Jihočeská univerzita v Českých Budějovicích

Přírodovědecká fakulta

Calculation of vibrational SFG spectra from molecular dynamics simulations

Master thesis

Bc. Patrik Musil

Supervisor: doc. RNDr. Milan Předota, Ph.D.

Consultant: Ing. Ondřej Kroutil, Ph.D.

České Budějovice 2022

Musil, P., 2022: Calculation of vibrational SFG spectra from molecular dynamics simulations, Mgr. thesis, in English. 51 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

Abstract

In this thesis, implementation of calculation vibrational Sum-Frequency Generation (SFG) spectra from classical molecular simulations of aqueous solutions in the O-H vibrations region is presented. The key part of the thesis was significant improvement of the original codes used by the group of Prof. Marie-Pierre Gaigeot to post-process *ab initio* molecular dynamics trajectories to predict vibrational SFG spectra and apply it to classical molecular dynamics simulations. Particular attention was paid to modifying and optimizing the code for processing much larger systems (number of molecules) and much longer trajectories (longer simulation times), allowing to take advantage of classical simulations to work on larges scales than the ab initio simulations. The code also integrates calculations of contributions from water O-H bonds and surface hydroxyls. As a new feature, calculation of cross-correlations among molecules was implemented. The performance of the resulting codes is verified and the influence of classical molecular simulation parameters, as well as parameters entering the SFG spectra calculations, on the calculated vibrational SFG spectra is studied for fluorite/water interface.

Prohlašuji, že svoji bakalářskou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění, souhlasím se zveřejněním své bakalářské práce, a to elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této

kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích dne 13.4.2022 ……

Patrik Musil

Acknowledgement

I would like to acknowledge especially these people:

Milan Předota for supervising my thesis and giving me solid theoretical background and useful programming tips.

Ondřej Kroutil for helping me with aplication part.

Marie-Pierre Gaigeot for helpful remote consultations about the topic.

Simone Pezzotti for providing us the original SFG codes and clarification of his code.

Contents

1 Introduction 1

1.1 Motivation 1

1.2 Non-linear optics background 1

1.3 Computational SFG vibrational spectroscopy 3

2 Methods 9

2.1 Reading input files 10

2.1.1 General implementation 10

2.1.2 BOXDATA parameters 10

2.1.3 Trajectories 11

2.2 Calculation of instantaneous surface 13

2.2.1 Improvement 14

2.3 Calculation of density function 15

2.3.1 Improvement 16

2.4 Calculation of SFG response function 17

2.4.1 Construction of D matrices and evaluation of v_z 18

2.4.2 Improvement 19

2.4.3 Inclusion of intermolecular terms 21

2.4.4 Application with the surface hydroxyls 22

2.5 Analysis of the SFG response function 23

2.5.1 Improvement 24

2.6 Program features 24

2.6.1 General improvements 24

2.6.2 Program options 26

2.6.3 Execution 28

3 Results and discussion 30
3.1 Reconstruction of previous experiments 30
3.2 Influence of simulation parameters 31
3.2.1 Length of simulation 32
3.2.2 Size of the system 33
3.2.3 Thermostat 35
3.3 Influence of the length of the SFG response function and filter 37
3.4 Effect of including contributions from surface hydroxyls 41
3.5 Effect of including intermolecular terms 43
4 Conclusions 45
5 References 47
6 Appendix 51

1 Introduction

1.1 Motivation

The specific microscopic organization of water molecules in proximity to air/water or solid/water plays role in many phenomena ranging from mineral dissolution and pollutant transport in ground water to peptide bond formation. [1]

Vibrational Sum-Frequency Generation (SFG) spectroscopy is very powerful and versatile tool to study interfaces because it is highly surface specific and sensitive to submonolayers of molecules, and it can be applied to all interfaces accessible by probing light. The technique is based on the different structure symmetries of bulk and the interface. Both the interface and bulk have different sets of response coefficients represented by the tensor elements of second-order non-linear susceptibilities ($\chi^{(2)}$). Discrimination of the bulk contribution can be achieved by detecting χ tensor elements with specific input and output polarization combinations. [2]

However powerful the vibrational SFG technique is, extracting a detailed understanding of molecular formation in the first water layers where the H-bond network reorients due to the atypical environment at the interface between water and solid (or vapour) phases, only from experimental data, remains challenging. [1]

1.2 Non-linear optics background

When a common light with an electrical field strength **E** irradiates the medium, the electric charge of molecule and atom in the medium occurs overall or relative displacement, inducing a secondary light electrical field, which is described by a physical quantity – electric polarization **P**. The magnitude of the induced polarization **P** depends linearly on the electrical field amplitude, however only in isotropic media the proportionality factor is scalar. [3] The general relationship between **P** and **E** is:

$$\mathbf{P} = \varepsilon_0 \boldsymbol{\chi}^{(1)} \cdot \mathbf{E} \qquad (1)$$

where $\boldsymbol{\chi}^{(1)}$ stands for the linear susceptibility, which is a complex number tensor for the anisotropic medium, ε_0 is the dielectric coefficient (permittivity) of vacuum.

If the incident light is not a common light, but light with much higher intensity (by several orders of magnitude), effects due to simultaneous interactions of several photons with the matter give birth to nonlinear optics phenomena, which started to be revealed since the 1960's [3]. We can expand the relationship of polarization in equation (1) into a power series of light electrical field amplitude, the relationship between **P** and **E** for the anisotropic medium is:

$$\mathbf{P} = \varepsilon_0\boldsymbol{\chi}^{(1)} \cdot \mathbf{E} + \varepsilon_0\boldsymbol{\chi}^{(2)}\mathbf{:EE} + \varepsilon_0\boldsymbol{\chi}^{(3)}\mathbf{:EEE} + \cdots \tag{2}$$

where $\boldsymbol{\chi}^{(1)}$ is linear susceptibility, $\boldsymbol{\chi}^{(2)}$ and $\boldsymbol{\chi}^{(3)}$ are the second- and third-order nonlinear susceptibilities, respectively. $\boldsymbol{\chi}^{(1)}$, $\boldsymbol{\chi}^{(2)}$ and $\boldsymbol{\chi}^{(3)}$ are second-, third- and fourth-rank tensors, with each of their indices being one of the 3D coordinates, respectively. The symbols "$\cdot$", ":" and "⋮" are denoted second-, third- and fourth-rank tensor multiplication operations, respectively. In the equation (2), the first item is linear polarization, the second item, third item, and so on are high-order nonlinear polarizations, from left to right, respectively. Equation (2) can be denoted by electric polarizations:

$$\mathbf{P} = \mathbf{P}^{(1)} + \mathbf{P}^{(2)} + \mathbf{P}^{(3)} = \mathbf{P_L} + \mathbf{P_{NL}} \tag{3}$$

where the first term is linear polarization, which is denoted by

$$\mathbf{P_L} = \mathbf{P}^{(1)} \tag{4}$$

the following items are high-order nonlinear polarization, which is denoted by

$$\mathbf{P_{NL}} = \mathbf{P}^{(1)} + \mathbf{P}^{(2)} \tag{5}$$

SFG (as one of the key representatives of second-order non-linear optics phenomena) originates from interaction of two optical beams of different frequencies ω_1 and ω_2 , giving rise to a beam of a frequency given by $\omega = \omega_1 + \omega_2$, to fulfill the condition of energy conservation (this equation gives rise to the name SFG). The second-order nonlinear polarization $\mathbf{P}^{(2)}$ and the second-order nonlinear susceptibility $\boldsymbol{\chi}^{(2)}$can be expressed as,

$$\mathbf{P}^{(2)}(\boldsymbol{\omega}) = \varepsilon_0\boldsymbol{\chi}^{(2)}(\omega;\omega_1,\omega_2)\mathbf{:E}(\omega_1)\mathbf{E}(\omega_2) \tag{6}$$

$$\boldsymbol{\chi}^{(2)} = \int_{-\infty}^{\infty} \boldsymbol{\chi}^{(2)}(t-t_1, t-t_2) e^{i[\omega_1(t-t_1)+\omega_2(t-t_2)]} dt_1 dt_2 \qquad (7)$$

where $\boldsymbol{\chi}^{(2)}(\omega)$ on the LHS is expressed as a double integral of the response function $\boldsymbol{\chi}^{(2)}(t-t_1, t-t_2)$ over time, capturing how electric fields $\mathbf{E}_{\text{vis}}(t_1)$ and $\mathbf{E}_{\text{IR}}(t_2)$ result in the nonlinear optics field $\mathbf{E}_{\text{SFG}}$ at time t. Note that from equation (6) it is evident why SFG is a non-linear optics phenomenon. The intensity of the resulting beam is proportional to the polarization $\mathbf{P}^{(2)}$, which is proportional to the product of the magnitudes of the two incoming beams. This is in contrast to linear optics, where the magnitude of the output signal is proportional to the magnitude of the input signal.

$\boldsymbol{\chi}^{(2)}$ representing a second-order nonlinear contribution is a third-order tensor with 27 tensor elements

$$\boldsymbol{\chi}^{(2)} = \begin{bmatrix} XXX & XYY & XZZ & XYZ & XZX & XXZ & XXZ & XXY & XYZ \\ YXX & YYY & YZZ & YYZ & YZY & YZX & YXZ & YXY & YYZ \\ ZXX & ZYY & ZZZ & ZYZ & ZZY & ZZX & ZXZ & ZXY & ZYZ \end{bmatrix} \qquad (8)$$

A detailed explanation of non-linear optics can be found e.g. in the book [3].

1.3 Computational SFG vibrational spectroscopy

A typical geometry of SFG measurement is shown in Fig. 1, with the three-layer model of two bulk phases and the interface characterized by the dielectric constants, $\varepsilon_1(\omega)$, $\varepsilon_1(\omega)$ and $\varepsilon'_1(\omega)$, respectively. We assume that the nonlinear polarization of the sum frequency is generated by $\boldsymbol{\chi}^{(2)}$ in the interface region. The angles $\theta_1(\omega_{vis})$ and $\theta_1(\omega_{IR})$ are incident angles of the visible light and the infrared light with the surface normal in the bulk phase 1. The sum frequency signal is emitted to the angle determined by the phase matching condition and the output angles of reflection and transmission are denoted with $\theta_i(\omega_{vis})$ (i = 1,2). Then, the observed SFG intensity in the phase i is formally represented as

$$I_i(\omega_{SFG}) = \frac{8\pi^2 \omega_{SFG}^2 \sec^2\theta_i(\omega_{SFG})}{c^3 \sqrt{\varepsilon_i(\omega_{SFG})\varepsilon_{i1}(\omega_{vis})\varepsilon_{i2}(\omega_{IR})}} \qquad (9)$$
$$|\mathbf{e}(\omega_{SFG} \cdot \boldsymbol{\chi}^{(2)} : \mathbf{e}(\omega_{vis})\mathbf{e}(\omega_{IR})|^2 I_{i1}(\omega_{vis}) I_{i2}(\omega_{IR})$$

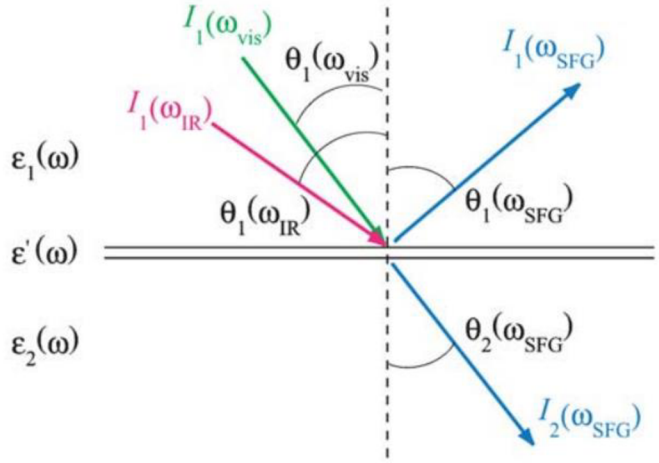

Figure 1: Typical geometry of the SFG measurement. Taken from [4]

where $I_i(\omega_{SFG})$ is the irradiance of sum frequency ω_{SFG} in the phase i (=1,2), $I_1(\omega_{vis})$ and $I_2(\omega_{IR})$ are the irradiances of the visible and infrared lights in the phase 1 and 2, respectively. The equation (9) is expanded to include some of the bulk polarization by replacing $\chi^{(2)}$ with an effective susceptibility including the bulk term. From the equation (9) we can assume that SFG signal is influenced by the following factors, i.e.

a) effective surface nonlinear susceptibility $\boldsymbol{\chi}^{(2)}(\omega_{SFG}, \omega_{vis}, \omega_{IR})$,
b) dielectric constants $\varepsilon_1(\omega)$, $\varepsilon_1(\omega)$ and $\varepsilon'_1(\omega)$, where $\omega = \omega_{SFG}$, ω_{vis}, ω_{ir},
c) geometry of the experiment (polarization of the input and output beam, and incident angles of the input beams $\theta_1(\omega_{vis})$ and $\theta_1(\omega_{IR})$)

$\mathbf{e}(\omega_{SFG})$, $\mathbf{e}(\omega_{vis})$ and $\mathbf{e}(\omega_{IR})$ in the equation (9) are the polarization vectors of output and input electric fields in the interface layer including the Fresnel coefficients, these vectors are determined by the dielectric constants and the geometry of the experiment. In the qualitative sense, these vectors are mostly influenced by the light polarizations in the bulk phase, denoted with three letters such as *ssp*, *spp*, *ppp* etc., where the letters (*s* and *p*) stand for the polarization (*s*-polarized electric field is perpendicular to the plane of incidence whereas *p*-polarized electric field is polarized parallel to the plane of incidence) of detected SFG signal, incident visible and infrared lights, respectively. Note that equation (9) is commonly valid for the Second-Harmonic Generation (SHG where $\omega_{vis} = \omega_{ir}$), though practical

computation of $\chi^{(2)}$for the vibrational SFG spectra is considerably different from the computation of SHG $\chi^{(2)}$ spectra [4].

SFG $\chi^{(2)}$ can be decomposed into the vibrationally resonant and non-resonant terms ($\chi^{(2)} = \chi^{(2),R} + \chi^{(2)N-R}$). The resonant part $\chi^{(2),R}$ is responsible for the $\chi^{(2)}$ dependency on ω_{IR}, and is the central quantity for the analysis of SFG spectra [4]. Morita [4] [5] has formulated $\chi^{(2),R}$ in two different ways, one of them is energy representation of $\chi^{(2),R}$ based on the perturbation theory and the second one is time-dependent representation, which can be implemented to the molecular simulation. The derivation detailed in [4] by Morita at the end leads to a simple general equation to calculate $\chi^{(2),R}$

$$\chi_{PQR}^{(2),R} = \frac{-i\omega}{k_B T}\int_0^{+\infty} e^{i\omega t}\langle \mathrm{A}_{PQ}(t) M_R(0)\rangle dt \qquad (10)$$

where $\chi^{(2),R}$ is the resonant part of second-order susceptibility tensor, (*P, Q, R*) are any directions of the laboratory frame, ω is the frequency of IR beam, A_{PQ} and M_R are respectively the components of the total polarizability tensor and the total dipole moment (more exactly their instantaneous values with their average values subtracted – a necessary condition for the correlation function to decay to zero). The correlation function of the two properties separated by a time difference *t* is averaged over all available time origins to provide good statistics, i.e.,

$$\langle \mathrm{A}_{PQ}(t) M_R(0)\rangle = \langle \mathrm{A}_{PQ}(t+\tau) M_R(\tau)\rangle_\tau \qquad (11)$$

Khatib et al. [6] later modified eq. (10) replacing the correlation function by correlation function of the time derivatives of the respective functions,

$$\chi_{PQR}^{(2),R} = \frac{-i}{k_B T\omega}\int_0^{+\infty} e^{i\omega t}\langle \dot{A}_{PQ}(t) \dot{M}_R(0)\rangle dt \qquad (12)$$

While this modification theoretically stems from the rule that the Fourier transform of a derivative can be expressed as n of $f'(t)$ is $i\omega F(\omega)$, it was introduced to simplify the numerical calculation of susceptibility, since the evaluation of time derivatives $\dot{A}_{PQ}$ and $\dot{M}_R$ using the information on atomic velocities is easy and the correlation function in eq. (12) converges faster than the one in eq. (11). Supposing that only O-H stretching of *M* molecules

with N_m O-H bonds has impact on the final spectra in the region of interest, total polarizability, and dipole moment of the system (A_{PQ}, M_R) can be decomposed into individual O-H bond contributions.

$$\dot{A}_{PQ}(t) = \sum_{m=1}^{M} \sum_{n=1}^{N_m} \dot{\alpha}_{mn,PQ}\,(t) \tag{13}$$

$$\dot{M}_R(t) = \sum_{m=1}^{M} \sum_{n=1}^{N_m} \dot{\mu}_{mn,R}\,(t) \tag{14}$$

In addition, thanks to basic geometry considerations, the dipole moment of the A-B bond can be projected from the A-B bond frame (μ_b) to the laboratory frame (μ_l):

$$\mu_l = \mathbf{D}\mu_b \tag{15}$$

where D is the direction cosine matrix projecting the A-B bond frame into the laboratory frame.

In the following equations two assumptions are applied. The first assumption is that the bond elongations are small enough to make first order Taylor expansion. The second assumption is that the stretching mode of the bond is much faster than the modes involving a bond reorientation – for example the liberation. Meaning of the second assumption is that $D_{Ri} \approx 0$ and that $\frac{dr_z}{dt} \gg \frac{dr_x}{dt} \approx \frac{dr_y}{dt}$ therefore $\dot{\mu}_R$ can be simplified into:

$$\dot{\mu}_R(0) \approx \sum_{i}^{x,y,z} D_{Ri}(0)\,\dot{\mu}_i(0) \tag{16}$$

$$\approx \sum_{i}^{x,y,z} D_{Ri}(0) \left(\sum_{j}^{x,y,z} \frac{d\mu_j}{dr_j} \frac{dr_j}{dt}\bigg|_{t=0} \right) \tag{17}$$

$$\approx \sum_{i}^{x,y,z} D_{Ri}(0) \frac{d\mu_i}{dr_z} v_z(0) \tag{18}$$

where $v_z(0) = \left.\frac{dr_j}{dt}\right|_{t=0}$ corresponds to the projection of the velocity on the bond axis.

Similarly, formula for the polarizability can be derived:

$$\dot{\alpha}_{PQ}(t) \approx \sum_{i}^{x,y,z}\left[D_{Pi}(t)\sum_{j}^{x,y,z}\left(\frac{d\alpha_{ij}}{dr_z}D_{Qj}(t)\right)\right]v_z(t) \qquad (19)$$

Where projection of the velocity on the O-H bond axis v_z and direction cosine matrices **D** can be evaluated from molecular dynamics trajectory, $\frac{d\mu_i}{dr_z}$ and $\frac{d\alpha_{ij}}{dr_z}$ can be parametrized [6].

The first water monolayer is characterized by specific H-bond networks that differ from the one in the bulk water. Generally speaking, the interfacial network is built to either maximize the water-water H-bonds within the first monolayer (in case of hydrophobic surfaces like air-water) or to maximize the interactions between the water molecules and the solid surface (in case of hydrophilic surfaces like quartz-water). SFG spectroscopy seems to be perfect tool to reveal the water arrangement at interfaces thanks to its interfacial specificity. It is not trivial to extract this information from SFG signals, especially the interface thickness is unknown. Thickness of the interface probed in the SFG experiments is probably the first monolayer only for neutral interfaces (i.e., at the isoelectric point), while the field generated by the charged surfaces reorients and polarizes water molecules in a thickness that is roughly estimated by the Debye length. The loss of inversion symmetry in the reoriented water slab thus increases the thickness probed in the SFG experiments, presumably up to microns in the case of low ionic strengths [7]. It was shown that any charge interface is composed only of two layers that are SFG active: Binding interfacial layer (BIL) and Diffuse layer (DL), corresponding to subsequent bulk water layer(s) reoriented by the surface field, which are SFG active because of the reorientation of water molecules along the surface field disrupts the liquid centrosymmetry, which is then recovered for the rest of the water bulk.

Note that the susceptibility $\chi^{(2)}$ (see eq. (12)) is a complex quantity. While originally only the magnitude $\left|\chi^{(2)}\right|$ was experimentally determined [8], significant advancement of the interpretation of SFG signals came with the phase-resolved SFG techniques, which allow independent determination of the real and imaginary contributions, $\mathrm{Re}(\chi^{(2)})$ and $\mathrm{Im}(\chi^{(2)})$

[9] [10] [11]. The $\mathrm{Im}(\chi^{(2)})$, that will be exclusively discussed in our presented results, is particularly suitable for interpretation, because its sign carries the information on the orientation of OH groups at the interface. The position of the spectral peaks, on the other hand, carries information on the strength of hydrogen bonding of OH groups [12]. Since the interface often invokes several layers of interfacial water with alternating alignment, the combination of these pieces of information, resulting in positive and negative peaks of $\mathrm{Im}(\chi^{(2)})$ at different wavenumbers, allows much more reliable interpretation of the experimental spectra as opposed just to magnitude of $\left|\chi^{(2)}\right|$, which is by definition non-negative for all frequencies.

Finally, lets mention that computer simulations provide both predicted SFG spectra and detailed information on the interfacial structure including orientations of molecules, thus allowing reliable discussion of the impact of various contributions (individual layers as described above, specific hydrogen-bonded configuration, etc.) to the total spectra. By examining these relationships, we learn how to interpret the spectra, allowing the experimentalists to more reliably predict the interfacial behavior based on SFG spectra – here the role of simulations can't be underrated.

2 Methods

This chapter will demonstrate in detail how the SFG spectra are calculated by the original programs [13], improvements and new features I have added to the original programs. Firstly, let me introduce the sequence of the programs (Fig. 2).

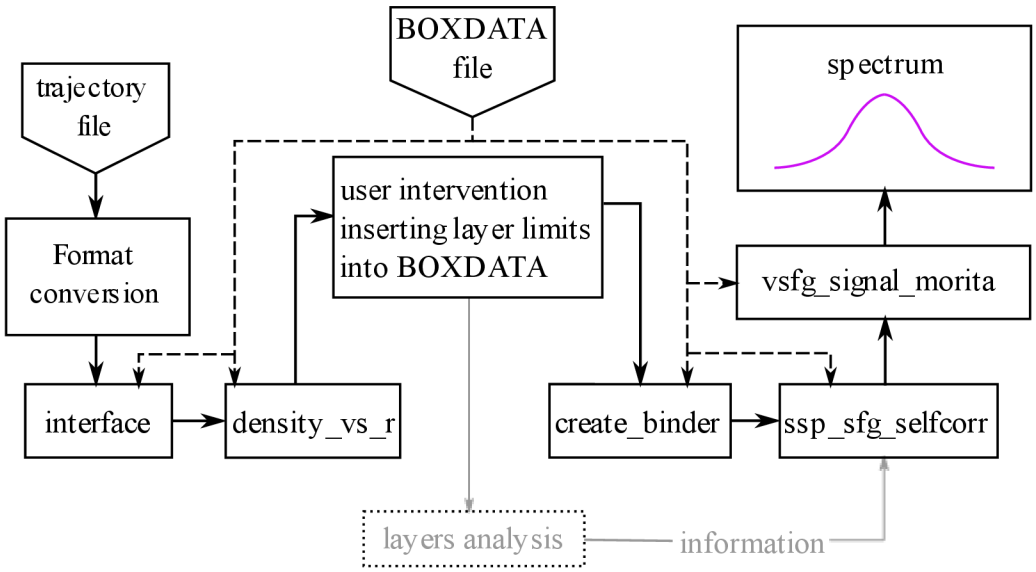

Figure 2: Workflow of the original programs by S. Pezzotti used by O. Kroutil.

Original programs [13] took as an input two .xyz files containing positions and velocities of water atoms (the files had to be strictly called "pos.xyz" and "vel_mol_rebuilt.xyz") and "BOXDATA" file, containing basic information about the system and some options for the calculations. In case of using CMD simulations (Gromacs) trajectory file .trr had to be converted to .gro file and .gro file had to be converted to two separate .xyz files using a simple code by M. Předota to fit the format used by M.P. Gaigeot's group in Paris when processing the *ab initio* trajectories.

With the prepared input files, the user had to run programs "interface" and "density_vs_r" respectively to obtain water density function. From the water density function the user had to specify (write to the "BOXDATA" file) the height of individual water layers (first 3 layers) from the water interface.

Set of binder files (files describing layer affiliation of each water molecule present in the system – each layer and their combinations had its own binder file) was created by executing the program "create_binder".

At the same point the O-H bond structure of each water layer could be analyzed [7] [14] by another set of programs (see Fig. 2). The information obtained from the O-H bond structure analysis was used to select which interfacial layers belong to BIL (binding interfacial layer), DL (diffuse layer) or BL (bulk layer) contributing to the final SFG spectra [15] [1] [7].

Program "ssp_sfg_selfcorr" was then executed (with the selected binder file) to calculate $\langle \dot{A}_{xx}(t)\dot{M}_z(0)\rangle$ autocorrelation function, which was then analyzed by "vsfg_signal_morita" program to obtain the SFG spectra. All the programs (excluding layers analysis) will be described in more detail later in this chapter.

2.1 Reading input files

In original programs [13] , reading of the input files was implemented by reading the BOXDATA parameters and input trajectories into general variables. For the diversity of input trajectory files (.xyz, .gro and .trr), a more general approach had to be implemented.

2.1.1 General implementation

Fortran 2003 introduced support for object-oriented programming. This feature allows to take advantage of modern programming techniques such as derived types and abstract derived types (polymorphism).

2.1.2 BOXDATA parameters

Reading the BOXDATA parameters is now implemented as a subroutine called "read_boxdata" in module "bdata" specifying the "boxdata" derived data type containing all the possible BOXDATA parameters. The main advantage is that adding a new parameter into "BOXDATA" file is easily made by expanding the "boxdata" and modifying the subroutine "read_boxdata". All the calculation parameters are stored in the derived data types, reading BOXDATA parameters does not need to be implemented over and over in the programs. An example of using "boxdata" derived type:

```
use bdata

type(boxdata) :: bd

call read_boxdata(20,bd) !file unit, boxdata

print*, bd%parameter !prints boxdata parameter
```

2.1.3 Trajectories

To read trajectory files frame by frame, abstract derived type “frame_reader” (Fig. 3) with procedures “open_file”, “read_frame” and “skip_frame” storing the whole input trajectory frame as an array of water molecules (derived type containing O, H1 and H2 atoms, which are another derived types containing position and velocity vectors) is introduced. That way the position and velocity of each atom can be easily accessed. The “frame_reader” is extended by “trr_frame_reader”, “gro_frame_reader” and “xyz_frame_reader” with their own implementation of “frame_reader” procedures for respective file types.

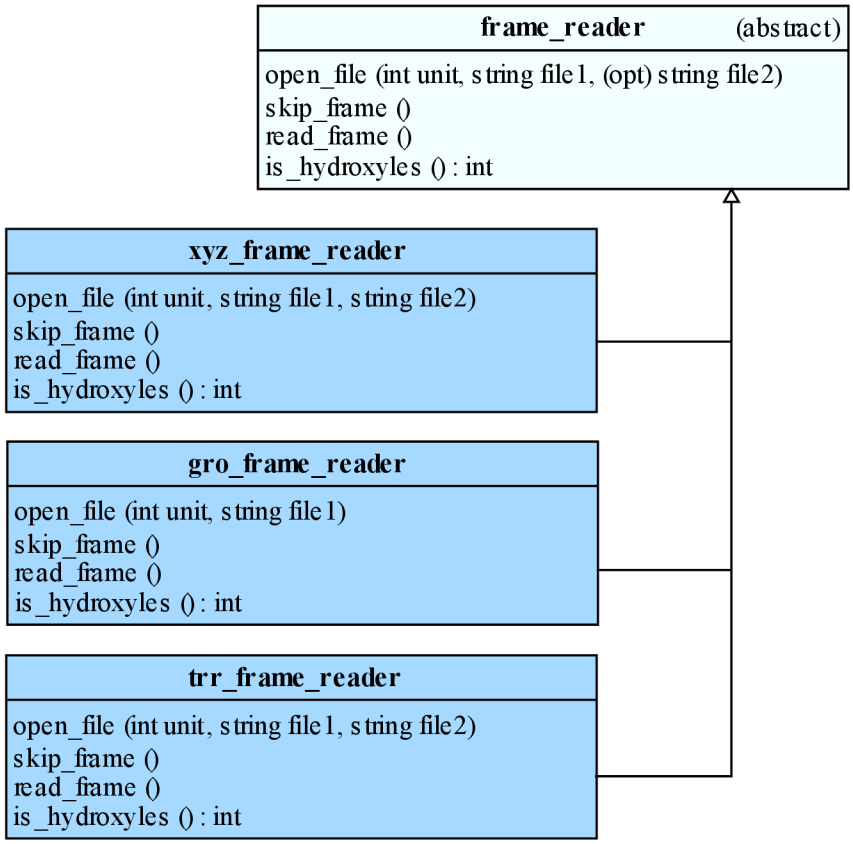

Figure 3: frame_reader class diagram.

Note that “trr_frame_reader” and “xyz_frame_reader” implementation of “open_file” takes two files as an argument (“.trr” trajectory file plus at least one frame “.gro” file providing the atomic identity of coordinates, or “position.xyz” plus “velocity.xyz”, respectively), “gro_frame_reader” takes only “.gro” trajectory file containing a series of frames as an argument.

Procedure “open_file” checks if the input trajectory contains only water molecules or only hydroxyl groups, according to that information it allocates array to store information of all water molecules (size of number of oxygens in the first frame).

Procedures “skip_frame” and “read_frame” leaves values or updates values of the molecule array, respectively. Each of the procedures has different implementation based on the file to be read. Frame readers “gro_frame_reader” and “trr_frame_reader” implement conversion of units (positions in nm are converted into Å, velocities in nm/ps are converted into Hartree atomic units)

Procedure “is_hydroxyles” returns 1 if the input file contains only hydroxyl groups, otherwise 0.

Example of usage (reading .trr file):

```
use SFG_FRAMES

class(frame_reader), pointer :: fr

type(trr_frame_reader), allocatable, target :: trr

type(gro_frame_reader), allocatable, target :: gro

type(xyz_frame_reader), allocatable, target :: xyz

allocate(trr) !allocate only trr_frame_reader

fr => trr !assign trr_frame_reader to frame_reader

call fr%open_file(30, “file.trr”, “file.gro”) !file unit, input files, opens the files

call fr%skip_frame() !skips the 1st frame

call fr%read_frame() !reads the 2nd frame

print*, fr%molecule(m)%o%position(:) !prints position vector of the m-th water oxygen
```

2.2 Calculation of instantaneous surface

For later assignment of water molecules to individual interfacial layers we need to calculate the instantaneous surface of the water slab. The original code [13] implemented calculation of instantaneous surfaces by calculation of coarse-grained density.

$$\bar{\rho}(\boldsymbol{r}, t) = \sum_i \phi(|\boldsymbol{r} - \boldsymbol{r}_i(t)|; \xi) \tag{20}$$

$$\phi(\boldsymbol{r}; \xi) = \frac{exp\left(-\frac{|r|^2}{2\xi^2}\right)}{(2\pi\xi^2)^{3/2}} \quad , \tag{21}$$

where $\boldsymbol{r}_i$(t) is position of i-th oxygen in the system, coarse graining length ξ = 2.4 Å, which is approximately equal to the molecular diameter of liquid water. As instantaneous surface we call all the points in space that satisfy the condition $\overline{\rho} = 0.016\text{Å}^{-3}$, which is approximately half the bulk density of water according to [16]. Volume of the simulation box has been discretized equidistantly in each direction, namely 0.5 Å for *X* and *Y* axes, and 0.25 Å for *Z* axis). Coarse-grained density was then evaluated as a function of height (*Z*), instantaneous surface was then identified by the local minimum of $|\overline{\rho} - 0.016|$ in positive (upper interface) and negative (bottom interface) *Z* coordinates (input trajectory had to be translated in *Z* axis to have 0 at approximately half of the height of the water slab), looping through each (*X,Y*) discretized point of the simulation box at given time (Fig. 4).

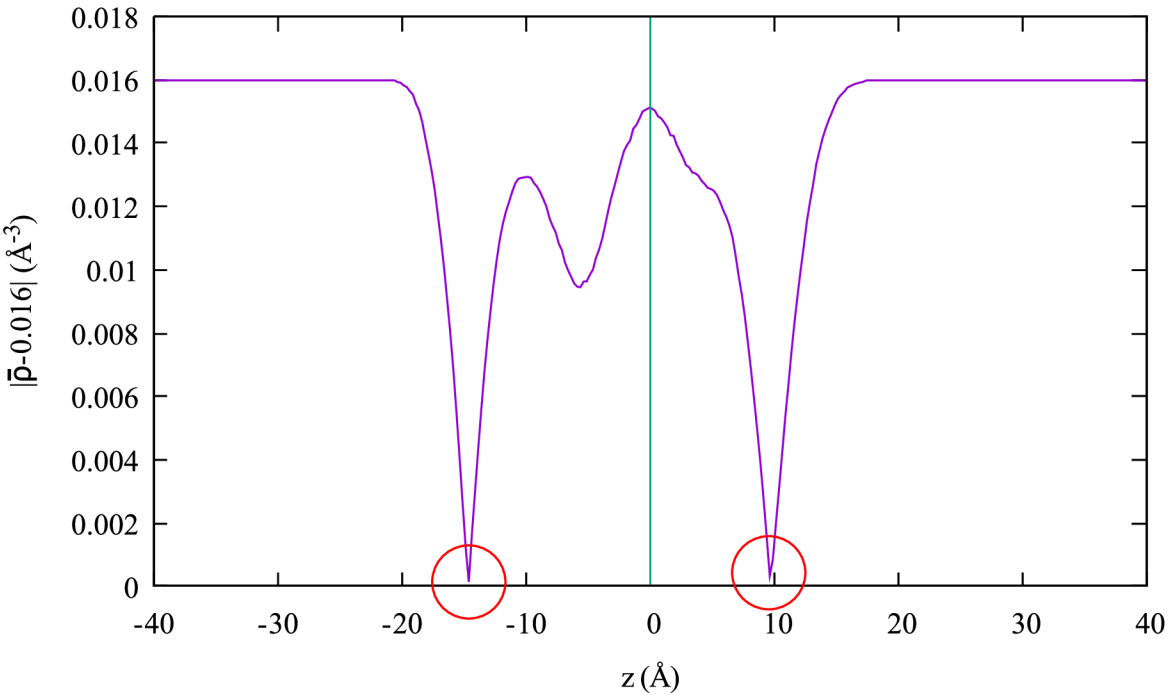

Figure 4: Original algorithm for searching instantaneous surface points.

Because of short input trajectory time step (0.4 fs) molecules do not have enough time for movement to significantly change the instantaneous surface of water bulk from one step to another. Therefore, instantaneous surface does not need to be evaluated in each time step (program switch `-nskip parameter` was present to skip `parameter` of timesteps to evaluate next instantaneous surface). The instantaneous surfaces were then saved in an ASCII file called "interface.xyz", which had typical .xyz file format storing every interface point found (if the `-nskip parameter` switch was set, the same interface was then copied `parameter` times since the further processing code required to read the stored interface for each molecular frame read).

2.2.1 Improvement

To speed up the process of finding instantaneous surface I have updated the algorithm. Instead of calculating the coarse-grained density over the whole height range, $|\overline{\rho} - 0.016|$ is evaluated from bottom of the simulation box until occurrence of positive non-zero derivative (Fig. 5), then the evaluation begins from the top of the simulation box (that way the input

trajectory does not need to be translated in Z axis to have 0 at approximately half of the height of the water slab)

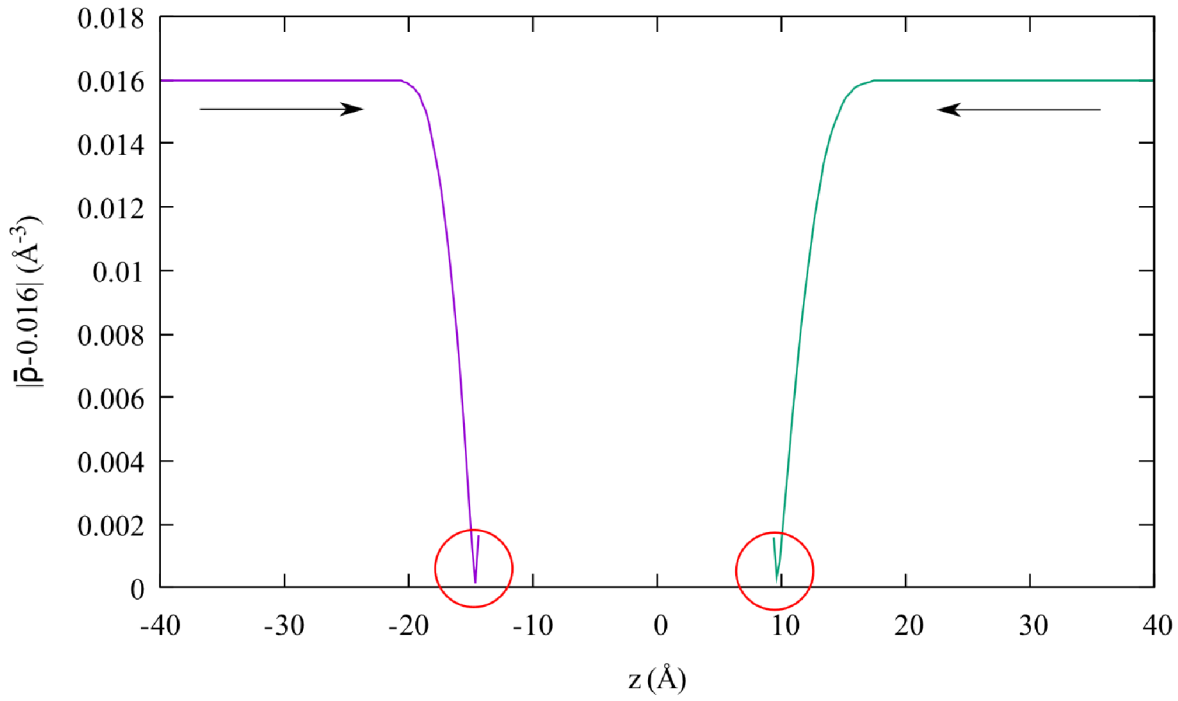

Figure 5: Improved algorithm for searching instantaneous surface points.

File "interface.xyz" is now generated only if the user wants it, for further needs the file is replaced by a file "interfacebin.xyz", which stores only one copy of each instantaneous surface in binary format ([8-byte integer] - number of points, number of points×3 [8-byte real] - x, y, z coordinate of each instantaneous surface point). Parameter $NSKIP was added to the "BOXDATA" file.

2.3 Calculation of density function

To further assign water molecules into interfacial layers [7] we need to calculate density profile as a function of the minimal distance of water oxygen atom from the interface (see Fig. 6) $r = |\mathbf{r}_{min}|$. The case when a molecule occurs under the interface (considering we are looking at the bottom interface, molecule containing oxygen O1 is considered to be under the interface) is distinguished by assigning a negative value $r = -|\mathbf{r}_{min}|$.

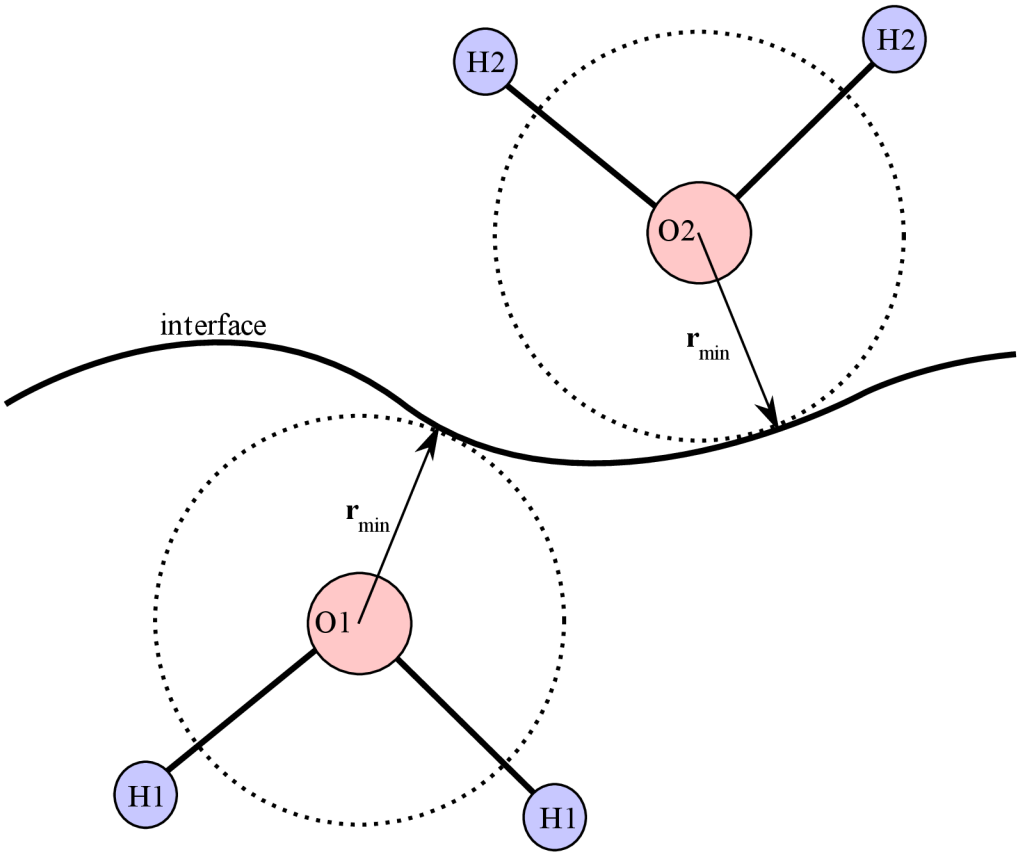

Figure 6: Schematic picture of the searching for the closest distance of a water molecule from the interface.

The density function is then calculated according to [13] as:

$$\rho(r) = \frac{N_{\text{bin}}}{V_{\text{bin}} \cdot n_{\text{step}} \cdot 0.035} \tag{22}$$

Where N_{bin} is the number of molecules belonging to a bin centered at r (r ranges from -4 Å to 25 Å with a step 0.2 Å), with width 2ε (from r-ε to r+ε) counted through all time steps (n_{step}) of the input trajectory file, ε = 0.7 Å. V_{bin} is the volume of the bin (box area in X-Y plane × 2ε), 0.035 is approximately bulk density of water (Å^{-3}). The density profiles (from bottom interface, upper interface, and average from both interfaces) were stored in 2 column ASCII files called "down_density_vs_r", "up_density_vs_r" and "density_vs_r".

2.3.1 Improvement

The function of original program [13] was to read the "interface.xyz" file generated by (2.2) and evaluate the density function over the whole input trajectory file. I have included

calculation of density function into the program calculating the instantaneous surfaces; extra reading of the "interface.xyz" file was therefore eliminated.

2.4 Calculation of SFG response function

The original program [13] applies the calculations presented in chapter 1.3. To begin the program calculates all the D matrices for all the O-H bonds in each time step of the trajectory file. After that, velocities v_z are evaluated for each O-H bond in each time step. Then the $\dot{\alpha}_{xx}$ and $\dot{\mu}_z$ is evaluated for each molecule in each time step of the simulation. Finally, the SFG response function $R^{(2)}_{xxz}(t) = \langle \dot{A}_{xx}(t)\dot{M}_z(0) \rangle$ is evaluated according to equation (up to the total simulation time $t = t_{sim}$), normalized by factor

$$\frac{average\ number\ of\ molecules\ in\ the\ selected\ layer}{X - Y\ area\ of\ the\ simulation\ box\ \cdot number\ of\ contributions(t)}$$

so, the result is independent of the sample box size and saved to the file to be processed in the next step. Binder file is taken into calculation selecting only molecules $M = M_{binder}$ (equation 13), that fulfills condition of being present at least part of the total trajectory time (BOXDATA file parameter $TARGETL0 – for the layer 0, and $TARGET – for all the other layers $\frac{\text{time to be present in layer}}{\text{total trajectory time}}$) in the selected layer.

2.4.1 Construction of D matrices and evaluation of v_z

D matrices are direction cosine matrices projecting O-H bond frames (Z axis is in direction of O-H bond, X axis is in the plane of water molecule, facing away of the second O-H bond and Y axis is perpendicular to the molecular plane) into laboratory frame (Fig.7).

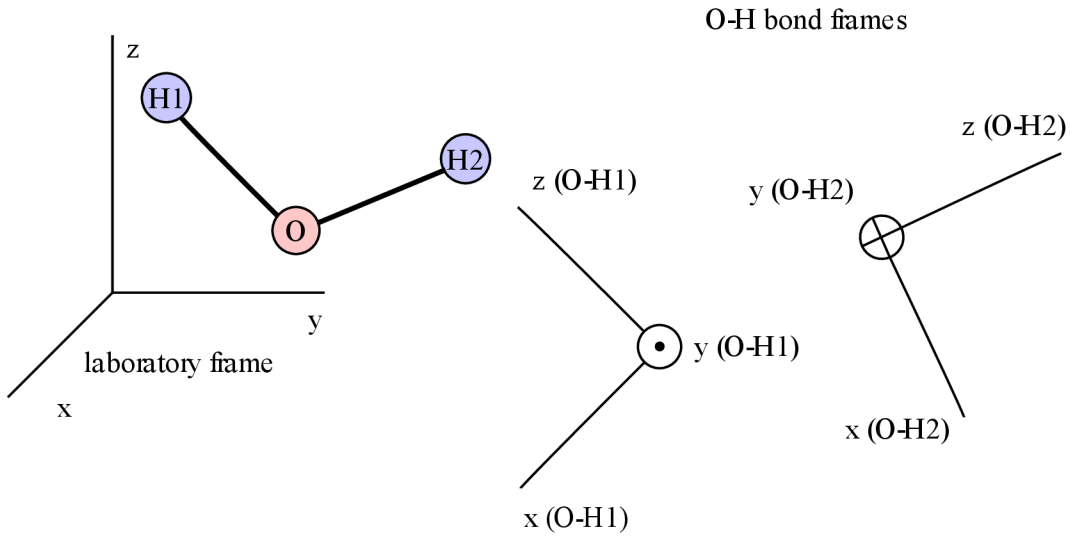

Figure 7: Comparison between laboratory frame and O-H bond frame.

The direction cosine matrix has form of:

$$\mathbf{D} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{23}$$

where the vectors $\mathbf{a}_{*1}$, $\mathbf{a}_{*2}$, and $\mathbf{a}_{*3}$ correspond to the laboratory coordinates of the O-H bond frame unit axes x, y, z, shifted to the origin, respectively.

Let the $\mathbf{r}_O$, $\mathbf{r}_{H1}$ and $\mathbf{r}_{H2}$ vectors be position vectors of atoms in water molecule in the laboratory frame and consider calculation of D matrix projecting ($O\text{-}H_1$) frame into the laboratory frame.

The $\mathbf{a}_{*3}$ vector in matrix **D** is a unit vector in the (0,0,0) point in the laboratory frame given by $\frac{\mathbf{r}_{H1}-\mathbf{r}_O}{|\mathbf{r}_{H1}-\mathbf{r}_O|}$.

To calculate $\mathbf{a}_{*1}$ vector we assume $\mathbf{u} = \mathbf{a}_{*3}$, $\mathbf{v} = \mathbf{r}_{H2} - \mathbf{r}_O$. Since the vector $\mathbf{u}$ is a unit vector, we can get projection of vector $\mathbf{v}$ on the vector $\mathbf{u}$ multiplying $\mathbf{u}(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\cdot|\mathbf{u}|\cdot|\mathbf{v}|\cdot\cos(\varphi) = \mathbf{u}\cdot|\mathbf{v}|\cdot\cos(\varphi)$. If we construct vector $\mathbf{w} = \mathbf{u}(\mathbf{u}\cdot\mathbf{v}) - \mathbf{v}$, it will be perpendicular to the vector $\mathbf{u}$ in the molecular plane (see Fig. 8). Finally, vector $\mathbf{w}$ must be normalized and assigned to $\mathbf{a}_{*1} = \frac{\mathbf{w}}{|\mathbf{w}|}$.

Vector $\mathbf{a}_{*2}$ is then cross product of unit vectors $\mathbf{a}_{*3}$ and $\mathbf{a}_{*1}$. Since both $\mathbf{a}_{*3}$ and $\mathbf{a}_{*1}$ are unit vectors and they are perpendicular to each other $\mathbf{a}_{*2} = \mathbf{a}_{*3} \times \mathbf{a}_{*1}$ will be by default a unit vector.

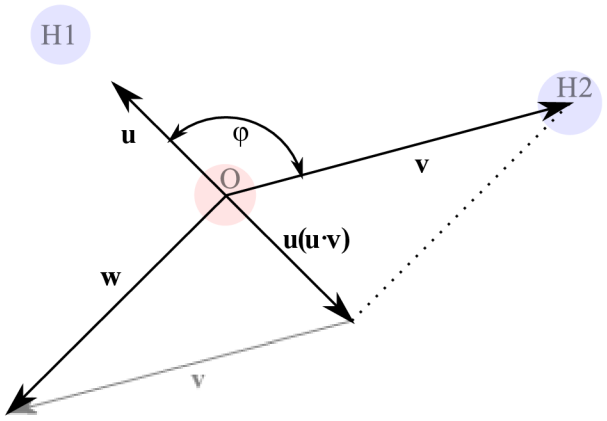

Figure 8: Geometric idea behind the D matrix calculation.

Properties of the D matrix ($D^T = D^{-1}$) allows us to easily transform vectors from O-H frame to the laboratory frame and vice versa. Projection of velocity on the O-H bond axis is simply obtained by $v_z = \mathbf{D}_{*3}\cdot(\mathbf{v}_H - \mathbf{v}_O)$, where $\mathbf{v}_H$ and $\mathbf{v}_O$ are velocities of H and O atoms in the laboratory frame.

2.4.2 Improvement

D matrices were calculated in advance for each O-H bond at each frame to later calculate $\dot{\alpha}_{xx}$ and $\dot{\mu}_z$, which led to unnecessarily large memory consumption (that proved to be a problem for larger systems and longer input trajectories). Now D matrices are calculated on the fly while calculating $\dot{\alpha}_{xx}$ and $\dot{\mu}_z$.

Since the original program [13] was created to process much shorter trajectories, the original idea of binder file had to be revisited. Longer trajectories led to the serious omission of many

molecule's contributions since the target layer occupation time (of some molecules) did not meet the requirements due to molecules being able to diffuse to multiple interfacial layers. New binder file stores layer affiliation of each molecule in each time step. Such a file in the previous binder file format would consume a lot of memory. Considering we sort molecules into four layers from each interface, each molecule can be labeled with 4-bit number (0-7) and its correspondence to the interfacial layer (L0-L3 either from the bottom or top interface) stored in binary file one label after each other since the number of oxygens (water molecules) in the system is known (user input to the BOXDATA file). In case of odd number of water molecules, the last nibble of each frame is redundant. In this way the new binder file stores one frame after each other. The length of the evaluated SFG response function has impact on the time requirements of its calculation. The original code evaluated the SFG response function up to $t = t_{sim}$, therefore a new parameter in BOXDATA file called $CORRLEN (in ps) was added to set calculation limit. As shown below, only limited length of the correlation function needs to be evaluated due to its decay, augmented by an applied filter to reduce the longer-time noise, producing undesired undulations of the resulting spectra.

2.4.3 Inclusion of intermolecular terms

Summation (14) from Khatib [6] can be rewritten as:

$$\dot{M}_R(t) = \sum_{m'}^{M} \sum_{n'}^{N_{m'}} \dot{\mu}_{m'n',R}(t) \tag{24}$$

The $\dot{A}_{PQ}\dot{M}_R$ product from the equation (10) can be then expanded:

$$\begin{aligned}
\dot{A}_{PQ}(t)\dot{M}_R(0) &= \left\{\sum_{m=1}^{M}\sum_{n=1}^{N_m} \dot{\alpha}_{mn,PQ}(t)\right\}\left\{\sum_{m'=1}^{M}\sum_{n'=1}^{N_{m'}} \dot{\mu}_{m'n',R}(0)\right\} \\
&= \left\{\sum_{m=1}^{M}\left[\sum_{n=1}^{N_m} \dot{\alpha}_{mn,PQ}(t)\right]\right\}\left\{\sum_{m'=1}^{M}\left[\sum_{n'=1}^{N_{m'}} \dot{\mu}_{m'n',R}(0)\right]\right\} \\
&= \sum_{m=1}^{M}\sum_{m'=1}^{M} \dot{\alpha}_{m,PQ}(t)\dot{\mu}_{m',R}(0) \\
&= \sum_{m=1}^{M} \dot{\alpha}_{m,PQ}(t)\dot{\mu}_{m',R}(0) \\
&+ \sum_{m=1}^{M} \sum_{\substack{m'=1 \\ m'\neq m \\ r_{\min}\leq r_{cut}}}^{M} \dot{\alpha}_{m,PQ}(t)\dot{\mu}_{m',R}(0) = T_{self} + T_{cross}
\end{aligned} \tag{25}$$

where $\dot{\alpha}_{m,PQ}$ and $\dot{\mu}_{m',R}$ represents the sum of the contributions (the square brackets above) over both bonds of a water molecule, yielding the contribution from a single molecule to the overall summation over all molecules. T_self stands for self-correlation terms and T_{cross} stands for cross-correlation terms.

The original programs [13] did not include calculation of SFG response function using these cross-correlation terms. The new program implements calculation of these cross-correlation terms.

Inclusion of the intermolecular cross-correlation terms is memory (program loads all the oxygen atom positions in each timestep of the trajectory into memory) and time inefficient (compared to the correlation function, one more loop through all the molecules is

implemented, evaluation of the truncating function also plays role in the time consumption, since the distances between the oxygen atoms needs to be calculated).

2.4.4 Application with the surface hydroxyls

The described approach focuses on calculation and interpretation of O-H vibrations, allowing evaluations of orientations O-H bonds and strengths of hydrogen bonds among molecules and/or hydrogen bonds with surfaces. The spectral interval of interest is approximately 3000–3800 cm^{-1}. While contributions from water molecules dominates this region, other O-H bonds, if present, can also contribute to the spectrum. If the surface becomes hydroxylated in contact with aqueous solution, the surface hydroxyls contribute to the total SFG spectrum. This is the case of e.g., SiO_2 terminated by surface hydroxyls called silanols (SiOH), similarly for TiO_2 (TiOH) or Al_2O_3 (AlOH).

The code "ssp_solid" for calculating SFG spectra of the surface hydroxyl groups was also included in the original programs [13]. Both "ssp_solid" and "ssp_sfg_selfcorr" implemented the same calculations [6] from section 1.3, "ssp_solid" program had different $\frac{d\mu_i}{dr_z}$ and $\frac{d\alpha_{ij}}{dr_z}$ parameters (see Table 1) and different approach to reading the input files. The hydroxyl groups were searched in the trajectory by finding closest metal atom to the oxygen and closest hydrogen to that oxygen (since *ab initio* MD simulations do not preserve bonds between atoms, the connectivity cannot be reliably stored in a topology file). The rest of the program calculated the SFG response function the same way same as "ssp_sfg_selfcorr" program, implementing equations from [6]. The binder file is irrelevant in this case because the surface hydroxyls are by nature contributing to the layer 0. Since in the CMD simulations atom bonds are specified, it is easy to reorganize the input file such way that the hydroxyl position and velocities can be read into water-like structures (metal-O-H instead of H-O-H), which is now implemented in the frame_reader.

	$\frac{d\mu_x}{dr_z}$	$\frac{d\mu_y}{dr_z}$	$\frac{d\mu_z}{dr_z}$	$\frac{d\alpha_{xx}}{dr_z}$	$\frac{d\alpha_{yy}}{dr_z}$	$\frac{d\alpha_{zz}}{dr_z}$	$\frac{d\alpha_{xy}}{dr_z}$	$\frac{d\alpha_{xz}}{dr_z}$	$\frac{d\alpha_{yz}}{dr_z}$
"ssp_sfg_selfcorr" parameters according to [6]	0.005	-0.047	1.028	1.168	-0.247	-1.331	0.319	0.281	3.444
"ssp_solid" according to Pezzotti code [13]	-0.62	0.4	3.9	0.4	0.24	2.1	-0.07	0.07	0.19

Table 1 Parameters used by "ssp__sfg_selfcorr" and "ssp_solid" programs.

2.5 Analysis of the SFG response function

The original code [13] implemented equation (12) in a modified version:

$$\chi_{PQR}^{(2),R} = \frac{i}{k_B T \omega} \int_0^{+\infty} f(t; t_{sim}) e^{i\omega t} \langle \dot{A}_{PQ}(t) \dot{M}_R(0) \rangle dt \tag{26}$$

where SFG response function is multiplied by filter function $f(t; t_{sim})$.

$$f(t; t_{sim}) = \exp\left(-\frac{100}{2}\left(\frac{t}{t_{sim}/2}\right)^2\right) \tag{27}$$

S. Pezzotti in his work never described the filter function and its parameters. He only described that he processed SFG spectra from 15-100 ps long trajectories [14] [1]. In the picture below, we can see influence of filter (assuming $t_{sim} = 15$ ps) on a SFG response function of fluorite/water interface BIL (simulation with 118 water molecules) and the shape of the filter function.

Approximating the prefactor 100/2 by 49, the filter can be expressed as

$$f(t; t_{sim}) = \exp\left(-\left(\frac{t}{t_{sim}/14}\right)^2\right) \tag{28}$$

and Fig. 9 indeed shows that at times around $t = 1$ ps the filter damps the function by a factor 1/e.

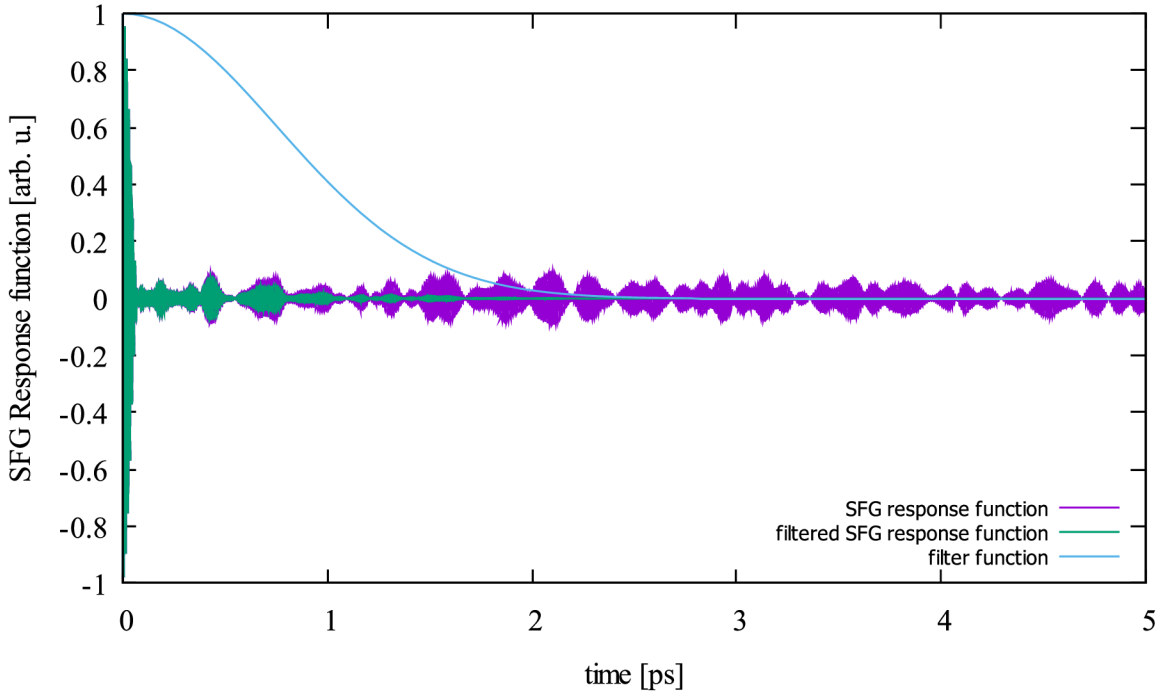

Figure 9: Influence of the filter on the SFG response function.

2.5.1 Improvement

The filter function has impact on the final SFG spectra (will be discussed later); therefore, the filter function was slightly modified:

$$f(t; t_{sim}) = \exp\left(-\frac{100}{2}\left(\frac{t}{p}\right)^2\right) \tag{29}$$

the user has an option to set the strength of the filter function by setting the parameter p.

2.6 Program features

2.6.1 General improvements

Frequently repeated parts of the original code were encapsulated in subroutines (for example frame_density_function – calculating frame's contribution to the density function), all the calculations with vectors are now made on the level of vector variables (the original programs [13] made all the vector calculations by vector components).

Input and output files of the programs have been revisited, necessary files for the calculation were optimized in size, generation of unnecessary files (that took up a lot of disk space) for the calculation is now optional. As demonstrated in Table 2 for the system of Fluorite-water-vacuum (with 118 water molecules and 50 000 frames in the trajectory), the necessary disk space for this task has been reduced 17 times, which becomes essential particularly for even bigger systems.

Original programs		New programs	
input files	size [kb]	input files	size [kb]
pos_rebuilt.xyz	884236	.xyz files	1612902
vel_mol_rebuilt.xyz	728666	.gro file	1196547
BOXDATA	1	.trr + 1frame.gro file	420736
		BOXDATA	1
output files	size [kb]	output files	size [kb]
interface.xyz	4550928	interfacebin.xyz	145586
grid_interface	1	grid_interface	1
density_vs_r file	7	density_vs_r file	7
16 × binder file	48	newbinder	2881
layers population	1	OH_REF	(optional) 3117090
OH_REF	3117090	tot_dAXX_dt	3
tot_dAXX_dt	3	tot_dMZ_dt	3
tot_dMZ_dt	3	z_vel_values	(optional) 420606
z_vel_values	420606	selfcorr file	2295
selfcorr file	2295	fft.dat	286
fft.dat	286	spectrum file	286
spectrum file	286	squarespectrum file	72
squarespectrum file	72		
total	9704529	minimum total	572157

Table 2: Comparison of files and their sizes used by the programs (original and new).

The CPU time needed for the whole set of programs to calculate the SFG spectra has been reduced roughly 2 times see Table 3 (the test runs were performed on the same machine, using trajectories of the same system described earlier in this section)

t_{sim}	200ps	20ps		200ps	20ps
original programs	t [h:mm:ss]	t [h:mm:ss]	new programs	t [h:mm:ss]	t [h:mm:ss]
interface	1:12:06	0:08:30	interface	0:54:26	0:05:28
density_vs_r	0:04:30	0:00:30			
create_binder	0:04:31	0:00:30	binder	0:09:18	0:00:52
ssp_sfg_selfcorr	3:36:54	0:02:31	SSP_CORR	2:23:38	0:01:30
vsfg_signal_morita	0:00:12	0:00:02	SFG_SIGNAL	0:00:13	0:00:02
total	4:58:13	0:12:03		3:27:35	0:07:52

Table 3 Comparison of CPU time used by the programs (original and new)

Programs (especially the SSP_CORR) implement the options to tweak the calculation parameters by the user using command line switches.

2.6.2 Program options

In this section all the program command line switches will be introduced.

Interface

- `-i arg1 (optional)arg2`

  Takes up to two arguments. Pay attention to the order of input files

  in case of .xyz files: position file + velocity file

  in case of .gro file: just .gro file containing whole trajectory

  in case of .trr file: .trr file + .gro file containing at least one frame of the trajectory
- `(optional) -vmdout`

  Program will also create interface.xyz file described in section 2.2

Binder

- `-i arg1 (optional)arg2`

  Takes up to two arguments. Pay attention to the order of input files

  in case of .xyz files: position file + velocity file

  in case of .gro file: just .gro file containing whole trajectory

  in case of .trr file: .trr file + .gro file containing at least one frame of the trajectory

SSP_CORR

- `-i arg1 (optional)arg2`

  Takes up to two arguments. Pay attention to the order of input files

  in case of .xyz files: position file + velocity file

  in case of .gro file: just .gro file containing whole trajectory

  in case of .trr file: .trr file + .gro file containing at least one frame of the trajectory

- (optional) -o arg

  Takes output file name, default output name is set to selfcorr.xyz

- -L arg1 arg2

  Selection of layers to be analyzed according to the following Table 4

layer	arg
upper layer 0	0
upper layer 1	1
upper layer 2	2
upper layer 3	3
bottom layer 0	4
bottom layer 1	5
bottom layer 2	6
bottom layer 3	7

Table 4: Meaning of new binder file layer labels.

  for example, we want to analyze only upper layer 1: `-L 1 1`

  yet another example, we want to analyze bottom layer 0 through bottom layer 2: `-L 4 6`

- `(optional) -refout`

  OH_ref file will be created, storing all the **D** matrices in ASCII format

- `(optional) -zvel`

  z_vel_values file will be created, storing all the v_z O-H bond velocities in ASCII format

- `(optional) -cross arg`

  Program will also calculate SFG response function including the cross-correlation terms with truncating radius `arg` (in Å)

- `(optional) -deriv arg`

  Takes name of file where $\frac{d\mu_i}{dr_z}$ and $\frac{d\alpha_{ij}}{dr_z}$ parameters are stored in ASCII file $\frac{d\mu_x}{dr_z}$, $\frac{d\mu_y}{dr_z}$, $\frac{d\mu_z}{dr_z}$, $\frac{d\alpha_{xx}}{dr_z}$, $\frac{d\alpha_{yy}}{dr_z}$, $\frac{d\alpha_{zz}}{dr_z}$, $\frac{d\alpha_{xy}}{dr_z}$, $\frac{d\alpha_{xz}}{dr_z}$ and $\frac{d\alpha_{yz}}{dr_z}$ in respective order, one line per parameter. If not selected, the parameters will be by default set to water molecule parameters according to [6].

SFG_SIGNAL

- `-i arg`

  Takes the correlation function file name

- `(optional) -filter arg`

  Sets the filter parameter to `arg` (in number of steps), by default the parameter is set to $p = \frac{\$CORRLEN}{2 \cdot timestep}$

- `(optional) -o arg`

  Takes output file name, default output name is set to spektrum.xyz

2.6.3 Execution

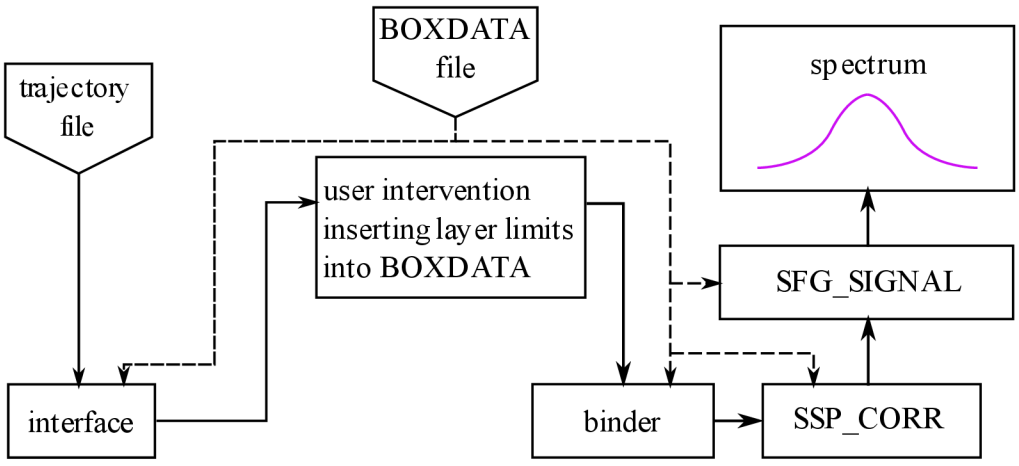

Figure 10: Workflow of the updated programs by author.

The sequence of execution the codes is now as follows (Fig. 10):

1) Calculation of the instantaneous surfaces

```
./interface.x -i trajectory.trr frame.gro
```

2) Assignment of water molecules to interfacial layers based on the user-defined boundaries

```
./binder.x -i trajectory.trr frame.gro
```

3) Calculation of correlation function

```
./SSP_CORR.x -i trajectory.trr frame.gro -L X X
```

4) Fourier transform yielding the spectra

```
./SFG_SIGNAL.x -i selfcorr.xyz
```

3 Results and discussion

3.1 Reconstruction of previous experiments

I have verified proper implementation and modification of the original code by calculating SFG spectra of the water molecules on fluorite(111)/water interface. Initial configuration of the system system was provided by Ondřej Kroutil. Results based on analysis of this particular system simulated by classical molecular dynamics (CMD) were reported in Ondřej Kroutil's work [15] together with experimental data and ab initio molecular dynamics (AIMD) based data published by Khatib et *al*. [6]. Simulated system is composed of the slab of fluorite(111) with dimensions 11.58 x 13.38 Å with thickness around 15.1 Å (5 layers), 118 water molecules placed above the surface leaving around 38 Å of vacuum above the surface of water (Fig. 14A). Positive surface corresponding to low pH conditions was prepared by moving one fluorine ion from the surface to approximately to the middle of the water slab. Trajectories used in this thesis were generated using Gromacs software [17] and running the Kroutil's initial configurations in NVE ensemble with 0.4 fs time step. This short time step is necessary to capture the O-H vibrations. To allow postprocessing of the trajectory to calculate the polarization tensor/dipole moment correlation function, the positions and velocities of water molecules were saved in each time step. NVE ensemble is routinely used in these kinds of simulations not to perturb velocities by the thermostat. I have made comparison between results of the original programs [13] and the new modified programs presented in this thesis.

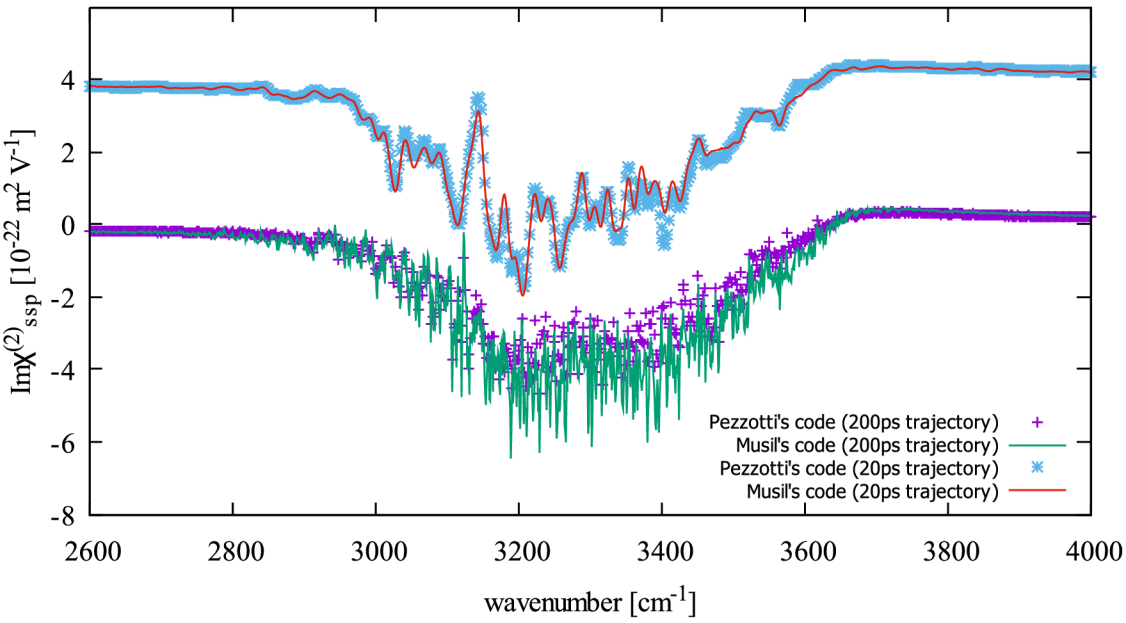

Figure 11: Comparison of the imaginary parts of the fluorite(111)/water SFG spectra computed with the original code and the modified one.

The Fig. 11 proves that the modifications made have no major impact on the final calculated spectra (a slight difference is expected due to the modified binder, which now assigns molecules to interfacial layer dynamically based on their positions at a given time). Note that parameters $CORRLEN and filter parameter (described in section 2.5.1) were set accordingly to the original program [13]. On the 200ps trajectory we can see more noise modulated on the signal, this can be eliminated by proper setting of $CORRLEN and proper filter parameter, as will be discussed later in the Results and Discussion section. Using the new binder, each molecule at each time step contributes to the calculations. The original code considered contributions only from molecules residing in the interfacial layer of interest for at least 70% of the simulation time.

3.2 Influence of simulation parameters

In the following section, I describe how different parameters of the simulations or the modified SFG code can affect the overall shape of the spectra. All the spectra in this section are calculated from the same trajectories as described in section 3.1.

3.2.1 Length of simulation

Ondřej Kroutil [15] presented influence of the length of the simulation on the final spectra (Fig. 12) using the original programs [13]. I have reconstructed his experiments using the modified programs. In the Fig. 12 – Kroutil, the lines are vertically offset by 3 and gray lines are obtained using running averages with the length of 25 samples. Spectra calculated by the modified program were obtained with $CORRLEN = 20 ps, and filter parameter 25000. Similarly, to the Kroutil's results, even in my results (see Fig. 13) the amplitude of the fluctuations in the spectrum decreases with the increasing simulation time, spectrum is less noisy with increasing analyzed length and the shape of the spectrum converge to the shape of the experimental curve.

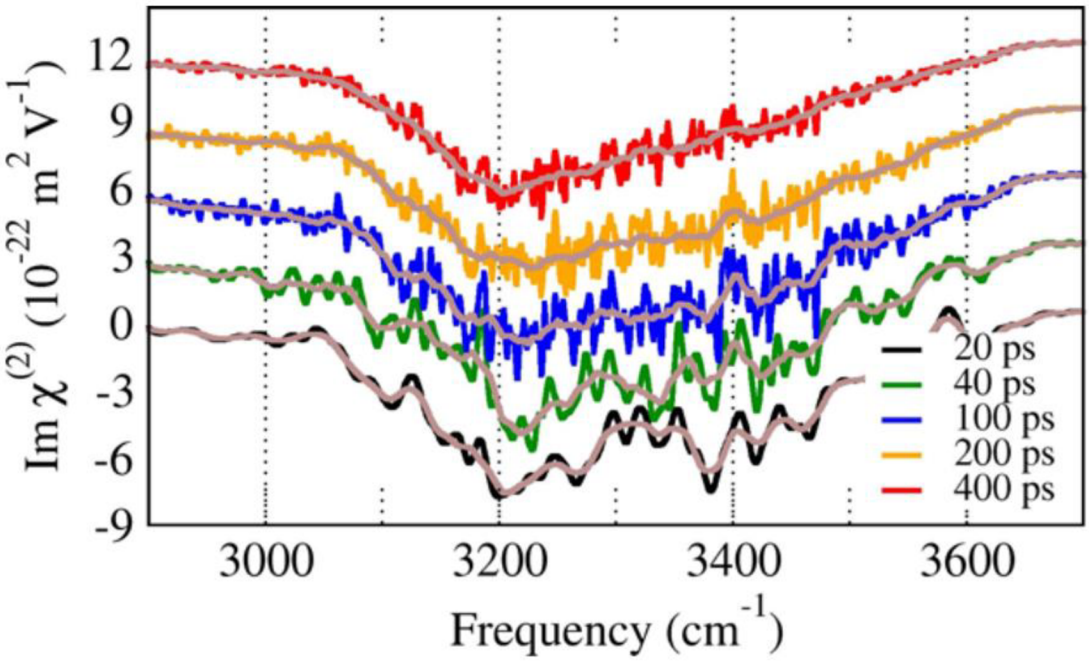

Figure 12: Impact of input simulation length on the final spectra using S. Pezzotti's programs. Taken from [15].

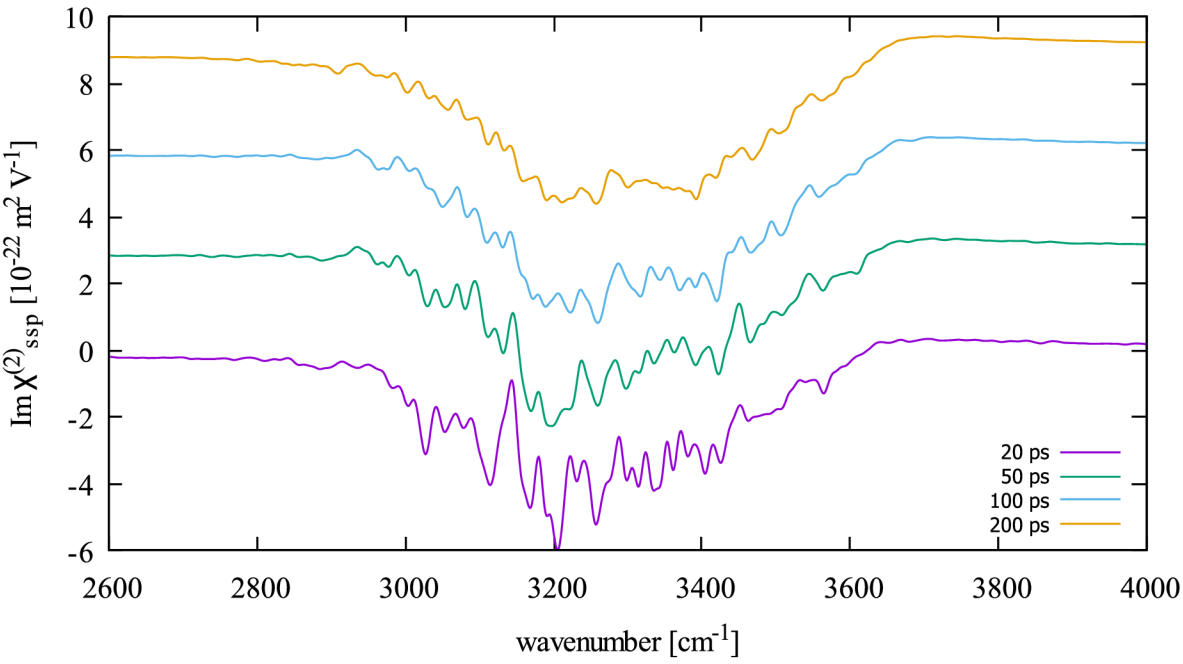

Figure 13: Impact of input simulation length on the final spectra using modified programs.

3.2.2 Size of the system

The major advantage of the CMD simulations over AIMD ones is accessibility of much larger systems (tens of nanometers) and much longer trajectories (hundreds of nanoseconds). For this reason, I have tested the new modified programs on the fluorite(111)/water system twice bigger in all directions (see Fig. 14B). Dimensions of this system were 23.16 × 26.76 Å and the number of water molecules was 471. For this bigger system, I have tested two values of the cut-off on electrostatics and van der Waals interactions: the first value 0.56 nm is the same as in the smaller system, and value 1.0 nm is the value usually used in simulations with standard force fields.

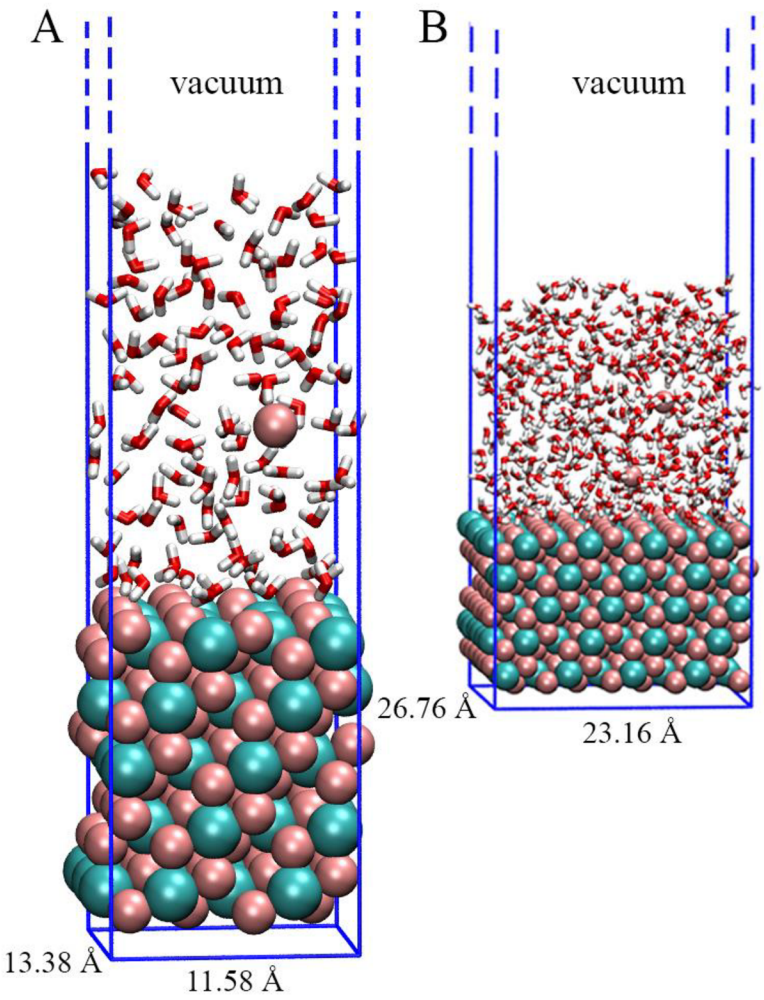

Figure 14: Fluorite(111)/water system setup. A) original system used in Kroutil's article; B) System twice bigger in all directions.

As can be seen from Fig. 15 the shapes of both spectra are very close to the one obtained in the smaller system, i.e., one negative peak that is experimentally observed around 3300 cm^{-1}. In the case of smaller system, the peak is located at ~3200 cm^{-1} while in the bigger system this peak is shifted to ~3400 cm^{-1}. One can notice small positive peak around 3100 cm^{-1} in the spectra generated from simulations of the bigger system. I have noticed that 2 out of 4 free fluorine atoms in the solution were incorporated back into the surface during the simulation time, thus decreasing surface charge from +4 to +2 and bringing it closer to the neutral pH form of the fluorite(111). As was shown in Khatib [6], imaginary part of the SFG spectrum of the neutral surface shows slightly positive and broad peak around 3000-3100

cm^{-1}. Thus, this peak could arise due to lower surface charge density and slightly different orientation of the water on the interface. While I confirmed that the processing of systems of different size produces similar spectra and the magnitude of the signal is properly normalized, the shift between the spectra determined for the small vs. system requires further inspections for other systems to determine if the shift is a size-dependent effect or artifact resulting from the change of the location of displaced fluor atoms during the simulation.

It can be also seen that different cut-off radius on electrostatic and van der Waals interactions does not influence the shape of the spectra, the positions of the peaks or the intensity of the signal.

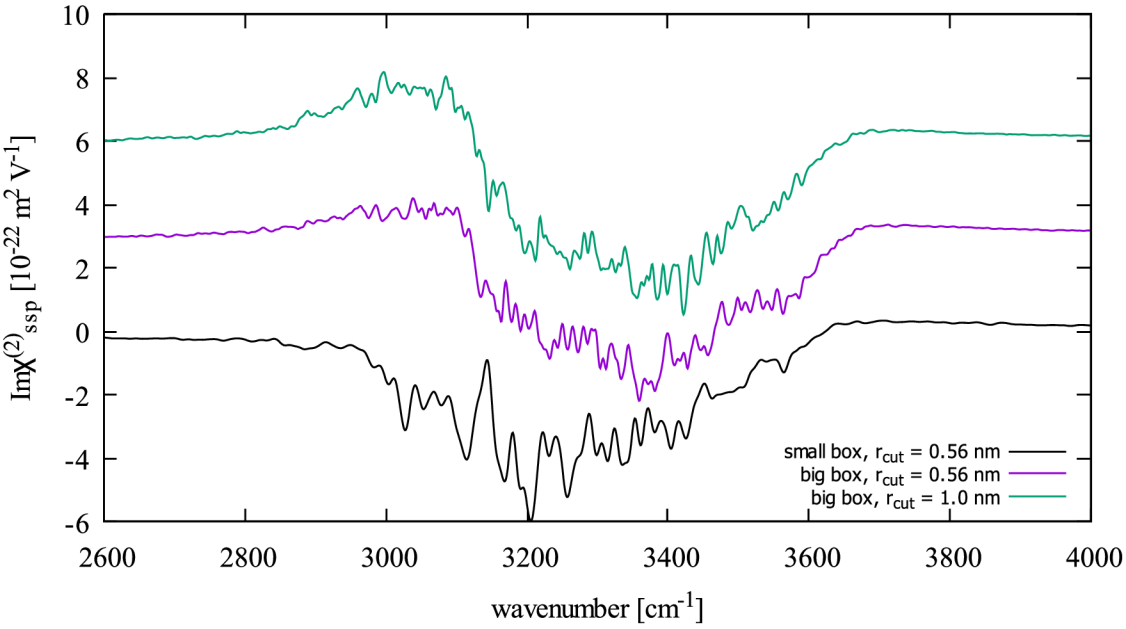

Figure 15: Influence of the system size and cut-off radius on electrostatic and van der Waals interactions.

3.2.3 Thermostat

As was stated elsewhere [15] [14] , velocities of water molecules or hydroxyl groups are crucial for the calculations of the SFG spectra in OH stretch region using eq. (12), and for this reason the NVE ensemble is routinely employed in the production runs. In general, thermostats act by modifying the velocities based on the difference of the instantaneous

kinetic temperature T_{kin} from the desired temperature T. While structural and thermodynamic properties are in general not sensitive to the choice of a thermostat and its parameters, it is known that thermostats can affect dynamic properties of the simulated systems. This is particularly relevant for the time correlation functions, as represented by eqs. (11) and (12). Note that the information on relative velocities of O and H atoms forming an intramolecular OH bond directly enter the eqs. (17-19). One must therefore suppose that thermostat can potentially change shape of the spectrum by perturbing velocities. To test this hypothesis, fluorite(111)/water system as described in section 3.1 was simulated in NVT ensemble with Nosé-Hoover thermostat [18] and different relaxation time constants for coupling, namely 1, 5, 10, 20 and 40 ps. Results are depicted on Fig. 16. Surprisingly, the overall shape of all NVT spectra follows the shape of the reference spectrum computed in NVE ensemble. Even the system with the shortest coupling constant does not shows any non-standard behavior. Although this topic should be examined more in depth (for example including other popular thermostats like v-rescale or Berendsen), these results suggest that running production runs in NVT ensemble with (at least) Nosé-Hoover thermostat should not influence final SFG spectra. This is a very positive message for the application of classical molecular dynamics to calculation of spectra. The *ab initio* molecular dynamics simulations (AIMD) typically last dozens of picoseconds due to the extreme time requirements of this technique. During such a short time the total energy is well conserved, and significant drift of the system temperature should not be observed. On the other hand, in classical MD (CMD) simulations lasting hundreds of picoseconds to dozens of nanoseconds, the temperature drift to numerical errors can be significant. There are however two positive pieces of information (i) the temperature calculated from CMD is less prone to deviations due to averaging over larger number of molecules compared to AIMD, and (ii) we are using time reversible Verlet-based integrator (Gromacs implements it via the leap-frog algorithm [17] [19]) which promotes energy conservation.

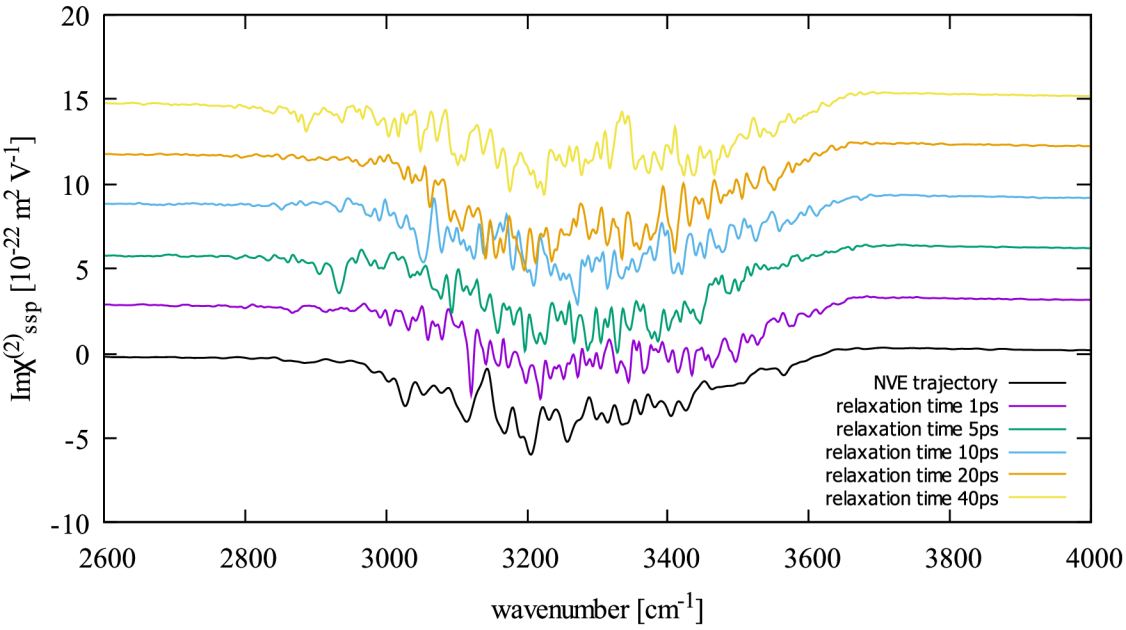

Figure 16: Comparison of the Im$(\chi^{(2)})$ for systems simulated in NVE and NVT ensembles, respectively. Different relaxation times are used in NVT simulations.

3.3 Influence of the length of the SFG response function and filter

All the spectra calculated in this section were calculated from 20 ps (if not stated other) trajectory file using the modified programs. The effect of $CORRLEN and filter parameters will be presented and discussed.

Firstly, according to the filter (presented in section 2.5) present in the original code [13], the SFG response function is dampened by the filter quite drastically. Calculation parameter $CORRRLEN was added to the calculation, allowing us not to calculate the SFG response function up to $t = t_{sim}$, the main reason was to speed up the calculation since the SFG response function is dampened by the filter and long-time contributions become negligible – either due to the underlying physics or due to the filter designed to reduce the noise. Of course, reduction of the time over which the correlation function is calculated becomes more essential in processing long CMD trajectories, as opposed to relatively short AIMD trajectories. Figure 17 shows the effect of the $CORRLEN parameter fixing the filter

parameter to 18750 as assumed in section 2.5. Note that the spectra are shifted vertically by 3.

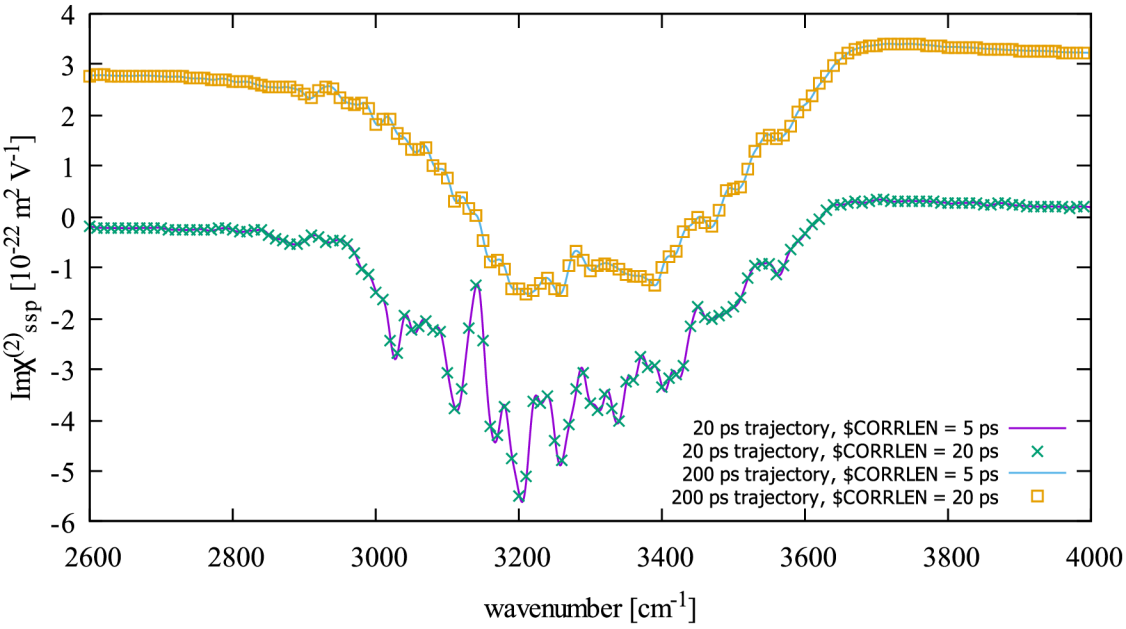

Figure 17: Effect of the $CORRLEN parameter on the final spectra assuming the same filter parameter.

We can see that the restriction made by $CORRLEN parameter makes sense to speed up the calculation, because the filter effectively cuts off the SFG response function and the length of the SFG response function does not play crucial role in the form of the final spectra.

Secondly, experiments with the filter itself were made with the $CORRLEN parameter fixed to the value 20ps; results are shown in Fig. 18 and Fig. 19.

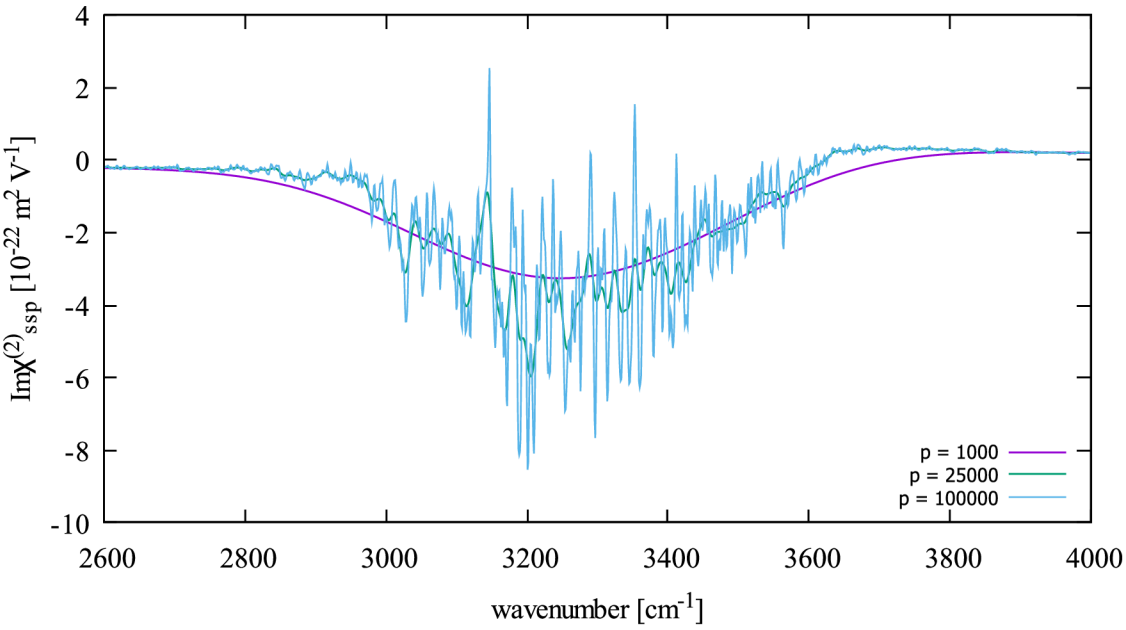

Figure 18: Effect of the filter on the final spectra (extreme values for demonstration).

In Fig. 18 demonstration of filter with extreme parameters is shown in comparison with the filter set accordingly to the parameters used in the original code [13] (-filter 25000). We can see that the filter with parameter 1000 (the smaller the value, the more severe damping, see eq. (29)) impacts the amplitude of the spectrum and smooths out the potential features of interest of the spectrum, on the other hand filter with parameter 100000 introduces too much noise to the spectrum, making it difficult to recognize the key characteristics of the spectrum.

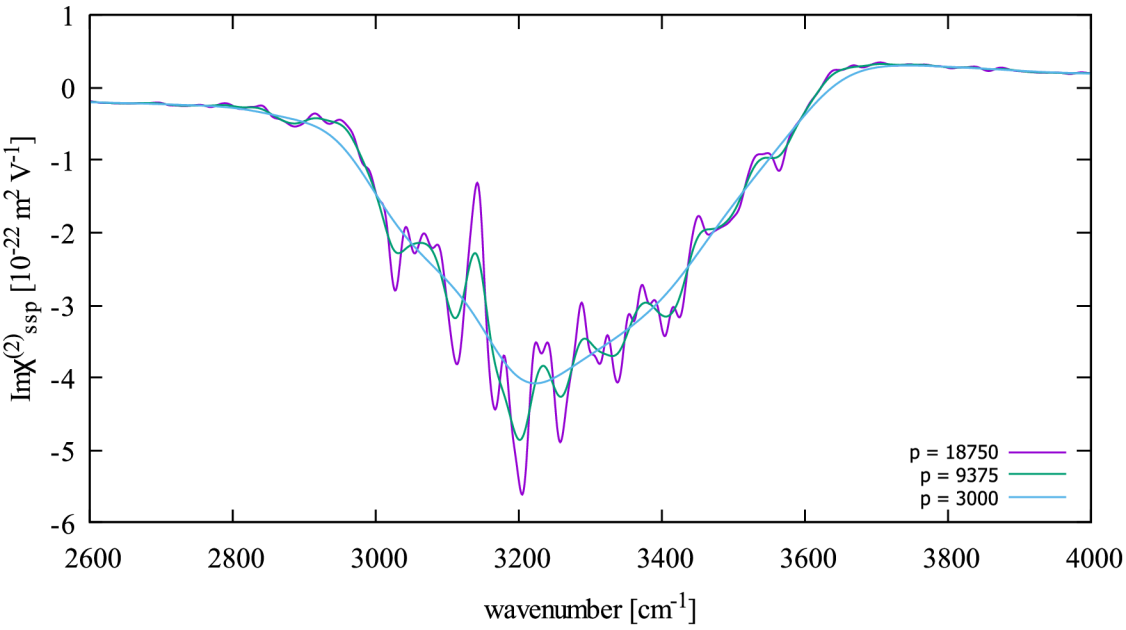

Figure 19: Effect of the filter on the final spectra (more precise value adjustments).

In Fig. 19 more gentle adjustment of the filter parameter is demonstrated. Filter with value 9375 simulation steps, i.e., 9375*0.4 fs = 3750 fs (equivalent of filter that would be used by the original program [13] processing 7.5 ps trajectory), returns the best results, the potential features of interest are conserved (an experienced user needs to distinguish the significant features from the noise artifacts), and the amplitude of the signal is not decreased dramatically. We can see that the filter parameter plays significant role on the calculated spectra. To discuss the filter values and get a rule of thumb on the essential time range to be evaluated on the correlation function, we must understand that the SFG response function reflects the vibrations of the O-H bond stretch region of the spectra centered at around 3200 cm^{-1}. This wavenumber corresponds to frequency of roughly 96 THz, and reciprocal value of that results in a period of vibrations roughly 10 fs (note that it corresponds to only 25 simulation steps of 0.4 fs). Due to the historical prefactor 100/2 in the filter (see eq. (22)), the filter parameter is made about 7 times more aggressive in damping, i.e., the preferred parameter about 9375 simulation steps, i.e. 3750 fs, effectively reduces any function to 1/e of its value within about 535 fs, which corresponds to about 50 vibrational cycles. We see that with this parameter significant number of vibrations are captured, while at the same time allowing to reduce the time span over which the correlation function is evaluated.

3.4 Effect of including contributions from surface hydroxyls

I have verified proper implementation and modification of the original code by calculating SFG spectra of the hydroxyls on quartz(0001) surface. Almost 40ps long *ab-initio* MD trajectory of the mentioned surface cut had been provided to us by Marie-Pierre Gaigeot's group. Results based on analysis of this particular trajectory are summarized in the article by Pezzotti et al. [14] System is composed of the slab of quartz(0001) with dimensions 9.820 × 8.504 Å and is surrounded by 64 water molecules (Fig. 20A).

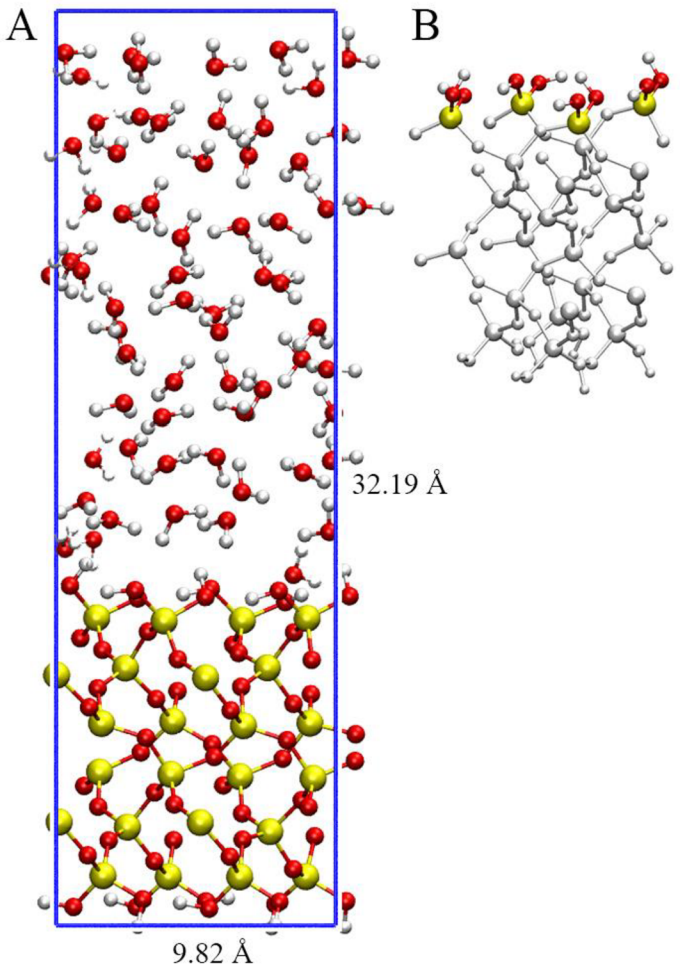

Figure 20: Quartz(0001)/water AIMD system setup. A) Whole system; B) Geminal silanols on upper surface highlighted.

Each face contains four geminal silanol groups (see Fig. 20B). I have analyzed O-H stretch of four upper silanols. SFG spectra calculated with Pezzotti's code and my modification using `-deriv` parameters according to Cyran [20] (Table 5):

$\frac{d\mu_x}{dr_z}$	$\frac{d\mu_y}{dr_z}$	$\frac{d\mu_z}{dr_z}$	$\frac{d\alpha_{xx}}{dr_z}$	$\frac{d\alpha_{yy}}{dr_z}$	$\frac{d\alpha_{zz}}{dr_z}$	$\frac{d\alpha_{xy}}{dr_z}$	$\frac{d\alpha_{xz}}{dr_z}$	$\frac{d\alpha_{yz}}{dr_z}$
0.005	-0.047	1.028	1.168	-0.247	-1.331	0.319	0.281	3.444

Table 5: Dipole moment (D.Å^{-1}) and polarizability (Å^2) derivative of the O-H bond in silanol groups. The results are given within the bond frame.

are shown on Fig. 21. During analysis, I have realized inconsistency between results I get from my code and Pezzotti's code (purple vs. green line). Careful analysis revealed that this inconsistency arises from translational movement of the surface slab in the simulation box. Compared to classical MD simulations, ab initio MD simulations are usually done without any restrains on movement of any atom and system as whole slowly translates in the simulation box. And this translational movement was not eliminated from the analyzed trajectory. After centering trajectory using command:

```
gmx trjconv -s trj.tpr -f trj.gro -o trj-centered.gro -n index-OH.ndx -pbc whole -center
```

and computing SFG spectra of the O-H groups again, I get exactly the same results using my code (purple line) and very similar one using Pezzotti's code (blue line). This shows that bug in Pezzotti's code was eliminated in my implementation. For comparison, I have included digitalized SFG spectrum of the OH stretch from Pezzotti et al. article [14]. It shows good agreement between currently computed and published data, although current data were not smoothed and just one surface was included in analyses (i.e., worse statistics).

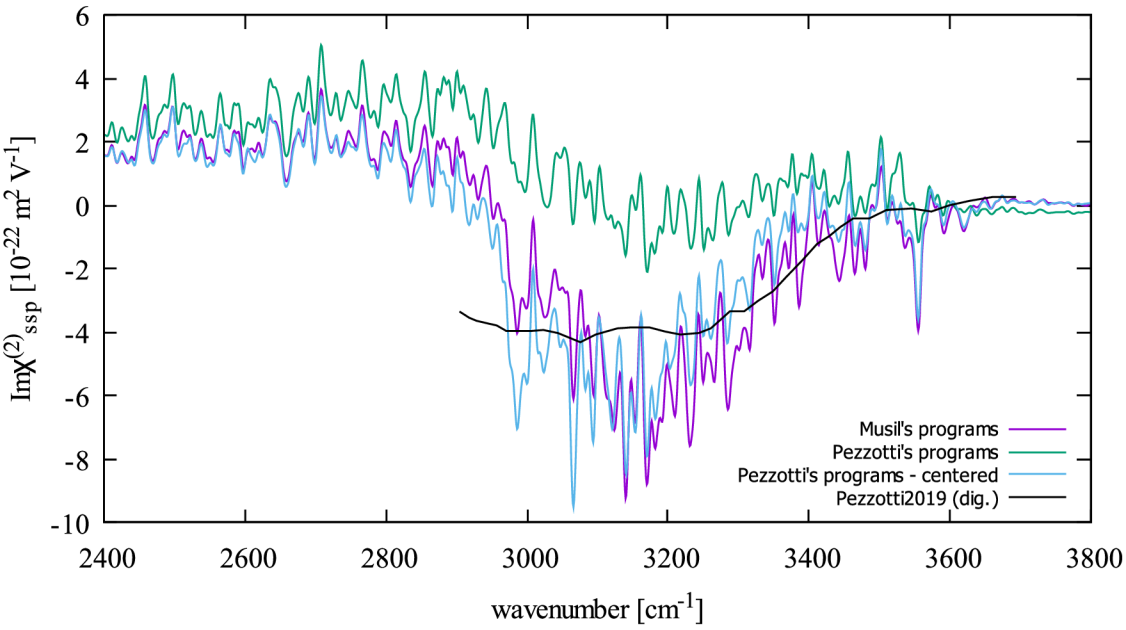

Figure 21: SFG spectra of the geminal hydroxyl groups on quartz(0001) surface. System simulated with AIMD.

3.5 Effect of including intermolecular terms

I have calculated SFG spectra from 20 ps long trajectories of fluorite(111)/water system (section 3.1) with newly implemented cross-correlation terms in the SFG response function that were not available in the original code. The truncating radii r_{cut} (as described in section 2.4.3) were set to 2, 3 and 4 Å. Note that r_{cut} > 4 Å has no meaning due to the size of the simulation box and possible inclusion of all the intermolecular interactions along the X-Y plane. On the other hand, calculations with parameter r_{cut} < 2 Å are not performed because of the O-O radial distribution of water [21]. For the r_{cut} below this value no water pairs are present. Effect of an inclusion of cross-correlation terms is depicted in Fig. 22. More detailed investigation of inclusion of the cross-correlation terms should be subject of further research and is beyond the scope of this theses.

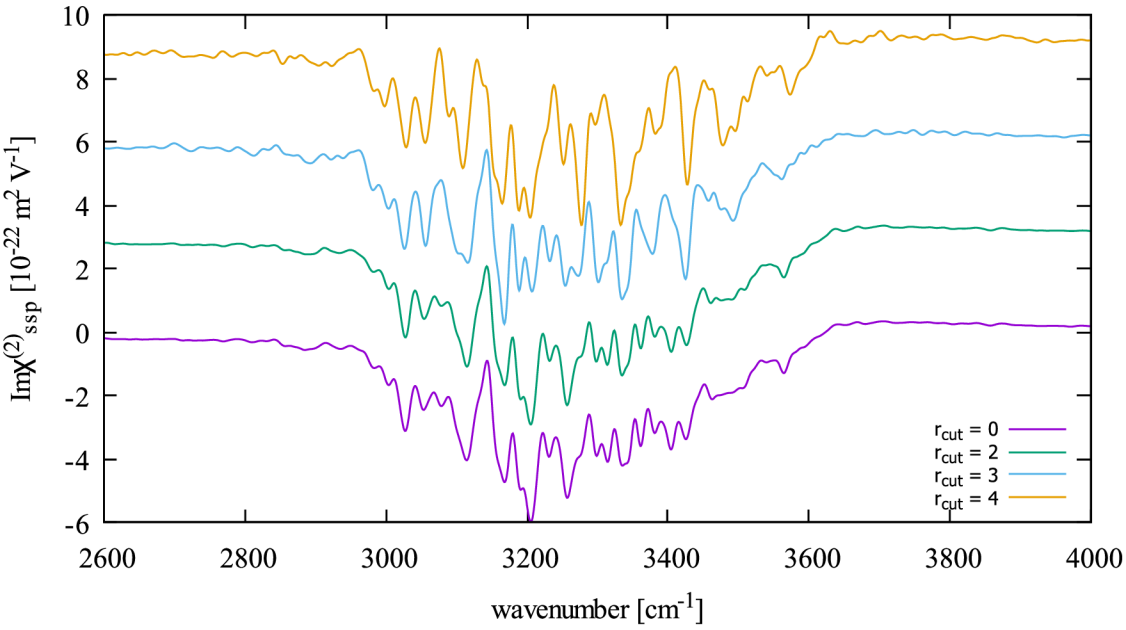

Figure 22: SFG spectra calculated from the cross-correlated SFG response function.

4 Conclusions

The goals set in the Assignment of master thesis form were:

> Refine and document the program developed by S. Pezzoti to calculate Sum Frequency Generation (SFG) spectra generated at aqueous solution interfaces (in contact with solid or gas phases) and evaluate the effect of added extensions on the resulting spectra. Specifically:
>
> 1) Optimizing the storage of working files in terms of data size and computational time.
> 2) Incorporating the contribution of surface hydroxyls into the determined spectra
> 3) Direct inclusion of the mutual polarizability of water molecules

All the goals were reached and representative results showing the functionality of the modified programs were obtained and presented. Significant part of the documentation is contained directly in the source codes, which moreover became much more legible and concise due to the use of objective FORTRAN features, structure datatypes and use of subroutines.

The scientific discussion of the resulting spectra was not subject of this thesis – moreover that discussion is present in the publication by Kroutil at al. [15], whose data I have reproduced.

I present however the discussion of the new features added to the program and their performance.

All the input and output files of the programs have been revisited, necessary files for the calculation were optimized in size and generation of unnecessary files is now optional. It was shown on particular fluorite(111)/water system that the necessary disk space for this task has been reduced 17 times, The CPU time needed to calculate the final spectra was reduced 2 times, which becomes essential especially for even bigger systems.

My new modification of the original “ssp_solid” code that is used to compute contributions of the surface hydroxyls to SFG spectra removed bug in the original code, included reading of Gromacs .trr trajectories and fairly reproduced published spectra.

The resulting code will be shared with the group of Prof. Marie-Pierre Gaigeot, where the classical molecular simulations are used now for systems unreachable by *ab initio* molecular dynamics (in addition to continuation of using the latter for smaller systems).

Finally (if not first of all) the performed modifications already fulfilled a partial goal of the Czech Science Foundation project 22-02972S "Computer modeling of nonlinear optics signals at interfaces", principle investigator Milan Předota, which promised to study systematically the roles of surface hydroxyls and intermolecular cross-correlations for a range of systems – a goal made now feasible with the help of the results of this thesis.

5 References

[1] PEZZOTTI, Simone, Daria GALIMBERTI, Y. SHEN a Marie-Pierre GAIGEOT. Structural definition of the BIL and DL: a new universal methodology to rationalize non-linear χ(2)(ω) SFG signals at charged interfaces, including χ(3)(ω) contributions: a new universal methodology to rationalize non-linear χ(2)(ω) SFG signals at charged interfaces, including χ(3)(ω) contributions. *Physical Chemistry Chemical Physics*. The Royal Society of Chemistry, 2018, **20**(7), 5190-5199. ISSN 1463-9076. Dostupné z: doi:10.1039/C7CP06110B

[2] SHEN, Y.R. Phase-Sensitive Sum-Frequency Spectroscopy. *Annual Review of Physical Chemistry*. Annual Reviews, 2013, **64**(1), 129-150. ISSN 0066-426X. Dostupné z: doi:10.1146/annurev-physchem-040412-110110

[3] LI, Chunfei. *Nonlinear optics*. 1. Singapore, Singapore: Springer, 2016, 386 s. ISBN 9789811014888.

[4] MORITA, Akihiro a Tatsuya ISHIYAMA. Recent progress in theoretical analysis of vibrational sum frequency generation spectroscopy. *Physical Chemistry Chemical Physics*. The Royal Society of Chemistry, 2008, **10**(38), 5801-5816. ISSN 1463-9076. Dostupné z: doi:10.1039/B808110G

[5] MORITA, Akihiro a James HYNES. A Theoretical Analysis of the Sum Frequency Generation Spectrum of the Water Surface. II. Time-Dependent Approach. *The Journal of Physical Chemistry B*. American Chemical Society, 2002, **106**(3), 673-685. ISSN 1520-6106. Dostupné z: doi:10.1021/jp0133438

[6] KHATIB, Rémi, Ellen BACKUS, Mischa BONN, María-José PEREZ-HARO, Marie-Pierre GAIGEOT a Marialore SULPIZI. Water orientation and hydrogen-bond structure at the fluorite/water interface. *Scientific Reports*. 2016, **6**(1), 24287. ISSN 2045-2322. Dostupné z: doi:10.1038/srep24287

[7] PEZZOTTI, Simone, Daria GALIMBERTI, Y. SHEN a Marie-Pierre GAIGEOT. What the Diffuse Layer (DL) Reveals in Non-Linear SFG Spectroscopy. *Minerals*. 2018, **8**, 305. Dostupné z: doi:10.3390/min8070305

[8] SHEN, Y. R. Surface properties probed by second-harmonic and sum-frequency generation. *Nature*. 1989, **337**(6207), 519-525. ISSN 1476-4687. Dostupné z: doi:10.1038/337519a0

[9] HOSSEINPOUR, Saman, Fujie TANG, Fenglong WANG et al. Chemisorbed and Physisorbed Water at the TiO2/Water Interface. *The Journal of Physical Chemistry Letters*. American Chemical Society, 2017, **8**(10), 2195-2199. Dostupné z: doi:10.1021/acs.jpclett.7b00564

[10] OSTROVERKHOV, Victor, Glenn A. WAYCHUNAS a Y. R. SHEN. New Information on Water Interfacial Structure Revealed by Phase-Sensitive Surface Spectroscopy. *Physical Review Letters*. American Physical Society, 2005, **94**(4), 046102--. Dostupné z: doi:10.1103/PhysRevLett.94.046102

[11] URASHIMA, Shu-hei, Anton MYALITSIN, Satoshi NIHONYANAGI a Tahei TAHARA. The Topmost Water Structure at a Charged Silica/Aqueous Interface Revealed by Heterodyne-Detected Vibrational Sum Frequency Generation Spectroscopy. *The Journal of Physical Chemistry Letters*. American Chemical Society, 2018, **9**(14), 4109-4114. Dostupné z: doi:10.1021/acs.jpclett.8b01650

[12] SMIRNOV, Konstantin S. Structure and sum-frequency generation spectra of water on neutral hydroxylated silica surfaces. *Physical Chemistry Chemical Physics*. The Royal Society of Chemistry, 2021, **23**(11), 6929-6949. ISSN 1463-9076. Dostupné z: doi:10.1039/D0CP06465C

[13] PEZZOTTI, Simone. *Programs*. provided through private correspondence, 2018.

[14] PEZZOTTI, Simone, Daria GALIMBERTI a Marie-Pierre GAIGEOT. Deconvolution of BIL-SFG and DL-SFG spectroscopic signals reveals order/disorder of water at the elusive aqueous silica interface. *Physical Chemistry Chemical Physics*. The Royal

Society of Chemistry, 2019, **21**(40), 22188-22202. ISSN 1463-9076. Dostupné z: doi:10.1039/C9CP02766A

[15] KROUTIL, Ondřej, Simone PEZZOTTI, Marie-Pierre GAIGEOT a Milan PŘEDOTA. Phase-Sensitive Vibrational SFG Spectra from Simple Classical Force Field Molecular Dynamics Simulations. *The Journal of Physical Chemistry C*. American Chemical Society, 2020, **124**(28), 15253-15263. ISSN 1932-7447. Dostupné z: doi:10.1021/acs.jpcc.0c03576

[16] WILLARD, Adam a David CHANDLER. Instantaneous Liquid Interfaces. *The Journal of Physical Chemistry B*. American Chemical Society, 2010, **114**(5), 1954-1958. ISSN 1520-6106. Dostupné z: doi:10.1021/jp909219k

[17] ABRAHAM, Mark, Teemu MURTOLA, Roland SCHULZ, Szilárd PÁLL, Jeremy SMITH, Berk HESS a Erik LINDAHL. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015, **1-2**, 19-25. ISSN 2352-7110. Dostupné z: doi:https://doi.org/10.1016/j.softx.2015.06.001

[18] FRENKEL, Daan a Berend SMIT. Rare Events. *Understanding Molecular Simulation*. Elsevier, 2002, s. 431-464. ISBN 9780122673511. Dostupné z: doi:10.1016/b978-012267351-1/50018-3

[19] FRENKEL, Daan a Berend SMIT. *Chapter 6 - Molecular Dynamics in Various Ensembles*. San Diego: Academic Press, 2002, s. 139-163. ISBN 978-0-12-267351-1. Dostupné z: doi:https://doi.org/10.1016/B978-012267351-1/50008-0

[20] D., Cyran, Donovan A., Vollmer DORIS et al. Molecular hydrophobicity at a macroscopically hydrophilic surface. *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 2019, **116**(5), 1520-1525. Dostupné z: doi:10.1073/pnas.1819000116

[21] SOPER, A.K. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chemical Physics*. 2000, **258**(2), 121-137. ISSN 0301-0104. Dostupné z: doi:https://doi.org/10.1016/S0301-0104(00)00179-8

6 Appendix

Fortran source codes:

- interface.f90
- Binder.f90
- SSP_CORR.f90
- SFG_SIGNAL.f90
- bdata.f90
- SFG_TYPES.f90
- SFG_FRAMES.f90

Sample input files for the calculations:

- Fluorite-water.trr
- Fluorite-water.gro
- BOXDATA

File containing commands to compile all the programs provided:

- compile_all