FROG: exploiting all-atom Molecular Dynamics trajectories to calculate linear and non-linear optical responses of molecular liquids within Dalton's QM/MM polarizable embedding scheme

Guillaume Le Breton, Oriane Bonhomme, Emmanuel Benichou, Claire Loison

FROG: exploiting all-atom Molecular Dynamics trajectories to calculate linear and non-linear optical responses of molecular liquids within Dalton's QM/MM polarizable embedding scheme.

Guillaume Le Breton,[1, a)] Oriane Bonhomme,[1] Emmanuel Benichou,[1] and Claire Loison[1]
University of Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-69622, Villeurbanne, France

(*claire.loison@univ-lyon1.fr)

(*guillaume_le_breton@live.fr)

(Dated: 12 April 2024)

Quantum mechanical/molecular mechanics (QM/MM) methods are interesting to model the impact of a complex environment on the spectroscopic properties of a molecule. In this context, `FROG` code is a tool to exploit molecular dynamics trajectories to perform QM/MM calculations of molecular optical properties. `FROG` stands for '**FRO**m molecular dynamics to second harmonic **G**eneration' since it was developed for the calculations of hyperpolarizabilities. These are relevant to model non-linear optical intensities, and compare them with those obtained from second harmonic scattering or second harmonic generation experiments. `FROG`'s specificity is that it is designed to study simple molecular liquids, including solvents or mixtures, from the bulk to the surface. For the QM/MM calculations, `FROG` relies on the `Dalton` package: its electronic-structure models, response theory and polarizable embedding (PE) schemes. `FROG` helps with the global workflow needed to deal with numerous QM/MM calculations: it permits the user to separate the system into QM and MM fragments, to write `Dalton`'s inputs, to manage the submission of QM/MM calculations, to check whether `Dalton`'s calculation finished successfully, and finally to perform averages on relevant QM observables. All molecules within the simulation box and several time steps are tackled within the same workflow. The platform is written in Python and installed as a package. Intermediate data such as local electric fields or individual molecular properties are accessible to the users in the form of Python object arrays. The resulting data are easily extracted, analyzed, and visualized using Python scripts that are provided in tutorials.

I. INTRODUCTION

Understanding the interplay between a molecule and its environment is one of the key objectives in physical chemistry, and a big challenge in theoretical modeling. Implicit solvent approaches can be used for many optical properties, but some cases require an explicit description of the solvent. For example, the non-linear optical properties of water strongly change between the gas and the liquid phases, and continuous implicit solvent models fail to describe this phenomenon[1–3]. Multiscale quantum-classical (QM/MM) approaches are methods of choice in this case. In particular, it has been used to model the non-linear response properties of molecular liquids[4,5], solubilized chromophores[6–8] and even more complex systems[9–13].

QM/MM schemes consider the system as a sum of two parts: one chemically relevant species or solute (QM fragment) and its environment or solvent, which is described using classical molecular mechanics models (MM fragment). In the specific domain of spectroscopy, they provide a framework for understanding environment (MM) polarization effects directly from electronic structure calculations of the QM fragment. The spectrum of QM/MM models has shown a wide expansion and diversification in the last couple of decades[14,15]. Among other QM software packages, `Dalton`[16,17] proposes a QM/MM implementation via the Polarizable Embedding scheme[18,19], and it is well established in the physical-chemistry community in the calculation of linear, quadratic, or cubic response properties[3,11–13].

If the observable strongly depends on the fragment configuration, the molecule conformation for instance, or on its local environment, as it is for the hyperpolarizability of water in the liquid phase[3,20], QM/MM approaches face the problem of configurational sampling. In a sequential scheme, a molecular dynamics trajectory is first performed to sample the configuration space. Then, the effect of the environment perturbation on the QM region of interest is evaluated for selected configurations at different time steps of the trajectory. Thanks to the development of computer capacities, it becomes possible to increase the number of configurations studied, or the size of the QM fragment. Yet, using a molecular trajectory to perform a QM/MM study over many different configurations using a given QM package is often technically challenging. One configuration (the fragment and its environment) represents one QM/MM calculation whose input file should be written, then run and finally analyzed to extract the relevant electronic property.

When dealing with non-linear spectroscopy of liquid, such as Second Harmonic Generation (SHG), additional care must be taken since the experimental intensity is not simply a time or spatial average over the response of a single molecular probe. The internal organization of the liquid may impact the intensity that is measured[21–23] via intermolecular correlations[24–28]. This sensitivity makes it possible to study the

a)Institut de Chimie Moléculaire de l'Université de Bourgogne (ICMUB, UMR-CNRS 6302), Université Bourgogne Franche-Comté, 21078 Dijon, France

structure of a liquid experimentally. But in the modeling, the non-linear intensity has to be reconstructed by taking into account both the individual molecule's hyperpolarizability values and the correlations between them[5,29]. As a consequence, the QM/MM calculations must be performed on several molecules within a single conformation of the system. Some fragmentation codes adapted for complex systems such as proteins have been published[30,31], but we felt the need for an open, user-friendly, and automated interface to tackle linear and non-linear optical properties of solvent or liquid mixtures.

In this article, we present our package FROG, which stands for "FROm molecular dynamics to second harmonic Generation". The motivation to develop FROG was to interpret the intensities measured during two experiences of non-linear optics: a setup to study bulk liquids using second harmonic scattering (SHS), and a setup to study liquid interfaces via surface second harmonic generation (S-SHG). FROG is a Python package started in 2019 at the Light and Matter Institute, France; its main goal is to compute the non-resonant molecular hyperpolarizability of molecules in solvents and at interfaces.. The text is organized as follows. Section II presents an overview of FROG: the motivation, the QM observables, and the global workflow. We compare it to similar fragmentation codes. Results on Second Harmonic properties of water in its liquid phase and at the liquid-gas interface obtained using FROG have been published elsewhere[5,32,33], but Section III briefly recalls some typical results from Ref. 32 to illustrate FROG's use. Other examples of applications are also listed with the tutorials.

Then, Section IV provides more details of FROG's implementation. It describes the different steps performed by FROG, its data structure, as well as the procedure to send and manage the numerous QM/MM calculations on a computing cluster. Finally, we briefly discuss the parallelization efficiency.

II. OVERVIEW

In this section, we first define the key assumptions to model the Second Harmonic (SH) response of liquids, and the molecular properties involved. Then, FROG's workflow is detailed, and a typical study performed with the use of FROG is briefly described. Finally, a quick detour presents other similar packages and FROG's specificity.

A. Quantum Mechanical Observables for Second Harmonic Generation of liquids

Second Harmonic Generation (SHG) is an even non-linear process where two photons at a frequency ω are converted into one photon at twice the frequency. The emitted photons can be detected and provide information on the structure of the liquid, from the bulk to the interface[27,28]. SHG is widely used to characterize equilibrium and dynamic physical–chemistry processes at interfaces with a high sensitivity[34]. But the interpretation of the SHG response of a liquid is not straightforward and has been the subject of numerous publications for decades[21,35]. For the special case of water - at a liquid interface or in bulk - numerical modelings have recently permitted to calculate various contributions to the signal in agreement with experimental data[5,36,37].

To model the experimental SH intensity generated by a liquid, our key hypothesis is that the emitted intensity can be deduced from the knowledge of the electric field emitted by each molecule within the condensed phase. Grossly, one can model the experimental signal in three steps: (i) define the electric perturbation felt by the molecules, (ii) predict the response of the molecules at the second harmonic frequency using a perturbative approach, (iii) compute the electromagnetic field emitted and measured experimentally at the second harmonic frequency. The first and third parts usually involve electromagnetism frameworks such as Fresnel coefficients or Clausius-Mossetti approximation and are not discussed here. In this article, we present how to perform the second part using FROG. For each molecule, the induced dipole are defined via a response tensor:

$$\boldsymbol{p}_i^{\omega} = \alpha_i(-\omega,\omega)\cdot \boldsymbol{e}_i^{\omega}, \quad \boldsymbol{p}_i^{2\omega} = \beta_i(-2\omega,\omega,\omega) : \boldsymbol{e}_i^{\omega}\boldsymbol{e}_i^{\omega}. \quad (1)$$

The linear order is given by: $\boldsymbol{p}_i^{\omega}$ the induced dipole moment of a molecule i, α_i its polarizability tensor and $\boldsymbol{e}_i^{\omega}$ the external field at the fundamental frequency. The second order is given by $\boldsymbol{p}_i^{2\omega}$ the induced dipole moment at the second harmonic frequency and β_i the hyperpolarizability tensor. Within the perturbative regime, $\boldsymbol{p}_i^{2\omega}$ is given by the tensorial product between the hyperpolarizability and two times the perturbative field $\boldsymbol{e}_i^{\omega}$, so that the hyperpolarizability matrix is a $3\times3\times3$ tensor. The previous molecular equations, expressed in the molecular frame, can be extended to a mesoscopic scale and in the laboratory frame through the use of surface susceptibilities, often noted $\chi^{(2)}$ for the second harmonic.

Instead of approaching the electronic degrees of freedom and the response properties of the whole liquid[36], one performs QM polarizability and hyperpolarizability calculations for single molecules[37]. This can be performed using the response theory implemented in Dalton[16]. Moreover, the condensed environment around the QM fragment can be described using the Polarizable Embedding (PE) scheme of Dalton[18,19]. The environment is described as a collection of sites, for instance at the neighbours' nuclei position, which can bear constant point charge, dipole or quadrupole moments. They impact the QM part by the static field they create (PE0). In addition, these sites can also be polarizable (PE1). In this case, they give rise to induced point dipoles as a linear response to the total field produced by the static multipoles in the environment and the QM electronic density. Thus, there is a mutual polarization between the QM and PE subsystems. The Hamiltonian acting on the electrons of the QM fragment $\hat{H}_{QM/MM}$ is the one of the vacuum phase $\hat{H}_0$ plus the spatial-dependent electrostatic potential generated by its surrounding $\hat{V}_{env}$ at the probe location. If there are polarizable sites, the MM electrostatic environment, and thus $\hat{V}_{env}$, depends on the electronic density as well. Once the ground state properties of the electrons are obtained for this environment-dependent $\hat{H}_{QM/MM}$, the frequency-dependent hyperpolarizability is predicted thanks to the response theory at the quadratic order,

using $\hat{H}_{QM/MM}$ for the unperturbed Hamiltonian. The PE scheme can also include the effects of the environment dynamics response through the polarizable sites to the molecular response[38]. This multi-step approach thus requires many parameters. Relevant choices for all these depend on the system, and are beyond the scope of the present article. The reader may refer to publications by the method developers on PE approaches[18,19], or to Ref. 39 that presents a pedagogical tutorial of the PE `Dalton` scheme. Different QM approaches and different partitionnings for various systems have also been compared[11,40–43]. We also present our recent works[32,33] based on `FROG` to compute the hyperpolarizabilities of water in Section III.

`FROG` was designed to help the user to perform such QM/MM calculations of the polarizability and hyperpolarizability tensors α_i and β_i for many molecules at dipolar and quadrupolar order, in the molecular frame and the laboratory one. Until now, `FROG` was used with an embedding containing solely point charges representing the charges of the corresponding MM model. But `FROG` was implemented so, that it is possible to attribute also multipoles, and polarizabilities to the different atoms of each molecule types.

B. `FROG`: a sequential QM/MM approach

In sequential QM/MM approaches, an ensemble of structures is obtained in a first step using a classical MD model, and the QM/MM is then applied in an independent, second step.

Within this frame, `FROG` allows the user to perform QM/MM response calculations starting from MD trajectories that were generated with other programs, such as `Tinker`, `Amber`, `openMM` or `Gromacs`. To perform the response calculations, `Dalton` is used since it can compute numerous response properties with versatile Polarizable Environment (PE) QM/MM schemes. Today, we plan to extend the accessibility of the functionalities and properties available in `Dalton` rather than make `FROG` compatible with other QM softwares. It requires three main steps (see Fig. 1): (I) generating the QM/MM `Dalton` input corresponding to the molecular configuration from the MD trajectory, (II) managing and performing the quantum calculations, and (III) reading the results and performing statistical analyses.

`FROG` is thought to be called as many times as needed using the same input file until the three steps are all performed. We recommend performing all these steps on the computing center where the `Dalton` calculations are done, to avoid data transfers.

The first step, (I), is performed within one single `FROG`'s call to prepare QM/MM `Dalton` input files, one for each selected molecule and each time frame.

For the second step, (II) `FROG` does not perform the QM/MM calculations directly: `Dalton` has to be run independently. However, `FROG` prepares the submission script needed to run the QM/MM calculations with `Dalton` on the cluster which shall be submitted by the user. The QM/MM calculations can be treated independently, leading to a robust and simple parallelization procedure. Once the submitted jobs are finished on the computing center, the user calls `FROG` again, with the same input as in the first step. `FROG` checks which QM/MM calculations have been performed successfully, and it creates submission scripts for the remaining QM/MM calculations to be performed. This second step is thus an iteration of `FROG`'s calls and HPC submission until all QM/MM calculations are performed.

If all QM/MM calculations are finished, the third step (III) can be performed: `FROG` is called once, it reads the results and performs the statistical analysis. The user can define groups of molecules over which the average must be done using several criteria.

C. (Hyper)polarizability analysis

`FROG`'s outcome, in particular molecular (hyper)polarizabilities, can be used to interpret various experiments such as static Electric Field Induced Second Harmonic generation (EFISH), dc Kerr effect, Surface Second Harmonic Generation, or Hyper Rayleigh Scattering (HRS)[44]. Typically, depending on the experimental setup, measurable quantities are different combinations of molecular or laboratory response tensors: α, β, and γ, the polarizability, first and second hyperpolarizability respectively.[45,46]. In particular, some care must be taken for the quantitative comparison of experimental and theoretical hyperpolarizabilities of molecules, and different contributions can be dominant depending on the system[3,46].

Concerning the average hyperpolarizability, it is first interesting to investigate which components are dominant, and what their ratios are. For example, this can permit the neglect of some hyperpolarizability components in the interpretation of surface-SHG[25]. Moreover, linear combinations of the hyperpolarizability tensors can - in some cases - be directly linked to the experimental depolarization ratio obtained by polarization-resolved second harmonic scattering of the liquid[46]. In other cases, like water, the intermolecular correlations of the second harmonic emitted electric fields have to be calculated explicitly[5]. Recently, the quadrupolar hyperpolarizability responses are also used to interpret surface SHG signal[37,47]. Thus, since the analysis depends on the experiment with which one wants to compare, the post-processing is not included in `FROG`, but delegated to the user. Some examples are provided in the Jupyter notebooks of the tutorials. Depending on the analysis, the user may prefer to collect the (hyper)polarizabilities in the molecular frame, or in the laboratory frame. Both are available in `FROG`.

D. Comparison with other fragmentation programs

Similar tools have been proposed previously to help users perform QM/MM calculations by fragmentation of a complex system.

Some are specific to the fragmentation of biomolecular systems. For example, the `AFNMR` package[30] provides a useful method to estimate chemical shifts in proteins, based on den-

sity functional calculations on fragments. The tool analyzes a PDB file of a protein (possibly with a ligand) and automatically generates fragments on which the QM calculations are performed using an embedding scheme. The entire fragment and a buffer region are treated by density functional theory (DFT). The surroundings of the QM region is described implicitly by a set of surface charges that are computed using the Poisson–Boltzmann equation. `AFNMR` generates inputs for the `GAUSSIAN09` package to calculate NMR observables: chemical shifts and/or scalar couplings.

The `PyFrame` package[31] that is contained in the `Dalton`'s Project[17] also performs automatized fragmentation and parameterization of complex molecular environments. `PyFrame`, additionally, helps to calculate the parameters necessary to create the polarizable embedding. This feature is not included in `FROG`, which will build the environment using charges and polarizabilities that are provided by the user for each molecule of the system. `FROG` also does not yet handle covalent bond fragmentation.

Another difference between `PyFrame` or `AFNMR`, and `FROG` is that we intended to study bulk liquids or interfaces, rather than a single or just a few probes in a complex environment. `FROG` was written to examine not only one molecule of interest per time step but all of them. With `FROG`, one can easily explore how the various conformations explored within an MD trajectory affect the optical properties of a molecule. Thanks to the `MDanalysis` python package, `FROG` can handle different formats of MD trajectories and is designed to deal with several time steps in a single job. We have tested input trajectories produced by `LAMMPS`, `NAMD`, and `Gromacs`. Moreover, it may also use a dynamical selection of molecules depending on their position relative to an interface (layer-based selection), for the various surface definitions implemented in the `pytim` package[48].

III. EXAMPLES

To illustrate a typical use of `FROG`, we briefly recall results published in Ref. 32, and we list the tutorials that are currently available. Other uses of `FROG` can be found in Ref. 5,33, where we notably manage to reproduce the experimental results of Second Harmonic Scattering on pure bulk water from the MD[5].

A. Hyperpolarizability variations of Water at the air/water interface

As mentioned already, for water molecules, the first hyperpolarizability is known to be extremely sensitive to their environment. In Ref. 32 we have investigated how the water molecular hyperpolarizability evolves as a function of the distance to an air/water interface. This is relevant to interpret Surface Second-Harmonic Generation (S-SHG) experiments performed at the air/water interface. Briefly, a strong laser field at a wavelength 800 nm is focused on an air/water interface while the emitted light at the harmonic frequency, 400 nm, is collected, see Fig. 2 A.

The input trajectories were generated using `LAMMPS`. A water film, about 10 nm thick was created and equilibrated in a simulation box ($5 \times 5 \times 40$ nm^3) containing 9000 rigid TIP4P/2005 water molecules. `FROG` was used to analyze 80 frames of the MD trajectory separated by 20 ps, and to manage about 50 000 QM/MM calculations dispatched over different molecules and time steps. These `Dalton` calculations have cost approximately 11 k-hours for a single 2.6 GHz CPU. Each QM system was composed of a single water molecule, surrounded by point charges representing the neighboring water molecules. If a neighbor molecule owns at least one atom in radius $Rc = 2.3$ nm around the center of mass of the QM fragment, the whole molecule is included in the electrostatic environment represented by the points charges of the MM model. Hyperpolarizability calculations of the static and dynamic hyperpolarizabilities of water molecules were performed at the density functional theory (DFT) CAM-B3LYP/d-aug-cc-pVTZ level.

Finally, `FROG` has permitted to calculate the profiles along the direction normal to the interface (noted Z) for the hyperpolarizabilities, and the local electric field created by the MM environment. It has averaged at each time step, over the molecules belonging to 0.1 nm-thick slices normal to the Z direction. In Figures 2 B and C, the profiles are plotted as a function of the altitude $\Delta Z = Z - Z_0$ relative to the average position of the interface Z_0. Thanks to `FROG`, we have demonstrated that the resulting average molecular first hyperpolarizability tensor depends on the distance relative to the interface. Such effects can then be included in the modeling of experimental results.

B. Tutorials

Several tutorials are already available with the source code, but still under development (see Table I). The wiki, available at `https://glb96.github.io/Frog`, contains further explanations for each one of the tutorials. The files necessary to perform locally the tutorials are available to download on the `Gitlab` `https://github.com/glb96/Frog`: MD trajectories as inputs, `FROG`'s input parameter files, `SLURM` submission templates, the QM/MM results if needed, and `jupyter` notebooks to perform the analysis.

IV. IMPLEMENTATION DETAILS

A. Python implementation

The code is available on `Zenodo`[49] or on `Github`: `https://github.com/glb96/Frog`. Its license (GNU GLPv3) implies that it can be downloaded, modified, and used freely. We yet ask the user to cite the present article if results obtained with `FROG` are published.

FROG is downloadable as a Python3 package. It depends on common Python packages such as Numpy[50], but also requires MDAnalysis[51,52] and pytim[48]. A simple installation procedure can be found in the wiki and on the Github page. FROG is designed to be used either on the computing center, to avoid data transfers, or on a local computer. To use FROG, one runs it through a Linux shell. A file containing input parameters must be available in the directory, and its name, for instance frog_input.py, is given as an argument in a command line as in Listing 1.

```
Frog frog_input.py
```

Listing 1. Example of bash shell typical commands to run FROG.

Examples of input files are provided with the source code and commented on in the tutorials. In Appendix, we have added some extracts of input files to illustrate FROG's use.

B. Structure of the data into GP and MT classes

The information needed by FROG is structured in classes and the user defines them in the Python input file.

a. GlobalParameter *class* (GP) The input file contains first the definition of the GlobalParameter (GP), which contains information shared by the whole system such as the MD trajectory to use or the directory where FROG should be executed (see Listing 3).

b. MolecularType *class* (MT) In a second part, the input file defines one or more objects of the class MolecularTypes (MT). This class is the backbone of FROG's architecture. This molecular-based implementation is motivated by the approach of Section II A based on Equations 1, which are defined for each molecule. It is also compatible with most MD softwares based on molecules as elementary building blocks for the topology. The user defines an MT object for each type of molecule contained in the MD trajectory. For instance, if the MD trajectory contains only water molecules, one MT would be defined pointing at all the water molecules. For a solution containing methanol, water, Na^+ and Cl^-, 4 MT should be defined. The input is therefore relatively small for systems with only a few types of molecules. The MT definitions depend neither on the size of the system nor the number of MD snapshots to treat. FROG can handle with MT all types of molecules used in a classical MD, neutral or ionic, rigid or flexible, small or large, provided they are defined in its internal library. Note that the user can upgrade the library by defining a new MT using available templates.

Each molecule of the system can be involved in a Dalton calculation: either in the QM part, if their optical properties are investigated, or in the MM part, if they belong to electrostatic embedding. The MTs therefore contain information about these two cases. If the molecule is in the QM region, the MT indicates the QM-level of the QM/MM calculation. If the molecule is in the MM region, the MT indicates the charge and polarization of each atom. These are fixed for each molecule type.

Moreover, the MT objects contain the user-defined analysis or calculations to be performed. In the input file, the user defines for each MT what should be computed (e.g. molecular density, Radial Distribution Functions, frequency-dependent polarizability or hyperpolarizability), and for which molecules. In the last step, after FROG has analyzed the QM/MM outputs, the MTs also store the results, such as the QM observables.

In the following, some details of the implementation of the three main steps of FROG are provided.

C. Step I.a : reading MD trajectory analysis and structural analysis

The MD trajectories are read via the MDanalysis package. Therefore, pieces of information on the topology are extracted by FROG in the MD files, in particular the different molecules.

To get insight into the molecular structures of the trajectory, some analyses have been implemented in FROG: molecular density, molecular orientation, hydrogen bonds, and radial distribution function. These structural analyses can be performed during one FROG run for every MT defined and are done during the first part of the run. Averages of these structural observables can be performed either on (1) all molecules, (2) on groups of molecules based on their position in a user-defined laboratory axis to get a profile, (3) on groups of molecules belonging to the same layer relative to an interface.

D. Step I.b : Creation of QM/MM inputs for Dalton

The input files for Dalton are written during this first step. The user may define which molecules of each MT should be part of the QM fragment, the "targets".

For all the time steps defined in the GP, FROG creates dynamically the QM fragments at the "target" molecule site, and its electrostatic environment which will later be converted by Dalton into $\hat{H}_0$ and $\hat{V}_{env}$, respectively. The atom positions of the MD trajectory are used, both for the MM and for the QM fragments.

The QM fragment can be composed of a target single molecule. It can also include the closest neighbors of the target molecule, whatever their MT. All the atoms of the neighboring molecules are included in the QM part. The parameters dealing with the QM description of the target molecule are contained in the QMParameter class.

The MM environment is constructed using the MD trajectory: FROG replaces each molecule in the neighborhood of the QM fragment by a set of charged and/or polarizable centers. The radius of this electrostatic environment, R_f, is typically of the order of a few nanometers. To build the environment, FROG uses the time-dependent Periodic Boundary Conditions (PBC) of the MD trajectory.

Every molecule composing the MD trajectory shall have an electrostatic description to be part of a polarizable embed-

ding environment. It is defined in the `MT`. For instance, in our works[5,32,33], we have used point charges equal to the charges of the classical MD model, and no polarizability.

The user may modify the electrostatic description of a `MT` or create a new one for its needs easily. One constraint in `FROG`'s implementation is that all the molecules of a given `MT` share the same electrostatic description. The electrostatic descriptions of all `MT`s have to be known in advance.

Moreover, we implemented in `FROG` the embedding scheme that we called PE+LI[33], illustrated in Fig. 3 : the environment at a distance between R_c and R_f can be included implicitly by the addition in the QM calculation of a homogeneous electric field. This field is an approximation of the electric field that is imposed on the QM center by the point charges of the environment at distances from R_c to R_f, which is calculated by `FROG`.

In our works on pure liquid water, with $R_f = 4$ nm, R_c can be diminished to about 1 nm without much loss of accuracy for the non-linear optical molecular properties. But these parameters R_f and R_c are expected to depend on the system and on the observables at hand: they should be optimized for any new application.

During this first step, for each target molecule and for each time step, `FROG` creates a directory containing three `Dalton` input files needed to perform the QM/MM calculation: (i) a `".dal"` for the type of calculation and the electronic parameters, (ii) a `".mol"` containing the atomic coordinate of what composes the QM fragment, and (iii) a `".pot"` containing the electrostatic environment description around this QM fragment.

E. Step II : QM calculations on HPC

Once the QM/MM inputs are written, the user has to perform `Dalton` calculations, independently of `FROG`. Typically, a High-Performance Computing (HPC) center is used, where a job queuing system is installed. `FROG` provides inputs adapted to the ubiquitous 'Simple Linux Utility Resource Management' (`SLURM`)[53]. `FROG`'s options permit to fine-tune the parallelization depending on the computing resources.

Figure 4 illustrates how a trajectory containing 2000 molecules could be analyzed every two time steps. For 3 frames (1, 3 and 5), there would be 6000 QM/MM calculations to perform. To avoid submitting thousands of jobs to the queuing system, a single job may consist of a bunch of QM/MM calculations, for example, one job would perform 200 QM/MM calculations (as depicted in Listing 4).

Since the calculation are often performed in computing nodes with several CPU available, `FROG`'s options also permit to parallelize over these. In particular, `FROG` proposes to use the Linux `parallel` functionality[54], and/or the Message Passing Interface (MPI) capability of `Dalton`. For example, for jobs that run on a single node with 32 CPU, one can choose to dispatch the bunches of QM/MM calculations into 8 parallel channels, and each channel runs one Dalton calculation parallelized on 4 MPI-workers (see Listing 5).

Despite the aggregation of QM/MM calculations into bunches, many jobs have to be submitted, either sequentially, or in parallel. `FROG` reads a template provided by the user to output a `SLURM` submission script for an array of jobs. In practice, a unique command line handles the submission of all QM/MM calculations with a user-defined scheduling strategy (see Listing 2).

```
sbatch Submission_script/submit_array_job.sh
```

Listing 2. Bash command to submit all the QM/MM jobs to a SLURM queing system. The array permits to submit in one command a series of similar jobs.

Once all the jobs have been performed, `FROG` is called again to check whether all the QM/MM calculations have ended successfully, and have provided the expected observables. If not, `FROG` creates the submission scripts for the remaining jobs only, and the user has to use the previous submission command again (Listing 2).

F. Step III: Analysis of QM/MM outputs and post-processing

Once all the expected QM/MM calculations ended successfully, the observables are extracted automatically from `Dalton`'s output files, and stored in a `pickle` format.

`FROG` typically reads the polarizabilities and hyperpolarizabilities of the QM systems. At the dipolar level, they are 3×3 and $3\times3\times3$ tensors respectively. At the quadrupolar level, some hyperpolarizabilities (dipole-quadrupole and quadrupole-dipole) will have $3\times3\times3\times3 = 81$ components instead of 27. The polarizability and hyperpolarizability tensors are available either in the molecular frame (α and β) or in the laboratory frame (ι and χ). Once `FROG`'s calculation ends, the user can have access to the individual frequency-dependent molecular (hyper)-polarizabilities or to averages performed over space and time (different MD snapshot). The individual values can be manipulated as `pandas` DataFrame or `numpy` Arrays for statistical analysis, see the tutorial about second-hyperpolarizability calculation for instance. This can also be useful in the case of Second Harmonic Scattering of liquids[5] where the individual values and their fluctuations matter.

For Surface-Second Harmonic Generation, the evolution of the average (hyper)-polarizability over space is relevant. `FROG` uses `Diagrams` to store this spatial evolution. For instance, Listing 7 defines how the 27 components of $3\times3\times3$ β and χ tensors are averaged over all molecules. A distribution within a range $[-35,35]$ a.u with 200 bins is used to describe the value of all the hyperpolarizability in their molecular frame ('beta') over the whole space. Similar distributions are defined for the hyperpolarizability in the laboratory frame ('chi') but depending on the molecule position with respect to the laboratory Z-axis: 50 different distributions are created. Using such space-dependent analysis, the user is able to see the evolution of the observables as a function of the molecule position in the MD trajectory.

G. Regular tests

Currently, few developers are involved in FROG. A pipeline has been built within our local Gitlab platform to perform regular tests upon request (before merging on the master branch). The goal of these tests is to prevent errors during FROG execution solely: the tests do not involve running Dalton or any MD software.

The test starts from scratch with the installation of FROG and its Python package dependencies. Then, all the possible inputs and many relevant combinations are tested. Two MD trajectories are used during these tests, a bulk mixture and a pure solvent at the liquid-gas interface, since they represent the current typical use of FROG. Today, this represents a set of 64 inputs that are run within a few minutes. The tests are defined as successful if no dramatic error stops FROG's execution. We intend to extend these tests with reference verification values soon. These regular tests are also included in the Github page and can be used as examples in addition to the tutorials.

H. Parallelization

Since FROG manages a lot of molecules and time steps in the same run, its execution time easily increases. In the following, we describe the scaling as a function of different parameters, and the possibilities to optimize FROG's efficiency.

As expected, for the fragmentation and QM/MM input writing, FROG's execution scales linearly both with the number of molecules (N_{mol}), and with the number of time steps (N_{ts}), see Fig. 5 A and B. Thanks to the multiprocessing Process-based package, FROG can exploit CPU parallelization, with different workers analyzing different time steps of the trajectory. This can done during the step I and step III. For instance, Listing 6 defines 15 workers to be used for such a parallelization.

As illustrated in Fig. 5 C, the parallelization is very efficient, as long as the number of workers is smaller than the number of time steps. As expected, it is useless to use more workers than the number of time steps.

The most time-expensive part of FROG is the first one where structural analysis and QM/MM fragmentation are performed. It is especially the case if large environments are required, for instance for Radial Distribution Function or large QM/MM electrostatic embedding. Yet, the time needed per analysis or QM/MM input writing per molecule is still negligible compared to the cost of a quantum calculation, even with one molecule per QM fragment.

I. Documentation

The wiki is composed of HTML files available at https://glb96.github.io/Frog. Its core was created automatically from the source code using the Sphinx program[55], which outputs an intelligent and well-structured documentation. It contains extensive cross-references with automatic links for functions, classes, etc., and a search tool that permits finding the source documentation on objects defined in FROG.

V. CONCLUSION

FROG is a software originally designed to help computing the linear and non-linear optical responses of molecular liquids, both in bulk and near an interface. To achieve this, it analyzes MD trajectories and creates inputs for QM/MM Polarizable Embedding calculations within the Dalton package. The calculations, performed independently of FROG, provide the optical responses of molecules within complex environments. Statistical averages over numerous conformations can be performed. FROG's main specificity are: (i) It can open many types of MD trajectory formats, and it is designed for pure liquid or mixture, in bulk or with interfaces. (ii) The user can easily define the parameters for a molecule description or analysis. (iii) It helps to manage and analyze thousands of QM/MM calculations done on a High Performance Computing center. Noticeably, it is developed to deal with various tensors such as the polarizabilities and the first hyperpolarizabilities at the dipolar and quadrupolar orders.

FROG would be adapted to compute optical properties of liquids at the QM/MM level for systems such as a mixture composed of small or medium-sized molecules in a molecular solvent. In particular, aqueous solutions containing surfactants or lipid layers would probably suit FROG, both in bulk or at interfaces. Some care would be necessary for anisotropic systems where long-range interactions may impact the optical responses. On the contrary, FROG would be unsuitable when its basic hypothesis could fail: (i) If there are strong intermolecular interactions or electronic delocalization that impede a QM/MM approach. (ii) If the electrostatic description for the MM part differs from one molecule to another. (iii) If the QM probe involves many atoms, for instance, a full protein. (iv) If the partitioning demands to cut covalent bonds.

FROG is still under development, and we envisage various outlooks. One could first test further QM levels and embedding schemes included in Dalton[56]. In particular, we wish to include the possibility to add the pseudo-potentials in the MM embedding, to avoid spill-out effects[57]. Moreover, it would be worth to include more species in the quantum description to deal with clusters, such as solvated ions. FROG can in principle be used to study resonant frequency in SHG, but may require some refinement regarding the Dalton input. FROG can, in principle, compute all observables that Dalton can compute within its QM/MM framework within a reasonable amount of code implementation.

We hope FROG will facilitate the quantitative calculations of the non-linear optical properties of various liquids and interfaces, and help the scientific community to take the best advantage of the available experimental setups.

APPENDIX : ILLUSTRATIVE EXTRACTS OF INPUT FILE

```
GP.MD_file_name_topology = 'water.data'
GP.MD_file_name_traj = 'water.dcd'
GP.MD_file_type = 'LAMMPS'
GP.nbr_time_step = 75 # number of analyzed time steps
GP.trotter_step = 40 # analyze only 1 every 40
```

Listing 3. Example of input lines to define the input trajectory definition and the timesteps that will be investigated.

```
GP.nbr_repetition_QM_perMT = [200] # nbr QM per job.
```

Listing 4. Example of input lines to define the setting for job Dalton management on HPC calculator.

```
GP.nbr_job_parr_QM = 8 # Dalton calculation per node
GP.nbr_mpi_dalton = 4 # Dalton MPI workers
```

Listing 5. Example of input lines to define the setting for job Dalton management on HPC calculator with nodes of 32 CPU.

```
GP.nbr_parra = 15  # CPU available for FROG
```

Listing 6. Example of input lines to define the parallelisation of `FROG` among different timesteps.

```
L_diagram_analysis_to_perform = [
    ['beta', 'Averaged', [1, 200], [-35, 35]],
    ['chi', 'Plane_xy', [50, 200], [-35, 35]] ]
```

Listing 7. Example of input lines to define two Diagrams to calculate the average, standard deviation and distribution of the 27 components of the tensors β and χ. The statistics are calculated over all molecules, and over all time steps.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the PSMN (Pôle Scientifique de Modélisation Numérique) of the ENS de Lyon for the computing resources and technical help.

DATA AVAILABILITY STATEMENT

The sources of the code are available on `Zenodo`[49]. The code, documentation, tutorials, and tests are available on `Github`: `https://github.com/glb96/Frog` and at `https://glb96.github.io/Frog` for an online version of the wiki.

REFERENCES

[1]K. V. Mikkelsen, Y. Luo, H. Ågren, and P. Jørgensen, "Sign Change of Hyperpolarizabilities of Solvated Water," J. Chem. Phys. **102**, 9362–9367 (1995).
[2]K. O. Sylvester-hvid, K. V. Mikkelsen, P. Norman, D. Jonsson, and H. Agren, "Sign Change of Hyperpolarizabilities of Solvated Water, Revised: Effects of Equilibrium and Nonequilibrium Solvation," J. Phys. Chem. A **108**, 8961–8965 (2004).
[3]C. Liang, G. Tocci, D. M. Wilkins, A. Grisafi, S. Roke, and M. Ceriotti, "Solvent Fluctuations and Nuclear Quantum Effects Modulate the Molecular Hyperpolarizability of Water," Phys. Rev. B **96**, 1–6 (2017), 1705.01649.
[4]M. H. Cardenuto, F. Castet, and B. Champagne, "Investigating the first hyperpolarizability of liquid carbon tetrachloride," RSC Adv. **6**, 99558–99563 (2016).
[5]G. Le Breton, O. Bonhomme, E. Benichou, and C. Loison, "Liquid water: When hyperpolarizability fluctuations boost and reshape the second harmonic scattering intensities," J. Phys. Chem. Lett. **14**, 4158–4163 (2023).
[6]C. Bouquiaux, P. Beaujean, T. N. Ramos, F. Castet, V. Rodriguez, and B. Champagne, "First hyperpolarizability of the di-8-ANEPPS and DR1 nonlinear optical chromophores in solution. An experimental and multi-scale theoretical chemistry study," J. Chem. Phys. **159**, 174307 (2023).
[7]C. Naim, R. Vangheluwe, I. Ledoux-Rak, B. Champagne, C. Tonnelé, M. Blanchard-Desce, E. Matito, and F. Castet, "Electric-field induced second harmonic generation responses of push–pull polyenic dyes: experimental and theoretical characterizations," Phys. Chem. Chem. Phys. **25**, 13978–13988 (2023).
[8]M. V. Damasceno, A. R. Cunha, P. F. Provasi, G. I. Pagola, M. Siqueira, V. Manzoni, R. Gester, and S. Canuto, "Modulation of the nlo properties of p-coumaric acid by the solvent effects and proton dissociation," J. Mol. Liq. **394**, 123587 (2024).
[9]C. Bouquiaux, C. Tonnelé, F. Castet, and B. Champagne, "Second-order nonlinear optical properties of an amphiphilic dye embedded in a lipid bilayer. a combined molecular dynamics–quantum chemistry study," J. Phys. Chem. B **124**, 2101–2109 (2020).
[10]T. N. Ramos, F. Castet, and B. Champagne, "Second Harmonic Generation Responses of Ion Pairs Forming Dimeric Aggregates," J. Phys. Chem. B **125**, 3386–3397 (2021).
[11]T. N. Ramos and B. Champagne, "Investigation of the second harmonic generation at the water–vacuum interface by using multi-scale modeling methods," ChemistryOpen (2022), 10.1002/open.202200045.
[12]C. Gustafsson, M. Linares, and P. Norman, "Quantum mechanics/molecular mechanics density functional theory simulations of the optical properties fingerprinting the ligand-binding of pentameric formyl thiophene acetic acid in amyloid-(1–42)," J. Phys. Chem. A **124**, 875–888 (2020).
[13]D. Grabarek and T. Andruniów, "What is the optimal size of the quantum region in embedding calculations of two-photon absorption spectra of fluorescent proteins?" J. Chem. Theor. Comput. **16**, 6439–6455 (2020).
[14]U. N. Morzan, D. J. Alonso de Armiño, N. O. Foglia, F. Ramírez, M. C. González Lebrero, D. A. Scherlis, and D. A. Estrin, "Spectroscopy in complex environments from qm-mm simulations," Chem. Rev. **118**, 4071–4113 (2018), pMID: 29561145.
[15]E. Brunk and U. Rothlisberger, "Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states," Chem. Rev. **115**, 6217–6263 (2015), pMID: 25880693.
[16]K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, and H. Ågren, "The dalton quantum chemistry program system," WIREs Comp. Mol. Sc. **4**, 269–284 (2014), https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1172.
[17]J. Magnus, H. Olsen, S. Reine, and O. Vahtras, "Dalton project: A python platform for molecular-and electronic-structure simulations of complex systems," J. Chem. Phys. **152**, 214115 (2020).
[18]J. M. H. Olsen and J. Kongsted, "Chapter 3 - molecular properties through polarizable embedding," (Academic Press, 2011) pp. 107–143.
[19]N. H. List, P. Norman, J. Kongsted, and H. J. A. Jensen, "A quantum-mechanical perspective on linear response theory within polarizable embedding," J. Chem. Phys. **146**, 234101 (2017).

[20]A. Osted, J. Kongsted, K. V. Mikkelsen, P. O. Åstrand, and O. Christiansen, "Statistical mechanically averaged molecular properties of liquid water calculated using the combined coupled cluster/molecular dynamics method," J. Chem. Phys. **124**, 124503 (2006).
[21]S. Brasselet and J. Zyss, "Multipolar Molecules and Multipolar Fields: Probing and Controlling the Tensorial nature of Nonlinear Molecular Media," J. Opt. Soc. Am. B **15**, 257–288 (1998).
[22]D. P. Shelton, "Long-range Orientation Correlation in Water," J. Chem. Phys. **141**, 224506 (2014).
[23]D. P. Shelton, "Long-range Orientation Correlation in Dipolar Liquids Probed by Hyper-Rayleigh Scattering," J. Chem. Phys. **143**, 1–6 (2015).
[24]M. N. Nasir, E. Benichou, C. Loison, I. Russier-Antoine, F. Besson, and P.-F. Brevet, "Influence of the tyrosine environment on the second harmonic generation of iturinic antimicrobial lipopeptides at the air–water interface," Phys. Chem. Chem. Phys. **15**, 19919–19924 (2013).
[25]C. Loison, M. N. Nasir, E. Benichou, F. Besson, and P.-F. Brevet, "Multiscale modeling of mycosubtilin lipopeptides at the air/water interface: Structure and optical second harmonic generation," Phys. Chem. Chem. Phys. **16**, 2136–2148 (2014).
[26]J. Duboisset and P. F. Brevet, "Salt-induced Long-to-Short Range Orientational Transition in Water," Phys. Rev. Lett. **120**, 263001 (2018).
[27]A. Pardon, O. Bonhomme, C. Gaillard, P. F. Brevet, and E. Benichou, "Nonlinear Optical Signature of Nanostructural Transition in Ionic Liquids," J. Mol. Liq. **322** (2021), 10.1016/j.molliq.2020.114976.
[28]O. Bonhomme, L. Sanchez, E. Benichou, and P. Brevet, "Multistep micellization of standard surfactants evidenced by second harmonic scattering," J. Phys. Chem. B **125**, 10876–10881 (2021).
[29]G. Tocci, C. Liang, D. M. Wilkins, S. Roke, and M. Ceriotti, "Second-Harmonic Scattering as a Probe of Structural Correlations in Liquids," J. Phys. Chem. Lett. **7**, 4311–4316 (2016).
[30]X. Jin, T. Zhu, J. Z. H. Zhang, and X. He, "Automated fragmentation qm/mm calculation of nmr chemical shifts for protein-ligand complexes," Frontiers in Chemistry **6**, 1–11 (2018).
[31]J. M. H. Olsen, P. Reinholdt, and contributors, "PyFraME: Python framework for Fragment-based Multiscale Embedding (version 0.4.0)," (2021), see https://gitlab.com/FraME-projects/PyFraME.
[32]G. Le Breton, O. Bonhomme, P.-F. Brevet, E. Benichou, and C. Loison, "First hyperpolarizability of water at the air–vapor interface: a qm/mm study questions standard experimental approximations," Phys. Chem. Chem. Phys. **23**, 24932–24941 (2021).
[33]G. Le Breton, O. Bonhomme, E. Benichou, and C. Loison, "First hyperpolarizability of water in bulk liquid phase: Long-range electrostatic effects included via the second hyperpolarizability," Phys. Chem. Chem. Phys. **24**, 19463–19472 (2022).
[34]X. Zhuang, P. Miranda, D. Kim, and Y. Shen, "Mapping molecular orientation and conformation at interfaces by surface nonlinear optics," Phys. Rev. B **59**, 12632–12640 (1999).
[35]P. Guyot-Sionnet and Y. Shen, "Bulk contribution in surface 2nd-harmonic generation," Phys. Rev. B **38**, 7985–7989 (1988).
[36]Y. Foucaud, B. Siboulet, M. Duvail, A. Jonchere, O. Diat, R. Vuilleumier, and J. F. Dufrêche, "Deciphering second harmonic generation signals," Chem. Sci. **12**, 15134–15142 (2021).
[37]Y. Foucaud, J.-F. Dufrêche, B. Siboulet, M. Duvail, A. Jonchère, O. Diat, and R. Vuilleumier, "Why local and non-local terms are essential for second harmonic generation simulation?" Phys. Chem. Chem. Phys. **24**, 12961–12973 (2022).
[38]N. H. List, H. J. A. Jensen, and J. Kongsted, "Local electric fields and molecular properties in heterogeneous environments through polarizable embedding," Physical Chemistry Chemical Physics **18**, 10070–10080 (2016).
[39]C. Steinmann, P. Reinholdt, M. S. Nørby, J. Kongsted, and J. M. H. Olsen, "Response properties of embedded molecules through the polarizable embedding model," International Journal of Quantum Chemistry **119**, e25717 (2019).
[40]Y. Tu, Y. Luo, and H. Agren, "Molecular dynamics simulations applied to electric field induced second harmonic generation in dipolar chromophore solutions," J. Phys. Chem. B **110**, 8971–8977 (2006).
[41]M. H. Cardenuto and B. Champagne, "Qm/mm investigation of the concentration effects on the second-order nonlinear optical responses of solutions," J. Chem. Phys. **141** (2014), 10.1063/1.4903545.
[42]P. Beaujean and B. Champagne, "Coupled cluster evaluation of the frequency dispersion of the first and second hyperpolarizabilities of water, methanol, and dimethyl ether," J. Chem. Phys. **145**, 1–9 (2016).
[43]P. Beaujean and B. Champagne, "Coupled cluster evaluation of the second and third harmonic scattering responses of small molecules," Theor. Chem. Acc. **137**, 50 (2018).
[44]P. Kaatz, E. A. Donley, and D. P. Shelton, "A comparison of molecular hyperpolarizabilities from gas and liquid phase measurements," J. Chem. Phys. **108**, 849–856 (1998).
[45]Y. Chen and S. Roke, "Generalized expressions for hyper-rayleigh scattering from isotropic liquids," (2017), arXiv:1705.04231 [physics.chem-ph].
[46]F. Castet, E. Bogdan, A. Plaquet, L. Ducasse, B. Champagne, and V. Rodriguez, "Reference molecules for nonlinear optics: A joint experimental and theoretical investigation," Journal of Chemical Physics **136** (2012), 10.1063/1.3675848.
[47]K. K. Ray, A. Limaye, K. C. Ng, A. Saha, S. Shin, B. Biswas, M.-P. Gaigeot, S. Pezzotti, A. P. Willard, and H. C. Allen, "Second-harmonic generation provides insight into the screening response of the liquid water interface," The Journal of Physical Chemistry C **127**, 14949–14961 (2023).
[48]M. Sega, G. Hantal, B. Fábián, and P. Jedlovszky, "Pytim: A python package for the interfacial analysis of molecular simulations," J. Comput. Chem. **39**, 2118–2125 (2018).
[49]G. Le Breton, O. Bonhomme, E. Benichou, and C. Loison, "Frog: From molecular dynamics to second harmonic generation," (2022), `https://zenodo.org/doi/10.5281/zenodo.5998192`, DOI: 10.5281/zenodo.5998192.
[50]C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," Nature **585**, 357–362 (2020).
[51]Richard J. Gowers, Max Linke, Jonathan Barnoud, Tyler J. E. Reddy, Manuel N. Melo, Sean L. Seyler, Jan Domański, David L. Dotson, Sébastien Buchoux, Ian M. Kenney, and Oliver Beckstein, "MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations," in *Proceedings of the 15th Python in Science Conference*, edited by Sebastian Benthall and Scott Rostrup (2016) pp. 98 – 105.
[52]N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, "Mdanalysis: A toolkit for the analysis of molecular dynamics simulations," J. Comput. Chem. **32**, 2319–2327 (2011), https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21787.
[53]A. B. Yoo, M. A. Jette, and M. Grondona, "Slurm: Simple linux utility for resource management," in *Job Scheduling Strategies for Parallel Processing*, edited by D. Feitelson, L. Rudolph, and U. Schwiegelshohn (Springer Berlin Heidelberg, Berlin, Heidelberg, 2003) pp. 44–60.
[54]O. Tange, *GNU Parallel* (2018).
[55]See https://www.sphinx-doc.org/.
[56]P. Reinholdt, J. Kongsted, and F. Lipparini, "Fast approximate but accurate qm/mm interactions for polarizable embedding," J. Chem. Theor. Comput. **18**, 344–356 (2022), pMID: 34951300.
[57]A. Marefat Khah, P. Reinholdt, J. M. H. Olsen, J. Kongsted, and C. Hättig, "Avoiding electron spill-out in qm/mm calculations on excited states with simple pseudopotentials," J. Chem. Theor. Comput. **16**, 1373–1381 (2020), pMID: 32031806.

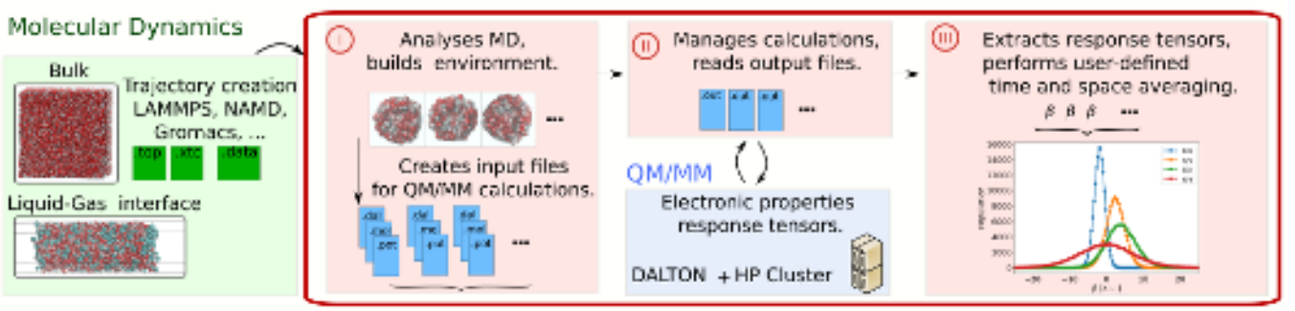

FIG. 1. Workflow to compute optical response properties of molecular assemblies such as molecular liquid in bulk or close to interfaces. `FROG` helps the user with analyzing the MD trajectories, and creating `Dalton` input files (I), managing the calculations (II) and finally extracting the average properties (III). In step (II), `FROG` does not perform or submit automatically jobs to Highly Parallelized Computing (HPC) center. The QM/MM calculations can be performed on the cluster independently over several days or weeks if needed without any risk of general error or crash.

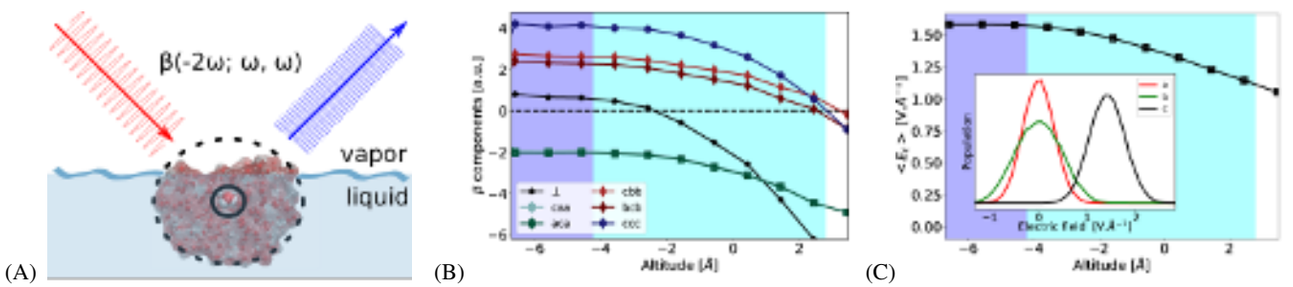

FIG. 2. A: Schematics of the system of interest and the S-SHG experiment, and of the QM/MM approach, where the black circle emphasizes one QM fragment, and the dashed circle the MM environment. B: Variation of the different hyperpolarizability tensor components as a function of the distance relative to the interface. The violet background depicts the bulk liquid region, and the light blue background is the interface region. C: Electrostatic field due to the MM environment along the C_{2v} molecular axis of QM water molecule, noted c. Inset: distribution of the electric field values in the molecular frame $\{a,b,c\}$ of the QM fragment. Figures from[32], under CC BY-NC 3.0 Deed license, without the need for RSC authorization in a publication by the authors.

Name of Tutorial	**Content**
Getting started	Description of first steps for beginners: install and run the first `FROG` job.
Space Discretization	Example of the spatial analysis and selection of molecules and how to plot profiles of the observables such as molecular density, orientation, hydrogen bonds....
Mixtures and Structural Analysis	Example of a water/methanol mixture where hydrogen bonds are analyzed.
Optical Analysis	Example of calculation of hyperpolarizabilities within the PE environment, and how to read and plot the related observables.
Quadrupolar Terms and Long Range QM/MM	Topics specialized for Surface Second Harmonic Generation, for experts, see Ref. 32,33.
Second Hyperpolarizability	Scripts to perform the calculation of the second hyperpolarizability using the Finite Field method, see Ref. 33
Loading Results	Tools and tips to open and plot `FROG`'s output files.
Molecular Type Module	Explanation of the definition of `MolecularTypes` and how one can define a new one.

TABLE I. List of tutorials available for `FROG`.

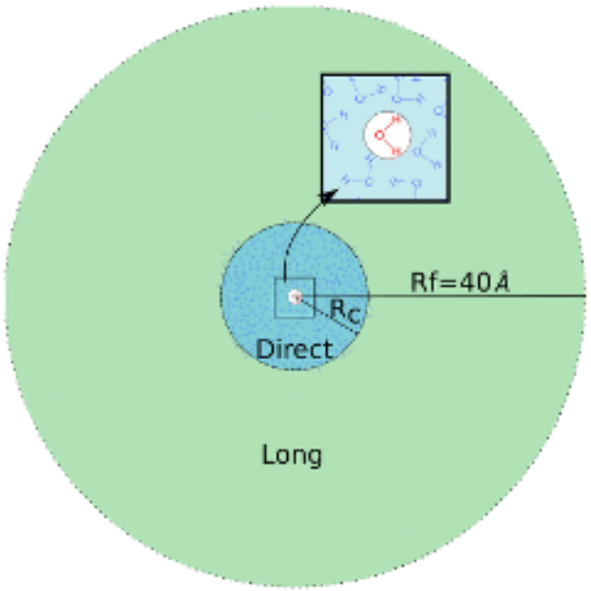

FIG. 3. Sketch of the PE+LI electrostatic embedding procedure. The QM box is defined for only one molecule, here, water. In the direct area, until R_c, all the neighbors are included in a QM/MM calculation with a discrete solvation procedure, such as the PE formalism. In the Long-range area, from R_c to R_f: the neighbors are included by the electrostatic field they generate at one specific point of the QM box. Figures from Ref 33, under CC BY-NC 3.0 Deed license, without the need for RSC authorization in publication by the authors.

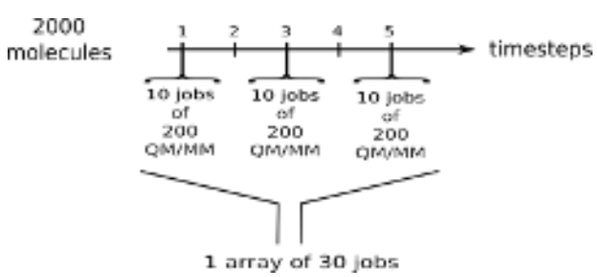

FIG. 4. Example of the management of a trajectory of 2000 molecules that is analyzed every 2 time steps. `FROG` dispatches the QM/MM calculations into jobs. Each job manages 200 QM/MM calculation. All the jobs that are submitted as a single `SLURM` job array on an HPC center.

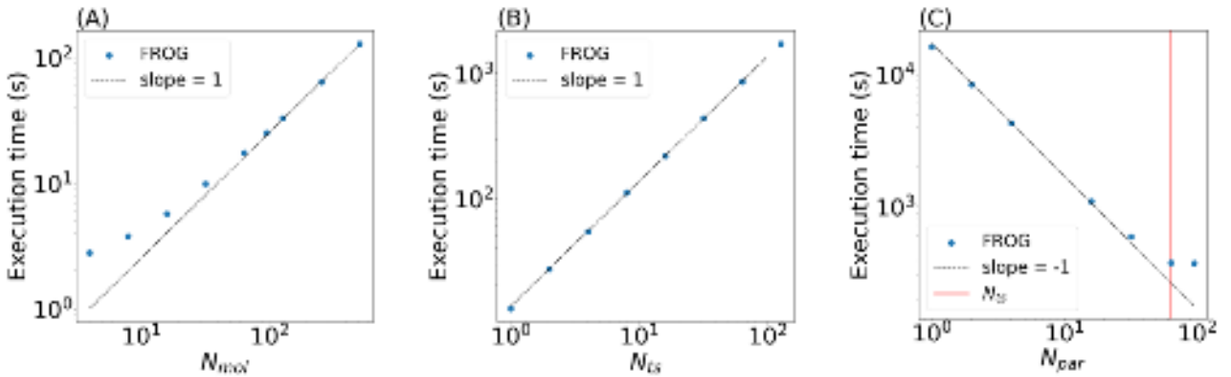

FIG. 5. Execution time of FROG (in sec.) for the structural analysis and writing the QM/MM inputs. (A) For the analysis of N_{mol} molecules, and for 2 time steps. (B) For the analysis of N_{ts} time steps, for 200 molecules. (C) Parallelized on N_{par} CPUs of the same computing node, for the analysis of 2000 molecules, and N_{ts} = 64 time steps. Black dashed lines with a slope of 1 or -1 have been added in the log-log representations.