framework the first- and third-order response functions (eqs. (3) and (5)) are

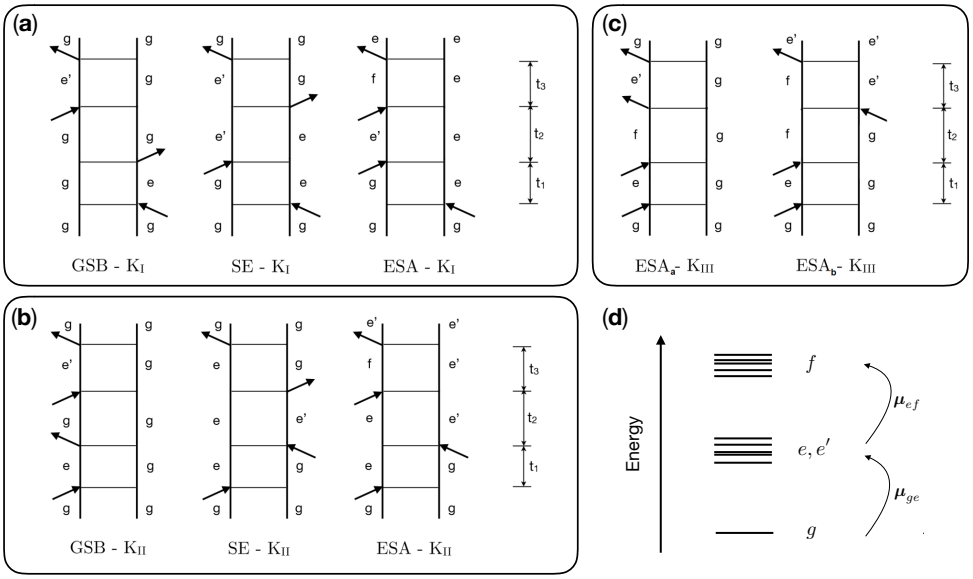



Figure 2: Ladder diagrams for the signal generated in the direction $\boldsymbol{K}_I$ (a), $\boldsymbol{K}_{II}$ (b) and $\boldsymbol{K}_{III}$ (c); GSB: ground-state bleaching; SE: excited-state emission; ESA: excited-state absorption. (d) Energy level diagram and allowed transitions between the states in the various manifolds.

expanded in terms of the system eigenstates. $R^{(1)}(t_1)$ thus becomes:

$$R^{(1)}(t_1) = \left(\frac{i}{\hbar}\right)\theta(t_1)\sum_e |\boldsymbol{\mu}_{eg}|^2 \exp\{-i\left(\omega_{eg} - i\gamma_{eg}\right)t_1\} + c.c. \tag{9}$$

We have added a phenomenological dephasing rate $\gamma_{eg}$ to the transitions. This accounts for the decay of the coherence and the associated signal broadening in an open system due to processes such as pure dephasing, internal conversion, etc., whose detailed and explicit description is omitted at this lowest[6] level of theory.

The linear absorption spectrum represents the real (absorptive) part of the Fourier transformed polarization $\mathrm{Re}\{\boldsymbol{P}^{(1)}(\Omega)\}$ (eq. (2)).

$$\boldsymbol{P}^{(1)}(\Omega) = \sum_e \frac{|\boldsymbol{\mu}_{eg}|^2}{\Omega - \omega_{eg} + i\gamma_{eg}} \tag{10}$$

It is given as a sum of transitions centered at $\omega_{eg}$, having Lorenzian line-shapes whose widths are governed by the dephasing rates $\gamma_{eg}$.

[6]"Lowest" in terms of the description of the signal line-shapes.

Similarly, the SE diagram of the non-rephasing third-order signal becomes:

$$R_{SE}^{(\boldsymbol{K}_{II})}(t_1,t_2,t_3) = \left(\frac{i}{\hbar}\right)^3 \theta(t_1)\theta(t_2)\theta(t_3) \sum_{ee'} |\boldsymbol{\mu}_{eg}|^2 |\boldsymbol{\mu}_{e'g}|^2 \times \exp\{-i\left(\omega_{ge} - i\gamma_{ge}\right)t_1 - i\left(\omega_{e'e} - i\gamma_{e'e}\right)t_2 - i\left(\omega_{e'g} - i\gamma_{e'g}\right)t_3\} \tag{11}$$

whereas the spectrum resulting from the Fourier transform of the third-order polarization (eq. (4))

$$\boldsymbol{P}_{SE}^{(\boldsymbol{K}_{II})}(\Omega_1,t_2,\Omega_3) = \sum_e \frac{|\boldsymbol{\mu}_{eg}|^2}{\Omega_1 - \omega_{eg} + i\gamma_{eg}} \cdot \frac{|\boldsymbol{\mu}_{e'g}|^2}{\Omega_3 - \omega_{e'g} + i\gamma_{eg}} \exp\{-i\left(\omega_{e'e} - i\gamma_{e'e}\right)t_2\} \tag{12}$$

has a more complex line shape with positive and negative regions (phase twists). Similar expressions can be derived for the remaining third-order diagrams[11,20]. Within the framework of eqs. (10) and (12), the necessary QM inputs to the spectroscopy simulation with Spectron are limited to the manifold of states energies (from which one builds the required energy differences $\omega_{eg}$ and $\omega_{fe}$) and the transition dipole moments ($\boldsymbol{\mu}_{eg}$ and $\boldsymbol{\mu}_{fe}$) between them. The dephasing rate $\gamma$ may be fitted to describe the observed line broadening.

### The line shapes: incorporating the coupling to a harmonic bath

Modelling the line shapes requires to describe the field-free dynamics of the system coupled to a bath. This can be realized at various levels of sophistication: a) the exact numerical solution of the Liouville equation in the reduced system space; b) semi-classical and mixed-quantum classical trajectory-based approaches which approximate or neglect altogether some quantum mechanical aspects (ranging from accurate path-integral methods, that can be made arbitrarily accurate, to crude phenomenological separations between quantum and classical degrees of freedom); c) methods based on parametrized Hamiltonians. The latter rely on the (strong) approximation that the potential energy surfaces of the electronic states of interest can be expressed by analytical expressions (e.g harmonic potentials) which can be parametrized by quantum mechanical calculations (energies, gradients, normal modes etc.). The appeal of this approach is boosted by the availability of analytical expressions for the signal line shapes. Therefore, facilitating line shape description based on the coupling between the electronic structure to a model Hamiltonian describing the intra-molecular and the environment vibrational degrees of freedom (DOF) is naturally the first strategy made

available through the iSPECTRON interface that goes beyond purely phenomenological line shapes.

Here we outline the displaced harmonic oscillator (DHO) model Hamiltonian used to describe the coupling to intra-molecular DOF and the resulting line shape functions. The model assumes that the potential energy surfaces of the electronic states are quadratic in the coordinates, with different equilibrium positions (i.e. electronic state $a$ is *displaced* by $\tilde{d}_{ak}$ along the $k-$th mass-weighted normal mode coordinate $\tilde{q}_k$). Different modes are assumed to be uncoupled (i.e., intramolecular vibrational energy redistribution cannot take place), and possible anharmonicities are not accounted for. It is also assumed that the different electronic states are described by the same normal modes and frequencies (which are thus computed only for a single electronic state, generally being the electronic GS)[7]

The displacements $\tilde{d}_{ak}$ of the various ES wells with respect to the GS can be computed with the vertical gradient approach, implemented in iSPECTRON, i.e. by projecting energy gradients for the states of interest onto the GS normal modes. Starting with the expression for the energy gap $\Delta E_{ab}(\tilde{q}_k)$ between states $a$ and $b$ in the framework of the DHO model

$$\Delta E_{ab}(\tilde{\boldsymbol{q}}) = \varepsilon_a - \varepsilon_b + \sum_k \frac{1}{2}\omega_k^2(\tilde{d}_{ak}^2 - \tilde{d}_{bk}^2) - \omega_k^2(\tilde{d}_{ak} - \tilde{d}_{bk})\tilde{q}_k \tag{13}$$

$\varepsilon_a - \varepsilon_b$ is the adiabatic excitation energy between states $a$ and $b$. The gradient of the energy gap $\nabla_{\tilde{\boldsymbol{q}}}\Delta E_{ab}(\tilde{\boldsymbol{q}})$ reads

$$\nabla_{\tilde{\boldsymbol{q}}}\Delta E_{ab}(\tilde{\boldsymbol{q}}) = \left[\nabla_{\boldsymbol{Q}} E_a(\boldsymbol{Q}) - \nabla_{\boldsymbol{Q}} E_b(\boldsymbol{Q})\right] \nabla_{\tilde{\boldsymbol{q}}}\boldsymbol{Q} = \mathbb{P}^\dagger \mathbb{M}^{-\frac{1}{2}} \left(\boldsymbol{f}_a - \boldsymbol{f}_b\right) \tag{14}$$

where $\mathbb{P}$ is a matrix whose columns are the normal modes $\tilde{\boldsymbol{q}}$ expressed in normalized mass-weighted Cartesian coordinates $\boldsymbol{Q}$, $\mathbb{M}$ is a diagonal matrix with the nuclear masses and $\boldsymbol{f}_{a/b} = \nabla_{\boldsymbol{Q}} E_{a/b}(\boldsymbol{Q})$ are the Cartesian energy gradient in electronic state $a$ and $b$. Taking the derivative with respect to $\tilde{\boldsymbol{q}}$ in eq. 13 and inserting the expression derived in eq. (14) gives

$$\tilde{\boldsymbol{d}}_{\boldsymbol{a}} = -\mathbb{W}^{-2}\mathbb{P}^\dagger \mathbb{M}^{-\frac{1}{2}} \left(\boldsymbol{f}_a - \boldsymbol{f}_b\right) + \tilde{\boldsymbol{d}}_b \tag{15}$$

where $\tilde{\boldsymbol{d}}_a$ and $\tilde{\boldsymbol{d}}_b$ are $k$-dimensional arrays of the displacements along all modes on states $a$ and $b$.

[7] Formulations for the linear response beyond this approximation, which incorporate so called Dushinsky rotation matrices, exist.[41]

The state-specific displacements along each normal mode $k$, i.e. $\tilde{d}_{ak}$, are directly related to spectroscopic observables like the dimensionless Huang-Rhys factors $S_{ak}$ or the reorganization energy $\lambda_{ak}$ through the following relations:

$$S_{ak} = \frac{\tilde{d}_{ak}^2 \omega_k}{2\hbar} \qquad \text{and} \qquad \lambda_{ak} = \frac{\tilde{d}_{ak}^2 \omega_k^2}{2\hbar}. \tag{16}$$

and are processed by the interface script to formulate the spectral density functions $J_{ab}(\omega)$ which is a convenient way to describe the coupling of an electronic state to a continuum of vibrational modes[11,20]

$$J_{ab}(\omega) = \pi \sum_k \tilde{d}_{ak} \tilde{d}_{bk} \omega_k^3 \; [\delta(\omega - \omega_k)] \tag{17}$$

The spectral density is the central quantity in the formulation of the line shape functions $g_{ab}(t)$[11,20]

$$g_{ab}(t) = -\frac{1}{2\pi} \int_0^{+\infty} d\omega \frac{J_{ab}(\omega)}{\omega^2} \left[ \coth\left(\frac{\beta\omega}{2}\right)(1 - \cos(\omega t)) + i(\sin(\omega t) - \omega t) \right] \tag{18}$$

that replaces the phenomenological broadening in the linear and nonlinear response. $g_{ab}(t)$ carries the complete information about the time-dependent coupling to the vibrational DOF including finite temperature effects. The precise form of the response function is derived via the cumulant expansion of Gaussian fluctuations (CGF)[11]. For linear techniques, the first-order response function becomes

$$R^{(1)}(t_1) = \left(\frac{i}{\hbar}\right) \theta(t_1) \sum_e |\boldsymbol{\mu}_{eg}|^2 \exp\left\{-i\left(\omega_{ge} - i\tau_{ge}^{-1}\right) - 2g_{ee}(t_1)\right\} + c.c. \tag{19}$$

$\tau_{eg}^{-1}$ here indicates the rate of coherence decay due to finite lifetime ($\tau_{eg}$) associated with internal conversion, energy or charge transfer (thus, representing part of the phenomenological broadening $\gamma_{ab}$) which causes a homogeneous broadening of the spectral lines.

For the third-order response function we again consider the same diagram shown in eq. (11), which now becomes

$$\begin{aligned} R_{SE}^{(\boldsymbol{K}_{II})}(t_1, t_2, t_3) = \left(\frac{i}{\hbar}\right)^3 \theta(t_1)\theta(t_2)\theta(t_3) \sum_{ee'} |\boldsymbol{\mu}_{eg}|^2 |\boldsymbol{\mu}_{e'g}|^2 \times \\ \times \exp\left\{-i\left(\omega_{ge} - i\tau_{ge}^{-1}\right) t_1 - i\left(\omega_{e'e} - i\tau_{e'e}^{-1}\right) t_2 - i\left(\omega_{e'g} - i\tau_{e'g}^{-1}\right) + \varphi_{ee'}^{SE}(t_1, t_2, t_3)\right\} \end{aligned} \tag{20}$$

and $\varphi^{SE}_{ee'}(t_1, t_2, t_3)$ is a multipoint line-shape function given by a specific combinations of $g_{ab}(t)$ evaluated at different times[11,20]

$$\begin{aligned}\varphi^{SE}_{ee'}(t_1, t_2, t_3) = &- g_{e'e'}(-t_2) - g_{ee}(t_1 + t_2 + t_3) - g_{ee'}(t_1) \\ &+ g_{ee'}(-t_2 - t_3) + g_{ee'}(t_1 + t_2) - g_{ee'}(-t_3)\end{aligned} \tag{21}$$

Similar expressions for the multipoint line-shape functions which enter in the remaining third-order diagrams are readily implemented in Spectron.[20]

The coupling to the environment (e.g. a protein cage, a solvent, a nearby molecule, etc.) may also be included in the form of spectral densities, and treated within the CGF framework presented above. While it can be computed from first principles based on molecular dynamics simulations[42,43], the motion of the environment can be also accounted for via model functions, as it is usually characterized by smooth spectral densities with significant contributions mainly in the low frequency regime. For example, the overdamped brownian oscillator (OBO) spectral density[44,45] available in Spectron has the form

$$J_{ab}(\omega) = 2\lambda_b \frac{\omega \Lambda_b}{\omega^2 + \Lambda_b^2} \tag{22}$$

and its parameters (the cutoff frequency $\Lambda_b$, which sets the time scale of the memory of the system-bath correlation, and system-bath coupling strength $\lambda_b$, which encode the degree of homogeneous signal broadening) can be fitted to reproduce the experimental line shapes. In general the spectral densities describing the coupling to the environment are state-specific, e.g. the energies of ionic state are more susceptible, i.e. strongly coupled, to the environment fluctuations.[42]. Nevertheless, to a first approximation a single OBO spectral density is commonly used for all states.

The decomposition of the spectral density into intra-chromophore and environment contributions is a reasonable approximation as energy fluctuations caused by inter- and intra-molecular vibrations (i.e. by short- and long- range interactions) can be considered uncorrelated.

### Beyond the adiabatic approximation: incorporating internal conversion, charge and energy transfer

The CGF approach is exact for fluctuations with Gaussian statistics and in the absence of electronic relaxation. Approximate expressions may be introduced to incorporate internal-conversion or energy transfer between different states. In the limit of decoupled evolution of the populations and coherence terms in the density matrix SE and ESA diagrams are separated in population (i.e. $e = e'$ during $t_2$ and coherence (i.e. $e \neq e'$ during $t_2$) terms. Population relaxation during $t_2$ is described by the Pauli master equation[20]:

$$\dot{\rho}_{ee}(t) = -\sum_{e'} K_{e'e',ee}\rho_{e'e'}(t) \tag{23}$$

with $\boldsymbol{K}$ the rate matrix with elements $K_{e'e',ee}$ depicting the population transfer rate from state $e$ into state $e'$. The solution of this equation is formally recast by the population Green's function $\rho_{e'e'}(t) = \sum_{e} \boldsymbol{G}_{e'e',ee}(t)\rho_{ee}(0)$ and the elements of the matrix $\boldsymbol{G}$ act as time-dependent weighting factors in the population pathways describing SE and ESA. For example, the population SE diagram of the $k_{II}$ signal becomes

$$\begin{aligned} R^{(\boldsymbol{K}_{II})}_{SE,pop}(t_1,t_2,t_3) = \left(\frac{i}{\hbar}\right)^3 \theta(t_1)\theta(t_2)\theta(t_3) \sum_{ee'} |\boldsymbol{\mu}_{eg}|^2 |\boldsymbol{\mu}_{e'g}|^2 \boldsymbol{G}_{e'e',ee}(t_2) \times \\ \times \exp\left\{-i\left(\omega_{ge} - i\tau_{ge}^{-1}\right)t_1 - i\left(\omega_{e'g} - i\tau_{e'g}^{-1}\right) + \varphi^{SE}_{ee'}(t_1,t_2,t_3)\right\} \end{aligned} \tag{24}$$

whereas the coherence SE diagram retains its from from eq. (20).

The rates $\tau_{ab}^{-1}$ entering the rate matrix $K_{e'e',ee}$ can be obtained from ab-initio computations that treat the coupling between different states (such as quantum-dynamics, semi-classical approaches equipped with surface hopping, Redfield/Förster transport methods in the context of excitons, etc.)[46] or may be fitted from available experimental information.[47] Finally, we note that Spectron offers variations of the multipoint line shape function $\varphi(t_1,t_2,t_3)$ depending on the relative time scales of transport, bath fluctuations and coherence dephasing in the system[20]. The different formulations are all expressed in terms of line shape functions $g_{ab}(t)$.

### Beyond impulsive excitations: incorporating pulse envelopes

So far we have shown how to compute the response function for linear and nonlinear (third-order) techniques at various levels of sophistication. The response function coincides with the signal itself in the limit of infinitely short laser pulses (known as the "impulsive limit"). Within this limit, the pulses are solely characterized by their wave-vector and mutual time delays.

To include the realistic pulse envelopes, the response function has to be convoluted with the fields, as prescribed by eqs. (2) and (4). The effect of the fields, which is entirely captured by the multiple integrals of these equations, can be accounted for in an approximate way by: (a) weighting the data along the frequency axes with the profile of the field (in the frequency domain) (b) convoluting the data along the time ($t_2$) to describe the effect of the finite time resolution that a realistic pulse (with finite duration) necessarily introduces. Exact equations that include the finite pulse-width effect when assuming temporally separated pulses and in absence of population transfer, are reported in Ref. [20].

Additionally, each pulse may have its own polarization: the polarization configuration of the pulse sequence weights differently the various contributions and affect the final spectra. The weights are given by the product of the field polarization $\hat{\boldsymbol{\nu}}_i$ ($i \in x, y, z$) and the transition dipole moment $\boldsymbol{\mu}_{ab}$, i.e. $\hat{\boldsymbol{\nu}}_i \cdot \boldsymbol{\mu}_{ab}$. A posteriori linear combinations of signals obtained with different pulse polarization configurations can help highlight weak features like cross-peaks which are often masked by the more intense but less interesting diagonal features in the spectra.[48,49]

## Interface Workflow(s)

iSPECTRON is a command-line interface written in Python 3 designed for Unix-based platforms[8]. It is composed of two parts: part (a) is designed to parse QC outputs and some parameters specified by the user to produce the necessary input files for the Spectron code; part (b) parses the Spectron output data and processes and manipulates the simulated spectra. The discussion that follows is thus developed in two sub-sections, one for each part of

[8]The code requires the NUMPY and SCIPY packages.

the interface.

### Interface - part a

The interface reads in inputs and outputs of the QM codes employed for the QM calculation to retrieve the following information: transition energies, transition dipole moments and vertical gradients for all states of interest (i.e. states of the $e-$manifold in case of linear techniques, states of the $e-$ and $f-$manifolds for nonlinear techniques). Alternatively, this information can be provided manually by the user (via communication files with well-defined format, see Supporting Information). The QM inputs and outputs can constitute a homogeneous set (i.e. all obtained with the same QM code) or coming from different sources (as, e.g., QM/MM calculations with COBRAMM). The interface interprets the provided inputs/outputs and tags them according to the software of origin and the type of data inside the file. In addition to the system-specific data, the user needs to specify several parameters for the 'virtual experiment" such as the energy window, the type of signal of interest (linear or transient absorption, $\boldsymbol{K}_I - \boldsymbol{K}_{III}$, quasi-absorptive) and the simulation time $t_2$. Optional parameters relevant to the system under scrutiny (e.g. phenomenological broadening or OBO parameters), the utilized pulses (e.g. polarization) and advanced simulation details (e.g., the selection of specific diagrams) can be also provided.

After all files have been collected and identified by the interface, the resulting input is analyzed for consistency. If there are no conflicts the code re-elaborates the data by ordering the manifold of energies and transition dipole moments and, if requested, computes spectral densities. Otherwise, if the inconsistencies are beyond remedy (e.g. no energies or transition dipoles are found or no transitions are present in the selected energy window) the code exits with an appropriate error message.

The code then prepares various intermediate files to be processed by Spectron. The minimal required information to run a basic Spectron simulation (in the SOS representation utilizing phenomenological broadening constants $\gamma_{ab}$ is as follows:

- The Spectron main input file, containing information about the chosen spectroscopic technique, the energy window, the number of states, etc.;

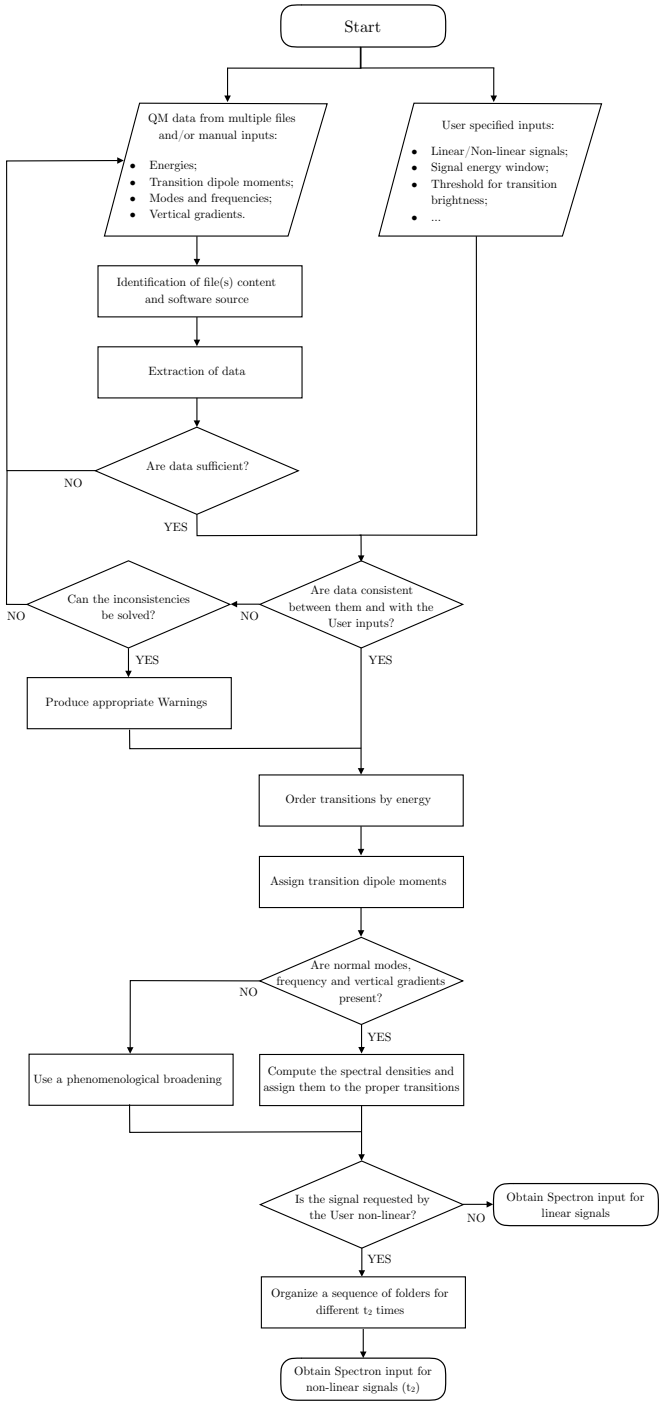



Figure 3: Workflow of the iSPECTRON program functioning for the collection and re-elaboration of quantum chemistry data

- The (single column) file with the list of transition energies relative to the GS, including the GS energy itself (zero by definition), in increasing order.

- The (five column) file with the transition dipoles; the first two indices represent initial and final states $a$ and $b$, while the last three numbers are the $x$, $y$ and $z$ components of the transition dipole $\boldsymbol{\mu}_{ab}$.

For simulations which go beyond the phenomenological broadening of the spectral lineshapes, making use of the CGF framework of Spectron, a few additional files are also created:

- The (single column) file containing, in consecutive order, spectral densities $J_{ab}(\omega)$ associated with different states;
- The coupling file, that defines the coupling between transitions and spectral densities (a sequence of lower triangular $N_{states} \times N_{states}$ matrices, one for each spectral density, in which the entry $ab$ is the coupling strength to the spectral density $J_{ab}(\omega)$);
- The transport rates file (square $N_{states} \times N_{states}$ matrix) which contains the rate matrix $\boldsymbol{K}$ (23) with the inverse of each state lifetime on the diagonal and the energy transfer rates between the states off-diagonal;

Finally, two additional files, named INFO.txt and VIB_INFO.txt, are also produced: the former contains the command line utilized by the user to launch the interface; the latter, summarizes the states reorganization energy $\lambda$ (eq. (16)). The workflow of the first part of the interface is reported in Scheme 3.

### Interface - part b

Once Spectron has finished simulating the linear or nonlinear signal, its output can be processed by the second leg of interface code.

In the case of linear techniques, the output provided by Spectron is a single two-column file (energy vs. intensity). For nonlinear techniques, instead, the output is organized in a directory tree structure:

- **top level**, named according to the simulated signal (e.g. $KI$). If a folder with the same name is already present (as in case of repeating a simulation with different parameters) an integer is attached to the tag (e.g. $KI_1, KI_2$);

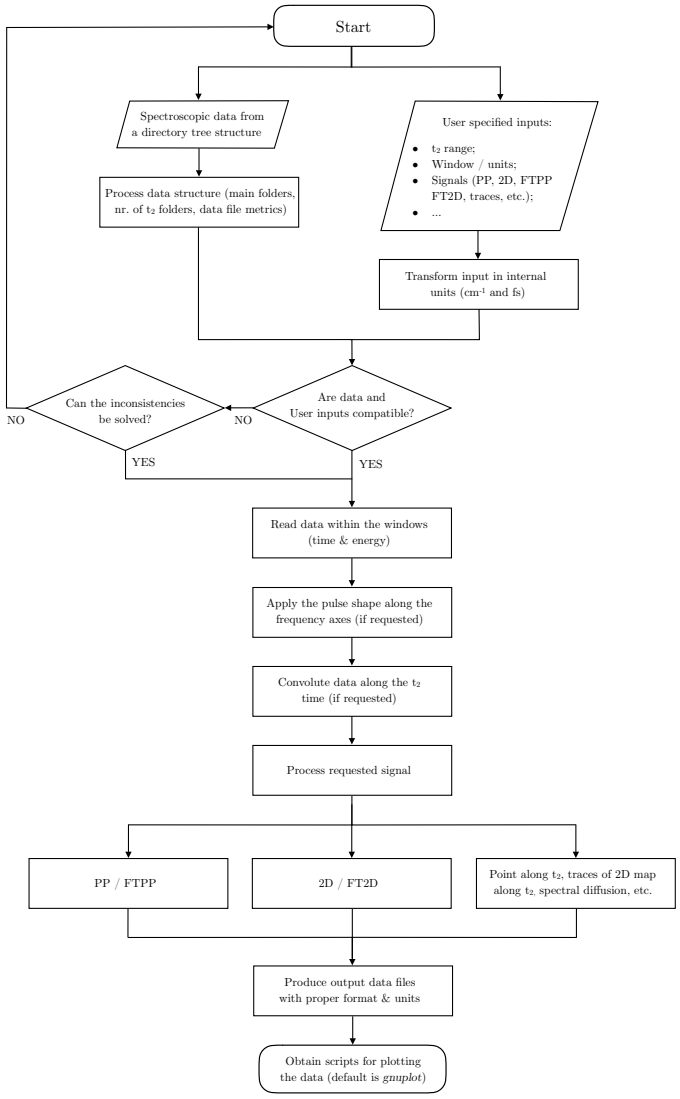



Figure 4: Workflow of the iSPECTRON program functioning for the production and analysis of a variety of spectra.

- **second level** (optional), with branches according to the diagrams, i.e. GSB, SE or ESA;
- **third level**, with branches associated with the waiting time $t_2$ (in case of rephasing, non-rephasing and quasi-absorptive signals) or $t_1$ or $t_3$ (in case of double quantum coherence), e.g. $t2_0$, $t2_4$ ... $t2_100$, in fs.

This allows to visualize and manipulate selectively one or multiple snapshots of individual diagrams constituting the overall signal.

In the post-processing the user can choose to further manipulate the 2D spectra by convoluting them with realistic pulses, by integrating them to obtain their marginal, the TA spectra (eq. (8)), or by Fourier transforming them along the $t_2$ time; one can request 1D and 2D spectral cuts or time traces and can adjust a number of additional optional parameters (e.g. the spectral window). Specifically, in the post-processing one has access to:

- One-dimensional transient absorption spectra (1DPP), that resolve the signal along a single frequency axis (the detection frequency $\Omega_3$). Single $t_2$ times snapshots and/or two-dimensional time/frequency heat-maps may be provided. The $t_2$ axis may be Fourier transformed, obtaining a frequency/frequency map that resolves the frequencies underneath the signal beating along $t_2$, that may be different at different detection frequencies. This new frequency axis is labeled as $\Omega_2$;

- Two-dimensional (2D) spectra, which resolve the signal along two frequency axes (excitation -$\Omega_1$- and detection -$\Omega_3$- frequency) are given as one or multiple $t_2$ snapshots (according to the user specifications). Maps at different $\Omega_2$ frequencies may also be obtained by Fourier transform of the data along $t_2$;

- Cuts of 2D maps along specific excitation or detection frequencies may be performed. The code was implemented to perform cuts along any specified couple of points in the $\Omega_1/\Omega_3$ plane (such as diagonal or anti-diagonal cuts);

- The ($t_2$) time trace of the value of a selected $\Omega_1/\Omega_3$ 2D map point (or the integrated value of a 2D map squared region) may be also extracted from the nonlinear data to visualize the beatings of the signals in different regions of the 2D map;

- Two approaches to extract the spectral diffusion of peaks of interest in a 2D map as a function of the waiting time $t_2$ are also implemented: i) evaluation of the slope of the line that connects the peak maxima along $\Omega_1$ at every fixed $\Omega_3$ value; ii) evaluation of the ratio between the widths of the diagonal and anti-diagonal cuts that goes through the center of the peak. In both cases, one expects to see how the observable approaches a value close to zero with $t_2$, a signature of the erasure of the system memory of the initial excitation energy, caused by the system-bath interaction.

Finally, output files with all requested features are created, together with *gnuplot* input files, with all the information necessary to plot these data.

The workflow of the second part of the iSPECTRON interface code is reported in Scheme 4. Note that this second part of the code is independent from the first one: indeed, the Spectron code can be used with other levels of theory implemented therein (excitons, quasi-particle approach, stochastic models for the system-bath coupling, etc.) and the third-order response, if organized in the folder structure presented above, may be processed by iSPECTRON.

iSPECTRON is available online at https://github.com/ispectrongit/iSPECTRON/ .

# EXAMPLES

In this section we demonstrate how the linear and nonlinear spectra generated with the iSPECTRON interface can be utilized to monitor the photo-induced vibrational dynamics and ultrafast internal conversion of pyrene. Pyrene is a widely used PAH owing to its interesting photophysical properties with a remarkably long excited state lifetime and high fluorescence yields. It presents well separated absorption bands with clear Franck-Condon progressions in the UV (at 320 nm and at 280 nm), ultrafast internal conversion to the lowest (dark) excited state and few characteristic ESA signatures, making it an excellent model case for assessing novel spectroscopic techniques.[5,50-53] The availability of high resolution experimental LA, PP and 2D maps naturally makes pyrene the system of choice to assess the capabilities of iSPECTRON.

First we present linear absorption of the *La* band - the lowest bright electronic excited state of pyrene absorbing in the near-UV - simulated at two levels of theory - RASSCF/RASPT2 (through OpenMolcas) and TDDFT (through NWChem) - and at two temperatures - at room temperature (i.e. 300 K) and at cryogenic temperature (i.e. 5 K). Subsequently, we report nonlinear spectra simulations at the RASSCF/RASPT2 level of theory. In all simulations coupling to the intra-molecular and environment vibrational DOF is described within the CGF framework (Section "The line shapes"). in its most general realization capable of treating bath fluctuations with arbitrary time scales and population

transfer[54]. The required spectral densities describing the coupling to the intra-molecular DOF are modeled based on the undamped DHO Hamiltonian utilizing the vertical gradient approach (eq. (15)). The spectral density describing the coupling to the environment DOF is obtained with the OBO model (eq. (22)) fitted to reproduce the width of the experimental spectra. The ultrafast $La \rightarrow Lb$ internal conversion is modeled on the basis of the rate eq. (23). Thereby, a lifetime $\tau_{eg}$ of 100 fs is assigned to the $La$ state, in agreement with experimental findings[53]. Energies, transition dipole moments, normal modes, frequencies, and gradients have been computed from first principles in vacuum[9]. $D_{2h}$ symmetry was enforced for the QC computations behind the nonlinear spectra. Additional information on the quantum chemical calculations, simulation parameters as well as the iSPECTRON commands used to generate the plots in Figures 5 and 6 are reported in the Supporting Information.

## Linear spectra

The first bright state of pyrene, dominated by a transition between the highest occupied orbital (HOMO) into the lowest unoccupied orbital (LUMO) and labeled $La$ following Platt's notation[60], shows a well resolved vibronic progression in the linear spectrum, with peaks at 334 nm, 318 nm and 302 nm with the fundamental accumulating the highest oscillator strength. Figure 5 documents the excellent agreement between theory and experiment for both aqueous solution room temperature data set[58], and Ne-matrix spectrum at 5 K[59]. The fine structure of the spectrum arises due to coupling to intra-molecular symmetry-conserving vibrational modes with frequencies $\sim$ 400 cm$^{-1}$ (ring breathing), $\sim$ 1100 cm$^{-1}$ and $\sim$ 1250 cm$^{-1}$ (in-plane bendings), $\sim$ 1450 cm$^{-1}$ and $\sim$ 1700 cm$^{-1}$ (carbon-carbon stretching, see Table 2 in the Supporting Information). Similar results have been reported by Freidzon et al. for LA simulations in water[56]. The 400 cm$^{-1}$ mode, clearly seen at 5K (black spectrum in fig. 5, right) and accurately reproduced by both at RASSCF/RASPT2 and TDDFT levels of theory, is responsible for the shoulder in the fundamental of the LA band at room temperature. Regarding the high-frequency region, RASSCF/RASPT2 predicts the

[9]The solvent has little effect on the line shapes, as different solvents: water, ethanol, polycrystalline powder, acetonitrile and benzene provide similar Raman spectra[55] which more over are in excellent agreement with gas-phase simulations.[56,57]

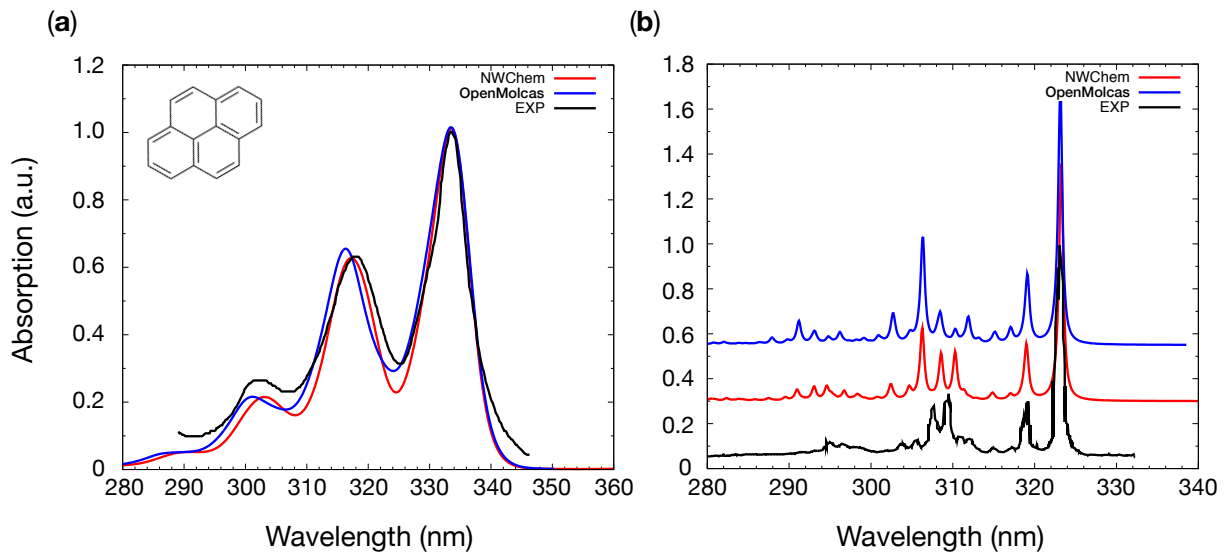



Figure 5: Comparison of the experimental (black line) and theoretical (TDDFT -red line- and RASSCF/RASPT2 -blue line-) linear absorption spectrum of pyrene recorded in (a) aqueous solution at room temperature[58] and in (b) a Ne-matrix at 5 K.[59] The theoretical and experimental spectra are normalized and centered at the fundamental (0-0) transition. Simulated spectra at 5 K have been vertically shifted to facilitate the comparison between them and with the experiment. The structure of the pyrene molecule is also reported (top left of panel (a)).

1700 cm$^{-1}$ stretching mode to dominate the line shape, whereas TDDFT predicts equal contributions from the 1250 cm$^{-1}$, 1450 cm$^{-1}$ and 1700 cm$^{-1}$ modes.

The excitation energy for the *La* state has been shifted by 2550 cm$^{-1}$ and 280 cm$^{-1}$ at room temperature, and by 3540 cm$^{-1}$ and 1250 cm$^{-1}$ at 5 K, for RASSCF/RASPT2 and TDDFT, respectively. Note that for both levels of theory, the difference between the applied shifts at room and cryogenic temperature is about 1000 cm$^{-1}$: this should therefore be ascribed to a solvatochromic shift induced by the aqueous environment of the 300 K experiment.

## Nonlinear spectra

Figure 6 summarizes some of the possible spectra obtained form general 2D data computed by Spectron. In Figure 6 (a) we report the quasi-absorptive (i.e. summing the $\boldsymbol{K}_I$ and $\boldsymbol{K}_{II}$ signals 2DES heat maps at different $t_2$ times (0, 50, 100 and 150 fs)). In particular, we note the dynamic multi-peak pattern of the (overlapping) GSB and SE contributions (shown in red) arising due to the vibronic structure of the absorption band. Furthermore, we can recognize ESA signatures of the $La$ state, a weak one at 400 nm and an intense one at 610 nm. The 610 nm ESA has been reported experimentally.[53,61]. It is established that the $La$ state decays through a non-adiabatic process mediated by a conical intersection to a lower lying dark excited state, denoted $Lb$ in Platt's notation, on a 100 fs time scale. In the spectra this population transfer is reflected in: a) decrease of the intensity of the diagonal bleach signal due to the decay of the SE; b) gradual decay of the spectral fingerprints of the $La$ state; c) the delayed build up of the $Lb$ signatures at 370 nm and 480 nm, both of which have been reported experimentally.[50,51,53,61] These observations are quantified by the time traces (reported in Figure 6 (b)) generated by iSPECTRON at wavelengths coinciding with the maximum of each spectral signature, highlighted with a green (SE), red ($La$ ESA) and violet ($Lb$ ESA) arrow, respectively, in the maps.

Figure 6 (c) shows two snapshots from the spectral dynamics of the GSB peak, highlighting in particular the elongated shape at early times (50 fs) and the circular shape at longer times (3 ps). This circularization of the 2D peaks is explained in terms of a memory erasure over time of the system initial excitation energy due to its interaction with the bath and of the finite life-time of system vibrations. This can be quantified by measuring the so called *flattening* parameter, which is the difference between the diagonal and the anti-diagonal peak widths, divided by the former. The magnitude of the flattening parameter is larger when the line shape significantly deviates from a circle, while it is zero when the two widths equate. The evolution of the flattening parameter is reported in Figure 6 (d), evidencing its progressive decay. The beating pattern is related to the long-living 400 $cm^{-1}$ intra-molecular mode activated in the $La$ state in the GSB process.

Broadband (300-650 nm) PP heat maps for the first 500 fs with both infinite resolution

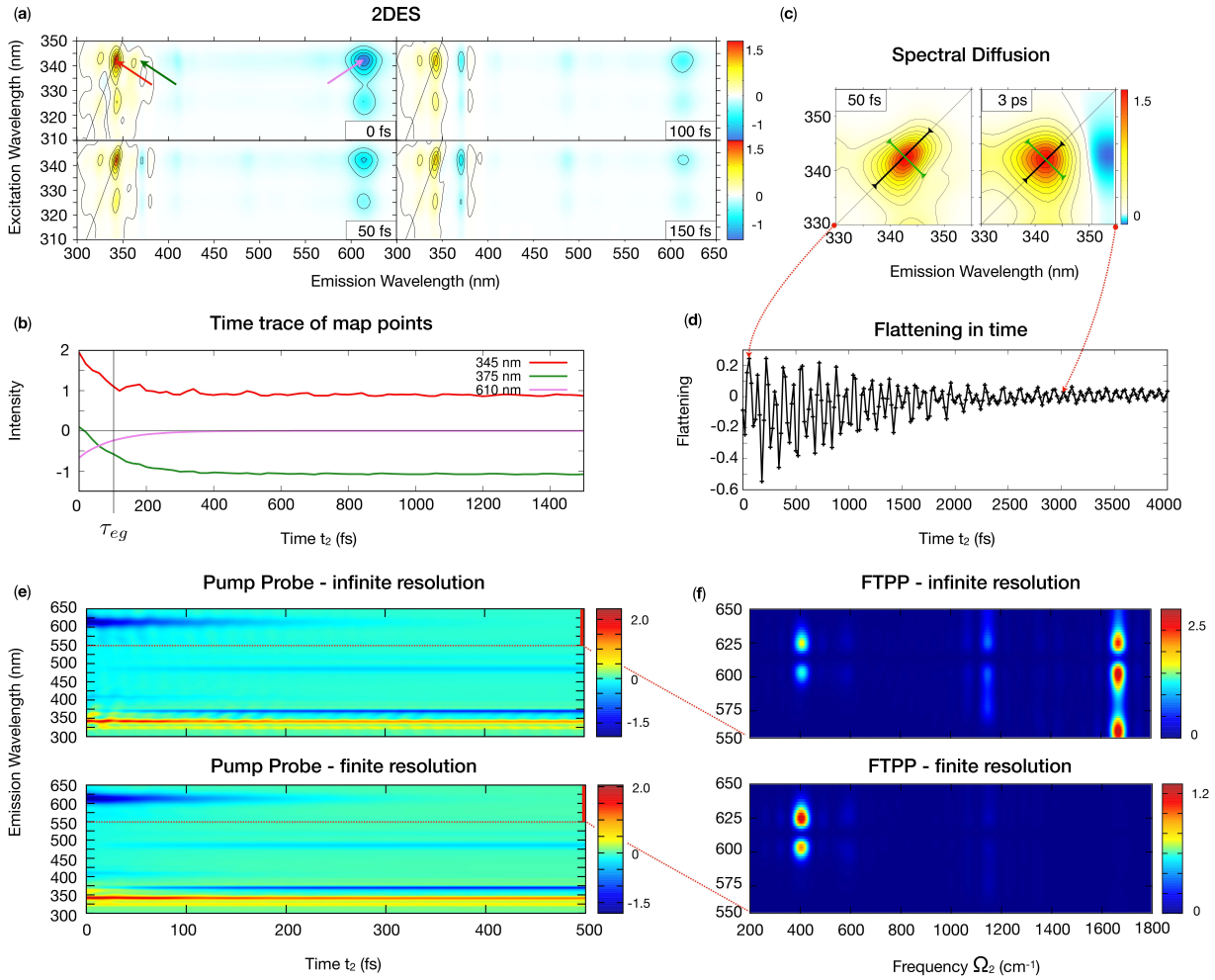



Figure 6: Various spectra and signal analysis obtained through iSPECTRON by processing 2DES simulation data produced by Spectron based on quantum chemistry computations performed at the CASPT2 level with OpenMolcas). (a) 2D maps at different $t_2$ times; positive peaks represent GSB and SE contributions, while negative peaks represent ESA contributions. The evolution of the various peaks due to the internal conversion between $La$ and $Lb$ states is made apparent by the time traces of selected points of the maps (indicated with arrows in (a)). (c) and (d) show the spectral diffusion and the evolution of the flattening parameter along $t_2$, for the main positive peak in the maps. (e) and (f) report PP heat maps (simulated from 0 to 500 fs) and the Fourier transform (along $t_2$) of the region between 550-650 nm

("impulsive limit") and finite resolution (employing pulses with realistic duration of $\sim 18$ fs) are reported in Figure 6 (e), together with a Fourier transform of the residual ESA in the 550-650 nm window (Figure 6 (d)) after subtracting the decaying part of the signal (due to the $La \rightarrow Lb$ population transfer). The observed intensity beating, giving rise to the peaks in the Fourier transform spectra, is to be ascribed to the $La$ ESA and as such it carries the information of the coherent wave-packet dynamics on the $La$ potential energy surface initiated by the pump pulse. The effect of introducing finite time-resolution is to wash out the high frequency features in the spectral dynamics of the ESA thereby uncovering low frequency modes, accentuating the necessity of ultra short pulses (preferably with a sub-10 fs duration) in order to thoroughly unravel the vibrational dynamics of photo-excited organic molecules[52].

# SUMMARY AND FUTURE DEVELOPMENTS

The simulation of linear and nonlinear spectra is important to support the interpretation of experimental data and identify the microscopic origin of macroscopic spectral features. iSPECTRON was developed to facilitate a) the calculation of spectra via the Spectron code via the automatised extraction of QM properties form a variety of quantum chemistry software and b) the re-elaboration of these data to produce and analyze different spectra (1DPP, 2DES, etc) from the Spectron third-order response output.

We have demonstrated the capabilities of the developed code by simulating linear and nonlinear spectra for the pyrene molecule at different levels of theory, and with different QC software.

The future development of iSPECTRON aims at expanding the readout capabilities to more routines of the presently interfaced codes (i.e. GASSCF/GASPT2, DMRG, etc.) and towards the integration of more QM codes. A second future development direction involves the calculation of the OBO parameters from first principles by integrating the calculation and evaluation of energy auto-correlation functions on top of MM simulations of the environmental DOF[42]. Finally, a third development direction aims at expanding the range of systems that can be treated through the interface. In this regard Spectron offers many more

appealing features. One of such features is the calculation of spectral densities from stochastic data (e.g. energy fluctuations along a given trajectory). Coupling molecular dynamics codes (such as COBRAMM for example) to this feature through iSPECTRON would allow one to describe vibrational dynamics beyond the DHO and CGF approximations. Another Spectron feature of interest, simulations in the exciton picture, would open the door to the spectroscopy simulation of systems which are currently beyond the capabilities of QM software.

## ACKNOWLEDGEMENTS

F.S., A.N., D.R.N., N.G., S.M, M.G. acknowledge support from the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division under Award Nos. DE-SC0019484, KC-030103172684. The Spectron code was developed with support from the National Science Foundation (Grant CHE-1953045). This research benefited from computational resources provided by EMSL, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at PNNL. PNNL is operated by Battelle Memorial Institute for the United States Department of Energy under DOE Contract No. DE-AC05-76RL1830.

## Supporting Information

The following sections are available in the Supporting Information: (1) Short review of the state-of-the-art of some electronic structure methods employed in this work; (2) Generalization of response function equations for arbitrary level schemes; (3) Details and format about auxiliary communication files that iSPECTRON is able to handle; (4) Quantum chemistry data that have been employed for linear and (5) nonlinear spectroscopy; (6) Input files for nonlinear spectroscopy; (7) iSPECTRON command line input examples. This information is available free of charge via the Internet at http://pubs.acs.org.

# References


1. A. H. Zewail and J. M. Thomas, *4D Electron Microscopy* (IMPERIAL COLLEGE PRESS, 2009), URL https://doi.org/10.1142/p641.

2. P. C. Chen, Applied Spectroscopy **70**, 1937 (2016), URL https://doi.org/10.1177/0003702816669730.

3. C. L. Smallwood and S. T. Cundiff, Laser & Photonics Reviews **12**, 1800171 (2018), URL https://doi.org/10.1002/lpor.201800171.

4. T. A. A. Oliver, Royal Society Open Science **5**, 171425 (2018), URL https://doi.org/10.1098/rsos.171425.

5. R. Borrego-Varillas, A. Nenov, L. Ganzer, A. Oriana, C. Manzoni, A. Tolomelli, I. Rivalta, S. Mukamel, M. Garavelli, and G. Cerullo, Chemical Science **10**, 9907 (2019), URL https://doi.org/10.1039/c9sc03871j.

6. A. Gelzinis, R. Augulis, V. Butkus, B. Robert, and L. Valkunas, Biochimica et Biophysica Acta (BBA) - Bioenergetics **1860**, 271 (2019), URL https://doi.org/10.1016/j.bbabio.2018.12.006.

7. Y. Song, A. Konar, R. Sechrist, V. P. Roy, R. Duan, J. Dziurgot, V. Policht, Y. A. Matutes, K. J. Kubarych, and J. P. Ogilvie, Review of Scientific Instruments **90**, 013108 (2019), URL https://doi.org/10.1063/1.5055244.

8. S. Pandey, R. Bean, T. Sato, I. Poudyal, J. Bielecki, J. C. Villarreal, O. Yefanov, V. Mariani, T. A. White, C. Kupitz, et al., Nature Methods **17**, 73 (2019), URL https://doi.org/10.1038/s41592-019-0628-z.

9. H. J. B. Marroux, A. P. Fidler, D. M. Neumark, and S. R. Leone, Science Advances **4**, eaau3783 (2018), URL https://doi.org/10.1126/sciadv.aau3783.

10. A. Wituschek, L. Bruder, E. Allaria, U. Bangert, M. Binz, R. Borghes, C. Callegari, G. Cerullo, P. Cinquegrana, L. Giannessi, et al., Nature Communications **11** (2020), URL https://doi.org/10.1038/s41467-020-14721-2.

11. S. Mukamel, *Principles of nonlinear optical spectroscopy*, vol. 523 (Oxford Univ. Press, 1995), ISBN 0195092783, URL http://www.gbv.de/dms/goettingen/152111840.pdf.

12. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, et al., *GaussianËIJ16 Revision C.01* (2016), gaussian Inc. Wallingford CT.

13. I. F. Galván, M. Vacher, A. Alavi, C. Angeli, F. Aquilante, J. Autschbach, J. J. Bao, S. I. Bokarev, N. A. Bogdanov, R. K. Carlson, et al., Journal of Chemical Theory and Computation **15**, 5925 (2019), URL https://doi.org/10.1021/acs.jctc.9b00532.

14. F. Aquilante, J. Autschbach, A. Baiardi, S. Battaglia, V. A. Borin, L. F. Chibotaru, I. Conti, L. D. Vico, M. Delcey, I. F. Galván, et al., The Journal of Chemical Physics **152**, 214117 (2020), URL https://doi.org/10.1063/5.0004835.

15. M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. Van Dam, D. Wang, J. Nieplocha, E. Aprà, T. L. Windus, et al., Computer Physics Communications **181**, 1477 (2010).

16. E. Aprà, E. J. Bylaska, W. A. De Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. van Dam, Y. Alexeev, J. Anchell, et al., The Journal of chemical physics **152**, 184102 (2020).

17. G. M. J. Barca, C. Bertoni, L. Carrington, D. Datta, N. De Silva, J. E. Deustua, D. G. Fedorov, J. R. Gour, A. O. Gunina, E. Guidez, et al., The Journal of Chemical Physics **152**, 154102 (2020), ISSN 0021-9606, 1089-7690, URL http://aip.scitation.org/doi/10.1063/5.0005188.

18. F. Santoro, R. Improta, A. Lami, J. Bloino, and V. Barone, The Journal of Chemical Physics **126**, 084509 (2007), URL https://doi.org/10.1063/1.2437197.

19. S. Jurinovich, L. Cupellini, C. A. Guido, and B. Mennucci, Journal of Computational Chemistry **39**, 279 (2017), URL https://doi.org/10.1002/jcc.25118.

20. D. Abramavicius, B. Palmieri, D. V. Voronine, F. Šanda, and S. Mukamel, Chemical Reviews **109**, 2350 (2009), pMID: 19432416, https://doi.org/10.1021/cr800268n, URL https://doi.org/10.1021/cr800268n.

21. W. Zhuang, T. Hayashi, and S. Mukamel, Angewandte Chemie International Edition **48**, 3750 (2009), URL https://doi.org/10.1002/anie.200802644.

22. B. O. Roos, *Ab Initio Methods in Quantum Chemistry: Part II* (Wiley, Chicester, UK, 1987).

23. B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, Chem. Phys. **48**, 157 (1980).

24. J. Olsen, B. O. Roos, P. Jo/rgensen, and H. J. A. Jensen, The Journal of Chemical Physics **89**, 2185 (1988), URL https://doi.org/10.1063/1.455063.

25. P. A. Malmqvist, A. Rendell, and B. O. Roos, The Journal of Physical Chemistry **94**, 5477 (1990), URL https://doi.org/10.1021/j100377a011.

26. K. Andersson, P. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, J. Phys. Chem. **94**, 5483 (1990).

27. D. Roca-Sanjuán, F. Aquilante, and R. Lindh, Wiley Interdisciplinary Reviews: Computational Molecular Science **2**, 585 (2011), URL https://doi.org/10.1002/wcms.97.

28. P. Verma and D. G. Truhlar, Trends in Chemistry **2**, 302 (2020), URL https://doi.org/10.1016/j.trechm.2020.02.005.

29. S. A. Fischer, C. J. Cramer, and N. Govind, Journal of Chemical Theory and Computation **11**, 4294 (2015), URL https://doi.org/10.1021/acs.jctc.5b00473.

30. D. R. Nascimento, E. Biasin, B. Poulter, M. Khalil, D. Sokaras, and N. Govind (2020), arXiv: 2010.03092 (physics.chem-ph), 2010.03092.

31. S. Ghosh, P. Verma, C. J. Cramer, L. Gagliardi, and D. G. Truhlar, Chemical Reviews **118**, 7249 (2018), URL https://doi.org/10.1021/acs.chemrev.8b00193.

32. H. Lischka, D. Nachtigallová, A. J. A. Aquino, P. G. Szalay, F. Plasser, F. B. C. Machado, and M. Barbatti, Chemical Reviews **118**, 7293 (2018), URL https://doi.org/10.1021/acs.chemrev.8b00244.

33. I. Conti, G. Cerullo, A. Nenov, and M. Garavelli, Journal of the American Chemical Society **142**, 16117 (2020), URL https://doi.org/10.1021/jacs.0c04952.

34. S. Mai and L. González, Angewandte Chemie International Edition **59**, 16832 (2020), URL https://doi.org/10.1002/anie.201916381.

35. F. Lipparini and B. Mennucci, The Journal of Chemical Physics **144**, 160901 (2016), URL https://doi.org/10.1063/1.4947236.

36. L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, et al., Chemical Reviews **115**, 5678 (2015), URL https://doi.org/10.1021/cr5004419.

37. O. Weingart, A. Nenov, P. Altoè, I. Rivalta, J. Segarra-Martí, I. Dokukina, and M. Garavelli, Journal of Molecular Modeling **24** (2018), URL https://doi.org/10.1007/s00894-018-3769-6.

38. R. Salomon-Ferrer, D. A. Case, and R. C. Walker, Wiley Interdisciplinary Reviews: Computational Molecular Science **3**, 198 (2013).

39. J. Segarra-Martí, S. Mukamel, M. Garavelli, A. Nenov, and I. Rivalta, in *Topics in Current Chemistry Collections* (Springer International Publishing, 2018), pp. 63–112, URL https://doi.org/10.1007/978-3-030-02478-9_3.

40. A. Nenov, S. Mukamel, M. Garavelli, and I. Rivalta, Journal of Chemical Theory and Computation **11**, 3755 (2015), URL https://doi.org/10.1021/acs.jctc.5b00443.

41. F. J. A. Ferrer, J. Cerezo, J. Soto, R. Improta, and F. Santoro, Computational and Theoretical Chemistry **1040-1041**, 328 (2014), URL https://doi.org/10.1016/j.comptc.2014.03.003.

42. J. Segarra-Martí, F. Segatta, T. A. Mackenzie, A. Nenov, I. Rivalta, M. J. Bearpark, and M. Garavelli, Faraday Discussions **221**, 219 (2020), URL https://doi.org/10.1039/c9fd00072k.

43. M. K. Lee, P. Huo, and D. F. Coker, Annual Review of Physical Chemistry **67**, 639 (2016), URL https://doi.org/10.1146/annurev-physchem-040215-112252.

44. W. B. Bosma, Y. J. Yan, and S. Mukamel, Physical Review A **42**, 6920 (1990), URL https://doi.org/10.1103/physreva.42.6920.

45. B. Li, A. E. Johnson, S. Mukamel, and A. B. Myers, Journal of the American Chemical Society **116**, 11039 (1994), URL https://doi.org/10.1021/ja00103a020.

46. R. Borrego-Varillas, D. C. Teles-Ferreira, A. Nenov, I. Conti, L. Ganzer, C. Manzoni, M. Garavelli, A. M. de Paula, and G. Cerullo, Journal of the American Chemical Society **140**, 16087 (2018), URL https://doi.org/10.1021/jacs.8b07057.

47. A. Nenov, R. Borrego-Varillas, A. Oriana, L. Ganzer, F. Segatta, I. Conti, J. Segarra-Marti, J. Omachi, M. Dapor, S. Taioli, et al., The Journal of Physical Chemistry Letters **9**, 1534 (2018), URL https://doi.org/10.1021/acs.jpclett.8b00152.

48. A. Nenov, I. Rivalta, S. Mukamel, and M. Garavelli, Computational and Theoretical Chemistry **1040-1041**, 295 (2014), URL https://doi.org/10.1016/j.comptc.2014.03.031.

49. A. Nenov, I. Rivalta, G. Cerullo, S. Mukamel, and M. Garavelli, The Journal of Physical Chemistry Letters **5**, 767 (2014), URL https://doi.org/10.1021/jz5002314.

50. N. Krebs, I. Pugliesi, J. Hauer, and E. Riedle, New Journal of Physics **15**, 085016 (2013), URL https://doi.org/10.1088/1367-2630/15/8/085016.

51. R. Borrego-Varillas, A. Oriana, L. Ganzer, A. Trifonov, I. Buchvarov, C. Manzoni, and G. Cerullo, Optics Express **24**, 28491 (2016), URL https://doi.org/10.1364/oe.24.028491.

52. A. Picchiotti, A. Nenov, A. Giussani, V. I. Prokhorenko, R. J. D. Miller, S. Mukamel, and M. Garavelli, The Journal of Physical Chemistry Letters **10**, 3481 (2019), URL https://doi.org/10.1021/acs.jpclett.9b01325.

53. R. Borrego-Varillas, L. Ganzer, G. Cerullo, and C. Manzoni, Applied Sciences **8**, 989 (2018), URL https://doi.org/10.3390/app8060989.

54. D. Abramavicius, L. Valkunas, and S. Mukamel, Europhysics Letters (EPL) **80**, 17005 (2007), URL https://doi.org/10.1209/0295-5075/80/17005.

55. H. Shinohara, Y. Yamakita, and K. Ohno, Journal of Molecular Structure **442**, 221 (1998), URL https://doi.org/10.1016/s0022-2860(97)00335-9.

56. A. Y. Freidzon, R. R. Valiev, and A. A. Berezhnoy, RSC Adv. **4**, 42054 (2014), URL https://doi.org/10.1039/c4ra05574h.

57. A. Nenov, A. Giussani, B. P. Fingerhut, I. Rivalta, E. Dumont, S. Mukamel, and M. Garavelli, Physical Chemistry Chemical Physics **17**, 30925 (2015), URL https://doi.org/10.1039/c5cp01167a.

58. F. P. Schwarz and S. P. Wasik, Analytical Chemistry **48**, 524 (1976), URL https://doi.org/10.1021/ac60367a046.

59. T. M. Halasinski, F. Salama, and L. J. Allamandola, The Astrophysical Journal **628**, 555 (2005), URL https://doi.org/10.1086/430631.

60. J. R. Platt, The Journal of Chemical Physics **17**, 484 (1949), URL https://doi.org/10.1063/1.1747293.

61. M. Raytchev, E. Pandurski, I. Buchvarov, C. Modrakowski, and T. Fiebig, The Journal of Physical Chemistry A **107**, 4592 (2003), URL https://doi.org/10.1021/jp027356c.