

### Second harmonic generation of non-resonant liquid: from the molecular response to the experimental measurement

Guillaume Le Breton

#### ► To cite this version:

Guillaume Le Breton. Second harmonic generation of non-resonant liquid : from the molecular response to the experimental measurement. Chemical Physics [physics.chem-ph]. Université Claude Bernard - Lyon I, 2022. English. NNT : 2022LYO10015 . tel-04022967

### HAL Id: tel-04022967 https://theses.hal.science/tel-04022967

Submitted on 10 Mar 2023  $\,$ 

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



### THESE DE DOCTORAT DE L'UNIVERSITE CLAUDE BERNARD LYON 1

École Doctorale N°52 Physique et Astrophysique de Lyon

Discipline : Physique & Chimie

Soutenue publiquement le 22/09/2022, par :

Guillaume Le Breton

Génération de Second Harmonique non-résonnant de liquides: de la Réponse Moléculaire à la Mesure Expérimentale

Second Harmonic Generation of non-resonant Liquid: from the Molecular Response to the Experimental Measurement

Devant le jury composé de :

NIEHAUS Thomas, Professeur des Universités, UCBL Lyon 1	Président du Jury
CHAMPAGNE Benoît, Professeur, Université de Namur	e Rapporteur
VUILLEUMIER Rodolphe, Professeur des Universités, École Normale Supérieur	Rapporteur
<b>CHARLAIX Elisabeth</b> , Professeur des Universités, Université Grenoble Alpes	Examinatrice
<b>DUBOISSET Julien</b> , Maître de Conférences, Université Aix-Marseille	Examinateur
<b>BENICHOU Emmanuel</b> , Professeur des Universités, UCBL Lyon 1	Co-Directeur de thèse
<b>BONHOMME Oriane</b> , Maître de Conférences, UCBL Lyon 1	Co-Directrice de thèse
LOISON Claire, Chargee de Recherche, UCBL Lyon 1	Co-Directrice de triese

À Madame Florence Gely qui a sut comment me mettre sur mon chemin. Comme promis maîtresse je suis devenu docteur (en physique)!

### Remerciements

Teamwork is not about the work nor the group composition: You can have all the perk and the right disposition, There will be no spark until everyone sees the same ship. Colleges are humans: we should starts with a smile at the beginning of a trip. And as passes time and miles, we will end up with friendship.

Weirldy enough (or like articles), this fist part is the last thing I am writting on this quite long document. Hence, I would like to address my final statement to any student how is considering starting a PhD with Claire, Oriane or Emmanuel: go ahead! They are, without any biais, the best supportive, humain and knowledgeable supervisors you can find. Performing a PhD is hard because you will care too much about what you want to do, what you do not know or what is beyound reach. They will help you reminding you what you have done, what is enough, and, what you should reasonably do. You will never feel alone, despite loockdown, despite lecture rushes, despite their own problems or deadlines, despite ours. With them, you will learn how to work by your own, to find and solve your problems and to explain your solutions. They will help you to trust in yourself, to dare, and will support you to travel and share your work. I have been excitated (almost) every day of my PhD before going to work. I wish you to have supervisors that want to be mentors.

J'ai également eu la chance de rencontrer de très nombreuses personnes qui m'ont beaucoup aidé dans des domaines très différents, et je ne vais en remercier ici que quelques-unes.

Merci Monsieur Nieto et Moreau professeurs au lycée qui m'ont fait découvrir la physique et les maths; Madame Corbel-Copin, Coheleach, Delebassée et Rogier, Monsieur Pippo et Saunois pour m'avoir poussé sans me casser lors de la prépa; à Prof. Chilà qui m'a permis de faire mon stage long à Hambourg, à Prof. Dufrêche et aux équipes de l'ENS de Lyon qui m'ont permis d'avoir ma bourse thèse.

I would like to warmly thank Dr. Cunha and Dr. Pr. Jansen for their guidance and help during my internship at Groningen. Merci pour ton aide, ton temps et tes très nombreux conseils Nicolas lors de mon stage à Hambourg. Je te suis reconnaissant pour ta patience et ta bonne humeur!

Je remercie Nicolas Reverdy pour ses conseils et sa gestion de ma première vraie expérience d'enseignement. Bravo encore pour la gestion en distanciel! Merci à Christophe Bonnet et à Cendrine Moskalenko de m'avoir permis de participer aux modules de Licence 3: c'était une super expérience qui me motive à continuer dans l'enseignement! Je remercie encore mes directeurs de thèses de m'avoir soutenu et conseillé vis-à-vis des enseignements.

Je remercie toute l'équipe administrative de l'ILM, aux agents d'entretien et aux agents de la restauration. Merci encore à Giles et Matthieu pour leur travail. J'ai beaucoup appris en bossant avec vous et vous avez toujours été à l'écoute pour faire au mieux! Merci Agnès pour ton aide, tes conseils et ta bonne humeur contagieuse! Merci Anne-Laure, Cécile et Marie pour votre disponibilité et votre aide tout au long du projet! Je remercie les équipes du PSMN qui ont toujours été disponibles.

Merci Laurent pour ton aide sur la MD et sur l'écriture d'article. J'ai apprécié travailler avec toi, même à distance. Tu m'as suivi tout au long de mon passage à l'ILM: j'essayerais d'appliquer tes conseils sur le monde de la recherche à l'avenir! Merci Kevin pour ta disponibilité et ta patience. J ai eu la tête dure pour comprendre le champs excité mais ça valait le coup!

Merci Estelle pour tes explications sur les lasers et Christophe pour ton support tout au long de la thèse. Merci de votre aide quand cela va mal et que l'on prend l'eau! Merci à Albdul Raman, Frank et Pierre pour nos discussions. Je n ai pas été très présent dans le groupe et je m'en excuse. Malgré cela vous m avez accueilli et je garderai de bons souvenirs de cette équipe théochem! Merci Pierre pour tes conseils sur l'optique et sur la recherche en général: on aurait tous aiméS te garder plus longtemps dans l'équipe!

Je remercie Docteur Zacharie Bahel pour son aide, ses explications et sa gentillesse. Encore félicitations à toi, et à Docteur Antonin Pardon. Je vous souhaite le meilleur pour la suite. Bravo également à Docteur Sébastien Crombez, et bon courage à Élisa, Fabien et Maxime qui sont les prochains!

Je tiens à remercier Pierre-François pour nos nombreuses discussions et son intérêt pour le travail que j'ai fait. J' ai apprécié ton effort de transparence quant au financement de l'équipe, au temps que tu as passé à expliquer certains rouages du monde académique et plus généralement pour ton franc parler.

Enfin, je remercie les membres du jury d'avoir accepté d'évaluer cette thèse et pour nos échanges lors de la soutenance. C'était pour moi un grand honneur d'être évalué par vous car j'estime beaucoup votre recherche et je sais votre place dans la communauté.

# Contents

Ι	Intro	luction	<b>21</b>
II	Conte	ext and Problematics	<b>25</b>
	II. A	Context	25
	II. B	Problematics and Roadmap	32
III	Descr	ibing the second harmonic generation of non-resonant liquid: from	
	QM/I	MM to Maxwell	37
	III. A	Introduction	38
	III. B	Microscopic response	42
		III. B.1 Exact Response Theory	43
		III. B.2 Approximate theory	58
	III. C	Microscopic-Mesoscopic links	76
		III. C.1 Fundamental electromagnetic field: refraction and exciting field $\ldots$	76
		III. C.2 From the individual response to the mesoscopic polarization $\mathbf{P}^{2\omega}$	81
		III. C.3 From $\mathbf{P}^{2\omega}$ to the detector: 3-layer model	88
	III. D	Conclusion: bulk and surface SHG specificities	93
$\mathbf{IV}$	How	to compute the hyperpolarizability in the liquid phase: FROG	97
	IV. A	Introduction	98
	IV. B	Overview of FRDG	100
		IV. B.1 Scientific presentation	100
		IV. B.2 Work flow and review	103
	IV. C	How to include the long-range electrostatic effect on $\beta$ ?	109
	IV. D	Conclusion and Perspectives	141
$\mathbf{V}$	Secon	d Harmonic Scattering	145
	V. A	Introduction: Experimental results & Context	146
		V. A.1 SHS experiment for neat water	146
		V. A.2 How to understand SHS results: scattering formalism	150
V. B Incoherent response: fluctuations matter			157
		V. B.1 Hypothesis 1: disregarding hyperpolarizability fluctuations	158
		V. B.2 Hypothesis 2: Taking into account independent fluctuations $\ldots \ldots$	158

		V. B.3 Hypothesis 3: Taking into account $\beta$ fluctuations and intramolecular	
		correlations $\ldots \ldots 1$	62
		V. B.4 Comparison between all $<<\mathscr{B}^2>>$ obtained with different hypotheses 1	63
	V. C	Short-Range Coherent Calculation	66
		V. C.1 Molecular based Coherent calculation	66
		V. C.2 Laboratory first hyperpolarizability spatial correlation function ${\cal C}(r)~$ . 1	70
		V. C.3 Impact of intermolecular hyperpolarizability correlations on the depo-	
		larization ratio $\ldots$	75
		V. C.4 First steps toward long-range correlations	78
	V. D	Conclusions	84
VI	Surfac	ze-SHG 18	87
	VI. A	Introduction	88
	VI. B	Neat water at the liquid-gas interface: how to compute $\chi$ at the dipolar level? . 1	92
	VI. C	Molecular Response at the Quadrupolar order	12
		VI. C.1 Vacuum and liquid phase	12
		VI. C.2 At the liquid-gas interface	20
	VI. D	Conclusion and Perspectives	26
VII	Towar	d Second Harmonic Generation in Confined Liquid 22	29
	VII. A	Introduction	30
	VII. B	Setup description	33
		VII. B.1 General presentation	33
		VII. B.2 Confinement procedure	39
		VII. B.3 SHG microscopy	44
	VII. C	Results	47
		VII. C.1 SHG at the lamella interface	47
		VII. C.2 $\Delta z$ measurement and mechanical stability $\ldots \ldots \ldots \ldots \ldots 2$	52
		VII. C.3 SHG of confined-like system	55
	VII. D	Conclusions	56
VIII	Concl	usions and Perspectives 2	57
IX	Appe	ndix 20	63
	IX. A	Chapter III	63
		IX. A.1 More on the Exact scheme	63
		IX. A.2 About Dunning basis set	71
		IX. A.3 About DFT framework	72
		IX. A.4 More on exciting field in the bulk phase	277
		IX. A.5 Maxwell equations and plane-wave formalism	80
	IX. B	Chapter IV	85

	IX. B.1	Appendix: More on FROG work flow	285
IX. C	Chapter	VI	293
	IX. C.1	How to get the electric field shined by a dipole at far distance: com-	
		parison between 2 formulas	293
	IX. C.2	How to compute $\langle \mathscr{B}^2 \rangle$ for a given $\beta$ ?	295
	IX. C.3	Are the $\beta$ component independent?	304
	IX. C.4	Y study	307
	IX. C.5	Deviation of $\mathscr{B}$ for a single frame $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	309
	IX. C.6	Coherent calculation with the same $\beta$ for all molecules $\ldots$	310
IX. D	Chapter	VI	312
	IX. D.1	Distribution of $\beta^{dq}$ in bulk	312
IX. E	Chapter	VII	313
	IX. E.1	Fundamental power control	313
	IX. E.2	Derivation of the interferometry:	314
	IX. E.3	Guanaco: numerical treatment of the spectral interferences:	316
	IX. E.4	Optical components characteristics	318

## Résumé

L'organisation des liquides aux échelles moléculaires joue un rôle primordial dans de nombreux phénomènes physiques, chimiques ou biologiques. La structure et la dynamique des liquides ont été intensivement étudiées, aussi bien expérimentalement, par des méthodes de diffusion de neutrons, rayons X ou de spectroscopie infrarouge, que théoriquement, via les théories de physique des liquides, ou bien numériquement par dynamique moléculaire par exemple. Dans ce contexte, nous avons appliqué la technique de Génération de Second Harmonique (SHG) pour sonder la structure de liquide de l'interface au volume. En effet, cette technique d'optique non linéaire est intrinsèquement sensible aux interfaces, mais également à l'organisation des espèces ou à leur environnement électrostatique. L'utilisation de la SHG est de plus en plus répandue même si de nombreuses questions subsistent quant à l'interprétation des résultats. En effet, la réponse individuelle des molécules est complexe et dépend fortement de son environnement.

L'apport principal de cette thèse est le développement et l'utilisation de modélisation multiéchelle dans le but de mieux comprendre les signaux SHG émit par des liquides, en volume ou aux interfaces. Cette thèse est le fruit d'une collaboration étroite entre plusieurs équipes de l'Institut Lumière Matière mêlant modélisation et expérience. Elle peut se décomposer suivant 2 parties: numérique et expérimentale.

La première partie s'intéresse au lien entre la réponse SHG des molécules individuelles et les signaux expérimentaux observés. Une méthode numérique multiéchelle a été développée et appliquée à l'eau pure, en volume et à l'interface liquide-gaz. Elle se base sur la résolution des degrés de liberté électroniques d'une molécule au niveau quantique tout en prenant en compte son environnement – c'est-à-dire l'effet de la phase condensée. Pour ce faire, j'ai effectué une implémentation robuste et automatique de cette méthode via le développement d'un code, nommé FROG. Ce travail a permis de déterminer de nombreuses valeurs microscopiques relatives à la réponse SHG de l'eau: la première hyperpolarisabilitée aux ordres dipolaire et quadrupolaire ainsi que la seconde hyperpolarisabilitée en phase liquide, et ceci aux fréquences optiques. Au-delà de la comparaison avec les résultats expérimentaux, nous avons testé et quantifié de nombreuses hypothèses couramment utilisées, parfois en les confirmant, en y apportant un nouvel éclairage ou en les réfutant. Enfin, nous avons pu reproduire quantitativement, et sans paramètre ajustable, des résultats expérimentaux de SHG en volume pour l'eau pure. La partie expérimentale propose un nouveau montage ayant pour but de confiner des liquides entre 2 surfaces de verre et de sonder ces systèmes par SHG. Ce montage se base sur le dispositif déjà bien établi d'Appareil à Force de Surface (SFA). L'apport de la microscopie SHG pourrait nous permettre d'obtenir de nouvelles informations quant à l'évolution de la structure des liquides lors du confinement. Aujourd'hui nous sommes capables de mesurer la réponse SHG de films de liquide d'épaisseur de l'ordre de la dizaine de micromètres. Ces résultats préliminaires sont prometteurs et confirment la faisabilité de ce montage.

## Abstract

Liquid organization plays a prior role in several physical, chemical or biological phenomena. Numerous studies address the liquid structure and dynamics, on the experimental side using neutrons scattering, X-rays or spectroscopic technic in general, or on the numerical side thanks to molecular dynamics or on the analytical side using theory of molecular fluids. In this context, we use Second Harmonic Generation (SHG) as a tool to probe the structure of liquids from interfaces to the bulk phase. Indeed, this non-linear optical process is intrinsically sensitive to interfaces and also very sensitive to the liquid structure or the electrostatic environment. For these reasons, SHG is more and more used in the literature. However, several problems and questions related to the interpretation of the experimental results are yet unsolved. The main reasons are that the individual molecular response is complex and fluctuates depending on its environment.

In that context, the major outcome of this thesis is the development and the use of a multi-scale numerical approach aiming at better understand the Second Harmonic signal emitted by liquids at interfaces or in the bulk phase. This thesis is the result of a close collaboration between several teams of the Institute for Light and Matter bringing experiments and simulation together. We can define two main sections for this manuscript: one numeric and one experimental.

The first part is devoted to the link between the individual response of molecules at the Second Harmonic and the signal collected experimentally. A multi-scale method has been developed and has been applied to both the volume and the liquid gas interface of pure water. The keystone of this approach is the calculation of the single molecule electronic response in SHG within a condensed phase: *i.e.* a Quantum mechanical calculation inside an electrostatic environment. To do so, I have coded a python software, named FROG, to achieve this task automatically. This work led to several results, one of them being the determination of several microscopic quantities relative to the SHG response of water molecules in the liquid phase: the first hyperpolarizability at the dipolar and quadrupolar level as well as the second hyperpolarizability at optical frequency. We have tested and quantified several widespread hypotheses. Sometimes we have asserted them, sometimes we have brought new insights and sometimes we have found them unjustified. Finally, we were able to reproduce some experimental results for pure water in the bulk phase without any fitting parameter.

The experimental part presents a new setup that aims to probe by SHG a liquid confined be-

tween two glass surfaces. This experiment is based on the well established Surface Force Apparatus (SFA). The goal of this SHG microscopy is to get new ways of probing the evolution of the liquid structure during such confinement. Today, we are able to collect Second Harmonic signal coming from liquid films of dozens of microns of width. These preliminary results are promising and confirm that this setup is experimentally accessible.

## Résumé étendu

Pour sonder la matière un physicien prend un marteau et frappe dessus: plus c'est fort mieux c'est! Un ancien (fabuleux) professeur dont nous tairons le nom

#### Introduction et Problématique:

Cette thèse s'intéresse à la structure de la matière en phase liquide et à ce que peut apporter la technique d'optique non linéaire de Génération de Second Harmonique pour la comprendre.

La phase fluide est aussi importante que complexe à appréhender: les interactions entre molécules sont suffisamment fortes pour induire une forte cohésion tout en présentant de larges fluctuations. Cette diversité permet un très grand nombre de conformations, ce qui va induire de grandes possibilités d'interactions et de réactions par exemple, mais rendant plus compliqué une vision "simple" de sa structure locale. De même, aux interfaces, la relation entre 2 phases joue un rôle central dans de très nombreux cas. On peut penser aux réactions en phase hétérogène ou aux transferts moléculaires aux membranes cellulaires par exemple.

Il existe de nombreux outils expérimentaux pour sonder la matière. Par exemple la diffusion de neutron permet d'obtenir la structure locale des fluides. Ou alors la mesure de tension de surface permet d'obtenir des informations sur les interfaces.

Depuis la démocratisation des lasers femtosecondes, l'optique non linéaire est également devenue un outil de plus en plus utilisé pour sonder les liquides. Pour le cas des phénomènes d'ordre 2 comme la Génération de Second Harmonique (SHG), la réponse est exaltée lorsque le système sondé est non centrosymétrique. Cela signifie que c'est une méthode intrinsèquement spécifique aux interfaces. De plus, cela rend également cet outil prometteur pour sonder les interactions intermoléculaires au sein de fluides: la moindre déviation à la centrosymétrie pouvant être mesurée.

Mesurer la réponse en SHG d'interfaces ou de liquides est un défi technique qui est maitrisé au

sein de l'équipe Optique Non-Linéaire et Interfaces de l'Institut Lumière Matière. Néanmoins, l'interprétation fine de ces signaux reste un problème pour toute la communauté. Aussi, cette thèse est le fruit d'une collaboration pluridisciplinaire qui inclut l'équipe de Physicochimie Théorique de ce même laboratoire permettant d'ajouter une approche de simulation moléculaire aux expériences. Ensembles, nous avons essayé d'apporter une vision multiéchelle de ce phénomène d'optique complexe dans les fluides. En premier lieu, l'objectif a été de mieux comprendre le lien entre structure et signal expérimental pour l'eau pure. Mais une motivation qui nous a également guidés était également de généraliser sur le long terme l'utilisation de la SHG pour comprendre des systèmes plus complexes.

Pour répondre à de telles questions, mon travail se situe à l'interface entre la physique et la chimie, la modélisation et l'expérience. L'apport principal de mon travail est le développement numérique, et plus particulièrement une approche multiéchelle allant de la réponse quantique des molécules à la mesure expérimentale. En complément, j'ai également participé activement au développement d'un nouveau montage expérimental.

#### Présentation des Chapitres:

Les Chapitres I et II sont une introduction au domaine de la SHG dans les liquides et la présentation des questions auxquelles nous allons essayer de répondre. Le Chapitre III est une présentation de la majorité des concepts théoriques utilisés lors de ma thèse. Notre approche est basée avant tout sur une description moléculaire de la réponse optique. Dans une première partie, la réponse d'une molécule unique en SHG doit être décrite en résolvant les degrés de liberté électroniques. Dans une seconde partie, la réponse collective est déterminée par les équations de Maxwell. À la fin de ce Chapitre III, nous aurons présenté les grandes lignes qui permettent de lier la réponse d'une molécule, représentée par son tenseur d'hyperpolarisabilité  $\beta$ , au signal expérimental en fonction du champ électrique fondamental incident.

Le Chapitre IV présente comment nous avons calculé le tenseur de première hyperpolarisabilitée,  $\beta$ , d'une molécule en phase condensée. En effet, dans le cas de l'eau, cette quantité moléculaire est très dépendante de l'environnement électrostatique autour de chaque molécule. Plusieurs travaux récents de la littérature proposent une approche multiéchelle mêlant Mécanique Quantique et environnement électrostatique classique (QM/MM) pour calculer cette quantité de façon précise et numériquement abordable. Nous présentons notre article dans lequel nous proposons d'éclairer la façon dont on peut prendre en compte cet environnement électrostatique, notamment à longue portée. Ce développement méthodologique nous a également permis d'améliorer notre façon d'obtenir le  $\beta$  en phase condensée dans notre propre code nomé FROG. En effet, nous avons conçu cet outil numérique afin d'automatiser le calcul de  $\beta$  à partir de Dynamiques Moléculaires (MD) classiques. FROG permet de calculer les propriétés moléculaires liées à la SHG aussi bien en volume qu'aux interfaces puisque ces 2 géométries sont très utilisées expérimentalement. Enfin, FROG a pour vocation d'être partagé en libre accès afin de rendre cette méthode QM/MM plus accessible à la communauté. Le Chapitre V décrit nos travaux sur la génération de second harmonique en volume de l'eau pure, aussi appelé Second Harmonic Scattering (SHS). Nous allons voir ce que l'utilisation de  $\beta$ calculé au niveau QM/MM change dans la compréhension du signal collecté expérimentalement. Pour comprendre ce lien, nous utilisons un formalisme de diffusion qui distingue naturellement 2 contributions: "incohérente" et "cohérente". Trois résultats principaux se dégagent de cette partie dans le cas de l'eau pure. (i) La partie incohérente est extrêmement impactée par les fluctuations du  $\beta$  moléculaire. Aussi la SHS est très sensible aux environnements électrostatiques et à son impact sur le  $\beta$  des molécules. (ii) Le facteur de dépolarisation, l'une des observables expérimentales, est sensible à la structure locale des liquides: les 3 premières couches de solvatation influencent la réponse cohérente de façon non négligeable. Ce résultat est assez important, car la prédiction quantitative du rapport de dépolarisation n'est pas facile, en particulier quand il n'est pas suffisant de s'appuyer uniquement sur la géométrie ou la symétrie moléculaire. (iii) J'ai proposé les bases d'un formalisme en vue de concilier deux approches assez différentes pour modéliser plus complètement le signal SHS: d'une part notre approche numérique QM/MM qui décrit des effets électrostatiques à courte portée, et d'autre part des développements analytiques qui décrivent la structuration du liquide à longue distance.

Le Chapitre VI présente l'état actuel de mon travail sur la Génération de Second Harmonique à l'interface liquide-gaz (S-SHG). Dans une première partie, nous présentons notre article dans lequel nous calculons le  $\beta$  à l'ordre dipolaire de l'eau à l'interface avec la phase gazeuse. Cette première utilisation de FROG dans un système hétérogène nous a permis de confirmer la nécessité d'avoir une méthode numérique robuste pour calculer  $\beta$ . En effet, nous observons que  $\beta$  varie de façon significative à cette interface et qu'utiliser uniquement la valeur moyenne n'est pas une bonne approximation pour calculer le tenseur impliqué dans la réponse macroscopique. Dans une deuxième partie, je présente notre travail préliminaire sur les tenseurs d'hyperpolarisabilitées à l'ordre quadrupolaire obtenus à cette même interface. En effet, de récentes études ont montré que la réponse en surface était très impactée par ce terme qui est dû notamment au gradient des champs électromagnétiques présents à l'interface liquide-gaz.

Le Chapitre VII fait état du développement d'une expérience mêlant confinement de liquide et SHG. Le nouveau montage est composé d'un SFA (Surface Force Apparatus, qui permet de confiner un liquide entre 2 surfaces) qui a été agencé de manière à pouvoir mesurer le signal SHG du liquide confiné. Le principal intérêt d'utiliser la SHG pour sonder cette structure est encore lié à la nature même de ce processus quadratique: si le confinement induit une structuration de la matière, nous devrions avoir une réponse très marquée en SHG. J'ai participé activement à la conception et au montage de cette expérience qui regroupe le savoir-faire de plusieurs équipes de l'Institut Lumière Matière. Ce Chapitre présente l'état actuel du montage: nous arrivons à confiner le liquide jusqu'à la dizaine de microns et à y observer un signal à la fréquence de second harmonique.

Nous concluons cette thèse par le Chapitre VIII qui présente des perspectives plus générales que celles présentées à la fin des chapitres IV, V, VI et VII.

Nous allons maintenant présenter plus en détail les chapitres III à VII.

#### Chapitre III: Rappels théoriques

Ce chapitre présente le formalisme nécessaire pour comprendre la génération de second harmonique (SHG) à l'échelle microscopique et mésoscopique. Cette partie rappelle les concepts et résultats théoriques sur lesquels se basent nos calculs numériques , et qui ne sont pas forcément très accessibles dans des articles scientifiques habituels. Ce Chapitre III n'est pas exhaustif, mais j'ai essayé de présenter le détail des points qui me semblent importants ou moins connu concernant le formalisme de la réponse en chimie quantique, et l'électromagnétisme appliqué à l'optique non linéaire.

En premier lieu, la partie III. B présente le calcul de l'hyperpolarisabilité au niveau quantique au sein du formalisme de la réponse. Elle s'appuie sur le livre [1] qui explique le formalisme utilisé dans le code DALTON qui permet pratiquement d'effectuer ces calculs. Elle est décomposée en 2 sous-parties: une présentation du schéma "exact" et du schéma "approché". Nous essayerons durant cette deuxième sous partie de faire le lien entre ces 2 mondes, exact et approché, et de vérifier si les hypothèses utilisées, nécessaires dans la pratique, sont bien fondées pour les systèmes d'intérêts. Puis, nous verrons dans la partie III. C comment définir la polarisation mésoscopique, qui est l'objet de base pour la mesure expérimentale, à partir de ces quantités microscopiques via les équations de Maxwell dans la matière.

#### Chapitre IV: FROG & Développement méthodologique

Dans ce chapitre nous présentons l'approche numérique nous permettant de calculer la réponse des molécules en génération de second harmonique au niveau quantique, mais au sein d'un environnement liquide. Pour ce faire, nous nous basons sur le formalisme dit "Polarizable Embedding" (PE) du code DALTON. Dans cette approche, nous rajoutons dans l'Hamiltonien quantique de chaque molécule un terme qui décrit son environnement électrostatique, c'est-à-dire le champ créé au niveau de la molécule par ses voisines. Aussi, nous utiliserons une trajectoire de Dynamique Moléculaire classique (MD) pour obtenir des images de la phase liquide à partir de laquelle nous pourrons créer ces environnements électrostatiques pour chaque molécule. Puis, en calculant pour chaque molécule les degrés électroniques, par exemple en utilisant la DFT, et il a été possible d'extraire la première hyperpolarisabilitée grâce au formalisme de la réponse présenté dans le précédent chapitre.

Cette approche est largement utilisée dans la littérature, mais nous n'avons pas accès à un code

qui permet de préparer ces calculs quantiques depuis une trajectoire de MD. De plus, nous voulons avoir la possibilité de traiter les informations structurelles et de réponse optique en même temps, par exemple en montrant la première hyperpolarisabilitée moléculaire en fonction de la densité à une interface liquide-gaz. Pour ce faire, nous avons codé sous Python le logiciel FROG: "FROm molecular dynamics to second harmonic Generation". Les objectifs, le fonctionnement, les forces et limites de ce logiciel sont décrites dans la première partie de ce chapitre. Il est à noter que nous avons la ferme intention de rendre accessible ce code à la communauté. Pour cela nous sommes en train de finir l'écriture un wiki complet ainsi que de nombreux tutoriaux.

La deuxième partie de ce chapitre présente un article accepté pour publication par le journal PCCP. Il vise à mieux comprendre comment décrire l'impact de l'environnement électrostatique sur l'hyperpolarisabilité de l'eau. Historiquement, cela a été rationalisé à l'aide du tenseur de seconde hyperpolarisabilité dipolaire,  $\gamma(2\omega, \omega, \omega, 0)$ , qui peut être estimé expérimentalement grâce à des techniques comme l'EFISHG (Génération de Second Harmonique induite par un Champ Éléctrostatique). Ainsi, en première approximation, l'hyperpolarisabilité d'une molécule,  $\beta^{env}$  lorsqu'elle est soumise à un champ externe statique,  $\mathbf{e}^{env}$ , serait:

$$\beta^{env}(2\omega,\omega,\omega) = \beta^{GS}(2\omega,\omega,\omega) + \gamma(2\omega,\omega,\omega,0) \cdot \mathbf{e}^{env}$$
(1)

Avec  $\beta^{GS}$  une valeur de référence, par exemple en phase gaz.

Cette étude comporte 2 messages principaux. Le premier est le calcul de  $\gamma(2\omega, \omega, \omega, 0)$  en phase liquide. Cette quantité est liée à l'évolution de  $\beta^{env}(2\omega, \omega, \omega)$  pour les molécules d'eau en phase liquide lorsqu'elles sont soumises à un champ électrostatique de faible amplitude et spatialement homogène. Nous observons que cette réponse purement électronique est plus faible en phase liquide qu'en phase gaz, et que ce tenseur présente peu de fluctuations d'une molécule à l'autre. Aussi, dans une bonne approximation l'on peut assigner à chaque molécule la même valeur de  $\gamma(2\omega, \omega, \omega, 0)$ , contrairement à la première hyperpolarisabilité.

La seconde partie de l'article discute des conditions de l'utilisation de l'Équation IV.2 pour tenir compte de l'environnement électrostatique en phase liquide. En effet, si l'on pouvait appliquer cette équation pour prendre en compte l'environnement électrostatique autour de chaque molécule de la phase condensée (à la place de notre approche QM/MM) nous aurions accès aux hyperpolarisabilités sans avoir à réaliser des calculs quantiques. Aussi, mesurer le champ créé par les autres molécules ( $e^{env}$ ) serait suffisant. Hélas, nous montrons que cela n'est pas possible, la raison principale étant que le champ crée par la phase condensée est très spatialement hétérogène au sein des molécules. Aussi, nous ne pouvons pas appliquer directement  $\gamma(2\omega, \omega, \omega, 0)$  qui par définition est la déviation à un champ spatialement homogène.

Néanmoins, nous pouvons tout de même utiliser ce formalisme pour prendre en compte l'environnement

autour d'une molécule si celui-ci est suffisamment éloigné. En effet, à de grandes distances le champ créé par les autres molécules devient de plus en plus homogène spatialement. Pour ce faire, nous avons décomposé les voisins autour d'une molécule d'intérêt en fonction de leur distance à celuici, voir Figure IV.5. Les voisins sont inclus différemment suivant leur distance par rapport à un rayon de référence  $R_c$  pour prédire l'hyperpolarisabilité de la molécule d'intérêt  $\beta^{PE+L}$  dans cet environnement:

$$\beta^{PE+L}(2\omega,\omega,\omega) = \beta^{PE}(2\omega,\omega,\omega)(R_c) + \langle \gamma(2\omega,\omega,\omega,0) \rangle \cdot \Delta \mathbf{e}(R_c).$$
<sup>(2)</sup>

 $\beta^{PE}(R_c)$  est le résultat de l'approche QM/MM lorsque l'on inclut l'environnement jusqu'à une distance de  $R_c$ , typiquement jusqu'à  $< \gamma >$  est la valeur moyenne 1 nm. de l'hyperpolarisabilité de second ordre en phase liquide, et  $\Delta \mathbf{e}(R_c)$  le champ crée par l'environnement à partir de  $R_c$ . En d'autres termes, pour calculer l'hyperpolarisabilité en phase condensée, il faut prendre en compte l'environnement explicitement jusqu'à une distance  $R_c$  ( $\beta^{PE}(R_c)$ ). Le reste de l'environnement peut être pris en compte en utilisant le champ total créé au niveau de la molécule d'intérêt ( $\Delta \mathbf{e}(R_c)$ ) et la valeur moyenne de l'hyperpolarisabilité de second ordre (<  $\gamma$  >). Nous avons comparé les résultats de cette approche pour différentes valeurs de  $R_c$  et obtenu d'excellents résultats pour  $R_c \approx$ 10 Å.

Cette décomposition permet de prédire le  $\beta$  de molécule en réduisant la taille de



Figure 1: Schéma de principe de la séparation de l'environnement électrostatique autour d'une molécule cible (en rouge) en "direct" (en bleu) et "long" (en vert). Les voisins situés avant une distance  $R_c$  sont considérés comme proche et sont inclus explicitement, les voisins plus lointains que  $R_c$  sont inclus implicitement. Dans cette situation l'environnement total est une sphère de rayon  $R_f = 40$  Å mais à priori cette démarche serait valable dans un environnement infini.

l'environnement explicite utilisé dans les calcules QM/MM. Néanmoins, elle nécessite de connaitre  $\gamma$ , ce qui est couteux et surtout peut évoluer dans certaines situations (par exemple à l'interface gaz-liquide). Aussi, nous avons poursuivi cette approche dans la dernière partie de ce travail en décrivant l'effet des voisins éloignés au niveau de l'Hamiltonien de la molécule:

$$\hat{H}_{PE+LI} = \hat{H}_0 + \hat{V}_{Rc} - \hat{\mu} \cdot \Delta \mathbf{e}(R_c) \tag{3}$$

On retrouve dans les deux premiers termes l'approche QM/MM déjà utilisée où l'environnement électrostatique est inclus explicitement jusqu'à une distance  $R_c$ , ajoutant un potentiel électrosta-

tique  $\hat{V}_{Rc}$ . Le dernier terme correspond à un champ électrostatique en approximation dipolaire, c'est-à-dire un champ constant à l'échelle de la molécule de valeur  $\Delta \mathbf{e}(R_c)$ . Ce champ "externe" est celui créé par l'ensemble de l'environnement électrostatique à partir d'une distance  $R_c$ . En d'autres termes, dans cette approche l'environnement à partir de  $R_c$  est inclus directement dans le degré de liberté électronique, alors que dans l'approche précédente l'environnement lointain était inclus à postériori, en corrigeant la valeur du  $\beta$  obtenue par QM/MM. Nous obtenons de meilleurs résultats en travaillant à l'échelle de l'Hamiltonien.

Cette étude montre les limites de l'utilisation de la seconde hyperpolarisabilitée pour prendre en compte l'environnement électrostatique. Néanmoins, elle amène aussi d'autres idées pour calculer de façon plus efficace des systèmes dans lesquels il existe des champs macroscopiques, par exemple aux interfaces chargées. Dans ce genre de cas, on peut s'attendre à devoir aussi décomposer l'environnement en deux parties: l'une où l'inclusion explicite doit être faite (lorsque les gradients de champs sont importants), et une autre où l'on pourra utiliser la seconde hyperpolarisabilité. Nous avons implémenté dans FROG la dernière technique qui ajoute le champ des voisins lointains dans l'Hamiltonien.

#### Chapitre V: SHG en phase liquide

Dans le Chapitre IV est présenté comment la réponse de molécule individuelle a été calculée via le code FROG. Dans ce chapitre, nous étudions comment utiliser ces quantités pour prédire la réponse SHG en volume de l'eau pure. Cette réponse est usuellement séparée en 2 parties dites incohérente et cohérente: le signal expérimental étant la somme (indissociable) des deux, voir Figure 2. Pour comprendre la partie incohérente, il faut rappeler le cas limite de la Diffusion Hyper Rayleigh (HRS).

Cette technique est particulièrement utilisée



Figure 2: Décomposition de la réponse SHG en volume: la partie incohérente dépend uniquement de la réponse individuelle des molécules alors que la partie cohérente est directement sensible à l'organisation des molécules les unes par rapport aux autres.

dans le cas de molécules à forte réponse en SHG qui sont diluées dans un solvant. Le signal SHG en volume est dû à la somme de la réponse de ces molécules, très éloignées les unes des autres. Dans ce cas, on suppose que ces molécules ne sont pas corrélées les unes par rapport aux autres et que donc le signal est purement incohérent. Avec ce cadre, on raccorde le coefficient de dépolarisation, une quantité expérimentale, à la symétrie de la molécule en supposant que toutes les molécules ont la même réponse individuelle.

La partie cohérente correspond plus directement à la structure du liquide et des corrélations entres les molécules émettrices. Si le système est ordonné, par exemple un cristal, cette contribution sera largement dominante par rapport à la partie incohérente. Dans le cas de solvant pur, cette réponse est plus difficile à prévoir: d'un côté on sait que ces systèmes sont centrosymétriques et il n'y a pas de corrélations fortes à longues portées ( $\mu$ m), mais d'un autre côté il y a une forte corrélation à courte portée (nm) due aux interactions entre molécules (par exemple les liaisons hydrogène). Aussi, nous parlerons de Second Harmonic Scattering (SHS) plutôt que d' HRS pour ces solvants purs.

Expérimentalement, nous avons montré que pour l'eau pure, il une forte déviation au modèle purement incohérent. Aussi, des modèles numériques proposant une inclusion de corrélation de faible intensité entre molécules d'eau sur des distances de l'ordre du nanomètre ont été développés pour expliquer ces résultats expérimentaux. Ces modèles se basent en général sur l'hypothèse que la réponse individuelle des molécules est la même pour toutes les molécules. Si cette hypothèse est justifiée pour de nombreux systèmes, nous avons vu que cela n'est pas le cas pour l'eau en phase liquide. Ainsi, nous avons étudié avec nos prédictions de réponse moléculaire en QM/MM les 2 parties de la réponse. Dans la partie V. B l'impact des fluc-



Figure 3: Évolution du coefficient de dépolarisation en fonction de la contribution cohérente. Pour R =0 uniquement la partie incohérente est incluse. Pour R = 7Å sont prises en compte la partie incohérente et la partie cohérente jusqu'à la troisième couche de solvatation de l'eau.

tuations de l'hyperpolarisabilité moléculaire sur la partie incohérente et dans la partie V. C comment calculer la réponse cohérente à partir des réponses individuelles.

Les résultats concernant la partie incohérente montrent que:

- Ajouter les fluctuations de l'hyperpolarisabilité moléculaire modifie énormément la réponse du liquide pour la partie incohérente. On observe que la réponse pour l'eau est dominée par les fluctuations des composants interdits par la symétrie de la molécule plutôt que ceux qui ont une valeur moyenne non nulle.
- Il y a des corrélations entre les valeurs des composantes du tenseur d'hyperpolarisabilité, et elles impactent la réponse incohérente.

Les résultats concernant la partie cohérente montrent que:

- Nous observons une corrélation à courte portée (<1nm) pour la réponse en second harmonique de molécule d'eau. La partie cohérente du signal est donc non-négligeable pour l'eau pure comme le montre la Figure 3.
- En incluant cette réponse cohérente, nous avons un excellent accord avec les résultats expérimentaux sur le coefficient de dépolarisation.

• Hélas, nous n'observons pas d'écart entre les 2 polarisations de sortie contrairement aux résultats expérimentaux.

En conclusion, nous proposons une voie pour allier nos calculs QM/MM au formalisme de la physique des liquides pour mieux prendre en compte la corrélation à longue portée. La clé de voute de cette approche est de décomposer la réponse cohérente en 2 parties: l'une où l'orientation des molécules et leurs tenseurs d'hyperpolarisabilité sont corrélées (courte portée), et l'autre où seules les orientations le sont (longue portée). Nous espérons ainsi décrire les différentes échelles intervenant dans la réponse SHS: incohérent et courte portée avec les calcules QM/MM, longue portée avec des formules analytiques.

#### Chapitre VI: SHG à l'interface liquide-air

Dans ce chapitre nous présentons l'application de la méthode multiéchelle pour décrire la réponse SHG à l'interface liquide-gaz de l'eau pure. Cette partie se décompose en 2 parties qui se suivent, mais qui correspondent au tout début et à la toute fin de ma thèse.

La première partie correspond au premier article que nous avons écrit dans le cadre de la thèse. Cet article illustre notre première utilisation de FROG pour obtenir des résultats complètement nouveaux. Il discute de l'hyperpolarisabilité dipolaire de l'eau à l'interface liquide-gaz et de comment utiliser ces valeurs individuelles pour calculer le tenseur de susceptibilité surfacique. En effet, il est commun de supposer que la réponse individuelle des molécules dans leur référentiel, le tenseur d'hyperpolarisabilité moléculaire  $\beta(2\omega, \omega, \omega)$ , est le même, quelle que soit la position de la molécule ou son orientation. Dans le cas des molécules d'eau, nous savons que ce tenseur est très sensible à son environnement électrostatique. Nous pouvons donc nous demander à quel point cette hypothèse est valable à l'interface liquide-gaz.

Pour ce faire, nous avons calculé  $\beta(2\omega, \omega, \omega)$ à cette interface en utilisant FROG. Cette description de l'environnement de façon explicite nous permet de calculer les variations de ce tenseur moléculaire dans l'interface. Ces évolutions sont représentées par exemple dans la Figure 4 pour les valeurs moyennes de quelques composantes de  $\beta$ . De façon surprenante, nous observons que dans la phase liquide comme dans l'interface la symétrie fréquentielle de Kleinman est respectée, alors qu'elle ne l'est pas en phase gaz.



Figure 4: Évolution de la moyenne de certains composants de  $\beta$  pour l'eau en fonction de la position des molécules à l'interface liquide-gaz. La zone violette avant -4 Åcorrespond à la phase liquide, et la zone bleu clair à l'interface.

Grâce à cette approche, nous avons aussi pu mettre en évidence que le signal S-SHG provenait du dernier nanomètre formant l'interface. Nous avons donc bien une réponse spécifique à la surface. De plus, nous avons testé l'hypothèse qui consiste à attribuer à toutes les molécules la même hyperpolarisabilité moléculaire pour calculer celle dans le référentiel du laboratoire. Nous avons comparé les résultats obtenus avec et sans cette approximation et la différence observée est importante et remet largement en question le formalisme habituellement utilisé pour comprendre le signal SHG de l'eau aux interfaces. En effet, cela met en avant le fait que la réponse de chaque molécule ne peut pas être expliquée uniquement par son orientation puisque sa réponse à la seconde harmonique dépend fortement de son environnement.

Hélas, en utilisant ce formalisme nous ne sommes pas parvenus à expliquer même qualitativement le signal obtenu expérimentalement. Après un temps de maturation, nous sommes revenus à cette approche moléculaire à l'interface en incluant la composante quadrupolaire. Aussi, la deuxième partie du chapitre décrit l'état actuel de notre travail en incluant ces réponses quadrupolaires. Pour parfaire notre approche multiéchelle, nous avons également poussé le formalisme côté électromagnétique qui est partiellement décrit dans le chapitre III. Actuellement, nous n'avons pas abouti, mais nous espérons prochainement pouvoir comparer avec les résultats expérimentaux.

#### Chapitre VII: Développement expérimental d'un prototype SFA-SHG

Le dernier chapitre présente le développement et l'utilisation d'un système expérimental inspiré d'un Surface Force Apparatus (SFA) où l'on peut mesurer la réponse SHG d'un liquide confiné.

Le SFA est un dispositif qui permet de confiner des liquides jusqu'à quelques couches moléculaires. L'évolution de la structure des liquides à ces échelles est un domaine très actif où de nombreuses questions restent en suspens. Habituellement, le SFA renvoie des mesures de force en fonction de la distance, en régime statique ou dynamique. Nous pensons que l'apport de l'optique non linéaire, et particulièrement de la SHG, peut permettre de formidables avancées. En effet, si la structure des liquides est fortement modifiée lors d'un confinement, nous nous attendons à une modification importante du signal SHG du fait de sa sensibilité naturelle à la brisure de centrosymétrie.

Cette expérience représente plusieurs années de développement au sein d'une équipe étendue du fait des nombreux défis soulevés par ce dispositif. En effet, si le confinement de liquide à l'échelle du nanomètre est déjà bien connu, il reste délicat à effectuer. De plus, l'ajout de la SHG a demandé de modifier la structure type SFA.

Après une remise en contexte, la partie VII. B décrit le montage et les procédures expérimentales. Un travail de développement expérimental très important a été réalisé pour mesurer la distance entre les 2 solides qui confinent le liquide par interférométrie spatiale et fréquentielle. Enfin notre calibration du signal optique non linéaire est présentée. Les résultats obtenus permettent de contrôler le confinement de l'ordre de la vingtaine de nanomètres. Nous avons également pu mesurer en SHG plusieurs systèmes, dont un composé des 2 solides approchés à quelques dizaines de microns. Néanmoins, il reste encore du travail pour aboutir à une bonne compréhension des signaux SHG obtenus.

Cette première utilisation de ce nouveau montage est prometteur. Nous avons pu montrer que nous étions capables de maitriser notre confinement, et d'en mesurer la génération de second harmonique. Le montage est toujours en cours de développement, et j'espère que les nouveaux instruments ainsi que les prochains tests pourront nous aider à caractériser plus précisément notre montage.

#### **Conclusion:**

L'objectif de cette thèse était de mieux comprendre le lien entre la réponse individuelle des molécules en SHG et les mesures expérimentales, ainsi que de proposer un nouveau montage mêlant liquide confiné et mesure SHG. De nombreux progrès ont été faits vis-à-vis de la nouvelle expérience: nous avons mis au point un protocole expérimental robuste qui permet d'approcher la sphère du plan à la dizaine de microns. De plus, la mesure de Second Harmonique est opérationnelle dans cette géométrie. Il nous reste des défis à surmonter, mais nous sommes proches de pouvoir mesurer l'effet du confinement sur des liquides.

L'approche numérique a abouti à la création d'un code maintenant bien établi et fonctionnel. Nous avons retrouvé de nombreux résultats de la littérature et nous eu un accord quantitatif pour le cas de la SHG dans de l'eau pure. De plus, nous avons apporté une nouvelle interprétation des signaux expérimentaux. De nombreux autres systèmes peuvent être traités de la même façon pour confirmer cette méthodologie.

Les résultats obtenus à l'interface liquide-air sont moins complets qu'en volume. La raison principale étant la difficulté associée aux termes quadrupolaires. Néanmoins, nous proposons des valeurs microscopiques des hyperpolarisabilités dipolaires et quadrupolaires de l'eau à cette interface liquide-air. La forte évolution du terme dipolaire souligne la nécessité de prendre en compte l'environnent dans le calcul de ces quantités microscopiques à ces interfaces afin de pouvoir interpréter les résultats expérimentaux.

En conclusion, cette thèse est, il me semble, un exemple de synergie entre le monde "expérimental" et "numérique" et de la nécessité de ce genre de collaboration. En effet, les quantités microscopiques calculées numériquement n'ont de sens qu'au regard des résultats expérimentaux, et les hypothèses utilisées pour comprendre ces signaux n'ont d'intérêt que s'ils peuvent être validés à l'échelle microscopique. Aussi, je souhaite de tout coeur que cette thèse motive plus encore ces études pluridisciplinaires et multiéchelles et encourage d'autres doctorants dans cette voie.

## I Introduction

In order to understand matters, a physician hits it with a hammer: the harder the better! From one of my previous teachers.

In every day's life, we all know linear optics: illuminating red light on a material will come out red. In some cases, the material can absorb a part of the light or emit another color, called fluorescence. In this case, the photons emitted have a smaller energy: with this phenomenon, one can induce the emission of a green or red photons by illuminating blue ones, but not the opposite.

Let us draw a simple picture of linear optics without equations. We illuminate matter with light, *i.e.* an electromagnetic wave oscillating at a given frequency in time, called the fundamental pulsation and noted  $\omega$ . This field interacts with the electrons composing the matter, for instance every electron of each molecule of a fluid, and creates a force on these electrons. This force, within linear optics framework, is small. The electrons have the time to adapt their position to the force applied by the light: they move back and forth at the same frequency as the force, *i.e.* the light. Hence, at the molecular level, this displacement of electrons creates oscillating dipoles: negative charges (the electrons) oscillating around positive ones (the nuclei or the molecule). According to Maxwell's equations, these oscillating dipoles also create light at the same frequency: the emitted light by the matter is at the same frequency than the incident light.

What does happen when we increase the light power?

This force becomes more and more important and the movement of the electrons is more complex than just an oscillation at one frequency, *i.e.* non linear. To describe it, we can try to decompose it using a combination of oscillations: at the fundamental frequency, at twice the fundamental, three times, etc. Doing so, this strong light creates a superposition of oscillating dipoles at the frequencies  $\omega$ ,  $2\omega$ ,  $3\omega$ ... This means that the matter emits **harmonics** of the fundamental light: red photons (800 nm) can create red photons (first harmonic wavelength: 800 nm), blue ones (second harmonic wavelength: 400 nm), ultraviolet ones (third harmonic wavelength: 266 nm) and so forth, see Figure I.1. Using such an approach, we can easily ensure the conservation of the energy: in other words, the incoming and outcoming lights hold the same amount of energy. Doing so, we obtain that two fundamental photons can be converted into one at twice the frequency, or three fundamental ones into one at the third harmonics...

When doing the maths, we expect different behavior of the harmonics emitted by the matter with respect to the light power. If one double the intensity of the incident fundamental light, this will lead to an emission by the matter twice more intense at the same frequency, but the second harmonic will be four times more intense, the third harmonic eight times... Since the intensity of these harmonics are not proportional with respect to the fundamental intensity, these processes are called "non-linear". For instance, the intensity of the second harmonic is quadratic with respect to the fundamental one:



Figure I.1: Linear and second harmonic response of a material under strong fundamental incoming light.

quadratic with respect to the fundamental one: this is a "second order optical process".

Let us focus more on this process called Second Harmonic Generation (SHG) at a molecular level. One can define a molecular response for the second harmonic process. At the dipolar level, the excitation (the electromagnetic field at the fundamental frequency) is composed of a vectorial electric field, and the response (the induced dipole moment at the second harmonic frequency) is also a vectorial dipole moment. The response of the molecule is given by a tensorial product between a tensor, called the first hyperpolarizability, and the excitation. These types of matrices are called response tensors: they describe how the molecule will react to an incoming fundamental light at the harmonic frequencies.

To go beyond the molecular point of view, one can link the molecular response to the mesoscopic response of matter. For example, concerning the linear optical response of an isotropic, homogeneous bulk material, the optical index is linked to the molecular polarizability. For non linear optical processes, the mesoscopic response is described by a tensor that depends on both the tensorial molecular response, and the structure of the liquid.

Moreover, the tensor describing the SHG of a medium has an intrinsic property: for centrosymmetric medium, for instance pure water, it averages to zero. This is a very interesting fact because this means that the SHG is **specific** to the centro-symmetry breaking. For instance, at a liquid-liquid interface between two non-miscible but centro-symmetric fluids (for instance water and oil), the second harmonic light should come only from the interface. In general, SHG is much more sensitive to the symmetry and the structure of the molecules compared to linear processes.

Hence, SHG is widely used to study the structure of matter both in liquid phases, and at interfaces. Many types of questions are tackled: the orientation of the molecules at interfaces, the electrostatic potential of solid surface or the structure of the liquid in volume. Experimentally, obtaining the second harmonic signal from non-resonant pure liquids require an high sensitivity, and this technique is in constant improvement. We can for instance mention the recent development of phase-resolved SHG at interfaces or commercial instruments for SHG microscopy. In this context, we have also tried to add some new experimental setups to the list. The combination of second harmonic measurement with a confinement apparatus to probe liquids under steric pressure is a good example.

If there is no doubt about the ability of SHG to study the structure of the liquid phase, there are many unsolved questions regarding the experimental data interpretation. The numerical community has worked a lot to model the non-linear response of matter, mainly in the gas or solid phases. The liquid phase is more complicated because the intrinsic fluctuations due to the fluid phase make the numerical cost skyrocket as one has to sample many configurations. Moreover, it has been proven that the electrostatic environment deeply impacts the individual molecular response. Hence, if there are many recent numerical works addressing the liquid second harmonic response, there are few which links the microscopic computed property to the experimental observables. Therefore, in this thesis, we propose a multi-scale approach to bring new answers to the open questions concerning the interpretation of SHG measurement on liquids.

This manuscript is organized as follows. Chapter II presents the scientific context, the problematics we will address and the method we will use. Chapter III reminds some theoretical objects, frameworks, and results while Chapter IV presents in detail how we have computed the first hyperpolarizability in the liquid phase. Chapter V and Chapter VI show the results we have obtained using this multi-scale approach for pure water in the bulk phase and at the liquid/gas interface respectively. Chapter VII presents our experimental development on a new setup which aims to probe by SHG a confined liquid. Finally, Chapter VIII draws some perspectives.

#### Personal thought:

To close this introduction, I would like to write a more personal note about how I see my work.

I am interested in solving problems related to liquid systems and, as said few sentences before, non-linear optics is an extremely promising tool to understand the structure of matter. However, many questions remain concerning the SHG signal interpretation. Hence, the main question of my thesis is **how** to use SHG to understand the structure of liquids, and not on acquiring new structural information. I have worked on the experimental and numerical tools, the theoretical frameworks and the hypotheses around SHG of liquids, and applied it to water.

During my thesis, the different questions I asked myself were: How to disentangle the response at the interface to obtain structural information? Can we predict the SHG signal that will be emitted by a material interacting with a laser field if we know its atomic structure? Is Second Harmonic Scattering a probe of a long-range orientation of liquid, or something else? Is SHG a good probe for electrostatic environment since the individual molecular response is difficult to define in the first place?

In the experimental lab, it is difficult to test some hypotheses or to weight the different terms in the total collected signal. For instance, one cannot "turn off" the quadrupolar term of the SHG response or characterize precisely the role of the electrostatic environment. We often cannot tune the parameters (sample compositions for instance) by small enough steps to be able to test all the hypotheses needed to understand in detail the SHG intensity measured. Experiments provide robust results, but not the explanations.

On the contrary, using a computer to test a hypothesis is simple: we can turn off some terms, order the molecular responses as a function of the molecule relative distances or include or not the electrostatic environment effect on the SHG response. The goal of this work is not solely to "reproduce" the experimental results <sup>1</sup>, but to discuss some import hypotheses of the framework. Hence, an important part of my thesis was devoted to build a numerical tool to better understand the SHG response of liquid. I hope that the methodology developed here will help other people to weigh their own hypotheses and to tackle relevant structural questions of liquid systems, hopefully, more diverse than just pure water.

I have performed both experimental and numerical studies around SHG. While the numerical part is more represented in this manuscript, it has really been inspired by the experimental team around me. I really think that such numerical approach deep-seated in a theoretical and experimental, chemist and physician, team is beneficial for both worlds.

 $<sup>^1\</sup>mathrm{Yet}$  you would better have to if you want to publish in Nature

## II | Context and Problematics

Once you have defined the question you want to address, it probably means that you have begun to really understand the problem. From one of my previous teachers.

### II. A | Context

#### Water liquid phase: essential but complex

Studying the water in the liquid phase and its properties is an evidence for numerous reasons: it is involved in most biological mechanisms, in a large number of chemical reactions, in lubrification or in depollution mechanisms. Yet, this phase is very complex due to its intermediate nature: fluid and dense. The intermolecular interactions make the matter cohesive enough so that it holds together, but the temperature is large enough to permit the structure to explore many conformations. We will present here several configurations to illustrate the diversity of structure.

In bulk water, around a given molecule will be a solvation structure, described numerically for instance by the normalized radial distribution function, see Figure II.1. In this figure, we can observe different oscillations corresponding to the different solvation shells. The first layer is cohesive thanks to Hydrogen bonds, which strongly bind the molecules together. These bonds remain until they switch to another acceptor/donor each dozens of picoseconds [2]. The molecules of the second solvation shell are involved in Hydrogen bonds with the one of the



Figure II.1: Radial Distribution Function (RDF) for water in the bulk phase. The red dots represent the first, second and third solvation shells.

first layer. Hence, the correlation between the position and the orientation of water molecules within the second solvation shell and the center one is still strong but smaller compared to the first shell. As the distance between molecules increases, the effect of the Hydrogen bond network is more and more diffuse. However, electrostatic interactions are still important. Indeed, water has a very strong permanent dipole moment and the electrostatic field interactions are dominant at large distances. Hence, at nanometric distances and beyond, we can still have a net preferential orientation of the water molecules between themselves [3, 4].

Moving to interfaces, the description of the structure is even more complex. This is the case of water at the liquid-gas interface. This surface can be seen as hydrophobic: the liquid water cohesion is strong due to Hydrogen bonds and these interactions cannot be made with the gas-phase. For this reason, the surface tension of the liquid-gas interface is large (72 mN.m<sup>-1</sup>). In order to maximize the water-water interaction, the molecule at the interface will tend to be aligned within the interface plane, in a so-called 2D-Hydrogen bond network [5]. Therefore, the structure at this surface is very different from the one of the bulk phase. An illustration is



Figure II.2: Evolution of the density, black line, and the number of (accepted or donated) H-bonds, square and dots, at the liquid-gas interface. The total amount of H-bonds per molecule in average is the sum of the two curves.

given in Figure II.2 showing the evolution of the H-bond numbers at the liquid-gas interface.

The structure of aqueous solutions are even more complex. For instance, there are at least  $10^{-7}$  Mol/L of ionic species due to the self-dissociation of water into H<sup>+</sup> and H<sub>3</sub>O<sup>-</sup>. In biological systems, this ionic concentration is about hundreds of mMol/L. The interactions between water molecules can be impacted by these ions and thus modify the liquid structure. At interfaces, the concentration of ions can be higher compared to the bulk phase [6], leading to a local charged interface [7, 8].

Moreover, many interfaces involve a mixture of several species. A prototypical system would be a water-methanol mixture at the liquid-gas interface. The methanol can be seen as a very small surfactant: it can be involved in Hydrogen bonds with its OH group, and as a hydrophobic part from its  $CH_3$  part. Hence, at the liquid-gas interface we can expect the methanol to be in higher concentration than the water at the very edge. Understanding this kind of fine structure is not only about scientific curiosity. In many industrial processes, for instance depollution or pharmacy, a big issue is to move one component from one liquid phase to another. We can design the liquid phases or the component to make this transport possible thermodynamically speaking. But if at the interface, there is a structure that completely prevent the component to cross it, no transport happens.

#### Probing the liquid structure at interfaces using Second Harmonic Generation

To probe the liquid phase, many experimental tools and methods are available. We do not aim at presenting them extensively, we will focus on the non-intrusive ones which can provide useful information on the structure of the liquid phase.

We can use for instance reflection-based experiment at several wavelengths. For instance, by measuring the interferences as a function of the wavelength to measure a width between 2 interfaces or to look at aggregation of surfactants at an interface using an ellipsometer [9, 10]. Linear optics are very efficient: the light sources are easy to use and the response of molecules is well understood in general. For instance, one can use the refraction at interfaces to measure the optical index of a liquid very easily: but this does not give any information on the molecular organization at the interface by itself. Indeed, linear optics is in general dominated by the response of the bulk phase: that is why we need to use tools specific to the interface.

One of them is the second-order non-linear optical techniques such as Surface Second Harmonic Generation (S-SHG) or Surface Sum Frequency Generation (S-SFG). In such experiments, the interface is illuminated by a strong electromagnetic field at a fundamental frequency  $\omega$ ,  $\mathbf{E}^{\omega}$ . This fundamental field generates an electromagnetic field at twice the frequency  $2\omega$ , see Figure II.5. For simplicity's sake, we assume that this process is non-resonant: the fundamental field induces a gentle oscillation of the electrons at twice the frequency of the exciting field, without any absorption. To describe this process at the macroscopic scale, we use the following equation:



Figure II.3: Scheme of Surface Second Harmonic Generation, in this case the liquid-gas interface for water. The operator illuminates the interface with the fundamental light (red) and collects and analyses the emitted second harmonic (blue).

$$\mathbf{P}^{2\omega} = \chi^{(2)}(2\omega, \omega, \omega) : \mathbf{E}^{\omega} \mathbf{E}^{\omega}$$
(II.1)

Where  $\mathbf{P}^{2\omega}$  is the induced polarization of the medium. The tensor  $\chi^{(2)}(2\omega, \omega, \omega)$  which links the excitation and the response of the matter is called the *susceptibility*. An important property is that  $\chi^{(2)}$  is null for centro-symmetric media such as the liquids and gas phases. Hence, if we measure the second harmonic signal generated by the liquid-gas interface, this signal should come solely from the "interface" where the centrosymmetry is broken. This property makes Surface-SHG very promising because this technic is intrinsically surface specific.

Hence, in the scientific community this quadratic optical process is used to probe biological materials at interfaces[11–13], and to investigate the molecular structure of liquid-gas [14–19], liquid-liquid [20–22] or liquid-solid [23–27] interfaces.

#### Surface-SHG experiment:

Figure II.4 presents the experimental setup used to perform Surface-SHG experiments. A femtosecond laser creates the fundamental electromagnetic field (800 nm) which illuminates the liquid interface. The electromagnetic field is rectilinear polarized. During the experiment, the fundamental polarization direction, described by an angle  $\gamma$ , is rotated using a halfwave plate: from 0 to 360 degrees with respect to the propagation direction. Then, a lens focalized the fundamental into the liquid surface to further enhance the electromagnetic field intensity. Combined with the large power, typically 1 W, and the pulse duration, about hundreds of femtoseconds with a repetition rate of 80 MHz,



Figure II.4: Scheme of the Surface-SHG experiment, from [28]. The fundamental beam (red) is created by a femtosecond laser and the second harmonic generated by the liquid interface (blue) is collected in the reflection direction. The lenses increase the second harmonic intensity created and collected, the half-wave plate and the analyzer control the fundamental and second harmonic polarization.

this is enough to produce a measurable second harmonic signal from our interface of interest, for instance neat water.

To further maximize the second harmonic generated, the fundamental arrives with an angle of 70 degrees with respect to the surface normal. The second harmonic collection is made in the reflection direction. We measure the second harmonic intensity also with respect to its polarization: either along the direction 'P', in the incidence plane, or 'S', perpendicular to the incidence plane. The collected intensity at the second harmonic frequency in function of the incoming and outcoming polarization is presented in Figure II.5 for pure water at the liquid-gas interface.

Experimentally speaking, it is very difficult to get the absolute value of  $\chi^{(2)}$ . Usually, we rather measure the ratio between several components of  $\chi^{(2)}$  for instance by using the following equations which link the intensities collected for the two outcoming polarizations as a function of the incoming fundamental polarization:

$$I_{P}(\gamma) \propto |\left(a_{2}\chi_{xxz}^{(2)} + a_{3}\chi_{zxx}^{(2)} + a_{4}\chi_{zzz}^{(2)}\right)\cos^{2}(\gamma) + a_{5}\chi_{zxx}^{(2)}\sin^{2}(\gamma)|$$
  

$$I_{S}(\gamma) \propto |a_{1}\chi_{yyz}^{(2)}\sin(2\gamma)|$$
(II.2)

Where the  $a_i$  coefficients are experimental parameters related to Fresnel coefficients, see for instance [29] for more details. Using the experimental results and the above formula, we can thus extract experimentally the surface susceptibility ratios:  $\chi_{xzx}/\chi_{zzz} \approx 0.4$  and  $\chi_{zxx}/\chi_{zzz} \approx 0.1$  for pure water at the liquid-gas interface.

#### What does Surface-SHG measure?

We have seen that the SHG technique is surface-specific and that is possible to extract relative values of the different elements of the susceptibility tensor at an interface. However, these quantities provide information only at the macroscopic scale. Hence, we will now present how to make the link between the experimental results and the microscopic properties: what kinds of information are buried in  $\chi^{(2)}$ ?

 $\chi^{(2)}$  is the macroscopic response of the system in SHG, written in the laboratory frame, and



Figure II.5: Second harmonic intensity of the water liquid-gas interface with respect to the fundamental polarization,  $\gamma$ , for the two possible outcoming polarization angle for the second harmonic, P or S.

involves the individual responses of the molecules and their orientations. To see it, we can write Equation II.1 at the molecular level:

$$\mathbf{p}_n^{2\omega} = \beta_n^{(2)}(2\omega, \omega, \omega) : \mathbf{e}^{\omega} \mathbf{e}^{\omega}$$
(II.3)

Where  $\mathbf{p}_n^{2\omega}$  is the induced dipole moment of a molecule *n* at the second harmonic frequency upon an exciting electromagnetic field at the fundamental frequency  $\mathbf{e}^{\omega}$ . The link between these two vectors expressed in the molecular frame is made by the *first hyperpolarizability* tensor  $\beta_n^{(2)}(2\omega, \omega, \omega)$ . At the mesoscopic scale, we can say that the response of the system is given by summing the response of each molecule, once written in the laboratory frame. To go from the molecule *n* frame to the laboratory frame, we can use the orientation of the molecule as follows:

$$\mathbf{P}^{2\omega} = \sum_{n} R_n \cdot \mathbf{p}_n^{2\omega} \tag{II.4}$$

Where  $R_n$  is the rotation matrix of the molecule which holds the information of the orientation of the molecule *n* in the laboratory frame. Hence, we have the following relation between the response of the whole interface (the surface susceptibility  $\chi^{(2)}$ ) and the individual molecules (the first hyperpolarizability  $\beta_n^{(2)}$ ):

$$\chi^{(2)}(2\omega,\omega,\omega) = \sum_{n} R_n R_n R_n \cdot \beta_n^{(2)}(2\omega,\omega,\omega)$$
(II.5)

Therefore, we can extract the averaged orientation of the molecule (the  $R_n$  matrix) by measuring  $\chi^{(2)}$  if we assume that the first hyperpolarizability is the same for each molecule, and by assuming some plausible ratio between the  $\beta$  components. But this approximation is not valid for water in such environment as the first hyperpolarizability varies strongly from one molecule to another [30].

#### Electrostatic field impact on $\beta$

Indeed, the first hyperpolarizability  $\beta$  is an electronic property of the molecule, and this quantity is strongly affected by the electronic environment around it. A picture of what is happening in the condensed phase is presented in Figure II.6. In a pure semi-classical and electrostatic point of view, a given molecule (inside the red circle) will see its environment as a col-



Figure II.6: Scheme of the electrostatic environment surrounding a molecule from its point of view in the liquid phase

lection of positive and negative charges. These charges will generate a strong electrostatic field at the molecule vicinity, in the range of Volt per Angstrom: it will modify the electronic degrees of freedom and thus the first hyperpolarizability. Because the electrostatic environment is different from one molecule to another, each water molecule will have in principle its own  $\beta$  value. Such electrostatic effects have been quantified experimentally throughout at least two kinds of experiments.

The first technique is the Electric-Field-Induced Second Harmonic Generation (EFISHG) of molecules in a gas phase [31]. During such experiments, a macroscopic static field  $\mathbf{E}^{DC}$  is applied and the SHG response of the system is described using second and third order susceptibility tensor:  $\chi^{(2)}(2\omega, \omega, \omega)$  and  $\chi^{(3)}(2\omega, \omega, \omega, 0)$ , respectively. The induced total dipole moment at the second harmonic frequency  $\mathbf{P}^{2\omega}$  expressed as a sum of two terms :

$$\mathbf{P}^{2\omega} \propto \chi^{(2)}(2\omega,\omega,\omega) : \mathbf{E}^{\omega} \mathbf{E}^{\omega} + \chi^{(3)}(2\omega,\omega,\omega,0) : \mathbf{E}^{\omega} \mathbf{E}^{\omega} \mathbf{E}^{DC}, \tag{II.6}$$

with  $\mathbf{E}^{\omega}$  the exciting electromagnetic field at the fundamental frequency. This electrostatic field modifies (i) the orientation of dipoles, and (ii) the electronic-dependent molecular intrinsic SHG response. Equation II.6 was also applied to condensed phases [32, 33], and the EFISHG is an established technique to determine the first and the second hyperpolarizabilities of compounds in solutions [34].

The second kind of experiments are studies on liquid-solid or liquid-air interfaces where have been reported the evolution of the Surface-SHG (S-SHG) response when the surface charge is modulated. Frameworks based on Equation II.6 for the fluid near the surface have been commonly used in the S-SHG community to extract a surface potential or an effective surface charge [19, 25, 35–37]. But the interpretation of the different terms in the S-SHG intensity generated by aqueous solutions is still the subject of many recent works [21, 38–40] requiring theoretical calculations at the molecular level [41].

#### Uses of SHG to probe bulk phases:

Until here, we have seen that SHG is applied to study interfaces. However, it is possible to apply this process to bulk phases, even without applying an external electrostatic field. Indeed, we can


Figure II.7: Left: scheme of the SHG experiment in the liquid phase, denoted as Second Harmonic Scattering (SHS). The fundamental and harmonic light polarizations are controlled. In this setup, the second harmonic is collected at 90 degrees with respect to the fundamental propagation direction. Right: SHG intensity collected for liquid water in function of the fundamental polarization direction. The V (blue circle) and H (orange stars) correspond to the two possible second harmonic polarization directions.

also use polarization-resolved SHG experiments to characterize the symmetry of molecules [42] or how centro-symmetric is a liquid at the microscopic scale. Such experiments are called either Hyper Raleigh Scattering (HRS) or Second Harmonic Scattering (SHS), see Figure II.7. Indeed, molecules inside a fluid interact and have preferential orientation one from another. When averaging the orientation on large sample and over a long time, the liquid is centro-symmetric, but not for any given time at a small enough scale. Hence, the SHG in bulk liquid can be a good tool to probe the structure, and the potential long-range orientational correlation between molecules [3, 43–45].

Similarly, using SHG to probe the structure evolution of confined liquid might be a turnover. Indeed, when a liquid gets confined between two solid surfaces at the scale of few molecular lengths, the liquid behavior changes drastically [46–50]. Using a simple picture, the molecules have no longer the place to orient themselves as in the bulk phase: the interaction with the solid became at least as important as the one with the other molecules of the liquid trapped. In other words, if the symmetry of the liquid structure changes, we may be able to detect it by SHG.

## II. B | Problematics and Roadmap

As mentioned previously, the prior goal of this thesis is to better understand how to extract from the experimental SHG results new information on the liquid structure. Indeed, the link between the SHG experimental results and the microscopic quantities is complex because it involves several scales and contributions. First of all, we can wonder what is the individual response of a molecule given by its first hyperpolarizability  $\beta$ : many works have addressed this question, in the gas phase [51, 52] or in bulk phase [30, 53–55], but very few at interfaces [22]. Second, we can question how to use



these individual quantities to compare with the experimental signal, in the bulk phase [43, 56] or at interfaces [20, 21, 41, 57, 58], and especially when including the  $\beta$  fluctuations?

Hence, in order to better understand the experimental results, we need to model the SHG response of liquids in the bulk and at interfaces. In this context, we will address the three following questions as a roadmap for this work:

- What value of  $\beta$  shall we use for a given molecule in the liquid phase or at interfaces?
- How to link the microscopic properties to the experimental?
- Can we build an experiment to probe confined liquid using SHG?

The first two questions are addressed thanks to numerical and methodological developments while the last one is purely experimental.

#### What value of $\beta$ shall we use for a given molecule in the liquid phase or at interfaces?

In order to compute the first hyperpolarizability of water in the liquid phase we use an explicit solvation scheme. Indeed, modeling the molecular environment as homogeneous dielectric material is not sufficient to reproduce the sign change of water hyperpolarizability from the gas phase to the liquid phase [59]. Here, we propose a sequential Quantum Mechanical/Molecular Mechanics (QM/MM) approach to compute the molecular hyperpolarizability. Such methods have been successful by applied to describe the impact of solvation on the nonlinear optical properties of dyes [60–66], or non-resonant molecules [30, 53, 55, 67, 68].

This numerical approach starts with the prediction of the liquid structure at the molecular scale using classical Molecular Dynamics (MD), see Figure II.8 A. The open-software LAMMPS [69] have been used to perform the MD calculations. The advantage of this method is that it is numerically

cheap and it can be used to model a large number of different liquid systems at any interfaces. However, it is heavily dependent on the force field used to describe the interactions. In this work, we will not discuss the method related to MD because the systems (pure water in the bulk phase or at gas-liquid interface) are well known in the community. More information regarding the MD can be found in our published papers [70, 71] presented in Chapter IV and VI.

Once the structure obtained, we can build the electrostatic environment of each molecule, Figure II.8 B. This neighborhood will generate an electrostatic field within the molecule vicinity and thus impacting the electronic degrees of freedom. Using the Polarizable Environment [72] (PE) formalism of the DALTON software [73], we can extract the response tensor of such embedded molecule at the Quantum level. The first hyperpolarizabilities are computed for an exciting field wavelength of 800 nm, typically used in our experiments[74, 75]. This QM/MM approach is very efficient because it includes the environment, from the liquid phase to the interfaces, with accuracy while conserving an acceptable numerical cost.

Unfortunately, I did not have access to an efficient software that performs the links between the MD simulations and the QM/MM calculations. Hence I developed a code in Python to fulfill this purpose. The outcome of several months of coding is named FROG for "FROm molecular dynamics to second harmonic Generation". This code is presented in more details in Chapter IV and will be in open-access. With this numerical tool, we were able to compute the hyperpolarizabilities of molecules at both the bulk phase and at the interfaces.

The work related to the calculation of the first molecular hyperpolarizability is presented in the first part of Chapter III for the methodological part, in Chapter IV for the FROG implementation and application in the bulk phase. Chapter VI presents the microscopic values obtained at the liquid-gas interface.

#### How to link the microscopic properties to the experimental?

Now that we have the knowledge of the molecules positions and first hyperpolarizabilities, we can reconstruct the experimental signals. This is symbolized in Figure II.8 C and D. The formalism used will be quite different depending on the experimental geometry (bulk phase or at interfaces).

In the bulk phase, there is no net induced mesoscopic polarization because the centro-symmetry of the bulk phase forbid it. Hence, to describe the signal collected experimentally, one has to use a scattering formalism. In the literature [3, 42–44], the  $\beta$  is always assumed to be constant and respects the molecular symmetry. We will discuss in the first part of Chapter V if such hypothesis holds for water in the bulk phase using our QM/MM results. Then, we will propose a calculation of the depolarization ratio, a key experimental observable, using our *ab-initio* results. We will also study the impact of the local structure of the liquid on this quantity.



Figure II.8: Scheme of the multiscale approach. Starting from an MD trajectory (A), we first compute the  $\beta$  at the QM/MM level (B). To do so, FROG creates the electrostatic environment for all the molecules (I). Then, the electronic degrees of freedom are approached using the DALTON software and the  $\beta$  values of this configuration are extracted (II). Then, the value obtained for the different molecules in the trajectory are organized and averaged spatially and temporally (III). Then, we use these ab-initio optical response tensor along with the position of the molecule (C) to simulate the experimental signal. This can be more or less difficult depending on the geometry of the system. Finally, we compare our simulation with the experimental results (D).

To compare the results obtained at the liquid-gas interface, more steps are needed. Indeed, as presented in the second part of Chapter III, the exciting field which induces the dipole moment of molecules at the second harmonic frequency varies with the distance to the interfaces. Thus, at the scale of the molecule, the electromagnetic field cannot be assumed to be constant anymore. Therefore, the dipolar approximation is no longer valid [41]. Hence, to go beyond the dipolar first hyperpolarizability, we have to compute the quadrupolar terms. This can be done using the same formalism as for the dipolar term, as explained in the first part of Chapter III. These microscopic values are presented in the second part of Chapter VI. We have also achieved a preliminary work regarding the use of these values at the liquid-gas interface, presented in the second part of Chapter VI. If we cannot yet compare our numerical prediction with the experimental results, the presented microscopic values are new and paves the way to a better understanding of the SHG intensity at the liquid-gas interfaces.

# Can we build an experiment to probe confined liquid using SHG?

Another goal of this thesis is to build a first experimental device which can detect the Second Harmonic Generation of confined liquid, presented in Chapter VII. To perform a controlled liquid confinement we have created a setup inspired by a Surface Force Apparatus (SFA)that we can couple with a SHG experiment. The development, building and characterization of this experiment has been the major experimental contribution of this thesis. The main challenge was to fulfill the requirements related to the confinement of a liquid below



Figure II.9: Overview of the confined setup probed by SHG. The fundamental frequency is represented in light purple, the second harmonic in green.

the nanometer and the SHG acquisition. Hence, we had to adapt the usual SFA experiments and start from almost scratch.

We came up with the setup presented in Figure II.9. The confinement is made using a sphere in the top of a lamella, both made in Pyrex glass. The gap between the two solid,  $\Delta z$ , is controlled by a piezzo stage and is measured using spectral interferometry. A microscope objective with a strong magnification and high numerical aperture injects the fundamental light from below. The same objective collects the second harmonic signal generated, which is then measured by a usual detection setup. During the SHG generation, we are able to change the position of the laser focal point and the confinement.

Today, we have promising results: we are able to confined liquid within dozens of micrometres and to collect the Second Harmonic generated. We still have some developments to do regarding the confinement stability and to deepen our understanding of the second harmonic signal collected. However, it clearly shows that such experiment can be performed.

An efficient and cheap approximate solution is better than an exact theoretical nightmare. Lazy Physicist Maxim.



### III. A | Introduction

Before going into the theoretical details, let us first describe the experimental routine at the lab and the different processes involved in SHG.

### Experimental reminders:

As previously introduced, there are two different setups available at the labs: Surface SHG (S-SHG) and volume SHG also called either Hyper Rayleigh Scattering (HRS) or Second Harmonic Scattering (SHS).

In S-SHG, a femtosecond laser at optical wavelengths (in our case red) emits a pulsed beam that is led toward the sample surface with an incident angle around 70 degrees with respect to the surface normal. The light generated at the second harmonic frequency (in blue) is collected with the same angle in the reflection direction. An experimental run for a given sample, for instance the water-air interface, consists in rotating the polarization of the fundamental light and collecting with two different polarizations the second harmonic one.

For volume SHG, two kinds of geometry exist. A first one consists in directing the fundamental with a fixed angle, and collecting the second harmonic with different diffraction angle (from 0 corresponding to the same direction as the fundamental to 180 degrees corresponding to the reflection direction). A second one works with a fixed angle between the fundamental and the second harmonic, 90 degrees, and rotating the fundamental and second harmonic polarization. We rather work with the second one in the lab, mainly for practical experimental reasons.

Hence, in a first glance, a Second Harmonic Generation experiment consists in selecting a geometry, a sample, and rotating the fundamental polarization applied to the sample while collecting the second harmonic with respect to its polarization at a given angle. After years of experimental development, the incoming field is well controlled as the outcoming detection. Hence, we will not discuss in detail the steps concerning the fundamental beam control nor the second harmonic field detection, but rather focus on the process occurring in the sample: how to link the fundamental field applied to the collected intensity at the second harmonic frequency?



Figure III.1: Multiscale description of the Second Harmonic Generation in liquids.

### What is going on in my sample?

Figure III.1 schematizes the different problems to solve in order to understand the second harmonic collected experimentally at a microscopic scale.

1. From  $\mathbf{E}^{\omega}$  to  $\mathbf{e}^{M}$ :

We can control experimentally the fundamental beam that is sent to the sample  $\mathbf{E}^{\omega}$ : its spectra, power, polarization and incidence angle. But not the one that is "felt" by a molecule inside the sample,  $\mathbf{e}^{M}$ . The effects of two phenomena should be taken into account: the optical index gaps, and the optical index inside the sample. The first one is specific to Surface-SHG, for instance at the air-water interface: the fundamental has a non-normal angle with respect to the surface plane. Hence, the fundamental light direction and polarization are modified according to the Fresnel coefficients once the surface crossed. However, what happens precisely inside at the interface is less obvious. The second one is due to the linear polarizability of the molecules. The SHG takes place in condensed phase, which is described by an optical index. This means that the electromagnetic field felt by the molecule ( $\mathbf{e}^{M}$ ) is not the total electromagnetic one because of the response of the molecule itself at the same frequency.

2. From  $\mathbf{e}^M$  to  $\mathbf{p}^{SHG}$ :

Once the exciting field  $\mathbf{e}^{M}$  defined for each molecule, we should describe the generation of the second harmonic at the molecular scale  $\mathbf{p}^{SHG}$ . To do so, we have to approach the electronic degrees of freedom of the molecules inside the sample: either in the bulk phase or at an interface. Then, one has to describe how their electronic cloud would oscillate at twice the fundamental frequency depending on the fundamental polarization. It should be noted that for non-resonant molecules, such as water for the frequencies used in this work, the function linking the fundamental electromagnetic field and the second harmonic generation (symbolized by  $\mathbf{p}^{SHG} = f(\mathbf{e}^{M})$  in Figure III.1) can be different for each molecule.

3. From  $\mathbf{p}^{SHG}$  to  $\mathbf{P}^{2\omega}$ :

According to the Maxwell equations, the source term of the electromagnetic wave that can be detected experimentally is the mesoscopic polarization oscillating at the second harmonic frequency:  $\mathbf{P}^{2\omega}$ . In a first glance,  $\mathbf{P}^{2\omega}$  is the space and time averaged of the individual induced dipole moment of each molecule at the second harmonic frequency  $\mathbf{p}^{SHG}$ . However, this term can also contain quadrupolar contributions, especially for S-SHG. Moreover, the links between the molecular sources of the second harmonic ( $\mathbf{p}^{SHG}$ ) and the mesoscopic response can be difficult due to the linear polarizability of the molecule themselves. Each molecule generates an electromagnetic field which can induce a dipole moment on the other molecules: the optical index of the matter at the second harmonic frequency is not 1. To address this question once can use the concept of "dressed dipole" for instance.

4. From  $\mathbf{P}^{2\omega}$  to  $\mathbf{E}^{2\omega}$ :

Once the mesoscopic polarization  $\mathbf{P}^{2\omega}$  defined, we can use the Maxwell equations to predict the electromagnetic field at the second harmonic frequency  $\mathbf{E}^{2\omega}$  that will be collected experi-

mentally. If this part is straightforward for volume systems, it is more difficult at interfaces. To deal with this issue, we can use the "three-layers" model – also called "polarization sheet" model.

#### What should be taken into account in the model:

Our experimental measurements do not give access to absolute values. The intensity collected experimentally is always renormalized using a reference. For instance, we can compare the evolution of SHG between samples (for instance the evolution of the SHG intensity as a function of NaCl concentration), or provides value with respect to a reference solvent (for instance compare the intensity of pure methanol to pure water). Hence, in the following we will not focus on the determination of the absolute SHG intensity, but on the prediction of the evolution of the second harmonic intensity with respect to the fundamental or second harmonic field polarization. Therefore, all phenomena which modify the fundamental or second harmonic intensities **regardless of the polarization direction** will not be studied or disregard because they are not needed in order to understand or predict the experimental measurement (polarization-resolved SHG) of a given sample.

#### What is new?

Many descriptions are available for the different systems studied [28, 76–79]. In this thesis, we try to propose a global approach starting from a new building block: the individual SHG response of the molecules in condensed phases which can be molecule dependent and can contain quadrupolar order. Indeed, with the knowledge of the individual molecular link between the felt fundamental field and the induced dipole moment, at the QM/MM level, we can rethink the whole approach.

The calculation of the molecular SHG response is already very well established. Hence, the first part of this Chapter recalls results already well established in the literature, even if the quadrupole part is more original and not often discussed in the literature at the molecular point of view. The second part aims to link the individual response to the experimental signal is more original. Each part was already established in previous works and I have been heavily inspired by the framework proposed by Morita and co-worker [79]. However, the articulations of the different parts and their adaptation for SHG, especially at the interface, has required some work, which is still ongoing. Hence, the second Section of this Chapter combines rock-solid models and home-made ones: it presents the multi-scale framework that we propose rather than the description of a well established one.

#### **References:**

Here is a list of some major references I have used to write this chapter.

This chapter has been heavily influenced by *Principles and Practices of Molecular Properties* [1] and the knowledge acquired at the Molecular Response Property Summer School 2021 at Stockholm, especially regarding the response theory framework and practical uses. I have used mainly *Electronic Structure: basic theory and practical methods* [80] for the DFT discussions. For elec-

tromagnetism reminders, see for instance *Electromagnétisme: fondements et applications avec 300* exercices et problèmes résolus [81] or Optique physique: Propagation de la lumière [82]. The formalism used for the exciting field calculations have been inspired by the approach made by Morita and co-worker, *Theory of Sum Frequency Generation Spectroscopy* [79]. Regarding the experimental or macroscopic SHG part, I have used *Nonlinear optics* [77] and *Surface second harmonic generation* [76] and also previous thesis of the team [28, 29, 78]. I have also been inspired by the thesis of Julien Guthmuller [83] for the theory section.

### Outline of this chapter:

In this chapter, we will go through the different theoretical concepts needed to build this multiscale approach.

First, we show how the second harmonic process is described at the quantum level. We start by an exact description followed by a practical approached.

Second, we describe how to use these microscopic properties to compare with the experimental results. In a first step, we present how to describe the electromagnetic field felt by a molecule, and then how to write the mesoscopic induced polarization. Finally, we will briefly present how to link the induced mesoscopic polarization to the electromagnetic field at the second harmonic frequency at interfaces.

We conclude with a sum-up of the specificities regarding the bulk and the surface SHG.

### Notes:

Throughout this chapter and the thesis, the atomic unit system is used in the equations. However, we use the SI system for the numerical applications of the length, density or sometime electric field. Hence, a.u. stands for atomic units and **not** arbitrary units in the numerical parts. There are some integrals in this thesis. Most of them goes over space. If nothing is specified on the upper and lower integration range,  $\int d\mathbf{r}$ , means that the integral goes over the whole space (the integrand  $\mathbf{r}$  goes over  $\mathbb{R}^3$ ). The laboratory frame is defined using the Cartesian frame  $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$  while the molecular frame is defined using  $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ .

# III. B | Microscopic response



Figure III.2: Multiscale description of the Second Harmonic Generation. Here is highlighted in orange the link between the felt electromagnetic field and the individual molecular response at twice the fundamental frequency, treated in this part.

We define the "microscopic response" to be the response of one molecule which represents our elementary building block. In the gas (or vacuum) phase, this definition is unambiguous: the different molecules are separated from one another. In condensed phases such as liquid water, Hydrogen bonds can be seen as weak chemical bonds and not only as inter-molecular interactions. In this sense, the smallest building block would be a cluster (or grasp) of molecules. However, we will still define molecular-based property, even in liquid. To do so, we use a mixed framework (QM/MM) under which the environment around a given molecule is included in a single-molecular quantum Hamiltonian.

To treat the interaction of a molecule with a strong electromagnetic field and the generation of the second harmonic, we have to deal with the electronic degrees of freedom, and thus with their quantum nature. In a first part, we will present the **exact** framework. However, such approach cannot be applied for our molecules, and thus we will then present the **approximate** framework we use in practice.

#### III. B.1 Exact Response Theory

In this section, we derive the fundamental equations based on the perturbation theory leading to the definition of the molecular hyperpolarizabilities. This formalism is used in the DALTON software which was used during this thesis to perform these calculations.

First, we will remind how to build the electronic Hamiltonian of a molecule without external electromagnetic field. Then, we introduce the light-matter interaction up to the quadrupole order. We will discuss the intrinsic difficulty due to the time-dependency and the need of a perturbative approach. Finally, we will present the response scheme and its corresponding objects: the response tensors. A more detailed presentation of the exact scheme and some of the calculations are given in Appendix IX. A.1.

Parts of this procedure undergo approximations mostly because we are not able to solve exactly the initial quantum problem for molecules made of several electrons. Some of these problems will be discussed in the Section III. B.2.

#### III. B.1.1 Non-interacting Hamiltonian

The fundamental framework of a non-relativistic quantum molecular Hamiltonian has been used within the Born-Oppenheimer approximation for a molecule composed by N electrons and M nuclei. The wave-function  $|\Psi\rangle$  describes the quantum state of this molecule.

Since the nuclei are at least 2000 time heavier than the electrons, the nuclei dynamics are much slower than the electronic ones. We assume the Born-Oppenheimer approximation: we can decouple the electronic and nuclear degrees of freedom. Practically, we assume the position of the nuclei:  $[\mathbf{R}'_1, \mathbf{R}'_2 \cdots \mathbf{R}'_k \cdots \mathbf{R}'_M]$  and then find the wave-function of the electrons:  $|\Psi_{[\mathbf{R}'_1, \mathbf{R}'_2 \cdots \mathbf{R}'_k \cdots \mathbf{R}'_M]}\rangle \equiv |\Psi\rangle$ . Note that in our numerical approach, the nuclei positions are provided by a classical Molecular Dynamics simulations. Therefore in the following,  $|\Psi\rangle$  designs only the electronic wave-function.

However, the nuclei degrees of freedom can dynamically affect the electronic degrees of freedom. Two major phenomena have to be considered in the general case:

- Zero Point Vibrational Correction (ZPVC): A minimal motion of the nuclei has to be accounted for since the nuclear vibrations are bosonic even in the zero temperature limit! This impacts the electronic structure since this induces several nuclei configurations.
- Vibrational electromagnetism coupling: the nuclei can also contribute to the light-matter interaction. For instance, the Sum Frequency Generation (SFG) is another second order non-linear process that targets precisely the molecular vibrational motions. For resonant frequency at the vibrational level, the electronic response is tremendously affected.

In the following, we will disregard these two effects: we will assume that the nuclei positions are completely frozen with or without an external electromagnetic field. Thus, the nuclei are treated classically: there are infinitesimally localized and fixed in space.

#### Evaluation of a property: $\Omega$

To compute a property of the system, we shall construct an operator describing it. The expected value of an observable  $\Omega$  is given by the evaluation of its operator over the wave-function:  $\Omega = \langle \Psi | \hat{\Omega} | \Psi \rangle$ . In particular, the dipole moment of the molecule is given by:

$$\boldsymbol{\mu} = \int \mathbf{r} q_e \rho(\mathbf{r}) d\mathbf{r} + \sum_k \mathbf{R}_k q_k = \langle \Psi | - \mathbf{r} | \Psi \rangle + \sum_k \mathbf{R}_k q_k$$
(III.1)

Where  $q_e = -e = -1$ , in atomic units, is the charge of an electron,  $\rho(\mathbf{r})$  the electronic density at a position  $\mathbf{r}$ ,  $\mathbf{R}_k$  the position and  $q_k$  the charge of the nuclei k. By identification, the operator associated to  $\boldsymbol{\mu}$  is  $\hat{\boldsymbol{\mu}} = -\mathbf{r}$  so that:  $\boldsymbol{\mu} = \langle \Psi | \hat{\boldsymbol{\mu}} | \Psi \rangle + \sum_k \mathbf{R}_k q_k$ .

The second term is due to the nuclei and thus does not depend on the electronic degrees of freedom within our Born-Oppenheimer approximation. Moreover, since the nuclei are fixed, they cannot oscillate at optical frequencies and thus do not contribute to the SHG response. Hence in the following, we will neglect the nuclear contribution to the dipole moment.

### The isolated Hamiltonian $\hat{H}^0$ :

According to Quantum mechanical postulate, the wave-function satisfies the time-dependent Schrödinger equation:

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\Psi\rangle (t) = \hat{H}(t) |\Psi\rangle (t).$$
(III.2)

 $\hat{H}(t)$  is the (time-dependent) Hamiltonian of the system. It contains all the interactions acting on the electrons – the one acting on the nuclei are disregarded because we are not treating the nuclei degrees of freedom. The (time-dependent) energy of the system is given by:  $E(t) = \langle \Psi | \hat{H} | \Psi \rangle (t)$ . First, let us consider the case without an external electromagnetic field. For an isolated molecule, the unperturbed (or isolated) Hamiltonian  $\hat{H}^0$  is:

$$\hat{H}^{0} = -\underbrace{\frac{1}{2}\sum_{i} \nabla_{i}^{2}}_{\hat{T}} - \underbrace{\sum_{i,k} \frac{q_{k}}{|\mathbf{r}_{i} - \mathbf{R}_{k}|}}_{\hat{V}_{ne}} + \underbrace{\sum_{i < j} \frac{1}{|\mathbf{r}_{i} - \mathbf{r}_{j}|}}_{\hat{V}_{ee}}.$$
(III.3)

The first  $\hat{T}$  term is the non-relativistic electronic kinetic operator, the second term  $\hat{V}_{ne}$  is the cohesive coulombic electron-nuclei interactions. The last term  $\hat{V}_{ee}$  is the repulsive coulombic electronelectron interactions.

This Hamiltonian is spin-independent since we are neglecting the spin-orbit coupling. Indeed, we are working with non-magnetic molecules such as water: the atoms (Hydrogen, Oxygen) present no net spin. Moreover, we are working with closed-shell electronic structure at the ground state: the total electronic spin is zero. Hence, in the following we also neglect the light-matter interaction terms involving the spin operators.

In Chapter IV, we will include the effect on the environment by adding a new potential on this unperturbed Hamiltonian  $\hat{H}^0$ , see .

#### Stationary state :

To solve the problem, the Hamiltonian is diagonalized on the Hilbert space of the wave-function, *i.e.* a set of wave-functions  $\{|\Psi^n\rangle\}$  with associated energy  $\{E_n\}$  is found so that:

$$\hat{H}^0 |\Psi^n\rangle = E_n |\Psi^n\rangle. \tag{III.4}$$

Where the convention is to arrange the energy from the lowest to the highest:  $E_0$  is the ground state energy while  $E_1, E_2, \cdots$  are the excited energies. The energy is defined up to a constant, hence only the difference in energy, or alternatively the associated **pulsation**,  $\hbar\omega_{ij} = E_i - E_j$ , really matters. These states are stationary because they respect the Schrödinger equation up to a phase factor: the time-dependent wave-function  $|\Psi\rangle(t) = |\Psi^n\rangle \exp\left[-i\frac{E_n}{\hbar}t\right]$  respects the Schrödinger equation with associated  $E_n$  eigenvalues.

### $\hat{H}^0$ resolution:

Solving the eigenvalue problem is tremendously complicated because electronic wave-function is the **joint** probability of the N electrons: even if the operator acts only on 1 or 2 electrons at the same time, the electronic degrees of freedom are correlated. In section III. B.2 we will see a way to solve this problem based on the electronic density: the Density Functional Theory (DFT). Within the Kohn-Sham framework, this provides a robust and "cheap" numerical method to find the electronic ground state  $|\Psi^0\rangle$  for the molecule of interest.

Then, in the following, we assume that we can find exactly these stationary states  $\{|\Psi^n\rangle\}$  for any time-independent  $\hat{H}^0$ . Moreover, we assume that we are in the zero temperature limit: only the state with the lowest energy is initially populated, or, in other words, that  $|\langle\Psi^0|\Psi\rangle| = 1$ .

#### III. B.1.2 Interaction with an electromagnetic field

#### Electromagnetic fields and interaction potentials:

A source creates an external electromagnetic field propagating toward the molecule vicinity. Two fields are felt by the molecule: the electric  $\mathbf{e}(\mathbf{r}, t)$  and magnetic  $\mathbf{b}(\mathbf{r}, t)$  fields. These fields, named exciting fields, are the ones felt by the molecule **without** its own reaction. We will assume here that these fields are not impacted by the molecule reaction and are the results of the propagation in the matter which will be discussed in Section III. C.1.

In the classical case, the interaction potential V(t), in the Hamiltonian convention, is given by:

$$V(t) = +q\phi(\mathbf{r}, t) - q\mathbf{v}(t) \cdot \mathbf{A}(\mathbf{r}, t).$$
(III.5)

Where q is the charge of the specie and v its velocity,  $\phi$  is the electric potential and A the potential

#### Page 45

vector. The relation between the electromagnetic fields and the potentials are:

$$\mathbf{e}(\mathbf{r},t) = -\nabla\phi(\mathbf{r},t) - \frac{\delta}{\delta t}\mathbf{A}(\mathbf{r},t) \qquad \qquad \mathbf{b}(\mathbf{r},t) = \nabla\wedge\mathbf{A}(\mathbf{r},t) \tag{III.6}$$

In the following, we assume that all these fields and potentials can be written as plane waves:

$$f(\mathbf{r},t) = f_0 \exp\left[-i(\omega t - \mathbf{k} \cdot \mathbf{r})\right].$$
(III.7)

Where  $\omega$  the time frequency and **k** the wave vector. The amplitude of the field  $f_0$  can be a complex number and/or a vector depending on the nature of the field.

The quantum operator  $\hat{V}(t)$  is defined so that the expected interaction energy obtained is the same as the classical case:

$$V(t) = \langle \Psi | \hat{V}(t) | \Psi \rangle = \int \left[ -\rho(\mathbf{r}, t)\phi(\mathbf{r}, t) - \mathbf{j}(\mathbf{r}, t) \cdot \mathbf{A}(\mathbf{r}, t) \right] d^3\mathbf{r}.$$
 (III.8)

Where  $\rho$  is the total density of electrons and **j** the current density. Using this implicit definition of the electromagnetic potential  $\hat{V}(t)$ , the new Hamiltonian of the system, called **interacting** Hamiltonian, is:

$$\hat{H}(t) = \hat{H}^0 + \hat{V}(t).$$
 (III.9)

If the potentials produced by the source are extremely space-inhomogeneous, the space-integral form given by Equation III.8 cannot be simplified. However, for exciting field with a wavelength of 800 nm, the spatial-dependency is limited relatively to the volume where  $\rho$  and **j** are non-negligible. For instance, Figure III.3 represents the electromagnetic field evolution with respect to the position. To observe the evolution within a molecule of interest, we have to zoom on a few Ångstroms. In the worst scenario where the space gradient is the largest, the evolution of the field across a water molecule (1 Å) is about 0.1 %<sup>-1</sup>. However, the evolution of the field across a molecule may be enhanced by the medium, for instance throughout an optical gap as presented in Sec III. C.1.

#### Taylor development of the fields: multipolar expansion

Hence, we assume that the external electromagnetic field can be considered as almost constant relatively to the position of the molecule of interest. Hence, using a Taylor development of the electric potential up to the quadratic order:

$$\phi(\mathbf{r},t) = \phi(\mathbf{0},t) - \mathbf{r}_i \mathbf{e}_i(\mathbf{0},t) - \frac{1}{2} \mathbf{r}_i \mathbf{r}_j \frac{\delta \mathbf{e}_i}{\delta \mathbf{r}_j}(\mathbf{0},t), \qquad (\text{III.10})$$

<sup>&</sup>lt;sup>1</sup>In the plane-wave approximation, the fields or potential propagating in the arbitrary direction z in the molecule vicinity can be written as:  $f(\mathbf{r}, t) = A \cos(n2\pi z/\lambda - \omega t)$ . The field oscillation space frequency is  $n2\pi/\lambda \approx 10^{-3} \text{Å}^{-1}$  in water. Hence, the maximal spatial gradient value is  $An2\pi\Delta/\lambda$  where  $\Delta$  is the extension of the molecule. The water molecule extension is about 1 Å. Hence, the electromagnetic field relative difference between 2 points in the molecule distant by 1 Å for any field or potential is less than 0.1 %.

Chapter III. Describing the second harmonic generation of non-resonant liquid: from QM/MM to Maxwell



Figure III.3: Spacial-evolution of the electromagnetic field for a wave length of 800 nm in vacuum. The orange dots represent a water molecule of size 1 Å. In insert is a zoom at the scale of a water molecule for the largest field evolution (when e(z) = 0).

we obtain this simplified interaction potential:

$$\hat{V}(t) = -\hat{\boldsymbol{\mu}} \cdot \mathbf{e}(t) - \frac{1}{2}\hat{Q} : \boldsymbol{\nabla}\mathbf{e}(t).$$
(III.11)

Where  $\mathbf{e}(t)$  is the "averaged" electric field felt by the molecule and  $\nabla \mathbf{e}(t)$  its spatial gradient. This potential involves two operators: the dipole moment  $\hat{\boldsymbol{\mu}} = -\hat{\mathbf{r}}$  and the quadrupole operator  $\hat{Q}$ . The quadrupole operator has 9 components so that:  $\hat{Q}_{ij} = -\hat{\mathbf{r}}_i \hat{\mathbf{r}}_j$ . Note that we use the Einstein summation convention and that more details can be found in Appendix IX. A.1.

For the molecule of interest, we will thus consider that the interacting Hamiltonian is:

$$\hat{H}(t) = \hat{H}^0 - \hat{\boldsymbol{\mu}} \cdot \mathbf{e}(t) - \frac{1}{2}\hat{Q} : \boldsymbol{\nabla}\mathbf{e}(t)$$
(III.12)

The interacting Hamiltonian present new terms which couples molecular properties,  $\boldsymbol{\mu}$  or Q, with an exciting external field,  $\mathbf{e}$  or  $\nabla \mathbf{e}$ . Hence, the eigenvector solutions of the isolated Hamiltonian  $\hat{H}^0$ are no longer a solution of the interacting one. In the static limit, a new set of eigenvectors  $\{|\Psi_{[\mathbf{e},\nabla\mathbf{e}]}^n\rangle\}$ is obtained. The external field modifies the molecule's Hamiltonian and thus the eigenvectors and eigenvalues and thus the expectation values of any observable.

#### III. B.1.3 How to solve the time-dependency?

#### Time-dependent problem:

Previously, we have assumed that we could solve any Hamiltonian: we are thus able to describe the electronic degrees of freedom and compute any properties. However, one major requirement was that the Hamiltonian has to be time-independent, which is not our case. Indeed, the electromagnetic field is oscillating in time at the frequency  $\omega$ . Even if we manage to solve the Hamiltonian at a given time  $t_0$ , the obtained solution would not be stationary. Let us suppose that the wave-function of the ground state at time  $t_0$  is:  $|\Psi\rangle(t_0) = |\Psi_{t_0}^0\rangle$  Applying the Schrödinger Equation just after  $t_0$ leads to :

at 
$$t_0^+$$
:  $i\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\Psi\rangle (t_0^+) = \hat{H}(t_0^+) |\Psi\rangle (t_0^+) = \hat{H}(t_0^+) |\Psi_{t_0}^0\rangle$ . (III.13)

But the state  $|\Psi_{t_0}^0\rangle$  is no longer an eigenstate of  $\hat{H}(t_0^+)$  because the fields have evolved! Hence, the wave-function cannot be expressed as pure state using only the eigenvalues found at a given time.

Therefore, it is much more difficult to solve the **explicit** time-dependent Hamiltonian. To overcome this situation, one can use at least three approaches: Finite Field, Real Time or Perturbative ones.

#### Finite Field (FF) method:

One way to solve this problem is to return to the static limit. In this case, the expectation value of any observable is known because the interacting Hamiltonian becomes time-independent, and thus can be solved. Thus, to obtain the molecular properties, such as the polarizability or the hyperpolarizability, one can use the *Finite Field* (FF) approach.

Briefly, one should apply an external field to the system by adding the interacting term in the Hamiltonian and extract the observable. Using several field amplitudes, we can obtain the molecular response within a linear or quadratic approximation. We present a concrete example in Sec III. B.2.3.

One major drawback of this method is that it completely neglects the dynamic response of the electron. In our case, even if the fundamental and the second harmonic frequencies are far from resonances the response of the molecule still depends on the electromagnetic field frequency. The frequency dispersion and its potential consequences in the SHG response, in particular at the liquid-gas interface (see Chapter VI), can be important. Hence, even in the non-resonant case, we would like to have a frequency dependent calculation method.

#### Real Time method:

Another way to solve this problem is by propagating in real time the wave-function using the

Schrödinger equation. Indeed, one can see it as a time propagator. At the first order:

$$|\Psi\rangle (t + \delta t) = |\Psi\rangle (t) + \frac{\delta t}{i\hbar} \hat{H}(t) |\Psi\rangle (t)$$
(III.14)

To obtain the wave-function at the time  $t + \delta t$ , one should know the initial wave-function at the time t, and the interacting Hamiltonian at time t. Once the wave-function is known during enough time, we can compute the time evolution of any observable:  $\Omega(t) = \langle \Psi(t) | \hat{\Omega} | \Psi(t) \rangle$ . For instance, to compute the polarizability, one should add a time-dependent  $\mathbf{e}(t)$  and measure the  $\boldsymbol{\mu}(t)$ . The ratio between the 2 time-dependent quantities should exactly be the polarizability.

This method is extremely powerful because it does not assume a particular form of the wavefunction at a given time t. Thus, the wave-function can explore a different part of the Hilbert space than the one of the isolated Hamiltonian. However, this approach is computationally demanding because it implies to solve the electronic dynamics with a resolution of about the Atosecond, during the duration of the electromagnetic oscillation (several femtoseconds at least).

#### Perturbative method:

In our experiment, we use a strong electromagnetic field with an average power of about 1 W over a spatial spot about 100  $\mu$ m<sup>2</sup>. The averaged power density, which is the unit of the Poynting vector  $\mathbf{\Pi}$ , is 10<sup>8</sup> W.m<sup>-2</sup>. To compare with something more familiar in real life, it is 100 000 times more intense than the light emitted by the Sun from Earth. Moreover, the laser field is not a continuous electromagnetic field, but consists in femtosecond pulses. A minimal duration for our pulse would be 100 fs with a repetition rate of 80 MHz. This means that the actual energy of a single pulse is  $1.25 \ 10^{-8}$  J, which corresponds to  $1.25 \ 10^{5}$  W during the pulse in averaged. Hence, the intensity of one pulse is about  $1.25 \ 10^{13}$  W.m<sup>-2 2</sup> To get the electric field, one can use the relation:  $|\mathbf{\Pi}| = |\mathbf{e}^2|c\epsilon_0$ in vacuum. Hence, the electric field amplitude is about  $1.25 \ 10^{8}$  V.m<sup>-1</sup>.

This extremely large external field used for Non Linear optical experiment corresponds to only  $4.3 \ 10^{-4}$  a.u. For comparison, the electric field produced by a nucleus to its core electron is about 1 a.u. Hence, the interaction potential presented in Equation III.12 are **orders of magnitudes** smaller than the isolated  $\hat{H}^0$ .

Therefore, a major hypothesis may be made: the Hilbert space explored by this time-dependent wave-function is the **same** as the one of the isolated molecules. In other words, the accessible electronic state is not very different under this electric field than without. Thus, the interaction potential can be treated as a perturbation. This allows us to write the time-dependent wave-function using the eigenstate of the isolated Hamiltonian for basis set. For instance, one can write:

<sup>&</sup>lt;sup>2</sup>Physical insights: let us compute the Earth-Sun distance needed to have the same surface power. The actual distance is about 150 million of km, the power receive on Earth is about 1000 W.m<sup>-2</sup>. To obtain the same power as the laser,  $1.25 \ 10^{13} \ W.m^{-2}$ , the Sun would be at 1 300 km from Earth. As a reminder, the international space station is at 400 km and the Moon at 300 000 km.

$$\left|\Psi\right\rangle(t) = \sum_{n} c_{n}(t) \left|\Psi_{\hat{H}^{0}}^{n}\right\rangle \tag{III.15}$$

Using a collection of time-dependent complex function  $c_n$  to weight the contributions of every eigenstate in the time-dependent wave-function.

#### Which approach?

Each of these methods has its advantages and drawbacks. In the literature, we can find the Real Time approach for Extreme Non Linear Optics, where the dynamic electronic degrees of freedom are primordial **and** the electronic structure is very different from the unperturbed case. This is the case in High Harmonic Generation (HHG) processes [84–86] for instance. Moreover, in the SFG community, the Real Time approach is also used to obtain the nuclear vibrational states with high accuracy [16, 37].

The Finite Field scheme and the perturbative ones are more widespread because they are way much cheaper numerically speaking. For NLO property, we can find FF approach for SFG [36, 57] and SHG [20, 41, 55, 87–89].

One major advantage of the FF approach is that it can be used to straightforwardly take into account the local field correction [41]. The perturbation approach is also used in SHG [30, 52, 90, 91]. We will use this last technique for several reasons:

- The Real Time approach would be much too expensive and quite overkill because we are in the non-resonant regime with weak external field.
- We want to perform frequency-dependent calculations to test the Kleinman symmetry hypothesis (presented later on).
- The Finite Field approach is more numerically demanding than the perturbation scheme.

Nonetheless, we will often compare the results obtained by the perturbative approach and the FF one at zero frequency. In this case, the FF method should be more reliable because it does not constraint the wave-function: comparing both approaches is a very good test of the perturbative hypothesis validity.

#### III. B.1.4 Response scheme

In this part we recall some important outcomes of the response framework as presented in the Chapter 5 of *Principles and Practices of Molecular Properties* [1] which is one application of the perturbative approach.

#### Fourier decomposition:

The external electromagnetic field is supposed to behave like a plane wave, or a superposition of

plane waves:

$$\mathbf{e}(\mathbf{r},t) = \sum_{\omega} \mathbf{e}^{\omega}(\mathbf{r}) \exp[-i\omega t], \qquad \nabla \mathbf{e}(\mathbf{r},t) = \sum_{\omega} \nabla \mathbf{e}^{\omega}(\mathbf{r}) \exp[-i\omega t], \qquad (\text{III.16})$$

where the wave frequency  $\omega$  can be positive or negative. Hence, the interaction potential, defined in Equation III.12, is written as:

$$\hat{V}(t) = -\sum_{\omega} \hat{\boldsymbol{\mu}} \cdot \mathbf{e}^{\omega}(\mathbf{0}) \exp[-i\omega t] - \frac{1}{2} \sum_{\omega} \hat{Q} : \boldsymbol{\nabla} \mathbf{e}^{\omega}(\mathbf{0}) \exp[-i\omega t]$$
(III.17)

Or in the general form:

$$\hat{V}(t) = \sum_{\alpha} \sum_{\omega} \hat{V}_{\alpha} \mathbf{F}^{\omega}_{\alpha} \exp[-i\omega t].$$
(III.18)

Where  $\hat{V}_{\alpha}$  is the potential operator,  $-\hat{\mu}$  or  $-1/2\hat{Q}$ , and  $\mathbf{F}_{\alpha}^{\omega}$  the external excitation,  $\mathbf{e}^{\omega}(\mathbf{0})$  or  $\nabla \mathbf{e}^{\omega}(\mathbf{0})$ .

#### Equation of motion of the time-dependent wave-function:

The time-dependent wave-function is decomposed on the eigenstates of the isolated Hamiltonian  $\{|\Psi^n\rangle\}$  noted  $\{|n\rangle\}$  with associated eigenvalues  $\{E_n\}$  – or frequencies  $\{\omega_n\}$  with  $E_n = \hbar\omega_n$ . We use the time dependent coefficients  $\{d_n(t)\}$  as follows:

$$\left|\Psi\right\rangle(t) = \sum_{n} d_{n}(t) \exp[-i\omega_{n}t] \left|n\right\rangle \tag{III.19}$$

When the interaction potential is small, the wave-function is decomposed in order of magnitude as:

$$|\Psi\rangle(t) = |\Psi^{(0)}\rangle(t) + |\Psi^{(1)}\rangle(t) + |\Psi^{(2)}\rangle(t) + \cdots$$
 (III.20)

From this point, one has to insert the previous expression of  $|\Psi\rangle(t)$  in the Schrödinger equation with the interaction potential  $\hat{V}(t)$ , develops in order of magnitudes and inverts in order to get the expression of  $|\Psi^{(n)}\rangle(t)$ , see Appendix IX. A.1 for more details. We obtain for the zeroth, first and second perturbation orders:

$$|\Psi^{(0)}\rangle(t) = \exp[-i\omega_0 t] |0\rangle, \qquad (\text{III.21})$$

$$|\Psi^{(1)}\rangle(t) = \frac{1}{\hbar} \sum_{n} \sum_{\alpha} \sum_{\omega_{1}} \frac{\langle n | \hat{V}_{\alpha} | 0 \rangle}{\omega_{1} - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega_{1}} \exp[-i(\omega_{1} - \omega_{n0} + \omega_{n})t] | n \rangle, \qquad (\text{III.22})$$

$$\left|\Psi^{(2)}\right\rangle(t) = \frac{1}{\hbar^2} \sum_{\alpha,\alpha'} \sum_{\omega_1,\omega_2} \sum_{n,m} \frac{\langle n | \hat{V}_{\alpha} | 0 \rangle}{\omega_1 - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega_1} \frac{\langle m | \hat{V}_{\alpha'} | n \rangle}{\omega_1 + \omega_2 - \omega_{m0}} \mathbf{F}_{\alpha'}^{\omega_2} \exp\left[-i(\omega_1 + \omega_2 - \omega_{m0} + \omega_m)t\right] | m \rangle.$$
(III.23)

These expressions involve transitions between the ground state  $|0\rangle$  and excited states  $|n\rangle$  or  $|m\rangle$ 

with the associated operator of the interaction  $\hat{V}_{\alpha}$ . Associated with these transitions, it appears at the denominator the differences in energy between these states  $\omega_{n0} = \omega_n - \omega_0$ . The zeroth order  $|\Psi^{(0)}\rangle(t)$  does not involve the perturbation (no  $\mathbf{F}_{\alpha}^{\omega_1}$ ) while the first order is proportional to the perturbations ( $\mathbf{F}_{\alpha}^{\omega_1}$ ) and the second order is quadratic with respect to the perturbation ( $\mathbf{F}_{\alpha}^{\omega_1}$  and  $\mathbf{F}_{\alpha'}^{\omega_2}$ ). For the quadratic term, two interacting potential,  $\hat{V}_{\alpha}$  and  $\hat{V}_{\alpha'}$ , associated with two external fields,  $\mathbf{F}_{\alpha}^{\omega_1}$  and  $\mathbf{F}_{\alpha'}^{\omega_2}$ , are involved. This means that:

- The same light-matter interaction can perturb twice the wave function, *i.e.*  $\hat{V}_{\alpha'} = \hat{V}_{\alpha}$  and  $\mathbf{F}_{\alpha}^{\omega_1} = \mathbf{F}_{\alpha'}^{\omega_2}$ .
- The same type of light-matter interaction can perturb twice the wave-function, but not with the same component. For instance, if an electric field lies along the *b* and *c* molecular axis, the dipole interaction bears two terms:  $-\hat{\mu}_b \mathbf{e}_b$  and  $-\hat{\mu}_c \mathbf{e}_c$ . Therefore, there can be  $\hat{V}_{\alpha'} = -\hat{\mu}_b$  and  $\hat{V}_{\alpha} = -\hat{\mu}_c$  and  $\mathbf{F}_{\alpha'}^{\omega_1} = \mathbf{e}_b^{\omega_1}$ ,  $\mathbf{F}_{\alpha}^{\omega_2} = \mathbf{e}_c^{\omega_2}$ .
- Different types of light-matter interaction can perturb once each the wave-function. For instance, if we have an electric field along the *c* direction and a gradient along the *c* direction, we get:  $\hat{V}_{\alpha'} = -\hat{\mu}_c$ ,  $\hat{V}_{\alpha} = -1/2\hat{Q}_{cc}$  and  $\mathbf{F}_{\alpha'}^{\omega_1} = \mathbf{e}_c^{\omega_1}$ ,  $\mathbf{F}_{\alpha}^{\omega_2} = \nabla_c \mathbf{e}_c^{\omega_2} = \delta \mathbf{e}_c^{\omega_2}/\delta c$ .

#### Kubo Relation:

From the time-dependent wave-function, we can compute any observable up to the second order. The observable  $\Omega$  is also decomposed in orders of magnitude:

$$\Omega(t) = \Omega^{(0)}(t) + \Omega^{(1)}(t) + \Omega^{(2)}(t) + \cdots$$
(III.24)

Combining with Equation III.20 and identifying the terms by order of magnitude:

$$\Omega^{(0)}(t) = \langle \Psi^{(0)} | \hat{\Omega} | \Psi^{(0)} \rangle (t)$$
(III.25)

$$\Omega^{(1)}(t) = \langle \Psi^{(1)} | \hat{\Omega} | \Psi^{(0)} \rangle (t) + \langle \Psi^{(0)} | \hat{\Omega} | \Psi^{(1)} \rangle (t)$$
(III.26)

$$\Omega^{(2)}(t) = \langle \Psi^{(2)} | \hat{\Omega} | \Psi^{(0)} \rangle (t) + \langle \Psi^{(1)} | \hat{\Omega} | \Psi^{(1)} \rangle (t) + \langle \Psi^{(0)} | \hat{\Omega} | \Psi^{(2)} \rangle (t)$$
(III.27)

Incorporating the values of  $|\Psi^{(m)}\rangle$ , Equations III.21 to III.23, these equations becomes the Kubo relations:

$$\Omega^{(0)}(t) = \langle 0 | \hat{\Omega} | 0 \rangle \tag{III.28}$$

$$\Omega^{(1)}(t) = \sum_{\alpha} \sum_{\omega_1} \langle \langle \hat{\Omega}; \hat{V}^{\omega_1}_{\alpha} \rangle \rangle \mathbf{F}^{\omega_1}_{\alpha} \exp[-i\omega_1 t]$$
(III.29)

$$\Omega^{(2)}(t) = \frac{1}{2!} \sum_{\alpha,\alpha'} \sum_{\omega_1,\omega_2} \langle \langle \hat{\Omega}; \hat{V}^{\omega_1}_{\alpha}, \hat{V}^{\omega_2}_{\alpha'} \rangle \rangle \mathbf{F}^{\omega_1}_{\alpha} \mathbf{F}^{\omega_2}_{\alpha'} \exp[-i(\omega_1 + \omega_2)t]$$
(III.30)

Where the  $\langle \langle \rangle \rangle$  are called the **response function** that we obtained using the DALTON software.

In Appendix IX. A.1 we provide more details about the calculations of these response functions. For the first order, the response functions for a given observable  $\Omega$  upon a perturbation  $\hat{V}_{\alpha}$  at the frequency  $\omega_1$  is:

$$\langle\langle\hat{\Omega};\hat{V}_{\alpha}^{\omega_{1}}\rangle\rangle = -\frac{1}{\hbar} \left[\sum_{n} \frac{\langle 0|\hat{\Omega}|n\rangle \langle n|\hat{V}_{\alpha}|0\rangle}{\omega_{n0} - \omega_{1}} + \frac{\langle 0|\hat{V}_{\alpha}|n\rangle \langle n|\hat{\Omega}|0\rangle}{\omega_{n0} + \omega_{1}}\right]$$
(III.31)

and at the second order, the response function for a given observable  $\Omega$  upon a perturbation  $\hat{V}_{\alpha}$  and  $\hat{V}_{\alpha'}$  at the frequency  $\omega_1$  and  $\omega_2$  respectively is:

$$\langle \langle \hat{\Omega}; \hat{V}_{\alpha}^{\omega_{1}}, \hat{V}_{\alpha'}^{\omega_{2}} \rangle \rangle = \frac{1}{\hbar^{2}} \mathcal{P}_{1,2} \sum_{n,m} \left[ \frac{\langle 0 | \hat{V}_{\alpha'} | m \rangle \langle m | \hat{V}_{\alpha} | n \rangle \langle n | \hat{\Omega} | 0 \rangle}{(\omega_{n0} + \omega_{1} + \omega_{2})(\omega_{m0} + \omega_{2})} + \frac{\langle 0 | \hat{V}_{\alpha} | n \rangle \langle n | \hat{\Omega} | m \rangle \langle m | \hat{V}_{\alpha'} | 0 \rangle}{(\omega_{n0} + \omega_{1})(\omega_{m0} - \omega_{2})} + \frac{\langle 0 | \hat{\Omega} | n \rangle \langle n | \hat{V}_{\alpha} | m \rangle \langle m | \hat{V}_{\alpha} | m \rangle \langle m | \hat{V}_{\alpha'} | 0 \rangle}{(\omega_{n0} - \omega_{1} - \omega_{2})(\omega_{m0} - \omega_{2})} \right]$$
(III.32)

Where  $\mathcal{P}_{1,2}$  is the permutation operator which permutes the index 1 to 2 and  $\alpha$  to  $\alpha'$ .

#### III. B.1.5 Calculating the molecular properties

The quantities of interest in our work are the dipole and the quadrupole moment of the molecule under an external field, and at the Second Harmonic frequency. First, we will present the dipolar approximation, and then how to deal with terms up to quadrupoles.

#### SHG within Dipolar approximation:

At the dipolar approximation and for a monochromatic wave, we assume that the interacting potential is:  $-\hat{\mu} \cdot \mathbf{e}^{\omega} \exp[-i\omega t]$ . For simplicity, we are not writing the  $-\omega$  frequency associated, and the optical rectification process ( $\omega - \omega = 0$ ) for the second order terms. The Kubo relations for the dipole moment of the molecule,  $\hat{\Omega} = +\hat{\mu}$  is:

$$\boldsymbol{\mu}_{i}(t) = \boldsymbol{\mu}_{i}^{0} - \langle \langle \hat{\boldsymbol{\mu}}_{i}; \hat{\boldsymbol{\mu}}_{j}^{\omega} \rangle \rangle \mathbf{e}_{j}^{\omega} \exp[-i\omega t] + \frac{1}{2!} \langle \langle \hat{\boldsymbol{\mu}}_{i}; \hat{\boldsymbol{\mu}}_{j}^{\omega}, \hat{\boldsymbol{\mu}}_{k}^{\omega} \rangle \rangle \mathbf{e}_{j}^{\omega} \mathbf{e}_{k}^{\omega} \exp[-i2\omega t]$$
(III.33)

Using the Einstein summation convention. This formula can be written using tensor  $\alpha(\omega, \omega)$ and  $\beta^{dd}(2\omega, \omega, \omega)$  so that:

$$\boldsymbol{\mu}(t) \equiv \boldsymbol{\mu}^{0} + \alpha(\omega, \omega) \cdot \mathbf{e}^{\omega} \exp[-i\omega t] + \frac{1}{2!} \beta^{dd}(2\omega, \omega, \omega) : \mathbf{e}^{\omega} \mathbf{e}^{\omega} \exp[-i2\omega t]$$
(III.34)

 $\mu^0$  is the permanent dipole moment of the molecule,  $\alpha(\omega, \omega)$  its polarizability at the fundamental frequency  $\omega$  and  $\beta^{dd}(2\omega, \omega, \omega)$  the first hyperpolarizability. By identification, the wanted molecular optical response properties are thus expressed with the response functions:

$$\alpha(\omega,\omega)_{ij} = -\langle \langle \hat{\boldsymbol{\mu}}_i; \hat{\boldsymbol{\mu}}_j^{\omega} \rangle \rangle, \qquad \beta^{dd} (2\omega,\omega,\omega)_{ijk} = +\langle \langle \hat{\boldsymbol{\mu}}_i; \hat{\boldsymbol{\mu}}_j^{\omega}, \hat{\boldsymbol{\mu}}_k^{\omega} \rangle \rangle. \qquad (\text{III.35})$$

#### Page 53

#### SHG at the quadrupolar order:

Now, we consider the light-matter interaction due to an external electric field with spatial gradients:  $-\hat{\mu} \cdot \mathbf{e}^{\omega} \exp[-i\omega t] - \frac{1}{2}\hat{Q} : \nabla \mathbf{e}^{\omega} \exp[-i\omega t]$ . Now there are two types of interacting potential. As for the dipolar case, we will only consider the second harmonic term for the non-linear contribution to the observable. At this order, we will not only compute the molecular dipole moment, but also the quadrupolar one. Therefore, applying the Kubo relation leads to :

$$\boldsymbol{\mu}_{i}(t) = \langle 0 | \, \hat{\boldsymbol{\mu}}_{i} | 0 \rangle + \sum_{j} \langle \langle \hat{\boldsymbol{\mu}}_{i}; -\hat{\boldsymbol{\mu}}_{j}^{\omega} \rangle \rangle \mathbf{e}_{j}^{\omega} \exp[-i\omega t] + \sum_{j,k} \langle \langle \hat{\boldsymbol{\mu}}_{i}; -\frac{1}{2} \hat{Q}_{jk}^{\omega} \rangle \rangle \nabla_{j} \mathbf{e}_{k}^{\omega} \exp[-i\omega t] \\ + \frac{1}{2!} \sum_{j,k} \langle \langle \hat{\boldsymbol{\mu}}_{i}; -\hat{\boldsymbol{\mu}}_{j}^{\omega}, -\hat{\boldsymbol{\mu}}_{k}^{\omega} \rangle \rangle \mathbf{e}_{j}^{\omega} \mathbf{e}_{k}^{\omega} \exp[-i(\omega + \omega)t] \\ + \frac{1}{2!} \sum_{j,k,l} \langle \langle \hat{\boldsymbol{\mu}}_{i}; -\hat{\boldsymbol{\mu}}_{j}^{\omega}, -\frac{1}{2} \hat{Q}_{kl}^{\omega} \rangle \rangle \mathbf{e}_{j}^{\omega} \nabla_{k} \mathbf{e}_{l}^{\omega} \exp[-i(\omega + \omega)t] \\ + \frac{1}{2!} \sum_{j,k,l} \langle \langle \hat{\boldsymbol{\mu}}_{i}; -\frac{1}{2} \hat{Q}_{kl}^{\omega}, -\hat{\boldsymbol{\mu}}_{j}^{\omega} \rangle \rangle \mathbf{e}_{j}^{\omega} \nabla_{k} \mathbf{e}_{l}^{\omega} \exp[-i(\omega + \omega)t]$$
(III.36)  
$$\boldsymbol{\mu}_{i}(t) = \boldsymbol{\mu}_{i}^{0} - \langle \langle \hat{\boldsymbol{\mu}}_{i}; \hat{\boldsymbol{\mu}}_{j}^{\omega} \rangle \rangle \mathbf{e}_{i}^{\omega} \exp[-i\omega t] - \frac{1}{2} \langle \langle \hat{\boldsymbol{\mu}}_{i}; \hat{Q}_{ik}^{\omega} \rangle \rangle \nabla_{j} \mathbf{e}_{k}^{\omega} \exp[-i\omega t]$$

$$+\frac{1}{2!}\langle\langle\hat{\boldsymbol{\mu}}_{i};\hat{\boldsymbol{\mu}}_{j}^{\omega},\hat{\boldsymbol{\mu}}_{k}^{\omega}\rangle\rangle\mathbf{e}_{j}^{\omega}\mathbf{e}_{k}^{\omega}\exp[-i2\omega t] +\frac{1}{2!}\langle\langle\hat{\boldsymbol{\mu}}_{i};\hat{\boldsymbol{\mu}}_{j}^{\omega},\hat{Q}_{kl}^{\omega}\rangle\rangle\mathbf{e}_{j}^{\omega}\nabla_{k}\mathbf{e}_{l}^{\omega}\exp[-i2\omega t]$$
(III.37)

Two new kinds of terms arise. One induces a molecular dipole oscillating at the frequency  $\omega$  due to an external field gradient. Another creates a molecular dipole oscillating at the frequency  $2\omega$  by the interaction with the electric field value (throughout the molecule dipole moment) and the electric field spacial gradient (throughout the molecule quadrupole moment). Introducing two new tensors, the quadrupole polarizability  $\alpha^{dq}(\omega, \omega)$  and the dipole-quadrupole first hyperpolarizability  $\beta^{dq}(2\omega, \omega, \omega)$ , we have:

$$\boldsymbol{\mu}_{i}(t) \equiv \boldsymbol{\mu}_{i}^{0} + \alpha(\omega, \omega) \cdot \mathbf{e}^{\omega} \exp[-i\omega t] + \alpha^{dq}(\omega, \omega) : \boldsymbol{\nabla} \mathbf{e}^{\omega} \exp[-i\omega t] \\ + \frac{1}{2!} \beta^{dd}(2\omega, \omega, \omega) : \mathbf{e}^{\omega} \mathbf{e}^{\omega} \exp[-i2\omega t] + \frac{1}{2!} \beta^{dq}(2\omega, \omega, \omega) : \mathbf{e}^{\omega} \boldsymbol{\nabla} \mathbf{e}^{\omega} \exp[-i2\omega t], \quad (\text{III.38})$$

And by identification with Equation III.37

$$\alpha^{dq}(\omega,\omega)_{ijk} = -\frac{1}{2!} \langle \langle \hat{\boldsymbol{\mu}}_i; \hat{Q}_{jk}^{\omega} \rangle \rangle \qquad \beta^{dq}(2\omega,\omega,\omega)_{ijkl} = + \langle \langle \hat{\boldsymbol{\mu}}_i; \hat{\boldsymbol{\mu}}_j^{\omega}, \hat{Q}_{kl}^{\omega} \rangle \rangle \qquad (\text{III.39})$$

Similarly, we can also compute the induced quadrupolar moment of the molecule, *i.e.*  $\Omega = Q$ :

$$Q_{ij}(t) = \langle 0|\hat{Q}_{ij}|0\rangle + \sum_{k} \langle \langle \hat{Q}_{ij}; -\hat{\boldsymbol{\mu}}_{k}^{\omega} \rangle \rangle \mathbf{e}_{k}^{\omega} \exp[-i\omega t] + \frac{1}{2!} \sum_{k,l} \langle \langle \hat{Q}_{ij}; -\hat{\boldsymbol{\mu}}_{k}^{\omega}, -\hat{\boldsymbol{\mu}}_{l}^{\omega} \rangle \rangle \mathbf{e}_{k}^{\omega} \mathbf{e}_{l}^{\omega} \exp[-i(\omega+\omega)t]$$
(III.40)

$$Q_{ij}(t) = Q_{ij}^0 - \langle \langle \hat{Q}_{ij}; \hat{\boldsymbol{\mu}}_k^{\omega} \rangle \rangle \mathbf{e}_k^{\omega} \exp[-i\omega t] + \frac{1}{2!} \langle \langle \hat{Q}_{ij}; \hat{\boldsymbol{\mu}}_k^{\omega}, \hat{\boldsymbol{\mu}}_l^{\omega} \rangle \rangle \mathbf{e}_k^{\omega} \mathbf{e}_l^{\omega} \exp[-i(\omega + \omega)t]$$
(III.41)

#### Page 54

Defining the quadrupolar polarizability  $\alpha^{qd}(\omega, \omega)$  and the quadrupole-dipole first hyperpolarizability  $\beta^{qd}(2\omega, \omega, \omega)$  so that:

$$Q(t) \equiv Q^0 + \alpha^{qd}(\omega, \omega) \cdot \mathbf{e}^{\omega} \exp[-i\omega t] + \frac{1}{2!} \beta^{qd}(2\omega, \omega, \omega) : \mathbf{e}^{\omega} \mathbf{e}^{\omega} \exp[-i2\omega t], \qquad (\text{III.42})$$

Where

$$\alpha^{qd}(\omega,\omega)_{ijk} = -\langle\langle\hat{Q}_{ij};\hat{\boldsymbol{\mu}}_k^{\omega}\rangle\rangle, \qquad \beta^{qd}(2\omega,\omega,\omega)_{ijkl} = \langle\langle\hat{Q}_{ij};\hat{\boldsymbol{\mu}}_k^{\omega},\hat{\boldsymbol{\mu}}_l^{\omega}\rangle\rangle. \tag{III.43}$$

Note that we only kept terms that involve one quadrupolar operator. Indeed for our molecules, the quadrupole is much smaller than the dipole.

#### Frequency symmetry:

As all the perturbations are at the same frequency, the fundamental one  $\omega$ , the different quadratic response tensors respect the *intrasinc* symmetry:

$$\langle\langle\hat{\boldsymbol{\mu}}_{i};\hat{\boldsymbol{\mu}}_{j}^{\omega},\hat{\boldsymbol{\mu}}_{k}^{\omega}\rangle\rangle = \langle\langle\hat{\boldsymbol{\mu}}_{i};\hat{\boldsymbol{\mu}}_{k}^{\omega},\hat{\boldsymbol{\mu}}_{j}^{\omega}\rangle\rangle \quad \langle\langle\hat{\boldsymbol{\mu}}_{i};\hat{\boldsymbol{\mu}}_{j}^{\omega},\hat{\boldsymbol{Q}}_{kl}^{\omega}\rangle\rangle = \langle\langle\hat{\boldsymbol{\mu}}_{i};\hat{\boldsymbol{Q}}_{kl}^{\omega},\hat{\boldsymbol{\mu}}_{j}^{\omega}\rangle\rangle \quad \langle\langle\hat{\boldsymbol{Q}}_{ij};\hat{\boldsymbol{\mu}}_{k}^{\omega},\hat{\boldsymbol{\mu}}_{l}^{\omega}\rangle\rangle = \langle\hat{\boldsymbol{Q}}_{ij};\hat{\boldsymbol{\mu}}_{k}^{\omega},\hat{\boldsymbol{\mu}}_{k}^{\omega}\rangle\rangle$$

In the limit of fully non-resonant case, or in the infinite wave-length limit, the Kleinman frequency symmetry also applies:

$$\langle \langle \hat{\boldsymbol{\mu}}_i; \hat{\boldsymbol{\mu}}_j^{\omega}, \hat{\boldsymbol{\mu}}_k^{\omega} \rangle \rangle = \langle \langle \hat{\boldsymbol{\mu}}_j; \hat{\boldsymbol{\mu}}_i^{\omega}, \hat{\boldsymbol{\mu}}_k^{\omega} \rangle \rangle \qquad \qquad \langle \langle \hat{\boldsymbol{\mu}}_i; \hat{\boldsymbol{\mu}}_j^{\omega}, \hat{Q}_{kl}^{\omega} \rangle \rangle = \langle \langle \hat{Q}_{kl}; \hat{\boldsymbol{\mu}}_i^{\omega}, \hat{\boldsymbol{\mu}}_j^{\omega} \rangle \rangle$$

However, in our notations of the first hyperpolarizabilities, the intrinsic symmetry is not fulfilled for the dipole-quadrupole term:  $\beta^{dq}$ . Indeed, we have defined  $\beta^{dq}$  so that the last term of Equation III.38 is:

$$\boldsymbol{\mu}_{i}^{dq} = \frac{1}{2!} \sum_{j,k,l} \langle \langle \hat{\boldsymbol{\mu}}_{i}; -\hat{\boldsymbol{\mu}}_{j}^{\omega}, -\frac{1}{2} \hat{Q}_{kl}^{\omega} \rangle \rangle \mathbf{e}_{j}^{\omega} \boldsymbol{\nabla}_{k} \mathbf{e}_{l}^{\omega} + \frac{1}{2!} \sum_{j,k,l} \langle \langle \hat{\boldsymbol{\mu}}_{i}; -\frac{1}{2} \hat{Q}_{jk}^{\omega}, -\hat{\boldsymbol{\mu}}_{l}^{\omega} \rangle \rangle \boldsymbol{\nabla}_{j} \mathbf{e}_{k}^{\omega} \mathbf{e}_{l}^{\omega}$$
(III.44)

$$\boldsymbol{\mu}_{i}^{dq} \equiv \frac{1}{2!} \beta^{dq} : \mathbf{e}^{\omega} \boldsymbol{\nabla} \mathbf{e} = \frac{1}{2!} \sum_{j,k,l} \beta_{ijkl}^{dq} \mathbf{e}_{j}^{\omega} \boldsymbol{\nabla}_{k} \mathbf{e}_{l}$$
(III.45)

Let us work on the second sum of the first definition in Equation III.44. Using the intrinsic symmetry and renaming the dummy index of the sum:

$$\sum_{j,k,l} \langle \langle \hat{\boldsymbol{\mu}}_i; -\frac{1}{2} \hat{Q}_{jk}^{\omega}, -\hat{\boldsymbol{\mu}}_l^{\omega} \rangle \rangle \boldsymbol{\nabla}_j \mathbf{e}_k^{\omega} \mathbf{e}_l^{\omega} = \sum_{j,k,l} \langle \langle \hat{\boldsymbol{\mu}}_i; -\hat{\boldsymbol{\mu}}_l^{\omega}, -\frac{1}{2} \hat{Q}_{jk}^{\omega} \rangle \rangle \mathbf{e}_k^{\omega} \boldsymbol{\nabla}_j \mathbf{e}_l^{\omega} = \sum_{j,k,l} \langle \langle \hat{\boldsymbol{\mu}}_i; -\hat{\boldsymbol{\mu}}_j^{\omega}, -\frac{1}{2} \hat{Q}_{kl}^{\omega} \rangle \rangle \mathbf{e}_j^{\omega} \boldsymbol{\nabla}_k \mathbf{e}_l^{\omega}$$

Hence, as proposed above:  $\beta_{ijkl}^{dq} = \langle \langle \hat{\mu}_i; \hat{\mu}_j^{\omega}, \hat{Q}_{kl}^{\omega} \rangle \rangle$ . This value is different from  $\beta_{ikjl}^{dq} = \langle \langle \hat{\mu}_i; \hat{\mu}_k^{\omega}, \hat{Q}_{jl}^{\omega} \rangle \rangle$  because we are exchanging an index of  $\hat{Q}^{\omega}$  with one of  $\hat{\mu}^{\omega}$ . Thus, here are a new set of frequency

symmetries for the first hyperpolarizabilities in the general case:

$$\begin{aligned} \beta_{ijk}^{dd} &= \beta_{ikj}^{dd}, \text{ intrinsic symetry} \\ \beta_{ijkl}^{dq} &= \beta_{ijlk}^{dq}, \text{ because } \hat{Q}_{lk} = \hat{Q}_{kl} \\ \beta_{ijkl}^{qd} &= \beta_{ijlk}^{qd}, \text{ intrinsic symetry} \\ \beta_{ijkl}^{qd} &= \beta_{jikl}^{qd}, \text{ because } \hat{Q}_{ij} = \hat{Q}_{ji} \end{aligned}$$

At zero frequency (Kleinman symmetry):

$$\beta_{ijk}^{dd} = \beta_{ikj}^{dd} = \beta_{jik}^{dd}, \ \beta_{ijkl}^{dq} = \beta_{jikl}^{dq} = \beta_{klij}^{qd} = \beta_{klji}^{qd}, \text{ intrinsic + Kleinman symetry}$$
$$\beta_{ijkl}^{dq} = \beta_{ijlk}^{dq}, \ \beta_{ijkl}^{qd} = \beta_{jikl}^{dq}, \text{ because } \hat{Q}_{lk} = \hat{Q}_{kl}$$

In top of these symmetries should be added the one due to the molecular structure. This point is discussed later on in Chapter V and more extensively in the second part of Chapter VI.

### III. B.1.6 Conclusive remarks:

Before going to the next part, let us review a bit what we have established and highlight some important details. Note that more details are presented in Appendix IX. A.1 regarding the quadrupole convention, the T or V conventions and the gauge variability.

### What are the approximations used?

Several approximations have been made at different steps:

- $\hat{H}^0$ : We have neglected all direct spin and relativistic effects interactions. The nuclei are supposed to be frozen within the Born-Oppenheimer approximation and do not participate in the light-matter interaction.
- Light-matter interaction: We have supposed that the electromagnetic field can be described using time-Fourier decomposition. Moreover, we have assumed that the field applied have only one frequency (monochromatic interaction). We also assumed that the molecule does not affect the shape and amplitude of the felt electromagnetic field. We neglect all magnetic contributions and focus on the electric one. Using a multipolar decomposition, two new terms arise: involving the value of the electric field and its spatial gradient. Finally, we have assumed that these interactions are largely smaller than the one of the unperturbed Hamiltonian to apply the perturbation theory.
- Time-dependency: We use the perturbation theory to develop the time-dependent wavefunction and any observable in function of the amount of interaction terms involved. The eigenstates of the unperturbed system are assumed to be a complete basis used to develop

the time-dependent wave-function. We neglect all non-optical relaxation terms – for instance vibrational.

• Response scheme: We are interested only in the SHG terms. Only terms involving one quadrupole operator are kept: the other ones are assumed to be smaller and then neglected.

### How to get the molecular response in SHG?

Within the response scheme, the Kubo relations link response functions to the molecular properties of interest such as the first hyperpolarizability. We defined several molecular response functions:  $\langle\langle \hat{\mu}; \hat{\mu}^{\omega}, \hat{\mu}^{\omega} \rangle \rangle$ ,  $\langle\langle \hat{\mu}; \hat{\mu}^{\omega}, \hat{Q} \rangle \rangle$  and  $\langle\langle \hat{Q}; \hat{\mu}^{\omega}, \hat{\mu}^{\omega} \rangle \rangle$ . They involve transition moments at the dipolar and quadrupolar levels between electronic states of the **unperturbed** Hamiltonian. These states are states of the whole electronic cloud, and not for only one electron. We will see in the next section what it implies for concrete and approximate calculations.

#### III. B.2 Approximate theory

In the previous section we have recalled how to compute the optical response properties of a molecule given the eigenstates of the unperturbed Hamiltonian. However, even though the Schrödinger equation is about a century old, no exact solution is known for systems with several electrons. One main reason is the electronic correlation: the electronic degrees of freedom are entangled. Hence, a major approximation made originally by Hartree is to work with an *uncorrelated* (or *non-interacting* or *independent*) wave-function. Physically, it means that the degrees of freedom of the several electrons are not entangled. Mathematically, it means that the electronic density,  $\rho(\mathbf{r}_1, \mathbf{r}_2 \cdots \mathbf{r}_N)$ , becomes  $\tilde{\rho}(\mathbf{r}_1)\tilde{\rho}(\mathbf{r}_2) \cdots \tilde{\rho}(\mathbf{r}_N)$ ; with  $\tilde{\rho}$  the independent one-electron density.

From this proposition, many frameworks have been developed. The major goal is to retrieve as best as possible the missed correlation while conserving a numerically affordable scheme. No method is efficient for all systems. But after decades of QM simulation, the community tends to understand more and more the dos and don'ts.

During my thesis I have used the Density Functional Theory (DFT) in the Kohn-Sham formulation which is presented in the next Section III. B.2.1 and in Appendix IX. A.3. DFT is one of the cheapest numerically speaking method, which was very needed because of the large number of calculations performed. A detailed comparison with other methods is made in the Section III. B.2.3 and more generally in the next numerical section. However, it should be noted that the usual Time-Dependent DFT scheme was not used to compute the response properties. For this thesis and as implemented in DALTON, scheme based on the wave-function and the eigenstates is presented in Section III. B.2.2. One of the major benefits of this method is that it allows to treat in the same way many frameworks, and especially DFT with hybrid functionals.

#### III. B.2.1 Solving the unperturbed Hamiltonian using DFT

#### From correlated wave-function to Slater Determinant

One major step is to define an uncorrelated wave-function of an N-electron system:

$$|\Psi\rangle = |\psi^1\rangle \otimes |\psi^2\rangle \otimes \cdots |\psi^N\rangle \equiv |\psi^1, \psi^2, \cdots \psi^N\rangle.$$
 (III.46)

 $|\psi^1\rangle$  described the degrees of freedom of the first electron,  $|\psi^2\rangle$  the one of the second and so forth. The sign  $\otimes$  transcribes that these degrees of freedom are not entangled. For a short-writing expression, we will use  $|\psi^1, \psi^2, \dots \psi^N\rangle$ . The functions  $\{\psi^i\}$  are called **Molecular Orbital (MO)** and the way they build the wave-function  $|\Psi\rangle$  are the new degrees of freedom. For convenience, the MO should be orthonormal  $(\langle \psi^i | \psi^j \rangle = \delta_{ij})$ .

Since electrons are fermions, the wave-function is odd when exchanging 2 electrons:  $|\psi^1, \psi^2, \cdots \psi^N\rangle = -|\psi^2, \psi^1, \cdots \psi^N\rangle$ . Hence, a useful mathematical tool which intrinsically respect this property, and thus fulfills the Fermi-exclusion principle, is the determinant. Object called Salter determinant is an extremely widespread object with organized MO to built wave-function of an N-electron system. It is the determinant of a NxN matrix with rows composed of the same MO, for instance  $\psi^i$  for

the row *i*, but applied for different electrons for each line, for instance the electron *j* for the line *j*. Hence, the wave-function is the sum of the product of  $\psi^i$  applied for all the electrons.

In our scheme, we will build the wave-function using only one Slater determinant, which is a strong approximation as we will see later on. Hence,  $|\psi^1, \psi^2, \cdots \psi^N\rangle$  represents in this document only **one** Slater determinant built for an N-electron system using the N single-electronic function  $\{\psi^1, \psi^2 \cdots \psi^N\}$ .

#### Basis set: Dunning's

Before going further, we have to discuss how is built these MO. Indeed, the  $\{\psi^i\}$  are a set of function defined in the whole space for a single electron. Since it is a building block of the wave-function, it is complex and we also assume that the wave-function is continuous and spatially derivable. Raise the question: how to construct this function defined in  $\mathbb{C}^3$ ?

One practical answer is to use a subset of function,  $\{\xi^j\}$ , called **Atomic Orbital (AO)** and to develop each MO on this basis:

$$\psi^{i}(\mathbf{r}) = \sum_{j} g_{ij} \xi^{j}(\mathbf{r})$$
(III.47)

Using an infinite set of  $\{\xi^j\}$  defined in the same space as the MO should provide all the degrees of liberty needed to build the MO. In practice, only finite sets of AO are used, and thus the choice of the basis is primordial for efficiency.

Several kinds of basis can be used, mostly depending on the system symmetry. For instance, for solids with 3D-periodic property, it is very efficient to use advantage of the Bloch Theorem and to define as elementary building blocks spatial-periodic functions. We are interested in non-periodic system (at the QM level) such as water in the liquid phase. Hence, the MO may be defined only within a finite amount of space.

For this reason, we used an analytical one based on Gaussian-polynomial expression  $^3$ :

$$\xi[\zeta, l_x, l_y, l_z](\mathbf{r}) = (x - X_0)^{l_x} (y - Y_0)^{l_y} (z - Z_0)^{l_z} \exp\left[-\zeta(\mathbf{r} - \mathbf{R}_0)^2\right]$$
(III.48)

Where  $\mathbf{R}_0 = (X_0, Y_0, Z_0)$  is the position of the nuclei. Such form depends explicitly on 4 parameters, but often the AO are rather labeled by the quantum numbers (n, l, m) they stand for – again we neglect here explicit spin-dependency. For instance, increasing the *n* leads to decreasing  $\zeta$  to make the AO more diffuse. The angular momentum *l* is an integer and *m* is the momentum direction: there are represented by the set  $(l_x, l_y, l_z)$ . For instance, the analogous of the 1s orbital is a large  $\zeta$  with  $l_x = l_y = l_z = 0$  while  $4p_z$  is a smaller  $\zeta$  with  $l_x = l_y = 0$  and  $l_z = 1$ .

Therefore, once the form of the basis set defined, here polynomial-Gaussian, one should provide which set of  $[\zeta, l_x, l_y, l_z]$  corresponds to a given n, l, m. We have chosen the parametrization proposed

<sup>&</sup>lt;sup>3</sup>One major numerical advantage of the Gaussian-polynomial expression is its integration behavior. During the numerical calculation, we need to compute integral of MO with another MO with maybe an extra function (for instance the dipole operator). This integral will become a sum over integrals over AOs. For Gaussian-polynomial expression, there are many analytical solutions, making these integral calculations very cheap!

by Dunning and coworker[92] for several reasons:

- It has been optimized not for the Hydrogen problem but for correlated ones atoms with as many electrons as protons. Hence, every type of atom of a molecule as its own (n, l, m) → (ζ, l<sub>x</sub>, l<sub>y</sub>, l<sub>z</sub>) mapping which is a better starting point than the uncorrelated problem.
- It has been made to provide an unambiguous way of improvement, which is known for converging very well see Section III. B.2.3.
- Many recent numerical works [20, 30, 52, 55, 57, 89] used this basis set: it is easier to compare our results using the same basis set.

There are other basis sets available, for instance the Sadlej's basis set [93] which was originally designed for (hyper)-polarizability calculations. Appendix IX. A.2 presents more details about the Dunning's basis set for water.

### Principle of DFT:

Using a single Slater Determinant and a set of AO, one can solve the Schrödinger equation for the electronic Hamiltonian defined in Equation III.3. Such approach is called Hartree-Fock and will not be discussed here. Rather, we will use the Density Functional Theory (DFT) approach which replaced the wave-function problem by a density one. A reminder of the Exact DFT is presented in Annexe IX. A.3 as long as details of the Kohn-Sham framework.

In a nutshell, within the Kohn-Sham (KS) scheme, the exact wave-function and density are replaced by the one of a non-interacting electronic one defined above using one Slater Determinant. The density of this scheme is noted  $n_{KS}$  and is assumed to be the same as the one of the real ground state density. The non-interacting wave-functions must fulfill the Kohn-Sham equations that provide a way to define the MOs:

$$\hat{H}_{KS}\psi^i = \epsilon_i\psi^i \tag{III.49}$$

With  $\epsilon_i$  the energy of each MO. The Kohn-Sham Hamiltonian  $\hat{H}_{KS}$  involves an exchange-correlation potential  $\hat{V}_{xc}$  which is not known and must be approximated. This term is approximated by a local functional of the density (and its space gradient), more details can be found in Annexe IX. A.3. More precisely, we have used a hybrid functional, CAM-B3LYP, which provides the best comparison with other reference calculations, see Section III. B.2.3.

This set of equations is still self-consistent: the  $\hat{H}_{KS}$  depends on the ground-state density  $n_{KS}$  which depends on the MOs. Hence, to find the MO one should proceed iteratively:

- Compute  $n_{KS}$  using the wave-function
- Compute the density dependent terms of  $\hat{H}_{KS}$ :  $\hat{V}_{Hartree}$  and  $\hat{V}_{xc}$

#### Page 60

- Solve the Kohn-Sham Equation to find the new MO
- Creating the new wave-function by selecting the energetically lowest MOs

After an initial guess of the wave-function (or density), this scheme should be repeated until the required convergence is met – for instance until the density at any position does not evolve from one iteration to another more than 1 %.

#### Regarding the excited states

We have defined the wave-function, the Hamiltonian and the equations that must fulfill the MOs. Using this, we can find for any molecule the ground-state density by solving iteratively the Kohn-Sham equation with sets of MOs build using the AOs. Once the convergence is reached, the ground state is defined to be a single Slater Determinant composed by the lowest energetic MO. Any ground-state property such as the permanent dipole moment can be calculated. However, we are interested in the molecular response to an electromagnetic field. According to the exact theory derived in Section III. B.1.4, such properties also involve the excited states. Thus, we can wonder how DFT can predict excited state properties and also what are these excited states.

One straightforward way of constructing the excited states would be to use the MOs in a different order to build the Slater Determinant. Let us order the MO by increasing energy:  $\epsilon_1 < \epsilon_2 \cdots < \epsilon_N < \epsilon_{N+1} \cdots$ . The ground state is given by the Slater Determinant  $|\psi^1, \cdots \psi^{N-1}, \psi^N\rangle$ . The MOs from 1 to N are called *occupied*. The MOs from N+1 are called *unoccupied* or *virtual*. To build the first excited state, we may try to build a Slater Determinant by exchanging the energetically highest occupied MO (N) by the lowest unoccupied one (N+1):  $|\psi^1, \cdots \psi^{N-1}, \psi^{N+1}\rangle$ 

However, the major problem of this method is that these virtual MOs are solutions of the Kohn-Sham equation, but not part of the  $n_{KS}$  used in the equation. Indeed, the  $\hat{H}_{KS}$  depends on the ground state Slater Determinant which does not involve any virtual MO by definition. Hence, the first virtual MO energy  $\epsilon_{N+1}$  does not represent an N-electronic excited state, but rather the electronic affinity: the energy required to add an electron – provided that the rest of the N-electron MOs is unchanged. Thus, the difference  $\epsilon_{N+1} - \epsilon_N$  is a poor estimation of the first electronic transition energy.

One way to go beyond is to redefine the wave-function itself: instead of working with a single Slater Determinant, we can use several. These approaches are often denoted as *post-Hartree Fock*: Configuration Interaction (CI), Couple Cluster (CC)... All of these techniques include explicitly (some) excited states using MOs: at the end of the ground state calculation they provide a better estimation of the ground state and the excited states. However, these methods are much more expensive compared to DFT: for instance the "golden standard" CCSD goes as  $N^6$  with N the number of electrons while hybrid functional goes as  $N^4$ .

#### III. B.2.2 Approximate response theory implementation in DALTON

Determining the excited state properties is thus more difficult. One can use the Time-Dependent DFT (TDDFT) formalism based on the Runge-Gross theorems. Yet, as for the Ground State DFT based on the Hohenberg-Kohn theorems, the Runge-Gross solely is not sufficient: one should also use many approximations. In TDDFT, the object developed using the perturbation theory is not the wave-function but the density:  $n_{KS}(t) = n_{KS}^0 + n_{KS}^1(t) + \cdots$ . This technic is extremely efficient and cheap for functional based on only on the density, but using such framework for hybrid functional is a bit more difficult.

Therefore, in DALTON the Time-Dependent Hartree-Fock (TDHF) like approach is preferred: it is based on the time-development of the wave-function as presented in Section III. B.1.4. Using this framework, the development of the response theory based on several approaches, for instance HF or DFT with a hybrid functional, is exactly the same: only the definition of the unperturbed Hamiltonian is different. In this section we will describe how it is achieved using the formalism developed in the DALTON software. As for the exact response theory, details can be found in the Chapter 7 of the textbook [1].

#### Time-Dependent parametrization:

As for the exact state response scheme, we will describe the time-dependent wave-function using a basis set obtained of the unperturbed Hamiltonian. In exact theory, the real excited states are used along with the ground state to define an assumed complete basis set to expand the time-dependent wave-function, see Equation III.19.

In our approximate approach, we will use all the MO in a single Slater Determinant. To do so, we are using the second quantization notation:  $|0\rangle$  is the ground-state wave-function represented by a single Slater Determinant:  $|0\rangle = |\psi^1, \psi^2, \psi^N\rangle = \hat{a}_1^{\dagger} \hat{a}_2^{\dagger} \cdots \hat{a}_N^{\dagger} |vac\rangle$ . Where  $\hat{a}_p^{\dagger}$  is the creation operator: promoting an electron in the MO  $p^4$ .  $\hat{a}_q$  is the annihilation operator: removes an electron of the MO  $q^5$ . Hence, the operator  $\hat{a}_p^{\dagger} \hat{a}_q$  transfers an electron from the MO q to the MO p provided that the q MO is occupied and the MO p is empty.

If we want to use the same framework as for the exact theory, we should write the time-dependent wave-function as:

$$|\Psi\rangle(t) = c_0(t)|0\rangle + \sum_{s>N} \sum_{i=1}^N \kappa_{si}(t)\hat{a}_s^{\dagger}\hat{a}_i|0\rangle + \sum_{s,t>N} \sum_{i,j=1}^N \kappa_{tsji}(t)\hat{a}_t^{\dagger}\hat{a}_s^{\dagger}\hat{a}_j\hat{a}_i|0\rangle + \cdots$$
(III.50)

Since the GS,  $|0\rangle$ , is occupied until the MO N by construction, the annihilation operator can go from 1 to N, noted i, j and the creation one from N + 1, noted s or t. We cannot use an infinite number of empty states in a real calculation so we set the number of empty states to M: s and t goes from N + 1 to N + M

The first term represents the case without perturbation. The second term represents the 'single

<sup>&</sup>lt;sup>4</sup>Two electrons cannot be promoted in the same MO because of the Fermi exclusion principle:  $\hat{a}_p^{\dagger} \hat{a}_p^{\dagger} |vac\rangle = 0$ 

<sup>&</sup>lt;sup>5</sup>If there is no electron in the orbital q, it returns 0:  $\hat{a}_q |vac\rangle = 0$ 

excited states': only one electron goes from the occupied to the unoccupied MOs. The third term represents the 'double excited states', and so forth. One of the problems with this explicit notation is to maintain the norm of  $|\Psi\rangle(t)$ . Thus, we will rather work with an exponential form:

$$|\Psi\rangle(t) = \exp\left[-i\hat{\kappa}(t)\right]|0\rangle = |0\rangle - i\hat{\kappa}(t)|0\rangle - \frac{1}{2}\hat{\kappa}(t)^{2}|0\rangle + \cdots ; \quad \hat{\kappa}(t) = \sum_{s>N}\sum_{i=1}^{N}\kappa_{si}(t)\hat{a}_{s}^{\dagger}\hat{a}_{i} \quad (\text{III.51})$$

Where we enforce that  $\kappa_{si} = \kappa_{si}^*$  so that the exponential is anti-hermitian <sup>6</sup>. As for the previous parametrization, the first term contains the unperturbed terms, the second contains the single excited state, the third contains the second excited states and so forth.

In this case, the perturbation is no longer used as a sum over perturbed wave-function, see Equation III.20, but within the exponential itself:  $\hat{\kappa}(t) = \hat{\kappa}^{(1)}(t) + \hat{\kappa}^{(2)}(t) + \cdots$ <sup>7</sup>. And as for the exact theory, we will use a time-Fourier decomposition for both the wave-function and the time-dependent perturbation  $\hat{V}(t)$ :

$$\hat{V}(t) = \hat{V}F^{\omega} \exp[-i\omega t] \qquad \qquad \hat{\kappa}(t) = \sum_{s>N} \sum_{i=1}^{N} \kappa_{si}^{(1)}[\omega] \hat{a}_{s}^{\dagger} \hat{a}_{i} \exp[-i\omega t] + \cdots \qquad (\text{III.52})$$

For simplicity we will stay at the linear order given by  $\kappa_{si}^{(1)}(t)$ , and we assume that there is only one type of perturbation at one frequency  $\omega$ . In this part we try to emphasize the path used in approximate theory and compare with the exact one: the remarks at the linear order are the same as for SHG calculation.

As for the exact scheme, we first have to find the time-dependent wave-function at the first perturbation order, *i.e.* obtained the  $\kappa_{si}^{(1)}(t)$  coefficient. Then, we can compute the time-dependent observable as:

$$\langle \Psi | \hat{\Omega} | \Psi \rangle (t) = \langle 0 | \exp [i\hat{\kappa}] \hat{\Omega} \exp [-i\hat{\kappa}] | 0 \rangle (t) = \langle 0 | \hat{\Omega} | 0 \rangle + i \langle 0 | \hat{\kappa} \hat{\Omega} | 0 \rangle - i \langle 0 | \hat{\Omega} \hat{\kappa} | 0 \rangle + \cdots$$
(III.53)

$$\langle \Psi | \hat{\Omega} | \Psi \rangle (t) = \Omega^{(0)} + i \sum_{s=N+1}^{N+M} \sum_{i=1}^{N} \langle 0 | \left[ \hat{a}_s^{\dagger} \hat{a}_i, \hat{\Omega} \right] | 0 \rangle \times \kappa_{si}^{(1)} [\omega] \exp[-i\omega t] + \cdots$$
(III.54)

Here we retrieve the perturbative development of the observable according to the time-dependent perturbation frequency: the linear term oscillates at the same frequency as the perturbation. The term  $\langle 0 | [\hat{\Omega}, \hat{a}_s^{\dagger} \hat{a}_i] | 0 \rangle$  corresponds to the transition between "ground" and "excited" states similarly to the  $\langle 0 | \hat{\Omega} | n \rangle$  term of the exact theory. However, for the approximated framework it represents transitions between Slater determinants. Later on, we will define a 2*NM* line vector  $\Omega^{[1]}$  in order to use matrix notation, where *N* is the number of occupied orbitals and *M* the number of virtual

 $<sup>^{6}</sup>$ The demonstration of the time-independent norm is made in p.256 of Ref [1].

 $<sup>^{7}</sup>$ The unperturbed term is given by the identity inside the very definition of the exp function used to build the time-dependent function.

ones considered.  $\Omega^{[1]}$  is composed of  $-\langle 0| \left[ \hat{a}_s^{\dagger} \hat{a}_i, \hat{\Omega} \right] |0\rangle$  for its first half, and  $\langle 0| \left[ \hat{a}_i^{\dagger} \hat{a}_s, \hat{\Omega} \right] |0\rangle$  for its second half. This vector is a ground-state property: once the set of MO obtained, this matrix is known.

#### **Response terms calculation:**

Let us compute the  $\kappa_{si}^{(1)}$  using the Ehrenfest theorem:

$$\frac{\mathrm{d}}{\mathrm{d}t} \left\langle \Psi \right| \hat{\Omega} \left| \Psi \right\rangle (t) = \frac{1}{i\hbar} \left\langle \Psi \right| \hat{H} \left| \Psi \right\rangle (t) + \left\langle \Psi \right| \frac{\mathrm{d}}{\mathrm{d}t} \hat{\Omega} \left| \Psi \right\rangle (t), \qquad (\mathrm{III.55})$$

applied to the operator:  $\hat{t}_{si} = \exp\left[-i\hat{\kappa}\right] \hat{a}_s^{\dagger} \hat{a}_i \exp\left[i\hat{\kappa}\right]$ . We will skip calculations detailed presented in Ref [1] to discuss the final result:

$$\kappa^{(1)}[\omega] = -i \left[ E^{[2]} - \hbar \omega S^{[2]} \right]^{-1} \cdot V^{\omega, [1]} \times F^{\omega}.$$
(III.56)

Where the different elements are commented in the following paragraphs.

 $\kappa^{(1)}[\omega]$  is a line vector of dimension 2NM: it contains all the  $\kappa_{si}^{(1)}[\omega]$  and the  $\kappa_{si}^{(1)}[-\omega]^*$  that we wish to obtain. The perturbation is driven by the operator:  $V^{\omega,[1]}$ , which can be the dipolar or the quadrupolar ones, it is also a 2NM line vector. The first half elements are given by  $\langle 0 | \left[ \hat{a}_s^{\dagger} \hat{a}_i, \hat{V}^{\omega} \right] | 0 \rangle$  and the second half by  $-\langle 0 | \left[ \hat{a}_s^{\dagger} \hat{a}_i, \hat{V}^{\omega} \right] | 0 \rangle^*$ . This is also a ground-state property: it can be computed after the MO optimization.  $F^{\omega}$  is proportional to the electric field amplitude (when  $V^{\omega,[1]}$  represents the dipolar interaction) or to the electric field spatial gradient (when  $V^{\omega,[1]}$  represents the quadrupolar interaction).

The remaining part involves an inversion of a  $2NM \times 2NM$  square matrix:  $E^{[2]} - \hbar \omega S^{[2]}$ . The generalized overlap matrix  $S^{[2]}$  is simple in the case of an orthonormal set of MOs. It is a diagonal matrix where the first half of the diagonals are +1 and the second half -1. The generalized Hessian,  $E^{[2]}$ , is more complicated: it is a dominant diagonal matrix given by the structure:  $E^{[2]} = \begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix}$ 

$$A = -\langle 0| \left[ \hat{a}_t^{\dagger} \hat{a}_j \left[ \hat{a}_i^{\dagger} \hat{a}_s, \hat{H}^0 \right] \right] |0\rangle = \langle 0_i^s | \hat{H}^0 | 0_j^t \rangle - \delta_{ij} \delta_{st} \langle 0| \hat{H}^0 | 0 \rangle$$
(III.57)

$$B = -\langle 0 | \left[ \hat{a}_{t}^{\dagger} \hat{a}_{j} \left[ \hat{a}_{s}^{\dagger} \hat{a}_{i}, \hat{H}^{0} \right] \right] | 0 \rangle = -\langle 0 | \hat{H}^{0} \hat{a}_{s}^{\dagger} \hat{a}_{i} | 0_{j}^{t} \rangle = -\langle 0 | \hat{H}^{0} | 0_{ji}^{ts} \rangle \left( 1 - \delta_{ij} \right) \left( 1 - \delta_{st} \right)$$
(III.58)

Where  $|0_j^t\rangle = \hat{a}_t^{\dagger} \hat{a}_j |0\rangle$  represents a "single excited state" where an electron from the orbital j is excited into the orbital t and  $|0_{ji}^{ts}\rangle$  is the double excited state: the MO i and j are depleted while the MO t and s are occupied.

To conclude the equation derivation, the first order perturbation of the observable is given by :

$$\Omega^{(1)} = -i \left[\Omega^{[1]}\right]^{\dagger} \cdot \kappa^{(1)}[\omega] \exp[-i\omega t] = -\left[\Omega^{[1]}\right]^{\dagger} \cdot \left[E^{[2]} - \hbar\omega S^{[2]}\right]^{-1} \cdot V^{\omega,[1]} F^{\omega} \exp[-i\omega t] \quad (\text{III.59})$$

The response function is thus:

$$\langle \langle \hat{\Omega}; \hat{V}^{\omega} \rangle \rangle = -\left[ \Omega^{[1]} \right]^{\dagger} \cdot \left[ E^{[2]} - \hbar \omega S^{[2]} \right]^{-1} \cdot V^{\omega, [1]}$$
(III.60)

#### Exact Vs Approximate response:

To better understand this result, let us discuss a bit about the nature of the excited states by comparison with the exact scheme.

In the exact scheme, we have developed the time-dependent wave-function using the true excited state. Let us imagine that we can keep the same equation, but where the single  $|0_{j}^{t}\rangle$  and double  $|0_{ji}^{ts}\rangle$  "excited state" are now eigenstate of the unperturbed Hamiltonian, which is not the case in our approximate scheme. In other words,  $|0_{i}^{t}\rangle$  and  $|0_{ii}^{ts}\rangle$  are now some  $|n\rangle$  and  $|n'\rangle$ .

The vector  $\Omega^{[1]}$  and  $V^{\omega,[1]}$  are not very different. In this case, the  $\langle 0 | \left[ \hat{a}_s^{\dagger} \hat{a}_i, \hat{\Omega} \right] | 0 \rangle$  is the same as transition moment  $\langle 0 | \left[ |n \rangle \langle 0 |, \hat{\Omega} \right] | 0 \rangle$  and similarly for  $V^{\omega,[1]}$ . The  $S^{[2]}$  matrix does not change: in both cases the basis used to develop the time-dependent wave-function are orthonormal (the MOs or the  $\{ |n \rangle \}$ ).

However, the matrix  $E^{[2]}$  is extremely simplified. The A term reads:  $A = \langle n | \hat{H}^0 | m \rangle - \delta_{nm} \langle 0 | \hat{H}^0 | 0 \rangle = E_m \langle n | m \rangle - \delta_{nm} E_0 = (E_n - E_0) \delta_{nm}$ . Therefore, the A sub-matrix becomes diagonal and is the difference in energy between the excited stated and the ground state:  $\hbar \omega_{n0}$ . The B term reads:  $B = -\langle 0 | \hat{H}^0 | n \rangle = 0$ .

Therefore,  $E^{[2]}$  becomes diagonal, and  $E^{[2]} - \hbar \omega S^{[2]}$  is diagonal with  $\hbar(\omega_{n0} - \omega)$  for its first half and  $\hbar(\omega_{n0} + \omega)$  for its second half. The inverse of a diagonal matrix is diagonal with for value the inverse. Performing the matrix product, we retrieve the same results as for the exact scheme:  $\langle\langle \hat{\Omega}; \hat{V}^{\omega} \rangle \rangle = -\sum_{n} \langle 0| \Omega | n \rangle \frac{1}{\hbar(\omega_{n0} - \omega)} \langle n| \Omega | 0 \rangle + \langle n| \Omega | 0 \rangle \frac{1}{\hbar(\omega_{n0} + \omega)} \langle 0| \Omega | n \rangle$ 

Therefore, if the single and double excited states were real eigenstates of the unperturbed Hamiltonian, the approximate framework would match the exact one. However, this is not the case, and the major difference is in the energy of these state:  $E^{[2]}$ . Using the MOs and the single and double excited state, the difference in energy arising in  $E^{[2]}$  is different from the pure difference of the MOs energy. As said above, the energy of an unoccupied MO is rather an electronic affinity than an excited state. Therefore, by inverting the matrix  $\left[E^{[2]} - \hbar\omega S^{[2]}\right]$  we are trying to improve the excited state description. This matrix inverse shows how an "excited state" behave: how strong the probability transition from one state to another is simply due to the unperturbed Hamiltonian. For exact states, this matrix is diagonal because all the excited states are orthogonal. In approximate theory, this will often increase the probability of transition from the ground state to an excited state – compared to the electron affinity energy. The main reason is that during the calculation of the unoccupied MOs, these MOs were not part of the Hamiltonian: the exchange term (which decreases the energy) is missing.

### Sum-up:

Before assessing the validity of our scheme, let us sum up the different steps:

- 1. For a molecule, create the Atomic Orbitals (AO) according to the basis set. These orbitals are centered around each atom.
- 2. Create a first set of Molecular Orbitals (MO) using these AO for all the N electrons.
- 3. Built the wave-function using a single Slater Determinant with the lowest N MOs.
- 4. Compute the density  $n_{KS}$
- 5. Compute the unperturbed Hamiltonian  $\hat{H}^0$
- 6. Solve the Kohn-Sham equation (matrix inversion): new set of MOs
- 7. Repeat steps 3 to 6 until convergence
- 8. Using the optimize set of MOs, compute  $\left[\Omega^{[1]}\right]^{\dagger}$ ,  $E^{[2]}$ ,  $S^{[2]}$  and  $V^{\omega,[1]}$ .
- 9. Invert  $\left[E^{[2]} \hbar\omega S^{[2]}\right]$
- 10. Compute the response function  $\langle \langle \hat{\Omega}; \hat{V}^{\omega} \rangle \rangle$

For the quadratic response, there is another matrix inversion to perform, which would represent more excited state transition probabilities. In total, the complexity is  $(N_{AO})^4$  for the ground state calculation (due to the "exact" exchange calculation of the hybrid functional), and  $(N \times M)^3$  for the response part (matrix inversion). Where N is the number of electrons,  $N_{AO}$  the number of atomic orbitals used and M the number of unoccupied orbitals.

#### III. B.2.3 Method details and comparison with literature

In this section we present a broad sum-up of the different comparison and verification we have performed on the method used to get the SHG response properties. First, we present the verification regarding the basis set used, then on the perturbation approach itself, and finally we compare our results obtained by the hybrid functional we used to the golden standard: CCSD.

#### Basis set convergence:

Solving the unperturbed Hamiltonian numerically implies to use a basis set to expand the molecular orbital. We have chosen the Dunning Gaussian-shaped basis set. Quantum approaches are based on variational principle: the self-consistent equations are solved in order to lower the total electronic energy. Nowadays, minimizing the total electronic energy and the ground state electronic state energies is not too difficult with respect to the basis set for simple systems like water for a DFT calculation. Indeed, the electronic energy is mainly carried by the high electronic density parts of space – for instance close to the nucleus. Therefore, for large enough basis set, the difference in the final density between two different ones is very small in absolute value, which leads to almost
no difference in the ground state energy.

However, the transition dipole moment between two states  $\langle n | \hat{\mu} | m \rangle$  is much harder to get since it involves explicitly the electronic density for each state in the whole space. Indeed, contrarily to the total energy, the transition dipole moment is particularly affected by electronic delocalization due to the positional operator within the dipole operator, even if the density at large distance is small. Thus, if small density fluctuations far away from the molecule center will have almost no impact on the energy, it can have an important one for the transition dipole moment. In top of that, the response function involves virtual MOs which are usually more delocalized than occupied ones. That is why, in order to converge the hyperpolarizability tensor, one has to pay attention to electronic delocalization.

Hence, in the basis set we will increase the number of both the "valence" orbitals ('pVnZ', where n=D,T,Q,5) and the delocalized one ('cc' or 'aug-cc' or had-aug-cc', or 't-aug-cc' for no augmentation, one, double and triple augmentation respectively). In the following a short hand notation is used: 'cc-pVnZ' is written 'VnZ', and the augmented 'XVnZ' where X=a,d,t for one, two or three augmentations respectively. In the Table III.1 the independent hyperpolarizability tensor component for water in vacuum for different Dunning's basis set is summarized to emphasis this effect.

Table III.1: Molecular hyperpolarizability of water in the gas phase for several basis sets. DFT/B3LYP has been used. The  $\beta$  is calculated from the ground state using the response scheme at zero frequency. The basis set evolution have to be understood in 2 steps: on the first 3 lines the most inner part of the basis set is increased, and then the next 3 basis set should be more delocalized.

Basis set	$\mu_z$	$\beta_{zxx}$	$\beta_{zyy}$	$\beta_{zzz}$	Basis set size	Calculation time:
VTZ	0.770	-17.76	-5.75	-13.43	58	55s
VQZ	0.764	-16.02	-6.80	-12.30	115	20m
V5Z	0.762	-14.60	-7.25	-11.07	201	38m
aVTZ	0.746	-14.90	-5.00	-11.75	92	20m
aVQZ	0.746	-13.62	-5.41	-13.00	172	31m
aV5Z	0.746	-13.07	-5.68	-14.50	287	10h30m
daVTZ	0.746	-11.85	-5.27	-14.66	126	17m
daVQZ	0.746	-11.91	-5.55	-15.27	229	65m
daV5Z	0.746	-11.96	-5.71	-15.53	373	23h21m
taVTZ	0.746	-11.96	-6.07	-15.12	160	26m
taVQZ	0.746	-11.98	-5.73	-15.43	286	8h41m

The permanent dipole moment along the z direction  $\mu_z$  converged extremely fast with respect to the basis set, as well as the total energy – data not shown. However, one can clearly see that the hyperpolarizability tensor component is much harder to converge and especially when increasing only the basis centered close to the nucleus (VTZ>VQZ>V5Z). Indeed, it is much more efficient to use delocalized basis sets (aug>d-aug>t-aug): for a similar number of atomic orbitals, the d-

Chapter III. Describing the second harmonic generation of non-resonant liquid: from QM/MM to Maxwell



Figure III.4: Evolution of average  $\beta$  components for several basis sets using DFT with the CAM-B3LYP functional in the liquid phase at 800 nm. The electrostatic environment is included up to 20 Å,  $|\beta|$  is defined by Equation III.61. The values are presented in atomic units. 55 000 configurations are used.

aug-cc-VTZ basis set provides better results than the cc-VQZ basis set. The d-aug-cc-VTZ basis provides good results within short calculation time. Using larger basis sets (like d-aug-cc-VQZ or t-aug-cc-VTZ) lead to a small increase in precision but for higher numerical cost.

The main subject of this thesis is not gas phase calculation but rather the liquid one using the PE scheme – see Chapter IV. In liquid water, the first hyperpolarizability is very different from one molecule to another: the relevant quantities are the average and standard deviation for each component. The impact of the basis set on the  $\beta$  components and on  $|\beta|$  for water in the bulk phase is depicted in Figure III.4, with

$$|\beta| = \left(\sum_{i,j,k} \beta_{ijk}^2\right)^{\frac{1}{2}}.$$
 (III.61)

As for the gas phase, polarization and diffuse functions in the basis set are necessary, and we have chosen to use the d-aug-cc-pVTZ in our work. The relative smooth convergence of the first hyperpolarizability as a function of the basis set permits to gain confidence in the absence of the "spill-out" effect which can appear using QM/MM approaches with large basis sets [94, 95].

As a conclusive remark about the basis set, many works also use the d-aug-cc-pVTZ basis set, or lower, for the first hyperpolarizability. For instance, in the gas phase [20] uses aug-cc-pVTZ and [52] compares many basis sets and show that d-aug-cc-pVTZ is a very good compromise. In the liquid phase [30, 55] use d-aug-cc-pVTZ. For the second hyperpolarizability in the gas phase [52, 55]

also use d-aug-cc-pVTZ basis. Finally, the quadrupolar first hyperpolarizability value we have found are also computed below this level, for instance [57, 89] used aug-cc-pVTZ. Hence, we use the d-aug-cc-pVTZ basis set for all the hyperpolarizabilities computed in this work for the water molecule.

#### **Response framework Vs Finite Field:**

One important approximation of our framework is the response scheme itself. The main argument is that the time-dependent wave-function can be decomposed using the ground state eigenstates (or MOs). In order to cross-validate this schema, we can compare the result obtained with the response scheme and the Finite Field (FF) method. Indeed, computing the response properties using the FF or the response frameworks for a static fundamental should return the same results.

In the FF framework, an external electrostatic field,  $e^0$ , is applied during the ground state calculation. Thus, the obtained ground state quantity, such as the permanent dipole moment is:

$$\boldsymbol{\mu}[\mathbf{e}^{0}] = \mu^{0} + \alpha(0,0) \cdot \mathbf{e}^{0} + \frac{1}{2!} \beta^{dd}(0,0,0) : \mathbf{e}^{0} \mathbf{e}^{0} + \frac{1}{3!} \gamma(0,0,0,0) : \mathbf{e}^{0} \mathbf{e}^{0} \mathbf{e}^{0} + \cdots$$
(III.62)

Applying several electric fields in all the directions provides the response quantity at the zero frequency by fitting the permanent dipole moment with respect to the applied  $\mathbf{e}^0$ . For instance, in Figure III.5 the induced dipole moment of water in the vacuum along the molecular c axis of water is plotted as a function of the external electrostatic field along the molecular a-direction:  $\mathbf{e}_a^0$ . At the quadratic order:

$$\boldsymbol{\mu}_{c}[\mathbf{e}^{0}] = \mu_{c}^{0} + \alpha(0,0)_{ca}\mathbf{e}_{a}^{0} + \frac{1}{2!}\beta^{dd}(0,0,0)_{caa}\mathbf{e}_{a}^{0}\mathbf{e}_{a}^{0}$$
(III.63)

Due to molecular symmetry,  $\alpha(0,0)_{ca} = 0$ . Hence:

$$\Delta \boldsymbol{\mu}_{c}[\mathbf{e}^{0}] = \frac{1}{2!} \beta^{dd}(0,0,0)_{caa} \mathbf{e}_{a}^{0} \mathbf{e}_{a}^{0}$$
(III.64)

Using this method, the obtained  $\beta^{dd}(0,0,0)_{caa}$  is 9.7 a.u. while the response scheme gives 9.8 a.u.

Similar tests were made on the other dipolar first hyperpolarizability components, on the quadrupole first hyperpolarizability, and on the dipole second hyperpolarizability – not shown. We find that the two method returns the same results within a very close error, about 0.1 a.u.

This FF technic is used in many works [20, 41, 55, 57, 89] while others use the response scheme [30, 52]. Here are some advantages for the response scheme compared to the FF:

• Cheaper: the ground state calculation should be made only once, and then followed by a response calculation. The response calculation is about the same order of magnitude compared to the ground state calculation. For FF method, the precision of the technic is (mostly) determined by the number of electrostatic field applied typically dozens for every molecular direction. This technique is more and more expensive as the order of the response tensor

Chapter III. Describing the second harmonic generation of non-resonant liquid: from QM/MM to Maxwell



Figure III.5: Evolution of the permanent dipole moment along the molecular c-direction with respect to the external electrostatic field along the a-direction.

increases.

• Frequency dependent: the FF method can only work in the infinite wave-length limit. This may not be very important for some molecule. However, in the case of water, the frequency dispersion between 800 nm and the infinite limit is about 10 % in the vacuum phase [90] for the first dipolar hyperpolarizability, and even more for the second hyperpolarizability. Yet, other approaches can be used in order to guess the frequency dispersion [96].

This last part concerning the frequency dependency is important for us because at the interface we are wondering about the origin of the Kleinman symmetry breaking. Indeed, experimentally we observed that the Kleinman symmetry is not respected at the water liquid-gas interface. Within the dipolar approximation, this behavior can be solely explained by the breaking of the Kleinman symmetry at the molecular level. Therefore, using a numerical approach which cannot take into account the frequency dispersion would not allow us to study this phenomenon. Hence, we will use the response scheme, especially in regard of this Surface-SHG problematic.

#### DFT/CAM-B3LYP Vs CCSD:

The major numerical error we made on the calculation of response property is due to the DFT itself. Indeed, even with an elaborated functional, we are still missing some correlation and/or exchange aspect of the real electronic system.

Nowadays, the "golden standard" is called Couple-Cluster (CC) with its increasing degree of precision (Single CCS, Double CCSD and Triple CCSD(T)). We do not use CC in this thesis but for us it is a numerical reference: we do not pretend to present it here. In an nutshell, CC uses several Slater Determinants combined in an exponential operator to take into account the virtual MOs in the Ground State calculation. Doing so, the obtained excited state are way more precise

than using DFT – in top of affecting the GS itself because the occupied MOs are also modified. It does not require any "functional" and, with a bit of algebra presented in Ref [1], the response scheme can also be used. This technic has a drawback: its numerical cost in  $N^6$ .

Hence, let us compare the error made by DFT/CAM-B3LYP compared to CCSD for the quantity at sake. Table III.2 presents the first hyperpolarizability in the vacuum phase for DFT and CCSD. We see that indeed the effect of the Coulomb Attenuated Method (the CAM of CAM-B3LYP) leads to quite a change in the predicted value compared to the case without (B3LYP). To compare with the CCSD results, we have to use  $\beta_{\parallel}$  defined as:

$$\beta_{\parallel} = \frac{\sum_{i} \beta_{i} \mu_{i}}{|\mu|}, \quad \text{with} \quad \beta_{i} = \frac{1}{5} \sum_{j} \beta_{ijj} + \beta_{jij} + \beta_{jji}, \quad (\text{III.65})$$

In Table III.3 is presented the dipolar hyperpolarizability in the water liquid phase at zero frequency. The results obtained with CAM-B3LYP compares well with the CCSD one for both the average value and the standard deviation. The largest error is found for the  $\beta_{cbb}$  by about 1 a.u. for the average value.

Table III.4 presents the dipolar first hyperpolarizability at the quadratic order  $(\beta^{dq})$  in vacuum at zero frequency. It should be noted that the Reference [89] uses another convention: the traceless one, see Appendix III. B.1.4. Hence, Table III.4 presents the dipolar first hyperpolarizability in this traceless convention: the results obtain with our convention is presented in Chapter VI.

In Table III.5 is presented the second hyperpolarizability of water in the vacuum at zero frequency, and Table III.6 for several frequencies. To compare with the CCSD results, we have used the same definition as the authors [52, 55]:

$$\gamma_{\parallel} = \frac{1}{15} \sum_{i,j} \gamma_{iijj} + \gamma_{ijij} + \gamma_{ijji}, \quad \gamma_{\text{THS}} = \sqrt{\Gamma_{zzzz}^2 + \Gamma_{zxxx}^2}, \quad \text{DR}_{\text{THS}} = \Gamma_{zzzz}^2 / \Gamma_{zxxx}^2, \quad (\text{III.66})$$

where  $\Gamma$  is the second hyperpolarizability in the laboratory frame.

For all of these quantities, an error about 10 % is observed. We have decided that the extra computational cost is not worth the improvement in accuracy and stick to DFT. However, can we increase the accuracy in DFT for water, and for other molecules – and for the bulk phase calculation by choosing a specific functional?

#### Choice of the functional:

First, the recent work by Besalú-Sala *et. al.* [98] shows that it is not obvious how to choose a priori the best functional for non linear optical property calculations, since the errors are much systemdependent. For second hyperpolarizability, the mean absolute errors of second hyperpolarizability components obtained by DFT relative to CCSD(T) are typically of 10 to 30 % for functional including exact exchange. Even worse, as shown by Hammond *et. al.* [99], for the polarizability of water clusters (n = 2 to 12), the ranking of different functional change as a function of the cluster size n.

Here, the QM calculations include solely a single water molecule. CAM-B3LYP may not be the most accurate functional in any case, but here are some results that may guide toward this functional in our case:

- While it is true that DFT/CAM-B3LYP is not accurate to compute cross section at resonant frequency, our work investigates SHG response for the fundamental frequency of 800 nm. Therefore, we assume to be in the non-resonant case, where no excited states are specifically targeted. For water, the excitation energies of the first states are about 7-10 eV [100], which would correspond to a fundamental wavelength about 300 nm. Even if the excitation wavelength of interest in our work is 800 nm, these states should still contribute to the SHG response. CAM-B3LYP seems a reasonable choice, since it performs best to predict the cross section and the frequency [100] compared to other DFT functionals.
- Our results in the Table III.2 show that CAM-B3LYP performs reasonably at 694 nm relative to CCSD for all the non-zero components. We did not try to compare to a very large set of functional since this had been done in the literature. The second hyperpolarizability frequency dispersion is also quite well reproduced by DFT/CAM-B3LYP, see Table III.6.
- Hammond *et. al.* [99] have calculated the polarizability of water cluster from 2 to 12 molecules, and compared CC calculations with DFT. They concluded that the PBE0 and CAMB3LYP functionals are globally most able to reproduce the CCSD isotropic polarizabilities of small water clusters.
- Finally, since we are working in electrostatic environment, the second hyperpolarizability plays an important role. Besalú-Sala *et. al.* [98] show that CAM-B3LYP is the best functional regarding the second hyperpolarizability for water. This is confirmed by the good agreement found in Table III.5 and III.6.

Chapter III. Describing the second harmonic generation of non-resonant liquid: from QM/MM to Maxwell

Table III.2: Comparison with literature of hyperpolarizability components for water molecule in gas phase (in atomic unit). All components are computed for a static field and for a wavelength of 694 nm. The  $\beta_{\parallel}$  is defined in the Eq. III.65. Reference notations : <sup>a</sup> is Ref. 20. <sup>b</sup> is Ref. 90. <sup>c</sup> is Ref. 55. <sup>d</sup> is the experimental value of Ref. 97 with its uncertainty. Results without references are our calculations. The absolute error (MAE) averaged of all non-zero  $\beta$  components is given for DFT/d-aug-cc-pVTZ relative to CCSD/d-aug-cc-pVTZ.

Wavelength ang-cc-pVTZ	-14.9	<sup>xaa</sup> 694 -17.4	-14.9 $\beta_a$	-17.4	$-5.0$ $\beta_c$	$\frac{bb}{694}$ -4.2	-5.0 8 B		$\frac{3b}{694}$	$\frac{3b}{694}$ $\beta_c$	$\frac{b}{694}$ $\beta_{ccc}$ $694$ $\infty$ $694$ $-10.2$ $-11.75$ $-14.5$	$\frac{3b}{694}$ $\beta_{ccc}$ $\beta_{ccc}$ $\beta_{cc}$ $\beta_{c$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$
aug-cc-pVTZ	-14.9	-17.4	-14.9	-17.4	י ק ס.0	-4.2	-5.0		0.2	0.2 - 11.75	0.2 - 11.75 - 14.5	0.2 - 11.75 - 14.5 - 19.0	0.2 -11.75 -14.5 -19.0 -24.1	0.2 -11.75 -14.5 -19.0 -24.1
aug-cc-pVTZ <sup>a</sup> d-aug-cc-pVTZ	-14.4 -11.8	-17.7 -14.2	-14.4 -11.8	-17.6 -14.0	5.3 5.3	-4.3	-5.0	5 L	0.0 9.5	0.0 -11.2 ).5 -14.7	$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	0.0 -11.2 -14.4 -18.3 0.5 -14.7 -18.8 19.1	0.0 -11.2 -14.4 -18.3 -24.1 ).5 -14.7 -18.8 19.1 -24.8	0.0 -11.2 -14.4 -18.3 -24.1 ).5 -14.7 -18.8 19.1 -24.8 1.1
				CA	M-B3	LYP								
aug-cc-pVTZ	-13.6	-15.7	-13.6	-15.7	-4.2	-3.5	-4.2		-8.4	-8.4 -10.3	-8.4 -10.3 -12.5	-8.4 -10.3 -12.5 -16.1	-8.4 -10.3 -12.5 -16.1 -21.0	-8.4 -10.3 -12.5 -16.1 -21.0
$aug$ -cc-p $VTZ^{a}$	-13.0	-15.5	-13.0	-15.5	-4.2	:3 :5	-4.2		-8.2	-8.2 -9.6	-8.2 -9.6 -12.0	-8.2 -9.6 -12.0 -16.1	-8.2 $-9.6$ $-12.0$ $-16.1$ $-20.5$	-8.2 -9.6 -12.0 -16.1 -20.5
d-aug-cc- $pVTZ$	-10.6	-12.5	-10.6	-12.4	-4.2	-5.0	-4.2		-7.4	-7.4 -12.1	-7.4 -12.1 -15.3	-7.4 -12.1 -15.3 -16.1	-7.4 -12.1 -15.3 -16.1 -20.6	-7.4 -12.1 -15.3 -16.1 -20.6 1.3
					CCSI	D								
d-aug-cc-pVDZ <sup><math>b</math></sup>												-15.0	-15.0	-15.0
d-aug-cc-pVTZ <sup><math>b</math></sup>												-17.9	-17.9	-17.9
d-aug-cc-pVQZ <sup><math>b</math></sup>												-18.3	-18.3	-18.3
d-aug-cc-pV5Z <sup><math>b</math></sup>												-18.2	-18.2	-18.2
d-aug-cc-pVTZ <sup><math>c</math></sup>	-9.9		-9.9		-5.7		-5.7			-14.0	-14.0	-14.0	-14.0	-14.0
d-aug-cc-pVTZ	-9.9	-11.5	-9.9	-11.4	-5.7	-6.7	-5.7		-8.5	-8.5 -14.0	-8.5 -14.0 -16.9	-8.5 -14.0 -16.9 -17.7	-8.5 - 14.0 - 16.9 - 17.7 - 21.7	-8.5 -14.0 -16.9 -17.7 -21.7 0
Exp. $ESHG^d$												-22.0 :	$-22.0 \pm 0.9$	$-22.0 \pm 0.9$

Chapter III. Describing the second harmonic generation of non-resonant liquid: from QM/MM to Maxwell

Table III.3: Molecular first hyperpolarizability  $\beta(-2\omega, \omega, \omega)$  for  $\omega = 0$  in the liquid phase of Liang et al. [55] using CCSD (left column), compared to our CAM-B3LYP results (right column), both using the d-aug-pVTZ basis set. Hyperpolarizabilities are in atomic units. Standard deviations of the  $\beta$  distribution are provided in brackets. The electrostatic environment has been included in the same way.

	Liquid					
λ	$\infty$ [55]	$\infty$				
$\beta_{caa}$	-2.06 [1.06]	-1.87 [1.01]				
$\beta_{cbb}$	$1.59 \ [1.83]$	$2.43 \ [1.74]$				
$\beta_{ccc}$	$3.00 \ [3.04]$	3.80 [2.82]				
$\beta_{bbb}$	$0.00 \ [5.74]$	$0.00 \ [5.21]$				
$\beta_{bcc}$	$0.00 \ [1.64]$	$0.01 \ [1.2]$				
$\beta_{baa}$	$0.00 \ [1.38]$	-0.01 [1.31]				
$\beta_{aaa}$	0.00  [1.56]	$0.01 \ [1.47]$				
$\beta_{abb}$	$0.00 \ [1.64]$	$0.00 \ [1.56]$				
$\beta_{acc}$	$0.00 \ [1.13]$	$0.01 \ [1.11]$				
$\beta_{abc}$	0.00  [0.78]	0.00  [0.85]				

Table III.4: Dipolar water first hyperpolarizability at the quadratic order. The convention used in this Table is not the one developed in the thesis: it uses the traceless quadrupole moment to match the one of the Ref [89], see Appendix IX. A.1. They computed this tensor using Finite Field framework at the CCSD level with the d-aug-cc-pVTZ basis set. Our results are obtained at the DFT/CAM-B3LYP level with the same basis using the response framework at zero-frequency.

$\beta^{dq}$ Component	CCSD [89]	DFT/CAM-B3LYP
aaaa	-92.15	-102.5
aacc	47.29	55.241
abab	-110.83	-124.4
acac	-104.49	-112.6
bbaa	92.96	103.25
bbcc	68.07	77.120
bcbc	-106.46	-117.2
ccaa	73.12	79.912
сссс	-125.17	-135.7

Table III.5: Water second hyperpolarizability in vacuum phase, in atomic unit. The frequency of calculation is in the infinite wave-length limit. Our results (DFT/CAMB3LYP) are compared to the ones obtained by Liang et. al. [55] (CCSD).

$\gamma$	Zero frequency				
	CCSD [55]	DFT/CAMB3LYP			
$\gamma_{aaaa}$	863	965			
$\gamma_{aabb}$	654	763			
$\gamma_{bbbb}$	2711	3156			
$\gamma_{caac}$	450	507			
$\gamma_{cbbc}$	724	828			
$\gamma_{cccc}$	1506	1678			
$\gamma_{\parallel}$	1747	1998			
$\gamma_{THS}$	1794	2090			
$\mathrm{DR}_{\mathrm{THS}}$	73	70			

Table III.6: Water second hyperpolarizability in vacuum, in atomic unit. Our results (DFT/CAMB3LYP) are compared to the ones obtained by Beaujean et. al. [52] (CCSD). Several frequency of calculation are shown, in nanometer. We also observe that  $\gamma_{THS}$  increases with decreasing frequency while the depolarization ratio decreases. For both work d-aug-cc-pVTZ basis set is used.

	$\operatorname{CCSD}$	[52]	DFT/CAMB3LYP	
Frequency [nm]	$\gamma_{THS}$ [a.u.]	$\mathrm{DR}_{\mathrm{THS}}$	$\gamma_{THS}$ [a.u.]	$\mathrm{DR}_{\mathrm{THS}}$
$+ \inf$	1821	76	2090	70
1064	2288	49	NA	NA
800	NA	NA	2640	47
694.3	3401	24	NA	NA

# III. C | Microscopic-Mesoscopic links

At this point, we know how to obtain the individual response of a molecule. In Chapter IV is presented how to extend this approach to molecules in the condensed phase using the structure obtained by classical simulations using the same perturbative framework. Yet, these values are only the building blocks of the optical response measured experimentally. It is then necessary to links this microscopic response to the mesoscopic one. For that, in the next three parts we will see:

- 1. How to link the electromagnetic field that comes to the sample to the one "felt" by a molecule
- 2. How to link the individual response to the mesoscopic one using Maxwell equations
- 3. How to link the mesosocopic polarization to the electromagnetic detected at the Second Harmonic frequency.

## III. C.1 Fundamental electromagnetic field: refraction and exciting field



Figure III.6: Multiscale description of the Second Harmonic Generation. Here is highlighted the link between the shined electromagnetic field at the fundamental frequency and the one felt by the molecule in the condensed phase.

In this section, we will present the tools used to compute the electromagnetic field felt by a molecule from the incident one (controlled by the experimental operator). To understand the different problems we should address, let us first describe the more general and simple case: the bulk liquid phase. At the end of this part, we will mention the issues related to the interfaces.

### Exciting field and electromagnetic field:

First of all, let us assume that there is an electromagnetic field,  $\mathbf{E}$ , in a homogeneous liquid phase. In the previous section, we have seen that we can predict the induced property under an external electromagnetic field at the molecular level. For instance, we have seen that the induced dipole moment at the fundamental frequency at the dipolar approximation can be expressed as:

$$\boldsymbol{\mu}_{M}^{\omega} = \alpha(\omega, \omega)^{M} \cdot \mathbf{e}^{M} \tag{III.67}$$

Note that we use the time-frequency space and that all the quantities are expressed in the laboratory frame.

The field  $\mathbf{e}^{M}$  involved in this molecular equation is the field "felt" by the molecule **before** any reaction of the molecule. This field is called *exciting field* in the rest of this thesis. In the most general case,  $\mathbf{e}^{M}$  is not the electromagnetic field  $\mathbf{E}(\mathbf{r})$  at the molecule position  $\mathbf{r} = \mathbf{r}_{M}$ . See for instance the example in Appendix IX. A.4 for neat water.

Indeed, the electromagnetic field is a measure of the total field at a position  $\mathbf{r}$ : including the response of all molecules. The definition of the response property, such as the polarizability  $\alpha$ , does not include the effect of the molecules on the electromagnetic field. This response tensor answers to the question: "What is the induced dipole moment if an external electromagnetic field  $\mathbf{e}^M$  is applied in the molecule vicinity?"

Hence, we first have to understand how to link the electromagnetic field  $\mathbf{E}(\mathbf{r})$  to the one "felt" by a molecule  $\mathbf{e}^{M}$ . To do so, we will use the molecule polarizability  $\alpha(\omega, \omega)$  and end up with a definition of the relative permittivity of the media,  $\epsilon_r$ .

#### Molecular properties for isotropic and homogeneous liquids: neat water

To compute the exciting field, we will discuss about neat water which is a homogeneous and isotropic fluid. We assume that there is a linear relation between the exciting field and the electromagnetic one for each molecule:

$$\mathbf{e}^M = \mathcal{T}^M \cdot \mathbf{E}(\mathbf{r}_M). \tag{III.68}$$

As we will see later on,  $\mathcal{T}^M$  describe the screening of the electric field due to the molecule reactions, characterized by their polarizabilities  $\alpha(\omega, \omega)$ .

In general,  $\mathcal{T}^M$  can be molecule dependent. However, using the system symmetry and the fact that the polarizability of the water molecule is almost the same everywhere in the liquid phase we can assume that  $\mathcal{T}^M$  is a diagonal 3x3 matrix, identical for each molecule, see Appendix IX. A.4.

For a subsystem smaller than the light spatial wavelength but large enough to respect the liquid symmetries, the molecules feel the same exciting field. In addition, we assume that each molecule as the same molecular polarizability  $\alpha(\omega, \omega)$  and that this tensor is diagonal. An induced dipole moment at the same fundamental frequency is created at each molecule, and, this dipole moment is the same for all molecules. These induced dipole moments will create an oscillating electromagnetic field which will be added to the one originally applied. The sum of these two fields, the source term **E** and the reaction of the whole sample, is the exciting field  $\mathbf{e}^{M}$ . Thus, we have to compute the total field produced by these dipoles to obtain  $\mathcal{T}^{M}$ .

#### Electric field produced by an induced dipole moment:

We compute the electric field produced by all the molecules on a target molecule noted i. The

electric field produced by a dipole  $\mu$ , in the static limit, at a position **r** from its center is given by:

$$\mathbf{F}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{3\left(\mathbf{u}_r \cdot \boldsymbol{\mu}\right)\mathbf{u}_r - \boldsymbol{\mu}}{r^3} - \frac{4\pi}{3}\delta(\mathbf{r})\boldsymbol{\mu}$$
(III.69)

Where  $\mathbf{u}_r$  is the unitary direction toward the position  $\mathbf{r}$ .

There are two terms in this field: inter-molecular and intra-molecular.

#### Inter-molecular field: radiation of dipoles

The first term is the well-known radiation of a dipole: the electrostatic field produced is in the same direction  $\mu$  in the polar direction and opposite at the equatorial ones. Hence, each molecule close to the molecule *i* produces an electric field due to their induced dipole moment. In the laboratory frame centered at the molecule *i* position, the total electric field produced by all the neighbors within a distance *R* and R + dR from the molecule *i* (contains in the shell denoted  $S_R$ ) is  $\mathbf{F}_i(R)$ :

$$\mathbf{F}_{i}(R) = \sum_{j \in \mathcal{S}_{R}} \frac{1}{4\pi\epsilon_{0}} \frac{3\left(\mathbf{u}_{R}' \cdot \boldsymbol{\mu}\right)\mathbf{u}_{R}' - \boldsymbol{\mu}}{R^{3}} = \frac{1}{4\pi\epsilon_{0}} \int_{0}^{\pi} d\theta \int_{0}^{2\pi} d\phi \frac{3\left(\mathbf{u}_{R}' \cdot \boldsymbol{\mu}\right)\mathbf{u}_{R}' - \boldsymbol{\mu}}{R^{3}} \sin(\theta)\sigma R dR \quad (\text{III.70})$$

Where we have assumed a uniform and continuous distribution of dipoles along the sphere of radius R around the molecule i.  $\sigma$  is the surface density (so that there is as many molecules in the sphere  $S_R$  and given  $\sigma R dR$ ) and  $\mathbf{u}'_R$  the unitary vector from the molecule j to the molecule i. Here we should emphasize again that within our hypothesis, the induced dipole moment of each neighbor  $\boldsymbol{\mu}$  is the same in the laboratory frame. Using this continuous approach, we can prove that the resulting field of the neighborhood at the molecule i position cancels out, see Appendix IX. A.4.

#### Intra-molecular field:

The second term represents the fact that a dipole is the infinite limit of two opposite charges q distant by  $\delta$ :  $\mu = \lim_{\delta \to 0} \delta q$ . In between these charges, the electromagnetic field produced is opposite to the direction of the dipole: that is why the second term is proportional to  $-\mu$ .

Since the first term cancels out, thanks to the liquid homogeneity, the difference between the electric field  $\mathbf{E}$  and the one felt by a molecule  $\mathbf{e}^M$  is its own reaction. Since the induced dipole moment is in the same direction as the felt electric field, because the polarizability is positive, the reaction of the molecule will lower the value of  $\mathbf{E}$ . One way to compute this reaction field is to compute the field produced by a single molecule within its vicinity. This electric field, named cavity field, can be calculated (see for instance Ref [1] p.87) and is equal to:  $\mathbf{E}^{cav} = -\frac{1}{3\epsilon_0}\boldsymbol{\mu}$ .

Hence, if one measured the "electric field"  $\mathbf{E}$  at the molecule position, it will measure the one felt by the molecule plus the molecule reaction:

$$\mathbf{E} = \mathbf{e}^{M} + \mathbf{E}^{\text{cav}}, \qquad \qquad \mathbf{E}^{\text{cav}} = -\frac{1}{3\epsilon_{0}}\boldsymbol{\mu}. \qquad (\text{III.71})$$

#### Clausius-Mossotti relation:

Using the previous Equations, Equation III.67 along with Equation III.68 we obtained that:

$$\mathbf{e}^{M} = \frac{1}{1 - \frac{\langle \alpha \rangle}{3\epsilon_{0}}} \mathbf{E} \qquad \qquad \mathcal{T} = \left[1 - \frac{\langle \alpha \rangle}{3\epsilon_{0}}\right]^{-1} \mathbb{1} \qquad (\text{III.72})$$

Working with these Equations, see Appendix IX. A.4, we can obtain the Clausius-Mossotti relation which links the permittivity of the liquid to the molecular polarizability:

$$\mathbf{P}^{\omega} = \epsilon_0(\epsilon_r(\omega) - 1)\mathbf{E}^{\omega} \qquad \epsilon_r(\omega) = \frac{3\epsilon_0 + 2 < \alpha(\omega, \omega) >}{3\epsilon_0 - < \alpha(\omega, \omega) >}.$$
 (III.73)

Where  $\mathbf{P}^{\omega}$  is the averaged mesoscopic polarization at the fundamental frequency.

Using our numerical results, for the water density and polarizability  $\langle \alpha(\omega, \omega) \rangle$ , we found for water at optical frequencies  $\epsilon_r \approx 1.73$  at 800 nm which leads to an optical index of 1.31. This result is very close to the experimental result, 1.33, showing that the hypothesis developed hold for pure water in the bulk phase.

#### Conclusion for the bulk phase:

We have seen that:

- The field involved in the response calculation is not the electromagnetic field  $\mathbf{E}$  but the exciting one  $\mathbf{e}^{M}$ .
- In the bulk phase we have: exciting field + reaction field = electromagnetic field, because the induced field created by the neighborhood cancels out.
- The relation between the exciting field and the electromagnetic field does not depend on the electric field polarization and/or propagation direction.

When we are interested to interpret liquid-SHG polarization-resolved experiment, we can neglect these effects. Indeed, the ratio between the exciting and electric field is the same regardless of the fundamental polarization. Hence, in Chapter V we will work as if the experimental operator controls directly the exciting field polarization inside the liquid phase:  $\mathbf{e}^M \propto \mathbf{E}$ .

#### Problems at the interface:

However, the link between  $\mathbf{E}$  and  $\mathbf{e}^{M}$  is more challenging at the liquid-gas interface. First of all, because the value of the electromagnetic field evolves at the interface, from the gas to the liquid value. The transition of the electromagnetic field value from the gas to the liquid phase is given by the Fresnel coefficients, see Appendix IX. A.5. Moreover, the definition of the exciting field would

be more complicated because the previous arguments based on centro-symmetry and homogeneity are no longer valid at the interface.

We have done a preliminary work based on the approach proposed by A. Morita and co-worker 101. In short, we can still obtain a linear dependency between the applied electromagnetic field in the gas phase,  $\mathbf{E}^{v}$ , and the one felt by the molecule at a given altitude z (z is the normal direction to the liquid interface),  $\mathbf{e}^{M}(z)$  using a 3x3 matrix:

$$\mathbf{e}^{M}(z) = \mathcal{T}(z) \cdot \mathbf{E}^{v}.$$
 (III.74)

 $\mathcal{T}(z)$  is still a diagonal matrix, but the values depend on the molecule position with respect to the surface. The evolution of  $\mathcal{T}$  at the interface is due to the fact that the density, and thus the local field effects, are not homogeneous along z. Moreover, we have  $\mathcal{T}_{xx} \neq \mathcal{T}_{yy} \neq \mathcal{T}_{zz}$  because of the Fresnel coefficients.

Using such approach we can define the exciting field inside the interface, and also the spatial gradient along the z direction. This work is still ongoing and is thus not presented in this manuscript.

### III. C.2 From the individual response to the mesoscopic polarization $P^{2\omega}$

Now that we have defined the links between the applied field and the one felt by the molecules, we can define the constitutive equations for SHG. These equations are the application of the Maxwell equations at the second harmonic frequency. A reminder of them in the gas and liquid phase is proposed in Appendix IX. A.5. As a starting point, the Maxwell equations for dielectric and non-magnetic systems without free charges or current are given by:

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \qquad \qquad \boldsymbol{\nabla} \cdot \mathbf{B} = 0 \qquad (\text{III.75})$$

$$\boldsymbol{\nabla} \wedge \mathbf{E} = -\frac{\delta}{\delta t} \mathbf{B} \qquad \qquad \boldsymbol{\nabla} \wedge \mathbf{B} = \frac{\epsilon_0}{c^2} \mathbf{J} + \frac{1}{c^2} \frac{\delta}{\delta t} \mathbf{E} \qquad (\text{III.76})$$

Where  $\rho$  and **J** are the **bonded** charge density and charge current respectively. We usually defined the mesoscopic polarization **P** so that :

$$\rho = -\boldsymbol{\nabla} \cdot \mathbf{P} \tag{III.77}$$

Coupled with the conservation equation of the electric charge,  $\delta \rho / \delta t = -\nabla \cdot \mathbf{J}$ , we also have that:

$$\mathbf{J} = \frac{\delta}{\delta t} \mathbf{P}.$$
 (III.78)

If we can define a linear susceptibility  $\chi^{(1)}$ , or a relative permittivity  $\epsilon_r$  so that  $\mathbf{P} = \epsilon_0 \chi^{(1)} \mathbf{E} = \epsilon_0 (\epsilon_r - 1) \mathbf{E}$ , then the dielectric displacement defined to be  $\mathbf{D} = \mathbf{E} + \epsilon_0 \mathbf{P}$  propagates in an homogeneous media as a plane wave with the dispersion relation given by:

$$k^2 = \frac{\epsilon_r \omega^2}{c^2} \tag{III.79}$$

Where  $\epsilon_r$  is given by the Clausius-Mossotti Equation III.73.

For the fundamental frequency, the source term is the applied electromagnetic field. For the second harmonic frequency, the source terms are provided by the molecules: a fundamental field induces  $\rho$  and **J** at twice the frequency. Hence, we have to link the microscopic properties (for instance the dipole moment  $\mu$ ) to the mesoscopic ones involved in the Maxwell equations.

#### Spatial averaging:

Indeed, the Maxwell equations should be understood at the mesoscopic scale: the quantities involved are homogeneous at a scale smaller than the wavelength. Hence, the property of the matter, the electric density  $\rho$  and the current **J**, are averaged in space. To do so, we use a based-molecular approach: we define the charge density at a point **r** to be the one produced by each molecule M:

$$\rho(\mathbf{r},t) = \sum_{M} \rho_M(\mathbf{r} - \mathbf{R}_M, t)$$
(III.80)

Where  $\rho_M$  is the molecular charge density in its frame and  $\mathbf{R}_M$  is the 'reference' position of the *M* molecule frame.

In our case, the water molecules are very small compared to the scale of the electric field evolution: the size over which the spatial averaging take place contains many molecules. However, the number of molecules involved in the charge density at a point  $\mathbf{r}$  is still finite: molecules far away from  $\mathbf{r}$  do not contribute because they have no electrons or nuclei far away from their reference position. Therefore, a smoothing function w (one can think of it as a smooth door-like function) is defined to reduce the number of molecules involved in the space averaging calculation:

$$<\rho>(\mathbf{r},t) = \int w(\mathbf{r}'-\mathbf{r}) \left[\sum_{M} \rho_{M}(\mathbf{r}-\mathbf{R}_{M},t)\right] d^{3}\mathbf{r}' = \sum_{M} \left[\int w(\mathbf{r}_{M}+\mathbf{R}_{M}-\mathbf{r})\rho_{M}(\mathbf{r}_{M},t)d^{3}\mathbf{r}_{M}\right]$$
(III.81)

Where  $w(\mathbf{r})$  goes to zero for large  $\mathbf{r}$ . Note that the last integral is done over each molecular frame  $-i.e. \mathbf{r}_M$  is expressed in the molecule's M frame.

#### Multipolar expression:

We have defined the smoothing function w so that it slowly varies at the scale of one molecule. Hence, we perform a Taylor development of  $w(\mathbf{r}_M + \mathbf{R}_M - \mathbf{r})$  near to the molecule center:  $\mathbf{r}_M = \mathbf{0}$ 

$$w(\mathbf{r}_M + \mathbf{R}_M - \mathbf{r}) = w(\mathbf{R}_M - \mathbf{r}) + \mathbf{r}_M \nabla_M w(\mathbf{R}_M - \mathbf{r}) + \frac{1}{2} (\mathbf{r}_M \nabla_M) (\mathbf{r}_M \nabla_M) w(\mathbf{R}_M - \mathbf{r}) + \cdots$$
(III.82)

where  $\nabla_M$  is the gradient in the frame of the molecule M, *i.e.* relative to  $\mathbf{r}_M$ . However, we can also define the spacial derivative using the laboratory frame and not the molecular one. Hence:

$$<\rho>(\mathbf{r},t) = \sum_{M} w(\mathbf{R}_{M}-\mathbf{r}) \int \rho_{M}(\mathbf{r}_{M},t) d^{3}\mathbf{r}_{M} - \boldsymbol{\nabla} \cdot \sum_{M} w(\mathbf{R}_{M}-\mathbf{r}) \int \mathbf{r}_{M} \rho_{M}(\mathbf{r}_{M},t) d^{3}\mathbf{r}_{M} + \frac{1}{2} \sum_{M} w(\mathbf{R}_{M}-\mathbf{r}) \int (\boldsymbol{\nabla} \cdot \mathbf{r}_{M}) (\boldsymbol{\nabla} \cdot \mathbf{r}_{M}) \rho_{M}(\mathbf{r}_{M},t) d^{3}\mathbf{r}_{M}$$
(III.83)

Using the same formalism from the microscopic description, see for instance Equation III.12 or IX.4, we recover the electric multipole expansion:

$$\mathbf{Q}_{M,j_1\dots j_n}^{[n]} = \int \mathbf{r}_{M,j_1} \mathbf{r}_{M,j_2} \dots \mathbf{r}_{M,j_n} \rho_M(\mathbf{r}_M) d^3 \mathbf{r}_M$$
(III.84)

Where  $\mathbf{Q}^{[0]} = 0$  for neutral molecules,  $\mathbf{Q}^{[1]} = \boldsymbol{\mu}$  is the molecular dipole moment and  $\mathbf{Q}^{[2]} = Q$  its quadrupole moment.

Using the definition of the mesoscopic polarization P, Equation III.77, we have at the quadrupo-

#### Page 82

lar order:

$$\mathbf{P}(\mathbf{r}) = \sum_{M} w(\mathbf{R}_{M} - \mathbf{r})\boldsymbol{\mu}_{M} - \frac{1}{2}\boldsymbol{\nabla} \cdot \sum_{M} w(\mathbf{R}_{M} - \mathbf{r})Q_{M}$$
(III.85)

Which is exactly the space averaging of the multipoles:

$$\mathbf{P}(\mathbf{r}) = <\boldsymbol{\mu} > (\mathbf{r}) - \frac{1}{2} \boldsymbol{\nabla} \cdot < Q > (\mathbf{r})$$
(III.86)

Now that we have defined how to write the mesoscopic polarization, we have to express how to link the microscopic response tensor and the mesoscopic susceptibility.

The source terms:  $\mu_M^{SHG}$  and  $Q_M^{SHG}$ 

In our experiments, there is no "external" electromagnetic field at the second harmonic frequency. The molecular induced multipolar moments are produced by a non-linear process with a field at another frequency. In the most general case, the induced dipole moment at the second harmonic frequency of a molecule M in the laboratory frame is given by:

$$\boldsymbol{\mu}_{M}^{SHG} = \frac{1}{2!} \mathscr{B}^{dd} (2\omega, \omega, \omega)^{M} : \mathbf{e}^{M} \mathbf{e}^{M} + \frac{1}{2!} \mathscr{B}^{dq} (2\omega, \omega, \omega)^{M} : \mathbf{e}^{M} \boldsymbol{\nabla} \mathbf{e}^{M}, \quad (\text{III.87})$$

where  $\mathscr{B}$  stands for the first hyperpolarizabilities in the laboratory frame. The quadrupolar moment at the second harmonic frequency is:

$$Q_M^{SHG} = \frac{1}{2!} \mathscr{B}^{qd}(2\omega, \omega, \omega)(\mathbf{r}) : \mathbf{e}^M \mathbf{e}^M.$$
(III.88)

Previously, we have seen that we can express the exciting field  $\mathbf{e}^M$  as a function of the electromagnetic one at the fundamental frequency using a linear relation:  $\mathbf{e}^M = \mathcal{T}^M \cdot \mathbf{E}$  and  $\nabla \mathbf{e}^M = (\nabla \mathcal{T}^M) \cdot \mathbf{E}$ . Here we should consider the case where the exciting field may evolve from one molecule to another. In this case, the spatial averaging should be made on the molecular response tensor (the  $\mathscr{B}$ s) and the exciting field ( $\mathcal{T}$ ) second case, the spatial averaging leads to:

$$\mathbf{P}^{SHG}(\mathbf{r}) = \frac{1}{2!} < \mathscr{B}^{dd}(2\omega, \omega, \omega) \cdot \mathcal{T} \cdot \mathcal{T} > (\mathbf{r}) : \mathbf{EE} \\ + \frac{1}{2!} < \mathscr{B}^{dq}(2\omega, \omega, \omega) \cdot \mathcal{T} \cdot \nabla \mathcal{T} > (\mathbf{r}) : \mathbf{EE} \\ - \frac{1}{2!} \frac{1}{2} \nabla \cdot \left[ < \mathscr{B}^{qd}(2\omega, \omega, \omega) \cdot \mathcal{T} \cdot \mathcal{T} > (\mathbf{r}) : \mathbf{EE} \right]$$
(III.89)

It should be noted that the last component contains two types of terms. One related to the spatial evolution of the molecular properties,  $\nabla \cdot \langle \mathscr{B}^{qd}(2\omega, \omega, \omega) \rangle$ , and the other related to the spatial evolution of the electromagnetic field:  $2\mathcal{T}\nabla\mathcal{T}$ . Hence, we can define a difference of first

hyperpolarizability tensor  $\Delta \mathscr{B}^q(2\omega, \omega, \omega) = \mathscr{B}^{dq}(2\omega, \omega, \omega) - \mathscr{B}^{qd}(2\omega, \omega, \omega)$  so that:

$$\mathbf{P}^{SHG}(\mathbf{r}) = \frac{1}{2!} < \mathscr{B}^{dd}(2\omega, \omega, \omega) \cdot \mathcal{T} \cdot \mathcal{T} > (\mathbf{r}) : \mathbf{E}\mathbf{E} + \frac{1}{2!} < \Delta \mathscr{B}^{q}(2\omega, \omega, \omega) \cdot \mathcal{T} \cdot \nabla \mathcal{T} > (\mathbf{r}) : \mathbf{E}\mathbf{E} - \frac{1}{2!} \frac{1}{2} \nabla \cdot \left[ < \mathscr{B}^{qd}(2\omega, \omega, \omega) > (\mathbf{r}) \right] \cdot \mathcal{T}(\mathbf{r}) \cdot \mathcal{T}(\mathbf{r}) : \mathbf{E}\mathbf{E}$$
(III.90)

Let us describe in more details the nature of these different terms that I called dipolar, gradient and quadrupolar for the first, second and last term.

#### 1. Dipolar term:

This is the best known in the literature. Considering only this term, we obtain:

$$\mathbf{P}^{SHG}(\mathbf{r}) = \epsilon_0 \chi(2\omega, \omega, \omega)(\mathbf{r}) : \mathbf{EE}, \quad \epsilon_0 \chi(2\omega, \omega, \omega) = \frac{1}{2!} < \mathscr{B}^{dd}(2\omega, \omega, \omega) \cdot \mathcal{T} \cdot \mathcal{T} > (\mathbf{r}). \quad (\text{III.91})$$

This expression is extremely powerful because we can take advantage of the system symmetry. In the bulk phase, we expect that the centro-symmetry leads to a zero value of  $\chi(2\omega, \omega, \omega)$ : there is no net mesoscopic  $\mathbf{P}^{SHG}$ . Hence, in Chapter V we will rather describe the electromagnetic field generated by the bulk phase using a scattering formalism. At the liquid-gas interface, we expect  $\chi(2\omega, \omega, \omega)$  to be non-zero only when the centro-symmetry is broken (only at the very edge of the interface). That is why Surface-SHG is said to be specific to the interface:  $\mathbf{P}^{SHG}$  is a sheet of polarization with an infinitely small width. In a first glance, if one assumes that the molecular first hyperpolarizability  $\beta^{dd}$  is the same for each molecule, we can link  $\chi(2\omega, \omega, \omega)$  to the orientation of the molecule at the interface. This is discussed in the first part of Chapter VI.

#### 2. Gradient term:

The second term involves the exciting field gradient, which is express thanks to the  $\mathcal{T}$  matrix. This gradient can be created by two phenomena: an evolution of the molecular structure or an explicit electromagnetic gradient.

Let us consider the bulk phase: we have demonstrated that within a good approximation the link between the exciting field and the electromagnetic one (at the fundamental frequency) is the same for each molecule ( $\mathcal{T} = \mathcal{T}^M$ ). But, if the operator applies an electromagnetic field which presents a spatial gradient, for instance by focalizing using a lens or a objective, this will still induce a gradient at the mesoscopic scale. At the liquid-gas interface, there will still be an evolution of the electromagnetic field even if there is no focalization due to the difference in the optical indices – between the two phases. Note that this second phenomenon leads to larger spatial gradient compared to the one due to the focalization. Finally, the femtosecond laser produces a Gaussian beam, which presents a spatial evolution. For simplicity, we neglect this effect due to the nature of the fundamental beam as the one due to the focalization: we assume a spatial and temporal plane-wave approximation. We can rewrite the product which generates the induced mesoscopic polarization: <sup>8</sup>

$$<\Delta \mathscr{B}^{q}(2\omega,\omega,\omega)\cdot\mathcal{T}\cdot\nabla\mathcal{T}>_{ijk} = <\sum_{J,K,l} \left(\mathscr{B}^{dq}(2\omega,\omega,\omega)_{iJKl} - \mathscr{B}^{qd}(2\omega,\omega,\omega)_{ilJK}\right)\mathcal{T}_{Jj}\frac{\delta\mathcal{T}_{Kk}}{\delta x_{l}} >$$
(III.92)

$$= <\sum_{J,K,l} \left( \langle \langle \hat{\boldsymbol{\mu}}_{i}; \hat{\boldsymbol{\mu}}_{J}^{\omega}, \hat{Q}_{Kl}^{\omega} \rangle \rangle - \langle \langle \hat{Q}_{il}; \hat{\boldsymbol{\mu}}_{J}^{\omega}, \hat{\boldsymbol{\mu}}_{K}^{\omega} \rangle \rangle \right) \mathcal{T}_{Jj} \frac{\delta \mathcal{T}_{Kk}}{\delta x_{l}} > \quad (\text{III.93})$$

For diagonal  $\mathcal{T}$  matrix, we have:

$$<\Delta \mathscr{B}^{q}(2\omega,\omega,\omega)\cdot\mathcal{T}\cdot\boldsymbol{\nabla}\mathcal{T}>_{ijk} = <\sum_{l} \left( \langle \langle \hat{\boldsymbol{\mu}}_{i}; \hat{\boldsymbol{\mu}}_{j}^{\omega}, \hat{Q}_{kl}^{\omega} \rangle \rangle - \langle \langle \hat{Q}_{il}; \hat{\boldsymbol{\mu}}_{j}^{\omega}, \hat{\boldsymbol{\mu}}_{k}^{\omega} \rangle \rangle \right) \mathcal{T}_{jj} \frac{\delta \mathcal{T}_{kk}}{\delta x_{l}} > \qquad (\text{III.94})$$

Here we should emphasize that this expression should be understood in the laboratory frame (ijkl). We can take two examples to understand how this term works.

First, we apply an electromagnetic field along the z direction, with a gradient along the same z direction. In this case, the induced mesoscopic polarization in the z axis for this gradient term will be given by the following expression:

$$<\left(\langle\langle\hat{\boldsymbol{\mu}}_{z};\hat{\boldsymbol{\mu}}_{z}^{\omega},\hat{Q}_{zz}^{\omega}\rangle\rangle-\langle\langle\hat{Q}_{zz};\hat{\boldsymbol{\mu}}_{z}^{\omega},\hat{\boldsymbol{\mu}}_{z}^{\omega}\rangle\rangle\right)\mathcal{T}_{zz}\frac{\delta\mathcal{T}_{zz}}{\delta x_{z}}>\tag{III.95}$$

In the zero frequency limit (Kleinman approximation),  $\langle \langle \hat{\boldsymbol{\mu}}_z; \hat{\boldsymbol{\mu}}_z^{\omega}, \hat{Q}_{zz}^{\omega} \rangle \rangle = \langle \langle \hat{Q}_{zz}; \hat{\boldsymbol{\mu}}_z^{\omega}, \hat{\boldsymbol{\mu}}_z^{\omega} \rangle \rangle$  and thus this term will not generate any net SHG. However, if there is some frequency dispersion, these terms can be different and lead to a coherent second harmonic generation.

Second, we apply an electromagnetic field along the x direction with a gradient along the z one. The induced mesoscopic polarization in the z axis is proportional to:

$$<\left(\langle\langle\hat{\boldsymbol{\mu}}_{z};\hat{\boldsymbol{\mu}}_{x}^{\omega},\hat{Q}_{zx}^{\omega}\rangle\rangle-\langle\langle\hat{Q}_{zz};\hat{\boldsymbol{\mu}}_{x}^{\omega},\hat{\boldsymbol{\mu}}_{x}^{\omega}\rangle\rangle\right)\mathcal{T}_{xx}\frac{\delta\mathcal{T}_{xx}}{\delta x_{z}}>$$
(III.96)

This term is not vanishing, even in the zero frequency limit as it involves the operators  $\hat{Q}_{zx}^{\omega}$  and  $\hat{Q}_{zz}$ .

#### 3. Quadrupolar term:

This term is often denoted in the literature to be the "pure" quadrupolar term: it can arise even if there is no spatial gradient in the exciting field. For the bulk phase, we expect this term to be vanishing. Indeed, even if the response value of the molecule can vary, *i.e.*  $\beta^{qd}$  is not the same for each molecule, this last contribution is the spatial gradient of a spatial averaged. However, such molecular property should be homogeneous in our liquid phase which explains the vanishing contribution for bulk phases.

<sup>&</sup>lt;sup>8</sup>We remind that the  $\nabla$  of the last term of Equation III.89 acts on the quadrupole: the summation over l for the spatial gradient is thus the same as one component of the quadrupole operator later on.

Nevertheless, at the interface this can be no longer the case. Either because the molecular property  $(\beta^{qd})$  or either due to the molecular orientation. In this case, the induced mesoscopic polarization will be generated at the same space where the property of the matter evolves. Hence, this quadrupolar term is still specific to the interfaces.

## Is there a difference between induced, $\mathbf{P}^{SHG}$ , and total, $\mathbf{P}^{2\omega}$ , mesoscopic polarization?

Equation III.90 leads in the most general case the mesoscopic polarization *induced* by the second harmonic generation process,  $\mathbf{P}^{SHG}$ , by an electromagnetic field at the fundamental frequency  $\mathbf{E}$ . However, this mesoscopic polarization **is not always** the one involved in the Maxwell equation  $\mathbf{P}^{2\omega}$ , which we will refer to the total mesoscopic polarization.

To understand the difference, we can consider a system composed of two molecules n and m with induced dipole moments at the second harmonic frequency  $\boldsymbol{\mu}_n^{SHG}$  and  $\boldsymbol{\mu}_m^{SHG}$ . The molecule n generates a field at the second harmonic frequency which will be felt by the molecule m. This molecule being polarizable, this will induced a new dipole moment, noted  $\boldsymbol{\mu}_m^n$ , for the molecule m:  $\boldsymbol{\mu}_m^{2\omega} = \boldsymbol{\mu}_m^{SHG} + \boldsymbol{\mu}_m^n$ . This polarization between the different molecules will modify their dipole moment and thus the total electromagnetic field generated by the sample.

If every induced moments,  $\boldsymbol{\mu}_{M}^{SHG}$ , are oriented in the same direction and with the same amplitude, one can use the exact same calculation made for the fundamental field. In this case, we shall find that there is no difference between  $\boldsymbol{\mu}_{M}^{SHG}$  and  $\boldsymbol{\mu}_{M}^{2\omega}$  in a homogeneous media in the static limit <sup>9</sup>. Hence, the electromagnetic field at the second harmonic frequency,  $\mathbf{E}^{2\omega}$  is the one induced by dipole moments, namely the reaction field. Therefore,  $\mathbf{E}^{2\omega} = -\frac{1}{3\epsilon_0} < \mathbf{P}^{SHG} >= -\frac{1}{3\epsilon_0} < \mathbf{P}^{2\omega} >$ . Note that this is not the same relation as the one found at the fundamental frequency:  $\mathbf{P}^{\omega} = \epsilon_0 \chi^{(1)}(\omega, \omega) \mathbf{E}^{\omega}$ .

#### Conclusions

We have seen that:

- For the second harmonic generation we have to compute the mesoscopic polarizability which can be expressed in terms of the molecular averaged quantities.
- Up to the quadrupole order, three terms can generate the second harmonic: dipolar, gradient and quadrupolar.
- The link between the induced mesoscopic polarization,  $\mathbf{P}^{SHG}$ , and the one involved in the Maxwell equation,  $\mathbf{P}^{2\omega}$ , can be difficult due to inter-molecular (linear) polarization.

For the bulk phase, we have chosen to use a scattering formalism which does not take into account such phenomena: the electromagnetic field measured experimentally is proportional to the one generated by the molecules,  $\mu_M^{SHG}$ . This approximation is supported by the fact that in the liquid phase there cannot be a net induced mesoscopic polarization within the dipolar approximation.

<sup>&</sup>lt;sup>9</sup>Note that if the induced moments are random from one molecule to another we also found that  $\mu_M^{SHG} = \mu_M^{2\omega}$  in averaged.

Therefore the link between the microscopic property and the experimental measurement require no more development for the bulk phase. We will sum up the different effect at the end of this chapter and go deeper in the next Chapter V about Second Harmonic Scattering.

The link between  $\mathbf{P}^{SHG}$  and  $\mathbf{P}^{2\omega}$  is complicated at the liquid-gas interface. We have done a preliminary work to tackle this issue as for the link between the external field and the excitation field at the fundamental frequency. In short, we ended up with another matrix relation:

$$\mathbf{P}^{2\omega} = \mathcal{M} \cdot \mathbf{P}^{SHG} \tag{III.97}$$

Where  $\mathcal{M}$  is a diagonal matrix which depends on the polarizability of the molecule at the second harmonic frequency and the density at the interface. We do not present in detail this work in this document. To conclude, for the surface-SHG we still need to connect the mesoscopic polarization,  $\mathbf{P}^{2\omega}$  located at the liquid-gas interface, to the signal observed experimentally in the reflection direction several meters away.





Figure III.7: Multiscale description of the Second Harmonic Generation. Here is highlighted the link between the individual molecular response at the second harmonic frequency and the electromagnetic field measured experimentally at the same frequency.

We have seen how to get the induced mesoscopic polarization  $\mathbf{P}^{SHG}$  once the microscopic quantity (the  $\mathscr{B}$ s) and the exciting field calculated using Equation III.90. Without going into the details, we assume that we can have the mesoscopic polarization of the interface at the second harmonic frequency,  $\mathbf{P}^{2\omega}$ , which contains the effect of the self-radiation between the emitters (the molecules at the interface) at the second harmonic frequency.

In the following section we will end this multi-scale approach by presenting the 3-layer models which allow us to make the link with the signal collected experimentally with  $\mathbf{P}^{2\omega}$ . Here are the main ingredients of this model.

We assume that we have an interface between a liquid and a gas phase as presented in Figure III.9. Due to an electromagnetic field at a fundamental frequency, a mesoscopic polarization is induced. The surface is invariant in the plane direction xy and the second harmonic is generated only within a thin layer at the interface. Hence, we assume that the mesoscopic polarization is given by:

$$\mathbf{P}^{2\omega}(z) = \mathbf{P}^{2\omega}\delta(z - z_0) \tag{III.98}$$

with  $z_0$  the position of the surface.

This model is presented in other theses of the team [28, 102, 103]. However, I would like to stress the beginning of this approach in this section as it is less often presented, and which answer to the question: "what is the electromagnetic field created by this extremely thin mesoscopic polarization  $\mathbf{P}^{2\omega}$ ?"

First, we will see what is the spatial phase of this mesoscopic sheet along the xy plane, and if it can be decomposed into a plane wave at the second harmonic frequency. Then, we will rewrite Maxwell's equations of this particular system, and then we will present the calculation of the electromagnetic field emitted by this source term  $\mathbf{P}^{2\omega}(z)$  at the second harmonic frequency in the gas

phase. This section is based on Ref [104] and Ref [76] Chapter 7.4.

#### Spatial phase of the source term: $\mathbf{P}^{2\omega}$

This mesoscopic polarization oscillates in time at the second harmonic frequency, but has also a spatial phase. Indeed, the field which induces this phenomenon is the fundamental one propagating in the gas phase with a wave vector noted  $\mathbf{k}^{v}$  and in the liquid phase with the wave vector  $\mathbf{k}^{l}$ . We remind that for a given position at the interface  $\mathbf{r}$  we have:  $\mathbf{k}^{v} \cdot \mathbf{r} = \mathbf{k}^{l} \cdot \mathbf{r}$  at the fundamental frequency. Hence, if we write again the exciting field at the interface with the spatial phase:

$$\mathbf{e}^{M}(\mathbf{r}) = \mathcal{T}(z) \cdot \mathbf{E}^{v} \exp\left[-i\mathbf{k}^{v} \cdot \mathbf{r}\right] \qquad \mathbf{\nabla} \mathbf{e}^{M}(\mathbf{r}) = (\mathbf{\nabla}\mathcal{T})(z) \cdot \mathbf{E}^{v} \exp\left[-i\mathbf{k}^{v} \cdot \mathbf{r}\right] \qquad (\text{III.99})$$

The other term arising from the spatial derivation, proportional to the spatial frequency, is neglected with respect to the one arising from the spatial derivation of  $\mathcal{T}$ .

Within the dipolar approximation, the mesoscopic polarization is given by:  $\mathbf{P}^{2\omega}(\mathbf{r}) \propto \mathscr{B}^{dd}(2\omega, \omega, \omega)$ :  $\mathbf{e}^{M}(\mathbf{r})\mathbf{e}^{M}(\mathbf{r}) >$ . Hence, the mesoscopic polarization has a spatial terms proportional to  $\mathbf{e}^{M}(\mathbf{r})^{2}$ . Therefore, its spatial phase evolves in the xy plane  $\exp[-i2\mathbf{k}^{v}\cdot\mathbf{r}]$ , which can be rewritten using another wave vector noted  $\mathbf{k}^{sheet}$ .  $\mathbf{k}^{sheet}$  is defined so that the phase evolution in the xy plane of  $\mathbf{P}^{2\omega}$  is  $\exp\left[-i\mathbf{k}^{sheet}\cdot\mathbf{r}\right]$ . Hence, we have:

$$\mathbf{k}^{sheet} \cdot \mathbf{r} = 2\mathbf{k}^{v} \cdot \mathbf{r} = \tilde{\mathbf{k}}^{v} \cdot \mathbf{r} \frac{n(\omega)}{n(2\omega)} = \tilde{\mathbf{k}}^{v} \cdot \mathbf{r}$$
(III.100)

where  $\tilde{\mathbf{k}}^{v}$  is the same wave vector as  $\mathbf{k}^{v}$ , but at the second harmonic frequency and  $\frac{n(\omega)}{n(2\omega)}$  is the ratio of the optical index in the gas phase at the fundamental and the second harmonic, which is 1.

Hence, we have shown that  $\mathbf{P}^{2\omega}$  has the same spatial phase as if it was the one of a plane wave propagating along the same direction as the fundamental excitation, but at the second harmonic frequency:  $\tilde{\mathbf{k}}^{v}$ . We can use the same argument also for the quadrupolar terms, and we found that:

$$\mathbf{P}^{2\omega}(\mathbf{r}) = \mathbf{P}^{2\omega}\delta(z - z_0) \exp\left[-i\mathbf{k}^{sheet} \cdot \mathbf{r}\right]$$
(III.101)

Hence, the second harmonic phenomenon creates a sheet of polarization which oscillates in time at the second harmonic frequency, and with a spatial phase corresponding to the reflection or propagation direction of the incoming fundamental field but at the second harmonic frequency. Therefore, we will try to solve the Maxwell equations for this system at the second harmonic frequency using plane waves.

#### New set of Maxwell equations:

The mesoscopic polarization of the whole system (in the sense that it is defined in each part of

space) at the second harmonic frequency is given by  $\mathbf{P}_t^{2\omega}$ :

[

$$\mathbf{P}_t^{2\omega} = \epsilon_0 \chi(2\omega, 2\omega) \mathbf{E}^{2\omega} + \mathbf{P}^{2\omega} \qquad \qquad \rho = -\boldsymbol{\nabla} \cdot \mathbf{P}_t^{2\omega} \qquad (\text{III.102})$$

While  $\mathbf{P}^{2\omega}$  is the source term: a  $\delta$ -Dirac function at the interface altitude. Rewriting the set of Maxwell equations, the Gauss-Maxwell reads:

$$\boldsymbol{\nabla} \cdot \mathbf{E}^{2\omega} = \frac{\rho}{\epsilon_0} = -\frac{1}{\epsilon_0} \boldsymbol{\nabla} \cdot \mathbf{P}_t^{2\omega} = -\chi(2\omega, 2\omega) \boldsymbol{\nabla} \cdot \mathbf{E}^{2\omega} - 4\pi \boldsymbol{\nabla} \cdot \mathbf{P}^{2\omega}$$
(III.103)

$$1 + \chi(2\omega, 2\omega)] \boldsymbol{\nabla} \cdot \mathbf{E}^{2\omega} = -4\pi \boldsymbol{\nabla} \cdot \mathbf{P}^{2\omega}$$
(III.104)

$$\boldsymbol{\nabla} \cdot \mathbf{E}^{2\omega} = -\frac{4\pi}{\epsilon_r(2\omega)} \boldsymbol{\nabla} \cdot \mathbf{P}^{2\omega}$$
(III.105)

Where  $\epsilon_r(2\omega)$  is the dielectric permittivity at the second harmonic frequency. Similarly, the Ampere-Maxwell equation becomes:

$$\nabla \wedge \mathbf{B}^{2\omega} = \frac{\epsilon_0}{c^2} \frac{\delta}{\delta t} \mathbf{P}_t^{2\omega} + \frac{1}{c^2} \frac{\delta}{\delta t} \mathbf{E}^{2\omega} = \mu_0 \frac{\delta}{\delta t} \mathbf{P}^{2\omega} + \frac{\epsilon_r (2\omega)}{c^2} \frac{\delta}{\delta t} \mathbf{E}^{2\omega}$$
(III.106)

Computing  $\nabla \wedge \nabla \wedge \mathbf{E}^{2\omega}$  leads to the propagation equation:

$$\boldsymbol{\nabla} \left( \boldsymbol{\nabla} \cdot \mathbf{E}^{2\omega} \right) - \Delta \mathbf{E}^{2\omega} + \frac{\epsilon_r(2\omega)}{c^2} \frac{\delta^2}{\delta t^2} \mathbf{E}^{2\omega} + \mu_0 \frac{\delta^2}{\delta^2 t} \mathbf{P}^{2\omega} = 0$$
(III.107)

For a given source term  $\mathbf{P}^{2\omega}(\mathbf{r})$ , Equation III.107 defines the electromagnetic field. At this point we should discuss the type of solution we are looking for by taking advantage of the system geometry.

The system is decomposed into three parts: on the top the air, on the bottom the liquid and in between the interface where lies the source term. Hence, the electric field form is also decomposed into three parts using the Heaviside function  $\theta(z)$ <sup>10</sup> and  $\delta$ -Dirac one:

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{v}(\mathbf{r})\theta(z-z_{0}) + \mathbf{E}^{l}(\mathbf{r})\theta(z_{0}-z) + \mathbf{E}^{sheet}(\mathbf{r})\delta(z-z_{0})$$
(III.108)

Note that for readability, we have dropped the indices  $2\omega$  for these electromagnetic fields.

We are looking for an electric field which can be detected in the reflection direction: the wave vector in the gas phase makes a angle about 70 degrees with respect to the surface normal. However, using such approach directly is difficult because the liquid has one optical index, the gas phase another one, and the optical index of the interface is not known. Therefore, the usual mathematical path is first to resolve the Equation III.107 with a  $\delta$ -Dirac polarization sheet where the optical index is the same for the whole system, *i.e.* only at the "interface". Then, we will use the results obtained at the interface to compute the field propagating toward the detector (in the gas phase).

 $<sup>{}^{10}\</sup>theta(z) = 1$  for z > 0, zero otherwise

#### Solutions for a polarization sheet inside a homogeneous media:

Hence, we assume a system (for instance a liquid phase) with a constant optical index. Considering that there is a sheet of polarization  $\mathbf{P}^{2\omega}$  at a given altitude, what is the electromagnetic field generated in the normal direction of this sheet?

We are looking for an electric field decomposed into three parts: one defined at the top of the sheet propagating in the +z direction  $\mathbf{E}^+$ , one defined below the sheet propagating in the -z direction  $\mathbf{E}^-$ , and one inside the sheet  $\mathbf{E}^{sheet}$ , see Figure III.8.  $\mathbf{E}^+$  propagates using the wave vector  $\mathbf{k}^+$  and  $\mathbf{E}^-$  with  $\mathbf{k}^-$ . The orthonormal basis set used to decompose the polarization of  $\mathbf{E}^+$  and  $\mathbf{E}^-$  use the same formalism as for the fundamental frequency: we defined the unitary vector in the sheet plane,  $\hat{s}$ , and the one orthonormal to the wave vector and  $\hat{s}$  noted  $\hat{p}$ . Since the optical index is the same above and below the sheet, the wave vector for  $\mathbf{E}^+$  and  $\mathbf{E}^$ and  $\mathbf{k}_z^+ = -\mathbf{k}_z^-$ .



Figure III.8: Scheme of the polarization sheet within a homogeneous media. A polarization sheet creates two fields propagating in the +z and -z directions:  $\mathbf{E}^+$  and  $\mathbf{E}^-$  with associated  $\mathbf{k}^+$  and  $\mathbf{k}^-$  wave vector. Along with these fields are defined their polarization basis:  $\hat{s}^{\pm}$  and  $\hat{p}^{\pm}$ . Note that  $\hat{s}^{\pm}$  is along the y-axis.

below the sheet, the wave vector for  $\mathbf{E}^+$  and  $\mathbf{E}^-$  have the same norm and  $\mathbf{k}_y^+ = \mathbf{k}_y^- = 0$ ,  $\mathbf{k}_x^+ = \mathbf{k}_x^-$  and  $\mathbf{k}_z^+ = -\mathbf{k}_z^-$ .

Using these definitions, we obtained for the electric fields values, see equations 7.50 to 7.53 of Ref [76]:

$$\mathbf{E}^{+}(\mathbf{r}) = A\left(\left[\hat{s}^{+}\cdot\mathbf{P}^{2\omega}\right]\hat{s}^{+} + \left[\hat{p}^{+}\cdot\mathbf{P}^{2\omega}\right]\hat{p}^{+}\right)\exp\left[-i\mathbf{k}^{+}\cdot\mathbf{r}\right]$$
(III.109)

$$\mathbf{E}^{sheet}(\mathbf{r}) = -\frac{1}{\epsilon_0 \epsilon_r} \hat{z} \cdot \mathbf{P}^{2\omega} \hat{z} \exp\left[-i\mathbf{k}^+ \cdot \mathbf{r}\right]$$
(III.110)

$$\mathbf{E}^{-}(\mathbf{r}) = A\left(\left[\hat{s}^{-}\cdot\mathbf{P}^{2\omega}\right]\hat{s}^{-} + \left[\hat{p}^{-}\cdot\mathbf{P}^{2\omega}\right]\hat{p}^{-}\right)\exp\left[-i\mathbf{k}^{-}\cdot\mathbf{r}\right]$$
(III.111)

Where A is a factor which does not depend on the fields directions. Note that since we are interested only in relative values, we will drop the A coefficient later on.

Using these Equations, we can link the mesoscopic polarization, which is created by the SHG processes at the molecular level, to the electromagnetic field generated to a particular direction,  $\mathbf{E}^+$  and  $\mathbf{E}^-$ . Importantly, the polarization of these fields, along the  $\hat{s}^{\pm}$  and  $\hat{p}^{\pm}$  directions depends directly on the mesoscopic polarization itself, for instance  $\mathbf{E}^+ \cdot \hat{s}^+ = \hat{s}^+ \cdot \mathbf{P}^{2\omega}$ . Therefore, collecting the outcoming electromagnetic field with respect to the polarization, *i.e.* the  $\hat{s}^{\pm}$  and  $\hat{p}^{\pm}$ , will indeed provide information about the mesoscopic polarization itself.

#### Solutions for a polarization sheet inside an interface:

With this ideal solution we can now find a solution of the system of interest composed of 3 parts with different optical indices. The trick is to give to the "interface" an infinitesimal thickness and a optical index as presented in Figure III.9. Inside this medium, we will assume that the source term

Chapter III. Describing the second harmonic generation of non-resonant liquid: from QM/MM to Maxwell

is still a  $\delta$ -Dirac function in order to compare to the calculation above. Therefore, we can define the electromagnetic field which propagates in the top and bottom direction ( $\mathbf{E}^+$  and  $\mathbf{E}^-$ ) which are generated by the polarization sheet  $\mathbf{P}^{2\omega}$ .

Once this is done, we are back to a very simple case to solve: propagating the electric field inside the interface through two surfaces: the air-interface and interface-liquid. We will not present the details here as this last part is done in many recent theses of the group [28, 102, 103]. In short, we are interested by the field propagating in the gas phase,  $\mathbf{E}^{v}$  in Figure III.9. This field is the sum of at least two fields. One corresponding to the electromagnetic field created by the polarization sheet propagating in the top direction,  $\mathbf{E}^+$ , and that gets transmitted to the gas phase. The other being the field which is reflected by the other surface (interface-liquid) and then gets transmitted to the air phase,  $\mathbf{E}^{-}$ . The calculation of the field propagating toward the detector,  $\mathbf{E}^{v}$ , is thus expressed using Fresnel transmission and reflection coefficients. It should be noted that one issue of this method is to define the optical index of the "interface" media.

#### **Conclusion:**

Using the Maxwell equations in matter solved for a source term corresponding to a polarization sheet,  $\mathbf{P}^{2\omega}$ , we have been able to define two propagating fields inside the "interface"  $\mathbf{E}^+$  and  $\mathbf{E}^-$ . Then, using a three-layer model and the Fresnel coefficients, the electromagnetic field propagating in the air in the reflection di-



Figure III.9: Scheme of the 3-layer model. The electromagnetic field is generated inside a "continuous and homogeneous media" called interface. A polarization sheet creates two fields propagating in the +z and -z directions:  $\mathbf{E}^+$  and  $\mathbf{E}^-$  respectively.  $\mathbf{E}^+$  is transmitted in the gas phase according to the Fresnel transmission coefficient from the "interface" to the gas. A part of the  $\mathbf{E}^-$  is reflected by the surface between the "interface" and the liquid phase, and then transmitted to the gas phase.

rection can be expressed as a function of the source term  $\mathbf{P}^{2\omega}$ . This approach has the advantageous to be very simple to use, but have an unknown parameter: the optical index of the interface medium. We are currently trying to use this approach at the liquid gas interface, and to find a robust way to define this optical index.

# III. D | Conclusion: bulk and surface SHG specificities

We have presented in this chapter the main equations that describe the electromagnetic waves and the second harmonic process at the misoscopic scale. To conclude, we will sum up the specificities of the two geometries: the bulk and at the liquid-gas interface.

#### SHG in the bulk phase:

In the bulk phase, the fundamental electric field arrives in the normal direction with respect to the sample. Hence, the transmitted electric field is the same regardless of the fundamental polarization because of this incident angle.

The link between the exciting field and the fundamental electromagnetic field is the same for all the molecules: we can use Equation III.71. Experimentally, the electromagnetic field is focalized inside the liquid phase using an objective. However, this objective should not induce an important electromagnetic intensity variation at the scale of the molecule. Therefore, we expect the spatial gradient of the fundamental to be small: the dipole-quadrupole term (second term of Equation III.87) can be disregarded for this experimental setup. Moreover, we are in the liquid phase: we do not expect a variation of the average first hyperpolarizability nor the electromagnetic field. Hence, we can also neglect the last term and stays in the dipolar approximation:

$$\mathbf{P}^{SHG}(\mathbf{r}) = \frac{1}{2!} < \mathscr{B}^{dd}(2\omega, \omega, \omega) > (\mathbf{r}) : (\mathcal{T} \cdot \mathbf{E}(\mathbf{r})) (\mathcal{T} \cdot \mathbf{E}(\mathbf{r}))$$
(III.112)

where  $\langle \rangle$  is a spatial averaging, note that in this equation all the quantities are expressed in the laboratory frame. However, if we compute the average of the first hyperpolarizability in the laboratory frame at the dipolar level,  $\langle \mathscr{B}^{dd} \rangle$ , we found that it should be strictly null for centro-symmetric medium. Hence, using this formalism, we found that  $\mathbf{P}^{SHG}$  is null. For this reason, this process is called "scattering" as there is no mesoscopic polarization at the second harmonic frequency in the bulk phase.

Hence, the second harmonic electromagnetic field is described using individual molecular sources: see Chapter V. As for the fundamental field, the emitted electromagnetic field is transmitted regardless of its polarization to the detector: the field propagates with a normal incidence with respects to the interfaces encountered.

Therefore, regarding the bulk geometry we should:

- 1. Compute the dipole-dipole first hyperpolarizability.
- 2. Compute the intensity collected at the second harmonic in this geometry as a function of the fundamental polarization using a scattering approach.

The method used is for the first point is presented in Chapter IV, and the second point is done in Chapter V.

#### SHG at liquid-gas interface:

Describing Second Harmonic Generation at interfaces, in particular at the liquid-gas interface, requires more steps.

First, the fundamental light arrives with a non-normal incident angle. Therefore, according to the Fresnel coefficients, see Equation IX.75, the electromagnetic intensity transmitted in the liquid phase depends on the incoming polarization. Hence, when describing the polarized-resolved experimental results, we should take into account explicitly the Fresnel coefficients.

Moreover, the link between the exciting field and the electromagnetic one is much more ambiguous. Indeed, the part of space that generates the most second harmonic is the 'interface" itself: where the centro-symmetric is explicitly broken. Therefore, if we can define a tensor  $\mathcal{T}$  which links the exciting field to the electromagnetic one, it may depend on the localization at the interface.

The mesoscopic polarization can be induced by a dipolar term because we are in a non-centrosymmetric medium (hence  $\langle \mathscr{B}^{dd} \rangle \neq 0$ ). But, within the interface there can have important spatial gradients. Indeed, since the electromagnetic field intensity evolves from the air to liquid within the interface, due to the optical indices difference between the gas and the liquid phases, there is an electromagnetic field gradient at the interface. Hence, the second term of Equation III.90 also generates an induced mesoscopic polarization. Moreover, the property of the liquid phase evolves at interfaces. Hence, we can have an evolution of the quadrupolar first hyperpolarizability which would generate a mesoscopic polarization at the second harmonic frequency, see the third term of Equation III.90. Therefore, the SHG processes at the surface can involve 3 terms which correspond to 3 different processes.

Once we have defined the induced mesoscopic polarization at the second harmonic frequency depending on the incoming fundamental polarization, we should wonder how it generates an electromagnetic field in the reflection direction. This problem contains two steps which can be difficult to disentangle. Indeed, we should recall that the induced mesoscopic polarization  $\mathbf{P}^{SHG}$ , from Equation III.90, is the **source** term of the Maxwell equations at the second harmonic frequency, not the mesoscopic polarization  $\mathbf{P}^{2\omega}$ . The difference between these two quantities is the radiation of the other molecules: the electromagnetic field at the second harmonic frequency created by molecules induces dipoles moment since these molecules are polarizable. Finally, once the mesoscopic polarization obtained, we have to describe the electromagnetic fields emitted: first inside the interface and then in the air toward the detector. This second part can be handled by the so-called "three-layer models" presented previously. Therefore, regarding the surface geometry we should:

- 1. Compute the electromagnetic field transmitted to the liquid phase using the Fresnel coefficients.
- 2. Find the link between the exciting field and the electromagnetic one at the interface:  $\mathcal{T}(z)$ .
- 3. Compute the dipole-dipole, dipole-quadrupole and quadrupole-dipole first hyperpolarizabilities and compute the induced mesoscopic polarization using  $\mathcal{T}(z)$  and its spatial gradient.
- 4. Find the link between the induced mesoscopic polarization and the mesoscopic polarization at the second harmonic frequency.
- 5. Compute the polarization dependent electromagnetic field emitted in the reflection direction in the air phase from this mesoscopic polarization at the interface.

Point 1 is briefly presented in Appendix IX. A.5. Point 2 and point 4 are ongoing works. Point 3 is presented in Chapter VI. Finally, point 5 is done thanks to the three-layer model presented in the Section III. C.3

# IV | How to compute the hyperpolarizability in the liquid phase: FROG

With great power comes great responsibilities. Spider-man Cluster rule

## The 5 Stages of Debugging

At some point in each of our lives, we must face errors in our code. Debugging is a natural healing process to help us through these times. It is important to recognize these common stages and realize that debugging will eventually come to an end.

## \_ Denial



This stage is often characterized by such phrases as "What? That's impossible," or "I know this is right." A strong sign of denial is recompiling without changing any code, "Just in case."



## **Bargaining/Self-Blame**

Several programming errors are uncovered and the programmer feels stupid and guilty for having made them. Bargaining is common: "If I fix this, will you please compile?" Also, "I only have 14 errors to go!"

## Anger 🔪



Cryptic error messages send the programmer into a rage. This stage is accompanied by an hours-long and profanity-filled diatribe about the limitations of the language directed at whomever will listen.





Following the outburst, the programmer becomes aware that hours have gone by unproductively and there is still no solution in sight. The programmer becomes listless. Posture often deteriorates.





The programmer finally accepts the situation, declares the bug a "feature", and goes to play some Quake.

Figure IV.1: From https://www.reddit.com/r/ProgrammerHumor/comments/1r0cw7/the\_5\_stages\_ of\_debugging/

## IV. A | Introduction

#### From pen and paper to the cluster:

In Chapter III, we have seen how to derive the response of a molecule at the second harmonic frequency using a perturbation framework together with a quantum method such as DFT. Here we present the electrostatic embedding concept to include the condensed phase effect on a single molecule Hamiltonian. Such an approach is an approximation of the real quantum system. The idea is to solve the electronic degrees of freedom on a Hamiltonian composed of one molecule, even if there are neighbors within few Angstrom. This method is efficient and quite cheap numerically speaking. The principle is to build for each molecule its electrostatic environment and then add the corresponding potential in the monomer Hamiltonian. For such calculation, we use the open-source software DALTON [73]. The Polarizable Environment (PE) scheme [95] allows including the electrostatic environment in many ways within this monomer approach and the DALTON software computes the hyperpolarizabilities relative to these embedded molecules.

However, we still have to make the link between the structure, either a bulk phase or an interface, obtained by classical Molecular Dynamics (MD), and the DALTON input files. This represents the writing of the corresponding input files of hundreds of time steps from a Molecular Dynamics trajectory containing thousands of molecules. Using these input files, we then have to perform all of the QM/MM calculations using DALTON, representing thousands of calculations which take about 15-20 min on a single core **each**. Finally, we have to read and analyze the hyperpolarizability tensors from each DALTON output files.

#### Home-made tool: FROG

At the beginning of this thesis, we had no access to any software performing the construction of these files. Hence, an important part of my thesis was devoted to the coding of the software FROG: "FROm molecular dynamics to second harmonic Generation". We would like to emphasize that we will share FROG in open access soon.

This Chapter presents the method used in FROG and Chapter V and VI will present an illustration of the results obtained. The code functionalities, numerical procedures, strengths and limits are presented in the first part of this Chapter.



Toward our journey to compute the SHG response of liquid-gas interfaces, we struggled to understand the "long-range" effect of the neighborhood. Indeed, it seemed that the neighbors at a distance more than 1 nm has an impact on the microscopic response. From this issue, we computed the dipolar second hyperpolarizability in the bulk phase and tried to use it to better understand the effect of the electrostatic environment at large distance. This study has been published in an article, accepted in PCCP. This work will be presented in the second part of this Chapter. Importantly, we propose an upgrade for the electrostatic embedding approach which aims to include the long-range neighborhood more efficiently.

# IV. B $\mid$ Overview of FROG

In this first section, we will briefly present our home-made code: FROG. Its development is closely related to the questions we addressed. Hence, we will first present the scientific part of the code and the objectives of this code. Then, we will go in a bit more details regarding the numerical procedure and briefly present the current stage of the code and the associated wiki. Note that more details are available in the code wiki available on Zenodo<sup>1</sup> or in Appendex IX. B.1.

## IV. B.1 Scientific presentation

In the previous Chapter, we have seen that the Second Harmonic process mainly involves the electronic degrees of freedom at optical frequency. The formalism developed in Section III. B based on the response scheme is unambiguous for a molecule in a given geometry in the vacuum phase. Typically, one starts by finding its optimized geometry at the quantum level. Then, using this single geometry we can estimate the ground state properties and extract the (hyper)polarizabilities. This represents one QM calculation. If the molecule presents many conformations, we should perform several calculations to explore the flexibility effects on the response tensor. For small molecules it still represents only few QM calculations.

However, for non-resonant molecules, such as water, it is known that the first hyperpolarizability is impacted by the electrostatic environment [30, 54, 55]. This raises many problems both at the numerical and experimental point of view. For instance, this means that each molecule in the liquid phase may have a unique first hyperpolarizability depending on its environment. Hence, the response of the molecule should be computed at the QM/MM level: including somehow the environmental effect on the electronic degree of freedom.

This increases dramatically the numerical cost and complexity of the approach in comparison with vacuum-like calculations. Indeed, to compute the SHG properties of the molecules in the liquid phase we have to sample the possible geometries along with their relative non-linear responses at the quantum level. Thus, we are left to at least two problems: (i) how to describe the structure of the liquid phase and (ii) how to compute the response property of such condensed structure?

## How to describe the condensed phase structure?

To answer the first problem, we will use Classical Molecular Dynamics (MD). In short, the intramolecular interactions are described using classical strings, and the inter-molecular interactions by Coulombic and Van der Walls interactions. The set of parameters used for each interaction is called the Force Field (FF) and represents the most important ingredient for a MD simulation with Periodic Boundary Conditions (PBC). In an MD simulation, we can describe the bulk phase or interfaces at fixed temperature and pressure. To do so, the Newton equations of motion are

<sup>&</sup>lt;sup>1</sup>https://doi.org/10.5281/zenodo.5998193

propagated starting from an initial structure with a thermostat (acting on the molecule velocities) and barostat (acting on the box simulation size). Once the MD performed, we have access to the structure of the condensed phase. A set of pictures with the position of each nuclei is provided. This collection of pictures gives us a pull of configurations to describe the condensed phase. We will use these structures as a starting point for the response property calculations.

#### How to describe the condensed phase at the Quantum level?

Now that we have a description of the liquid phase structure, we can compute its SHG response. We can either decompose the system into clusters of molecules, or use PBC and include everything.

Using the PBC approach, we build a cell composed of a small number of molecules (typically hundreds of water molecules) and describes the wave-function of the whole system. Using such approach, we can tackle liquid or solid systems in the bulk phase or at interfaces. For instance, such approaches have been applied to calculate SFG [16, 37] or SHG [41] responses at liquid interfaces.

However, this approach has at least two drawbacks:

• The numerical cost:

For DFT with hybrid functional, the ground state calculations cost grows with the  $4^{th}$  power of the number of electrons. Describing the condensed phase using PBC still requires a certain number of configurations and number of molecules inside the direct space because of the disorder of the liquid phase.

• Identification of the molecular response:

Once the ground state calculations performed, we can compute the SHG response of the whole system, using a Finite Field method for instance. Yet, we are more interested in a molecular-based understanding: we would like to attribute to each molecule its "individual response". Hence, once the electronic states calculated in this explicit condensed phase, we have to re-localize the property at the molecular level. For instance, using method such as the Maximally Localized Wannier Functions, see Ref 41.

Using such method is possible and should provide a very good description of the short-range intermolecular interaction. Here, by contrast, we have chosen to use a cluster approach, and, in particular, the simplest one: a monomer QM calculation. In this case the quantum Hamiltonian is defined for only one molecule, and the Hamiltonian is modified to include the effect of the condensed phase.

#### QM/MM calculations of monomer:

To do so, we can use an **explicit solvatation** scheme such as the Polarizable Embedding (PE) scheme. This method is a QM/MM approach: it mixes a classical description of the neighborhood with a Quantum one for the molecule of interest. The interaction between the neighborhood and the monomer is described by an electrostatic interaction: the neighbors generate a **space-heterogeneous** electrostatic field in the vicinity of the molecule:  $\mathbf{E}^{\text{env}}(\mathbf{r})$ . The electric potential

associated  $V^{\text{env}}(\mathbf{r})$  is computed and added to the "isolated" Hamiltonian as follow:

$$\hat{H}^{\text{env}} = \hat{H}^0 + \hat{V}^{\text{env}}(\mathbf{r}) \tag{IV.1}$$

Hence, we can keep the same response formalism: the ground state property of the molecule are computed not for the vacuum phase Hamiltonian,  $\hat{H}^0$ , but for the condensed phase one,  $\hat{H}^{\text{env}}$ . Then, the response tensor calculation follows the same scheme. This new interaction will thus modify the electronic degrees of freedom, and hence the electronic first hyperpolarizability.

Such method is called explicit solvation because the neighborhood inclusion depends explicitly on the local structure: each molecule of the liquid phase will have its own  $\hat{V}^{\text{env}}(\mathbf{r})$  created by its own neighborhood. It is also often called "electrostatic embedding" as the Quantum box is placed inside a classical condensed phase made of charges, dipoles... This is depicted in Figure IV.2: the surroundings of a "target" molecule (in red) is converted into an electrostatic description: here a collection of positive and negative charges. To compute the



Figure IV.2: Conversion of the condensed phase environment around a "target" molecule in red into an electrostatic environment. The positions of the neighbors on the left are converted into a collection of (partial) positive and negative charges around the target molecule on the right.

electrostatic field created by the neighborhood, we can use the same electrostatic description as the one used for the Force Field of the MD. We have the structure of the neighborhood, thanks to the MD, and for each neighbor around the molecule of interest, we attribute a point change description. For instance, a neighboring Hydrogen will be described as a point charge of +0.5564 e. The electrostatic field generated inside the QM box by a given neighbor is the one of a point charge at the distance given by the MD. Performing such attribution for each neighbor define the electrostatic potential felt  $V^{\text{env}}(\mathbf{r})$  by the monomer in the QM box.

Hence, we have decided to use a monomer approach: we solve the electronic degrees of freedom of each molecule, but one at a time. Such methods are already established, for instance in the DALTON [73] community. A robust method has been developed to include the electrostatic environment around a molecule thanks to the Polarizable Embedding (PE) scheme [62, 72, 105–107]. DALTON also has a very powerful and versatile response theory framework. We have been able to compute the dipolar and quadrupolar first hyperpolarizability and the second hyperpolarizability of water in the liquid phase at optical frequency.

#### Why a new numerical tool?

However, at our knowledge, we do not have access to an efficient software/environment which
builds from the MD simulations the input files needed to perform these Quantum calculations by DALTON <sup>2</sup>. Moreover, since we are particularly interested in the inclusion of the first hyperpolarizability fluctuations we will have to perform **an important number** of calculations. This means that we have built the electrostatic environment of a large number of molecules, and write the input files for DALTON, and perform the calculations on a cluster. Then we have to load the results from the DALTON output files. Such considerations may seem in a first glance to be "details", but it is really not: what is not made automatic should be done by humans. For instance, we have performed more than 150 000 QM/MM calculations in the liquid phase (Chapter V), and 100 000 QM/MM calculations at the liquid-gas interface for (Chapter VI), and, only for the production runs.

Hence, we have built a software which aims to make this QM/MM approach for the calculation of SHG-related molecular properties efficiently on a practical point of view, starting with a MD trajectory, and ending up with the molecular hyperpolarizabilities at the quantum level within the PE scheme using DALTON. This code should run on a cluster, and also manage the quantum calculations that should be performed.

We want to be able to work at least in two kinds of geometries: in the bulk and at liquid-gas interfaces. Hence, the building of the electrostatic environment should be able to work with periodic boundary conditions for these two types of systems.

Finally, an important goal is the user accessibility: in the short-mid term, we want to distribute this software in open access with an efficient set of tutorials and a complete documentation.

The result is a Python code named FROG for "FROm molecular dynamics to second harmonic Generation". Let us review its work flow, strengths and limitations on today's implementation.

### IV. B.2 Work flow and review

#### Work flow:

FROG work flow is described in Figure IV.4. We shortly present it in this chapter (more details can be found in the Appendex IX. B.1).

FROG is written in Python and can be run on a cluster in parallel (Single-CPU). It is an inputscript based software, the typical number of entries for a FROG run is about fifty lines. FROG is divided in four parts: initialization, part I, II and III.

In the initialization the input file is read, FROG opens the MD trajectory that should be investigated and built its core objects. FROG can read several time steps and compute the SHG properties of all or parts of the molecules within the frames. Importantly, even if we present in this thesis results

<sup>&</sup>lt;sup>2</sup>Recently, the Dalton Project, https://www.daltonproject.org has been released. We had already started to code FROG to especially tackle interfaces, and it seems to us that such feature was not present inside this project. Hence we have continued with FROG during the thesis.

obtained only for water, FROG can work for any liquids.

Then, the part I of FROG performs all the "structural" analysis (density, molecular orientation, H-bonds, Radial Distribution Function and electrostatic field environment). If any QM/MM calculations should be performed, it will also write the input files for DALTON. The second part of FROG does not actually performs the QM/MM calculations. It generates submission files intended for a cluster so that the user can launch (with many options) the many QM/MM calculations that should be

FROG	Release:         1.0.0           Date:         Mar 14, 2022
	FRG is a software originally designed to compute the second harmonic generation response of liquids from Molecular Dynamics (MD) at the quantum level. To achieve this, it uses the PE scheme of the <u>DALTON</u> software which provides optical response of molecules within an electrostatic embedding.
FROm molecular dynamics to second harmonic Generation	Here are the main FROG functionalities:
Navigation	<ul> <li>Perform usual structural analysis such as density, molecular orientation, Redial Distribution Function or H bond</li> </ul>
<ol> <li>How to Install Frog?</li> <li>Tutorials</li> <li>Global Parameter</li> </ol>	<ul> <li>Compute automatically the <b>optical</b> response (polarizability, first and second hyperpolarizability) of molecules at the quantum level in an embedded environment.</li> </ul>
<ol> <li>Molecule Type</li> <li>Diagram</li> </ol>	Its main advantages are:
6. Optical Analysis 7. Mathematical and Numerical Index 8. Frog package Outick search	<ul> <li>Can open many types of MD, works for pure liquid or mixture.</li> <li>Can deal with 2D liquid interfaces, but also work in bulk.</li> <li>The user can quite easily define the parameters for a molecule description or analysis.</li> <li>Has a whole part designed to help you to treat thousands of QM calculations in</li> </ul>
Go	a cluster. And for honesty sake, its drawback:
	• Not that easy to use at first.

performed, automatically. The user can loop *Figure IV.3: Screen shot of the FROG wiki home-page.* FROG call in this second part to perform few QM/MM calculations at a time, or all at once. Finally, once all the QM/MM calculations performed, FROG will first read the results obtained from the DALTON output files. Then, it will perform statistical averaging, in time and in space if required. The final result is a relatively small file (from few MB to hundreds of MB) that can be stored in a personal computer.

In order to compute the first hyperpolarizability for a given frame and molecule, FROG first finds the environment around it up to a given distance (Part I). Then, it will convert all the neighbors into a charge distribution and write the file needed for the PE DALTON calculations. This procedure is repeated for all the "target" molecules and frames required. All of these QM/MM configurations (one target molecule + its electrostatic environment) are then centralized and a script file is written to help the user to perform these calculations on a cluster (Part II). We are dealing with many QM/MM calculations where each calculation can be done on a single core. FROG can write cluster job submission files so that several QM/MM calculations can be made on a single CPU. Special effort has been made to deal with this important number of calculations, see Appendex IX. B.1 for more details. Finally, once all the QM/MM calculations have been performed, FROG reads the output files and store the first hyperpolarizability calculated (Part III). Moreover, FROG can perform statistical analysis, for instance the evolution of the average with respect to the distance to an interface or distribution analysis.

## Today's tutorial and wiki:

Here is a review of the current stage of FROG wiki and tutorials. FROG wiki is an HTML files and, as the code, is written in English. The wiki is created thanks to Sphinx <sup>3</sup>, see Figure IV.3. Its contains several sections:

<sup>&</sup>lt;sup>3</sup>https://www.sphinx-doc.org



Figure IV.4: Scheme of FROG workflow. FROG can use as raw material Molecular Dynamics trajectories from several software and for different geometries: pure liquid or at interfaces. During step I, FROG creates electrostatic environment for every molecule and writes the relative input files for the QM/MM calculation performed by DALTON. Then, the user has to perform the QM calculation on a cluster: FROG helps the user by generating files and by giving feedback in step II. Once all the QM calculations done, FROG reads the obtained values in the DALTON output files during the step III. The returned data contain the distributions of the computed quantities or the spatial evolution of averages.

## Installation:

How to install FROG and to run it. We also present some numerical-efficiency-related information.

## Global Parameter, Molecule Type, Diagram and Optical Analysis:

These 4 other sections describe in detail the main objects of FROG and how to define them or use them. The last one regarding the Optical Analysis is a mixed of pure numeric and physical insights.

## Tutorials:

9 tutorials present how to use FROG from the simplest to the most complex functionalities.

- Overview and Get started are the beginner tutorials.
- *Space Discretization* explains how FROG can assign molecules to a particular part of the system, and how to perform spatial-based analysis.
- *Mixture and structural analysis* uses a methanol-water mixture to explain all the possibilities available in FROG regarding the structural analysis.
- *Optical Analysis* presents the QM/MM framework to compute the first hyperpolarizabilities within the PE environment.
- *Quadrupole and Long range QM/MM* and *Gama calculation* are for expert users who want to use the formalism developed in the following articles.
- *How to read results* sums up some tools and tips to open and analyze FROG results.
- *Molecular Type Module* presents how to define a new Molecular Type to treat a new type of molecule within FROG.

Finally, in order to run these tutorials, MD trajectories are available, as the input files for FROG. In the case where QM/MM calculations are needed, the results are also provided so that the user can go throughout all the procedures even without a cluster. A Jupyter Notebook files associated to each tutorials are also provided to help new users to read and analyze FROG results.

Another important point to mention is the possibility for new users to develop FROG for their own use. One key feature is the definition of a so-called "Molecule Type": this object will define how FROG should deal with a molecule at a Quantum level (how many electrons, which charges for the nuclei...) or as an electrostatic environment (what charge distribution to use). Today, few Molecule Types are defined: I have worked on a reduced number of molecules during my thesis, and I have tried for the sake of examples just a few more. Hence, if a new user wants to work with new molecules without a corresponding molecule type, he can quite easily. One would have to define a new Molecule Type in a library directory: several pages of the wiki and a tutorial are devoted to this new definition procedure. The code, the wiki and the tutorials are available on Zenodo. We will soon move the code into a real open access mode. FROG is currently under the Lesser GPL copyright.

#### Strengths:

To conclude this section, the following items present some of the FROG functionalities and accomplishments so far:

- FROG can work with several MD trajectories thanks to the MDAnalysis package [108]. We confirmed that it works with LAMMPS and NAMD.
- The composition of the liquid can be complex: as for any MD software one has to define for every molecule its "type of molecule" to deal with mixtures. Today the list of molecules already implemented FROG is not large, water, methanol, some ions, but we made it simple to extend.
- FROG is designed to perform QM/MM calculations for the SHG properties, but it can be used for other optical properties (for instance polarizability). Moreover, we have added some structural analysis such as the H-bond number or the electric field felt by a molecule from its neighborhood.
- The electrostatic embedding framework is user-friendly and works with liquid mixture in the bulk and at liquid-gas interface. We are currently working on more advanced embedding scheme: cluster of several molecules or polarizability of the environment, see Appendix IX. B.1. We have also implemented the implicit long-range inclusion of the environment as described in the following article.
- FROG has several ways to define on which molecule the QM/MM calculations should be performed. For instance, at the liquid-gas interface we are interested only in the molecules at the "surface". Hence, using the Pytim package [109], we can define the set of molecules of interest using the layer definition. We can compute for instance the hyperpolarizabilities of the molecules located only at the last liquid layer: close to the surface edge.
- Many analysis tools have been developed especially for interfaces. For instance, FROG returns the averaged value of many quantities with respect to the molecule altitude.

## Limitations:

However, there are still some limitations:

- FROG would still be difficult to run for a new user. We are working on the tutorials and wiki to improve this.
- Even if very similar QM/MM calculations have to be performed, for instance the response of water molecules in the bulk phase, the QM calculations start from scratch. It would be better to start with a reference calculation to speed up the Ground State calculation.

• Today the QM calculations can be performed only in a limited number of ways (for instance only in DFT). Yet, it would not require a huge amount of coding to implement new options.

# IV. C | How to include the long-range electrostatic effect on $\beta$ ?

Ce travail a été publié en 2022 dans la revue PCCP.

#### Résumé et mise en contexte:

Afin de décrire la Génération de Second Harmonique à l'échelle moléculaire, le tenseur d'hyperpolarisabilité du premier ordre dipolaire  $\beta(2\omega, \omega, \omega)$  est couramment utilisé. Dans le cas dit non résonant, cette réponse peut être fortement impactée par l'environnement électrostatique, par exemple pour l'eau en phase liquide. Historiquement, ces changements ont été analysés à l'aide du tenseur de second hyperpolarizabilité dipolaire,  $\gamma(2\omega, \omega, \omega, 0)$ , qui peut être estimé expérimentalement grâce à des techniques comme l'EFISHG (Génération de Second Harmonique induite par un Champ Électrostatique). Ainsi, en première approximation, l'hyperpolarizabilitée d'une molécule,  $\beta^{env}$  lorsqu'elle est soumise à un champ externe statique,  $e^{env}$ , serait:

$$\beta^{env}(2\omega,\omega,\omega) = \beta^{GS}(2\omega,\omega,\omega) + \gamma(2\omega,\omega,\omega,0) \cdot \mathbf{e}^{env}$$
(IV.2)

Avec  $\beta^{GS}$  une valeur de référence, par exemple en phase gaz.

Cette étude comporte 2 messages principaux. Le premier est le calcul de  $\gamma(2\omega, \omega, \omega, 0)$  en phase liquide. Pour ce faire, nous utilisons et FROG l'approche QM/MM décrite Il est à noter que le calprécédemment. cul de  $\gamma(2\omega, \omega, \omega, 0)$  a été réalisé en utilisant un formalisme mixte de réponse et de Finite Field (FF), voir Section SI.1.1. Nous observons que cette réponse purement électronique est plus faible en phase liquide qu'en phase gaz, et que ce tenseur présente peu de fluctuations d'une molécule à l'autre, contrairement à l'hyperpolarisabilité. Aussi, on peut raisonnablement assigner à chaque molécule la même valeur de  $\gamma(2\omega, \omega, \omega, 0)$ , contrairement à la première hyperpolarisabilité.



Figure IV.5: Schéma de principe de la séparation de l'environnement électrostatique autour d'une molécule cible (en rouge) en "direct" (en bleu) et "long" (en vert). Les voisins situés avant une distance  $R_c$  sont considérés comme proches et sont inclus explicitement, les voisins plus lointains que  $R_c$  sont inclus implicitement. Dans cette situation l'environnement total est une sphère de rayon  $R_f = 40$  Å mais à priori cette démarche serait valable dans un environnement infini.

La seconde partie de l'article discute des conditions de l'utilisation de l'Équation IV.2 pour tenir compte de l'environnement électrostatique en phase condensée dans la détermination de l'hyperpolarisabilité  $\beta$ . En effet, si il était possible d'appliquer cette équation pour prendre en compte l'environnement électrostatique autour de chaque molécule de la phase condensée (à la place de notre approche QM/MM) nous aurions accès aux hyperpolarizabilités moléculaires sans calculs quantiques. Mesurer le champ créé par les autres molécules ( $e^{env}$ ) serait suffisant. Hélas, nous montrons dans que cette équation n'est pas valable. La raison principale est que le champ créé par la phase condensée est très hétérogène spatialement. Aussi, il n'est pas possible d'appliquer directement  $\gamma(2\omega, \omega, \omega, 0)$  qui par définition correspond à la déviation de  $\beta$  sous l'effet d'un champ spatialement homogène.

Néanmoins, nous pouvons tout de même utiliser ce formalisme pour prendre en compte l'environnement autour d'une molécule si celui-ci est suffisamment éloigné. En effet, à de grandes distances le champ créé par les autres molécules devient de plus en plus homogène spatialement. Pour ce faire, nous avons décomposé l'environnement autour d'une molécule d'intérêt en deux parties, voir Figure IV.5. Les voisins sont inclus différemment suivant leur distance par rapport à un rayon de référence  $R_c$  pour prédire l'hyperpolarizabilité de la molécule d'intérêt  $\beta^{PE+L}$  dans cet environnement:

$$\beta^{PE+L}(2\omega,\omega,\omega) = \beta^{PE}(2\omega,\omega,\omega)(R_c) + \langle \gamma(2\omega,\omega,\omega,0) \rangle \cdot \Delta \mathbf{e}(R_c).$$
(IV.3)

 $\beta^{PE}(R_c)$  est le résultat de l'approche QM/MM lorsque l'on inclut l'environnement jusqu'à une distance de  $R_c$ , typiquement jusqu'à 1 nm.  $\langle \gamma \rangle$  est la valeur moyenne de l'hyperpolarizabilité de second ordre en phase liquide, et  $\Delta \mathbf{e}(R_c)$  le champ crée par l'environnement à partir de  $R_c$ . En d'autres termes, pour calculer l'hyperpolarizabilité en phase condensée, il faut prendre en compte explicitement l'environnement jusqu'à une distance  $R_c$  ( $\beta^{PE}(R_c)$ ), le reste de l'environnement pouvant être pris en compte en utilisant le champ total créé au niveau de la molécule d'intérêt ( $\Delta \mathbf{e}(R_c)$ ). Nous avons comparé les résultats de cette approche pour différentes valeurs de  $R_c$  dans les Figures 3, 4 et 5 de l'article et obtenu d'excellents résultats pour  $R_c \approx 10$  Å.

Cette décomposition permet de prédire l'hyperpolarisabilité  $\beta$  de molécules en réduisant la taille de l'environnement explicite utilisé dans les calculs QM/MM. Néanmoins, elle nécessite de connaitre  $\gamma$ , ce qui est couteux numériquement et surtout peut évoluer dans certaines situations (par exemple à l'interface gaz-liquide). Aussi, nous avons poursuivi cette approche dans la dernière partie de ce travail en considérant l'effet des voisins éloignés au niveau de l'Hamiltonien de la molécule:

$$\hat{H}_{PE+LI} = \hat{H}_0 + \hat{V}_{Rc} - \hat{\mu} \cdot \Delta \mathbf{e}(R_c) \tag{IV.4}$$

On retrouve dans les deux premiers termes l'approche QM/MM déjà utilisée où l'environnement

#### Page 110

électrostatique est inclus explicitement jusqu'à une distance  $R_c$  (le potentiel électrostatique  $\hat{V}_{Rc}$ ). Le dernier terme correspond à un champ électrostatique en approximation dipolaire, c'est-à-dire un champ constant à l'échelle de la molécule de valeur  $\Delta \mathbf{e}(R_c)$ . Ce champ "externe" est celui créé par l'ensemble de l'environnement électrostatique à partir d'une distance  $R_c$ . Dans cette approche l'environnement à partir de  $R_c$  est inclus directement dans le degré de liberté électronique, alors que dans l'approche précédente l'environnement lointain était inclus dans  $\beta$ . Les résultats obtenus avec cette méthode sont présentés dans la dernière Figure de ce papier. Nous obtenons de meilleurs résultats à la fois en termes de précision et de cout numérique en travaillant à l'échelle de l'Hamiltonien.

Cette étude montre donc les limites de l'utilisation de la seconde hyperpolarizabilité pour prendre en compte l'environnement électrostatique. Néanmoins, elle amène aussi d'autres pistes pour calculer de façon plus efficace la réponse SHG de systèmes dans laquelle il existe des champs macroscopiques (par exemple des interfaces chargées). Dans ce cas, on peut s'attendre à devoir aussi décomposer l'environnement en deux parties: l'une pour laquelle l'inclusion explicite doit être faite (lorsque les gradients de champs sont importants), et une autre où l'on pourra utiliser la seconde hyperpolarizabilité. Nous avons implémenté dans FROG la dernière technique qui ajoute le champ des voisins lointains dans l'Hamiltonien.

# PCCP

# PAPER

Check for updates

Cite this: Phys. Chem. Chem. Phys., 2022, 24, 19463

Received 17th February 2022, Accepted 30th June 2022

DOI: 10.1039/d2cp00803c

rsc.li/pccp

# 1 Introduction

Non-Linear Optical (NLO) technique(s) are increasingly used to probe structural properties of matter. In Second Harmonic Generation (SHG) or Hyper-Rayleigh Scattering (HRS), two photons with the same fundamental frequency interact with a nonlinear material that generates a new photon with twice the energy of the initial photons.<sup>1</sup> SHG-based technologies have been developed to investigate simple and complex liquids,<sup>2,3</sup> biomimetic systems,<sup>4-6</sup> or even biological materials.<sup>7,8</sup> One interesting property of non-resonant SHG is the sensitivity of the response to the electrostatic environment.<sup>9</sup> Historically, the prototypical application is the electric-field-induced second harmonic generation (EFISHG) of molecules in a gas phase.<sup>10</sup> During such experiments, a macroscopic electrostatic field  $\mathbf{E}^{DC}$ is applied and the SHG response of the system is described using second and third order susceptibility tensor:  $\chi^{(2)}(2\omega, \omega, \omega)$ and  $\chi^{(3)}(2\omega, \omega, \omega, 0)$ , respectively. For a mesoscopic unitary

University of Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-69622, Villeurbanne, France. E-mail: claire.loison@univ-lyon1.fr † Electronic supplementary information (ESI) available: Methodological details; summetry, avarages and standard devistions of water second humarpolariashility.

# First hyperpolarizability of water in bulk liquid phase: long-range electrostatic effects included *via* the second hyperpolarizability<sup>†</sup>

Guillaume Le Breton, (D) Oriane Bonhomme, (D) Emmanuel Benichou (D) and Claire Loison (D) \*

The molecular first hyperpolarizability  $\beta$  contributes to second-order optical non-linear signals collected from molecular liquids. For the Second Harmonic Generation (SHG) response, the first hyperpolarizability  $\beta(2\omega, \omega, \omega)$  often depends on the molecular electrostatic environment. This is especially true for water, due to its large second hyperpolarizability  $\gamma(2\omega, \omega, \omega, 0)$ . In this study we compute the electronic  $\gamma(2\omega, \omega, \omega, 0)$  and  $\beta(2\omega, \omega, \omega)$  for water molecules in liquid water using QM/MM calculations. The average value of  $\gamma(2\omega, \omega, \omega, 0)$  is smaller than the one for the gaz phase, and its standard deviation among the molecules is relatively small. In addition, we demonstrate that the average bulk second hyperpolarizability  $\langle \gamma(2\omega, \omega, \omega, 0) \rangle$  can be used to describe the electrostatic effects of the distant neighborhood on the first hyperpolarizability  $\beta(2\omega, \omega, \omega)$ . In comparison with more complex schemes to take into account longrange effects, the approximation is simple, and does not require any modifications of the QM/MM implementation. The long-range correction can be added explicitly, using an average value of  $\gamma$  for water in the condensed phase. It can also be easily added implicitly in QM/MM calculations through an additional embedding electric field, without the explicit calculation of  $\gamma$ .

volume, the induced total dipole moment at the second harmonic frequency  $P^{2\omega}$  expressed as a sum of two terms:

$$\mathbf{P}^{2\omega} \propto \chi^{(2)}(2\omega, \, \omega, \, \omega): \mathbf{E}^{\omega} \mathbf{E}^{\omega} + \chi^{(3)}(2\omega, \, \omega, \, \omega, 0): \mathbf{E}^{\omega} \mathbf{E}^{\omega} \mathbf{E}^{\mathrm{DC}},$$
(1)

where  $\mathbf{E}^{\omega}$  is the exciting electromagnetic field at the fundamental frequency, and  $\mathbf{E}^{DC}$  is the external electrostatic field that potentially modifies (i) the orientation of dipoles, and (ii) the electronic-dependent molecular intrinsic SHG response. More recently, EFISHG has received renewed interest as a nonintrusive tool for probing electric fields in diluted media with spatial and temporal definition, using laser pulses.<sup>11</sup>

Eqn (1) was also applied to condensed phases,<sup>12,13</sup> and the EFISHG is an established technique to determine the first and the second hyperpolarizabilities of compounds in solutions.<sup>14</sup> Even more, studies on liquid/solid or liquid/air interfaces have reported the evolution of the Surface-SHG (S-SHG) response when the surface charge is modulated. Frameworks based on eqn (1) for the fluid nearby the surface have been commonly used in the S-SHG community to extract a surface potential or an effective surface charge.<sup>15–19</sup> But the interpretation of the different terms in the S-SHG intensity generated by aqueous solutions is still the subject of many recent works<sup>20–23</sup> requiring theoretical calculations at the molecular level.<sup>24</sup>

View Article Online View Journal | View Issue

symmetry, averages and standard deviations of water second hyperpolarizability in the liquid phase; heterogeneity of the electrostatic field generated by the neighborhood; impact of second hyperpolarizability fluctuations on the  $\gamma$  long correction. See DOI: https://doi.org/10.1039/d2cp00803c

To provide an interpretation at the microscopic scale, eqn (1) can also be rewritten at the molecular level:

$$\mu^{2\omega} = \frac{1}{2!} \beta^{(2)}(2\omega, \omega, \omega) : \mathbf{e}^{\omega} \mathbf{e}^{\omega} + \frac{1}{2!} \gamma^{(3)}(2\omega, \omega, \omega, 0) : \mathbf{e}^{\omega} \mathbf{e}^{\omega} \mathbf{e}^{\mathrm{DC}}$$
(2)

where  $\mu^{2\omega}$  is the induced dipole moment of the molecule at the second harmonic frequency,  $\mathbf{e}^{\omega}$  the exciting electromagnetic field in the molecular frame,  $\beta^{(2)}(2\omega, \omega, \omega)$  and  $\gamma^{(3)}(2\omega, \omega, \omega, 0)$  are respectively the first and second molecular hyperpolarizabilities, noted  $\beta$  and  $\gamma$  later on. The second hyperpolarizability  $\gamma$  describes the variations of the first hyperpolarizability value under a static external electrostatic field  $\mathbf{e}^{\text{DC}}$ , present in the molecular frame. This approach permits to link the EFISHG macroscopic measurements for the gas phase to quantum calculations of the molecular first hyperpolarizability  $\beta$ .<sup>25–30</sup>

The application of field expansions such as eqn (2) for condensed phases is very useful for the interpretation of S-SHG of aqueous solutions, and the values of  $\beta$  and  $\gamma$  for water in liquid water become parameters of the equation. To obtain values of  $\beta$  and  $\gamma$  for water in liquid water, one can use Quantum Mechanical/Molecular Mechanic (QM/MM) calculations.<sup>9,24,26,31,32</sup> In such approaches, the electrostatic effect of the environment is included within the QM calculations of a molecular hyperpolarizability  $\beta^{env}$  for a given electrostatic environment, so that

$$\mu^{2\omega} = \frac{1}{2!} \beta^{\text{env}} : \mathbf{e}^{\omega} \mathbf{e}^{\omega}, \qquad (3)$$

where  $\beta^{\text{env}}$  depends on the electric field exerted by the environment  $\mathbf{e}^{\text{env}}$  (T convention<sup>33</sup>). Using the values of  $\beta^{\text{env}}$ , one can also calculate  $\gamma^{\text{env}}$  from its definition as  $\gamma_{ijkl} = \frac{\delta \beta_{ijk}}{\delta e_l}$ , where  $\beta_{ijk}$  ( $e_l$ ) is the *ijk*-component of the first hyperpolarizability of a molecule on which an extra electrostatic field  $e_l$  is applied along the *l* molecular direction. In the present work, we first provide the  $\beta^{\text{env}}$  and  $\gamma^{\text{env}}$  tensors of individual water molecules in the liquid phase using QM/MM approaches at an optical wavelength of 800 nm (typically used for experiments<sup>34,35</sup>). Relatively to previous works, <sup>9,26,32</sup> we detail all the components of both tensors, and discuss the distribution of  $\gamma$  among water molecules.

In a second step, possessing a database of { $\beta^{\text{env}}$ , $\gamma^{\text{env}}$ } for individual molecules in liquid water, we can additionally scrutinize the validity of the field perturbation approach represented by eqn (2). Similar field expansions for the calculation of solvent effects have been tested for observables such as nuclear magnetic shieldings,<sup>36</sup> or linear and non linear optical properties.<sup>37</sup> However, the precision of the field expansion for the non-linear optical properties were limited,<sup>37</sup> and more terms are needed in the perturbation expression.<sup>38</sup> Concerning the first hyperpolarizability of water, Liang *et al.*<sup>9</sup> have compared  $\beta^{\text{env}}$  of water within the liquid phase obtained by QM/MM calculations to the linear field expansion of eqn (2), where  $\mathbf{e}^{\text{DC}}$  was the electrostatic field generated by the surrounding water molecules, noted  $e^{env}$ , and  $\beta$  and  $\gamma$  were the gas-phase values: a quantitative difference emerged. It is therefore expected that the strong solvation of water molecules within liquid water cannot be simplified to eqn (2). Nevertheless, in this work, we show that the perturbation field Taylor expansion can be useful to predict long-range electrostatic effects on the hyperpolarizability of water molecules in their liquid phase.

In the following, Section 2 describes the numerical details. In Section 3, the individual values of  $\gamma^{\text{env}}$  and  $\beta^{\text{env}}$  are first reported. Then, we show that the electrostatic field generated by the environment, namely  $e^{\text{env}}$ , is strongly heterogeneous in space. Based on this result, we propose to separate the effects of the electric field generated by the environment into a short- and a long-range part. Using this separation, we show that, for a water molecule in the bulk phase, the effect of the long-range electrostatic environment on  $\beta^{\text{env}}$  can be included using a linear correction proportional to  $\gamma$ .

# 2 Method

The calculations are based on a sequential QM/MM approach in two steps. First, we use a classical Molecular Dynamics (MD) simulation of bulk water to obtain typical structures of the liquid. Then, we compute the first and second hyperpolarizabilities of individual water molecules at the Density Functional Theory (DFT) level, within an electrostatic embedding framework. To investigate fluctuations due to the changing environment, statistics are performed over numerous configurations of the MD simulation.

#### 2.1 Molecular dynamics

LAMMPS,<sup>39</sup> V.11.08.2017 is used to perform the MD simulation along with the rigid TIP4P/2005 water force field.<sup>40</sup> 15625 rigid TIP4P/2005 water molecules are placed in a simulation box (approx 7.8 × 7.8 × 7.8 nm<sup>3</sup>) to form a 3D-periodic bulk system. We have used an isothermal, isobaric ensemble (NPT) with Nose–Hoover thermostat at 300 K ( $\tau_{\rm T}$  = 0.4 ps) and Nose–Hoover barostat at 1 atm ( $\tau_{\rm P}$  = 2 ps).<sup>41</sup> 1 ns of equilibration is performed before the 1 ns production run, both with a time step of 2 fs. Both electrostatic and Lennard-Jones intermolecular interactions are computed using the long-range Particle-Particle-Particle-Mesh (PPPM) formalism.<sup>42,43</sup> Neighbor lists are updated every time step within a radius of 10 Å. This simulation leads to a density of about 0.996 kg L<sup>-1</sup>. Such a large system was necessary to investigate environmental effects up to 40 Å.

#### 2.2 First hyperpolarizability in the liquid phase

To compute the first hyperpolarizability of water molecules within the liquid phase, an explicit environment composed of point charges is used: the Polarizable Embedding model at the zero<sup>th</sup>-order (PE0),<sup>44,45</sup> implemented in the DALTON software,<sup>46</sup> release 2018.2 package. As in our previous recent work,<sup>32</sup> the QM calculations were carried out on individual water molecules. Point charges represent the surrounding water molecules,

and the same electrostatic description is used as for the MD (TIP4P/2005 model). The MD trajectories are used directly in the QM calculations without further optimization, in contra. We define the parameter  $R_c$  as the maximal distance upon which neighbors are included in the PE formalism. The electrostatic field generated by this environment is spatially-heterogeneous in the vicinity of the target molecule and creates a new potential for the target molecule electronic Hamiltonian.

The QM/MM approach could be refined by optimizing geometries,<sup>47</sup> or by adding dipole, quadrupole or polarizability of the environment,<sup>45,48–51</sup> but we shall see in the results that eqn (2) becomes relevant only for long-range effects, where the point-charges electrostatic field is dominant. We have therefore chosen to keep the environment description as simple and robust as possible. To optimize parts of the QM/MM routines, the home-made software FROG is used.‡

The whole procedure is composed of the following steps: (1) perform MD simulation using LAMMPS. (2) Read the MD trajectories using FROG to build the electrostatic environment for each water molecule defined into a sphere of radius  $R_{\rm c}$ centered on the mass center of the target molecule. If a neighbor possesses at least one atom in the sphere, the whole molecule is considered in the environment. The output files for DALTON are created for 300 molecules for 8 selected MD configurations. (3) For each of the 2400 target molecules, calculate the electrostatic field, and its gradients, generated by the environment using FROG. (4) For each of the 2400 target molecules, calculate the value of  $\beta$  at the fundamental wavelength of 800 nm using DFT with the functional CAM-B3LYP<sup>52</sup> and the basis d-aug-cc-pVTZ,53 using the quadratic response formalism as implemented in DALTON.<sup>54</sup> (5) Read all DALTON output files and analyze the data using FROG.

The largest inaccuracy of our approach is due to the QM method and the corresponding functional. For the first hyperpolarizability of water molecules, the values in vacuum obtained with our method (DFT/CAM-B3LYP) are typically overestimated by about 10%32 in comparison to the golden standard for QM/MM calculation of small molecules (CCSD<sup>55-57</sup>). Both calculation schemes use the TIP4P/2005 rigid force field, the MM size  $R_c = 15$  Å and the basis is d-aug-pVTZ. The basis set d-aug-cc-pVTZ<sup>53</sup> used is widespread across the community<sup>26,55-57</sup> and presents a good balance between accuracy and cost. In a previous work,<sup>32</sup> we have observed a smooth convergence of the first hyperpolarizability relative to the basisset size for water in bulk water, and similar absolute values of the nuclear and electronic contributions to the electrostatic and polarization energies. Therefore, we estimate that spill-out effects are not dominant,<sup>50</sup> even if we use the large d-aug-ccpVTZ basis set. We have not taken into account the nuclear degree of freedom (vibrational effects) since Beaujean et al.<sup>58</sup> have shown that this effect is small for the second hyperpolarizability of water, especially at small wavelengths.

#### 2.3 Second hyperpolarizability in the liquid phase

In this work, we are interested in the evolution of the first hyperpolarizability under a static and homogeneous electrostatic field, which is reflected by  $\gamma(2\omega, \omega, \omega, 0)$  noted here  $\gamma$ . For each individual molecule, the second hyperpolarizability  $\gamma$  can be computed from the first hyperpolarizability  $\beta$  using the Finite Field (FF) framework. For that, we use the PE formalism described above to compute  $\beta$  when an extra homogeneous electrostatic field is applied, and then compute  $\gamma$  according to

$$\gamma_{ijkl} = \frac{\delta\beta_{ijk}}{\delta e_l} \simeq \frac{\beta_{ijk}(e_l) - \beta_{ijk}(0)}{e_l},\tag{4}$$

where  $\beta_{ijk}$  (e<sub>l</sub>) is the first hyperpolarizability of a molecule on which an extra static electrostatic field  $e_l$  is applied along the lmolecular direction. Numerically, the  $\gamma_{ijkl}$  is obtained for each molecule as the slope of a linear fit performed on seven calculations with  $e_l$  from 0 to  $1.5 \times 10^{-3}$  a.u. Noticeably, during the  $\gamma$  calculation, the target molecule is embedded in a total electrostatic field composed of a spatially-heterogeneous contribution from its PE environment plus the spatiallyhomogeneous  $e_l$ . More details regarding the FF procedure can be found in the ESI,† Section S1.1. The first and second hyperpolarizabilities are given in atomic units, our conventions are defined by eqn (2) and (4). They are expressed in the molecular frame  $\{a, b, c\}$  with the axis c along the dipole moment pointing from the Oxygen towards the Hydrogens, and the atoms within the  $\{a, c\}$  plane, see insert in Table 1. Noticeably, one could alternatively have used the response scheme to the third order, as implemented in Dalton, to obtain the second hyperpolarizability. For a given QM/MM conformation, the numerical cost of the response scheme is about ten times one of the finite field (FF) scheme, for similar results (see ESI,<sup>†</sup> Table S1). The comparison of the FF and the response

**Table 1** Water second hyperpolarizability  $\gamma^{(3)}$  (2 $\omega$ ,  $\omega$ ,  $\omega$ ,0) components, either in the vacuum ( $\gamma^{vac}$ ) or in its liquid phase ( $\gamma^{env}$ ) for an exciting wavelength of 800 nm, in atomic units. For the bulk phase, the average value ( $\gamma^{env}$ ) and the standard deviation  $\sigma$  [ $\gamma^{env}$ ] are reported. The tensors components are expressed in the molecular frame {*a*, *b*, *c*}, see inserted figure. Comparing with response calculations, the error on  $\gamma$  due to the finite field differentiation is estimated at about 20 a.u. (see ESI, Table S1)

	Vacuum	Liquid		
ijkl	$\gamma^{vac}$	$\langle \gamma^{\rm env} \rangle$	$\sigma[\gamma^{env}]$	
aaaa	1110	740	140	
bbbb	4090	2940	650	
сссс	2000	1400	230 a	
aabb	930	640	200	
bbaa	1060	710	150	
baab	1010	680	220 b	Ċ
abba	940	660	140	~
aacc	590	370	60	
ссаа	610	380	80	/
caac	610	380	60	<b>í</b>
асса	590	370	80	$\mathbf{N}$
bbcc	1120	770	170	$\mathbf{X}$
ccbb	1010	720	180	
cbbc	1010	730	150	•
bccb	1060	750	200	

This article is licensed under a Creative Commons Attribution 3.0 Unported Licence

<sup>&</sup>lt;sup>‡</sup> This code is deposited on Zenodo platform https://doi.org/10.5281/ zenodo.5998193, and available on demand to the authors.

schemes provides an estimation of the numerical error due to the FF procedure: about 5 to 20 a.u.

Frequency dispersion can be important for  $\gamma$ : for example, Beaujean *et.al.*<sup>58</sup> have shown a variation of about 25% between an excitation wavelength of 800 nm and at the infinite wavelength limit for  $\gamma(3\omega, \omega, \omega, \omega)$ . Here, even if we are using a Finite Field approach, the frequency dispersion of  $\beta$  related to the fundamental frequency (here 800 nm) is correctly reproduced since we use the first hyperpolarizability obtained using the frequency dependent response scheme.

#### 2.4 Embedding environment size: $R_c$

One of the objectives of this study is to understand the impact of the electrostatic environment on the first hyperpolarizability, distinguishing the effects of close and far neighbors. Therefore, different values of  $R_c$  are used for  $\beta$  calculations. In the following, the notation  $\beta^{\text{PE}}(R_c)$  denotes the hyperpolarizability obtained by a QM/MM calculation at a given  $R_c$  and  $e(R_c)$  the electric field generated by this PE environment. In this notation,  $R_c = 0$  corresponds to a gas phase calculation. According to previous works,32,57,59 a large environment (typically between 10 and 20 Å) is necessary to obtain a good convergence of the  $\beta$  values. Here, for each molecule, a reference calculation is performed using the PE formalism with an extremely large radius of  $R_{\rm f} = 40$  Å, and we consider that  $\beta^{\rm env} = \beta_{\rm PE}$  (40 Å). Similarly, we consider that the electrostatic field  $e^{env} = e$  (40 Å). Note that such a large radius still does not represent the complete MD liquid phase, described using periodic boundary conditions.

On the contrary, the variation of  $\gamma$  with the size of the environment is not investigated in details here. Therefore, the PE radius used to compute the second hyperpolarizability  $\gamma$  is set to 10 Å. Yet, we have verified that the  $\gamma$  does not evolve much using larger radius in agreement with Osted *et al.*<sup>59</sup> (data not shown) and have noted it  $\gamma^{\text{env}}$ .

#### 2.5 Statistical averaging

The first and second hyperpolarizability in the liquid phase are obtained from 2400 configurations of the PE environment. These configurations are extracted from 8 MD snapshots, with 300 water molecules randomly selected in each frame. These frames are separated by 100 ps to ensure time decorrelation.<sup>32</sup> We have verified that this number of configurations is sufficient to reach convergence; indeed, the numerical error due to the configuration number is about 1% (see ESI,† Fig. S2 and S3). This is smaller than the error due to the DFT Finite field calculation relative to CCSD calculations (see ESI,† Tables S2 and S3). The  $C_{2v}$  symmetry of the first and second hyperpolarizability obtained by QM/MM confirms this good convergence (see the Section 3 and ESI,† Section S2.1).

# **3** Results

In the following, Section 3.1 first presents the second hyperpolarizability tensor  $\gamma$  calculated using QM/MM approaches for single water molecules within the liquid bulk phase, and the standard deviation of the calculated values. Then, Section 3.2.1 presents some difficulties encountered when applying eqn (2) for condensed phases. We attribute them to the properties of the electrostatic field generated by the environment,  $e^{\text{env}}$ , described in 3.2.2. Based on these results, and on eqn (2), in Section 3.2.3, we propose a correction to  $\beta$  that permits to include long-range interactions. Finally, we show that this correction enhances the precision of standard QM/MM calculations.

# 3.1 QM/MM results: second hyperpolarizability of water in the bulk phase

Table 1 presents the values of selected  $\gamma$  components calculated either in the vacuum or in liquid phase – the other non-zero components of  $\gamma$  are presented in ESI,† Table S4.

For the vacuum phase, the molecular  $C_{2\nu}$  symmetry is fulfilled: the only non-vanishing components are the  $\gamma_{iiii}$  and  $\gamma_{iijj}$  for *i* and *j* in {*a*, *b*, *c*}, plus their permuted terms<sup>1</sup> (see ESI,† Table S4). In the liquid phase, this symmetry is not valid for individual values, but is recovered for the average  $\langle \gamma^{\text{env}} \rangle$ . Our results agree with the ones by Osted *et al.*<sup>59</sup> within about 15%. We impute the discrepancies to both the difference of methods (DFT *vs.* CCSD, different MD models) and the difference of excitation wavelengths (800 *vs.* 1080 nm). More detailed comparisons with literature are provided in ESI,† Section S1.1.

Noticeably, when transferred into the laboratory frame, the average second hyperpolarizability becomes centro-symmetric (see ESI,† Table S5), indicating that our sampling of molecular orientation in the liquid phase is sufficient. Interestingly, the average value of all  $C_{2v}$ -authorized  $\gamma^{env}$  components are positive. Moreover, in the liquid phase, the  $\gamma^{env}$  components are about 30% smaller than the ones in the vacuum – see ESI,† Fig. S4.

Beyond the average values, the environment induces fluctuations of  $\gamma^{\text{env}}$ , and of  $\beta^{\text{env}}$ . A typical and relevant case is illustrated by Fig. 1, displaying the join probabilities of  $\{\beta_{ccc},\gamma_{cccc}\}$  obtained for the 2400 water molecules within their environment. Projections on the axes display the distributions of  $\beta_{ccc}$  and  $\gamma_{cccc}$  separately. As already discussed in the literature, <sup>32,57,59</sup> the water first hyperpolarizability fluctuations in the liquid phase are strong, with standard deviations similar to the absolute values of the averages – for instance  $\langle \beta_{ccc}^{\text{env}} \rangle = 4.1$  a.u. and  $\sigma [\beta_{ccc}^{\text{env}}] = 3.1$  a.u. For individual molecules, some  $\beta$  components, null in average, can be larger than the one with a net average (see ESI,† Fig. S4).

In complement to the results published by Osted *et al.*,<sup>59</sup> we show that the second hyperpolarizability  $\gamma$  also fluctuates, but in a smaller extent than  $\beta$ : the standard deviations of the different components  $\sigma$  [ $\gamma^{\text{env}}$ ] are about 20% of their average values  $\langle \gamma^{\text{env}} \rangle$  – see Fig. 1 for  $\gamma_{cccc}$ , and ESI,† Table S4.

Fig. 1 moreover shows that the values of  $\beta$  and  $\gamma$  are not strongly correlated. Hence, in the following, we neglect the  $\gamma^{\text{env}}$  dispersion and attribute the same second hyperpolarizability  $\langle \gamma^{\text{env}} \rangle$  to all the water molecules.

#### 3.2 Using $\gamma$ to calculate individual $\beta$ in condensed phases

After describing the second hyperpolarizability  $\gamma^{env}$  in the liquid phase, we will question if it can be used to predict the



**Fig. 1** Join population of  $\{\beta_{ccc}^{env}, \gamma_{cccc}^{env}\}$  obtained by QM/MM approaches for bulk liquid water. The independent ones are presented on the top and right for the  $\beta_{ccc}^{env}$  and  $\gamma_{cccc}^{env}$  respectively. The values are normalized with respect to their average ones:  $\langle \beta_{ccc}^{env} \rangle = 4.1$  a.u. and  $\langle \gamma_{cccc}^{env} \rangle = 1400$  a.u.

first hyperpolarizability of water molecules in the liquid phase using different approximations inspired by eqn (2). We shall compare the  $\beta^{\text{env}}$  obtained by QM/MM to linear expansions relative to a perturbing electric field.

**3.2.1** Limitations of the gas-phase reference model. Comparing eqn (2) and (3), one straightforward approach to include the effect of the environment on the first hyperpolarizability  $\beta$  is to use the second hyperpolarizability  $\gamma$  and the field created by the environment  $e^{env}$ :

$$\beta^{\rm GR} = \beta^{\rm vac} + \gamma^0 \cdot \mathbf{e}^{\rm env},\tag{5}$$

where a linear perturbation is done relative to the Gas-phase Reference (GR), for which  $e^{env} = 0$  and and the reference first and second hyperpolarizability would be the one of the vacuum for instance. For pedagogical reasons, we will first highlight here the limitation of this gas-phase reference model, that was already mentioned in the literature.<sup>36,57</sup>

If the linear proportionality factor  $\gamma^0$  is known, and the same for all molecules, this approach would be very helpful: the liquid first hyperpolarizability depends only on  $e^{env}$ . An MD simulation can provide  $e^{env}$  for each molecule, and the liquid phase first hyperpolarizability is calculated using eqn (5), so that no expensive QM/MM approach is needed. Moreover, such an approximation is commonly used in the Surface-SHG community (for example ref. 18 and 19).

Liang *et al.*<sup>57</sup> have reported that eqn (5) using the vacuum value for the second hyperpolarizability  $\gamma^0 = \gamma^{\text{vac}}$  is a poor estimator for  $\beta^{\text{env}}$ . We have thus applied eqn (5) using the average bulk second hyperpolarizability tensor  $\gamma^0 = \langle \gamma^{\text{env}} \rangle$  for all the molecules (values from Table 1 and from ESI,† Table S4), and the electrostatic field generated by the environment  $\mathbf{e}^{\text{env}}$ . For liquid water, the distribution of the largest component of  $\mathbf{e}^{\text{env}}$ , along the water molecular axis *c* is represented on Fig. 2: it has typical values around  $3 \times 10^{-2}$  a.u., *i.e.* 1.6 V Å<sup>-1</sup>. The other



**Fig. 2** Distribution of the electrostatic field generated by the environment  $e^{env}$  along the molecular *c*-axis. The electrostatic field is measured at 2 positions: at the oxygen atom position (O: orange squares), or at the negative charge one (T: blue dots) according to the TIP4P/2005 force field. The O and T positions are distant by 0.15 Å.

components average to zero, with approximately the same standard deviations (data not shown).

The predicted average values  $\langle \beta^{\text{GR}} \rangle$  are compared to the QM/ MM values  $\langle \beta^{\text{env}} \rangle$  in Table 2. First, the addition of the linear correction in eqn (5) allows recovering signs for  $\beta^{\text{GR}}$  components in agreement with  $\beta^{\text{env}}$ . However, the  $\beta^{\text{GR}}$  model largely over-estimates the reference values: the correction  $\gamma^0 \cdot \mathbf{e}^{\text{env}}$ should be smaller. This overshooting is even worse when the vacuum value  $\gamma^{\text{vac}}$  is used for  $\gamma^0$ , as Liang *et al.* had tested.<sup>9</sup> Thus, even if it contains some relevant physics, eqn (5) describing a correction that is proportional to  $\mathbf{e}^{\text{env}}$  is not workable: it predicts an average hyperpolarizability  $\langle \beta^{\text{GR}} \rangle$  one order of magnitude too large.

Moreover, the value of  $\mathbf{e}^{\text{env}}$  in eqn (5) is ill-defined because of its strong spatial heterogeneity. To highlight this fact, Fig. 2 presents the electrostatic field created by the environment along the molecular *c*-direction,  $\mathbf{e}_c^{\text{env}}$ , calculated at 2 positions within the embedded molecule: either at the position of the negative charge of the TIP4P/2005 model, or at the position of the oxygen atom. While the 2 distributions look very similar, a difference on the average values of about  $10^{-3}$  a.u. appears. Given the large value of  $\gamma_{cccc}$  ( $\approx 1400$  a.u., see Table 1), the component  $\beta_{ccc}$  predicted by  $\beta^{\text{GR}}$  changes by about 40%, depending on where  $\mathbf{e}^{\text{env}}$  is calculated. Therefore, we attribute

**Table 2** Comparison between relevant QM/MM components of the first hyperpolarizability,  $\beta^{env}$ , the vacuum ones,  $\beta^{vac}$ , and  $\beta^{GR}$  predicted by eqn (5) with  $\gamma^0 = \langle \gamma^{env} \rangle$  from Table 1. All values are in atomic units and are calculated for fundamental wavelength of 800 nm. The values of  $\mathbf{e}^{env}$  obtained at the center-of charge position have been used

ijk	$\langle \beta^{ m vac}  angle$	$\langle eta^{ ext{env}}  angle$	$\langle \beta^{ m GR}  angle$
ссс	-15.3	4.1	28.5
саа	-12.5	-2.0	-0.6
aca	-12.4	-2.0	-0.8
cbb	-5.0	2.5	17.7
bcb	-7.4	2.2	16.9

the failures of eqn (5) to the strong intensity and large gradient of the embedding field  $e^{env}$ , which are incompatible with a simple linear expansion. Such limitations of perturbation field expansions to model solvation effects had already been discussed in the literature. When studying the nuclear magnetic shielding tensors of the atoms of *N*-methylacetamide solvated by liquid water, Kjaer *et al.*<sup>36</sup> have shown that the field experienced by the nuclei is very inhomogeneous, so that considering one averaged field from all configurations and/or across all nuclei is not a good approximation.

**3.2.2** Separation into short and long-range neighbors. To go beyond previous works,<sup>57,59</sup> we propose to rewrite the pertubation field expansion so that the perturbing electric field and electric field gradients are weak, and the linear regime is verified. The evolution of the electrostatic field  $\mathbf{e}(R_c)$  generated by the environment around the target molecule of the QM/MM calculations with the environment size  $R_c$  is presented on Fig. 3. The main contribution to this electrostatic field is created in the first and second solvation shells, at distances between 2 and 5 Å from the target molecule. This neighborhood region also creates the largest electrostatic field gradients, see ESI,<sup>†</sup> Fig. S6.

Given that the closest shells are responsible for both the intensity and the heterogeneity of  $e^{env}$ , we divide the neighborhood into two parts, as illustrated on Fig. 4.

The direct area contains the closest neighbors of the target molecule (at the center), up to a distance  $R_c$ : they create a strong and heterogeneous electric field, noted  $\mathbf{e}(R_c)$ . The long-range area, due to neighbors beyond  $R_c$  from the target molecules, creates a less intense and more homogeneous electric field on the target molecule.

The total embedding electrostatic field  $\mathbf{e}^{\text{env}}$  is thus separated into a short-range contribution  $\mathbf{e}(R_c)$ , and a correction,  $\Delta \mathbf{e}(R_c)$ due to the long-range electrostatic interactions:

$$\mathbf{e}^{\mathrm{env}} = \mathbf{e}(R_{\mathrm{c}}) + \Delta \mathbf{e}(R_{\mathrm{c}}). \tag{6}$$

15 20 25 30

20

25

30

 $R_c$  [Å]



15

 $R_c [Å]$ 



**Fig. 4** Sketch of the electrostatic embedding procedure. The QM box is defined for only one molecule of water at the center. Direct area, until  $R_c$ : all the neighbors are included in a QM/MM calculation with discrete solvation procedure, such as the PE formalism. 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.

where all the fields are evaluated at the position of the negative charge in the TIP4P/2005 model (T-position, see Fig. 2 inset). The evolution of  $\Delta \mathbf{e}$  ( $R_c$ ) as a function of  $R_c$  is available in the insert of Fig. 3. At  $R_c = 10$  Å, the electrostatic field generated by the long-range neighbors is very small compared to the one created by close neighbors: the average correction  $\Delta e (R_c = 10 \text{ Å})$ is close to  $10^{-3}$  a.u. The finite field calculations reported in Section 3.1 have shown that for such amplitudes of the perturbation, the dependence of the QM/MM  $\beta^{env}$  relative to an external, homogeneous field is linear (see ESI,† Fig. S1). Moreover, we have investigated whether the correction  $\Delta e(R_c)$  is spatially homogeneous by measuring the correction to the electric field gradient  $\Delta \frac{\delta e}{\delta x}(R_c)$  (see ESI,<sup>†</sup> Fig. S1). Generally, the corrections to the gradients are small (typically  $10^{-4}$  a.u) for neighbors further away than 10 Å. Since the correction  $\Delta \mathbf{e}(R_c)$  is more homogeneous than  $e^{env}$ , it hardly depends on the point on which it is calculated. In the following, we use the value of  $\Delta \mathbf{e}(R_{\rm c})$  evaluated at the T-position.

**3.2.3** Long-range correction to  $\beta$  using  $\gamma$ . The further the neighbor is from the target molecule, the better its contribution to  $e^{env}$  can be modeled as a weak and homogeneous electrostatic field. A linear  $\gamma$  correction to  $\beta$  could be tested to describe the long-range electrostatic effects.

For the first hyperpolarizability, as for  $e^{env}$ , we describe the impact of the environment in two different ways depending on the neighbors distance. The direct neighbors within the sphere of radius  $R_c$  are included explicitly in the QM/MM calculation through the PE approach. The long-range neighbors, beyond  $R_c$ , are included implicitly *via* the homogeneous field they produce on the target molecule,  $\Delta e(R_c)$ , with a correction

PCCP

3.0

2.5

2.0

1.5

1.0

0.5

0.0

Ó

< |e<sub>i</sub>| > × 100 [a.u.]

19468 | Phys. Chem. Chem. Phys., 2022, 24, 19463-19472

10

10

5



**Fig. 5** Comparison of the reference individual molecular  $\beta_{ccc}^{env}$  component and the one obtained using PE (TOP) and PE + L (BOTTOM) approximation with  $R_c = 5$  Å. The dashed line corresponds to the ideal case where a correction matches perfectly the expected value. Each dot represents a molecule.

proportional to  $\langle \gamma^{\text{env}} \rangle$ . This approximation can be viewed as a linear expansion relative to the value of first hyperpolarizability already perturbed by the direct neighbors,  $\beta^{\text{PE}}(R_c)$ :

$$\beta^{\rm PE+L}(R_{\rm c}) = \beta^{\rm PE}(R_{\rm c}) + \langle \gamma^{\rm env} \rangle \cdot \Delta \mathbf{e}(R_{\rm c}). \tag{7}$$

In the following, we compare this approximation of  $\beta$ , noted  $\beta^{\text{PE+L}}(R_c)$  and the usual QM/MM calculation where only the PE scheme up to a distance  $R_c$  is used, noted  $\beta^{\text{PE}}(R_c)$ . To estimate the accuracy of the two approximations PE and PE + L as a function of  $R_c$ , we compare them to our reference values  $\beta^{\text{env}}$ .

As a typical example, we compare on Fig. 5 the  $\beta_{ccc}$  computed with  $R_c = 5$  Å either using PE approximation (top) or the PE + L one (bottom) to the reference value  $\beta_{ccc}^{env}$ . The diagonal



**Fig. 6** Average of the  $\beta_{ccc}^{X}(R_c)$  component as a function of  $R_c$  for the PE (blue disks) and PE + L (orange squares, see eqn (5)) approximations. The dashed line is the reference value  $\beta^{env}$  obtained for  $R_f = 40$  Å.

dashed lines represent the case where the approximation reproduces exactly the reference.

For the usual calculation  $(\beta^{\text{PE}})$ , there is a systematic error leading to inaccuracy: at  $R_c = 5$  Å, the average value of  $\beta_{ccc}^{\text{PE}}$  is smaller than the one of  $\beta_{ccc}^{\text{env}}$ . Moreover, the error fluctuates a lot, it can reach tens of a.u., depending on the molecule. On the contrary, the PE + L approach almost quantitatively reproduces the reference, even for  $R_c$  as small as 5 Å

To illustrate the effect of  $R_c$ , Fig. 6 depicts the evolution of  $\langle \beta_{\rm PE} \rangle$  and  $\langle \beta_{\rm PE+L} \rangle$  as a function of  $R_c$  for the *ccc* component (similar trends are obtained for other components). As expected, the inaccuracy of the PE and PE + L approaches decreases when  $R_c$  increases.

When the correction is added, the convergence is reached faster ( $R_c = 4$  Å) compared to the usual  $\beta^{\text{PE}}$  calculation ( $R_c \simeq 12$  Å). The effect of the neighbors further than 5 Å can thus be very efficiently included by a second hyperpolarizability corrective term. Beyond the description of the average first hyperpolarizability, Fig. 5 also illustrates that the PE + L approach improves the precision of the individual predictions. This can be quantified through the mean absolute error (MAE), averaged over all the N = 2400 molecular configurations and all the 27 components, noted  $\Delta\beta_{\text{T}}$ :

$$\Delta \beta_{\rm T}^{\rm X} = \frac{1}{27N} \sum_{ijk} \sum_{n=1}^{N} \left| \beta_{ijk}^{\rm env(n)} - \beta_{ijk}^{\rm X(n)} \right| \tag{8}$$

where X stands for the approximation PE, or PE + L, and the exponent (*n*) indicates that the values of the *n*-th molecule are used. Fig. 7 reports the evolution of  $\Delta \beta_{T}^{X}$  as a function of  $R_{c}$ .

The long-range correction provides very good results and reaches an error below 0.1 a.u. at  $R_c = 12$  Å. To obtain the same degree of accuracy, one would need to include explicitly neighbors up to  $R_c \simeq 30$  Å in the usual PE scheme without long-range correction. For  $R_c$  larger than 10 Å, the error obtained with the long range correction (PE + L) is about one order of magnitude lower than the one with the traditional PE approach. This shows that – once the strong short-range effects are explicitly



**Fig. 7** Mean absolute error of the individual first hyperpolarizability tensor defined by eqn (8) as a function of  $R_c$  for the PE (blue disks), PE + L (orange squares) and PE + LI (green triangles) approximations. The dashed line corresponds to an error of 0.1 a.u. at which we define convergence.

included – it is possible to use the knowledge of  $\gamma$  to predict accurately the long-range electrostatic effects on  $\beta$ . Note that the long-range correction is proportional to the average value  $\langle \gamma^{\text{env}} \rangle$ , *i.e.* Eqn (9) neglects  $\gamma^{\text{env}}$  fluctuations. As discussed further in ESI,<sup>†</sup> Section S4, including the fluctuations of  $\gamma^{\text{env}}$ in eqn (9) only slightly improves the results. However, it implies to compute  $\gamma^{\text{env}}$  for every molecule, which greatly increases the numerical cost. Therefore, using an average  $\gamma^{\text{env}}$  is a very good compromise between computational cost and accuracy, at least for bulk liquid water.

**3.2.4 Implicit long-range correction.** The linear correction is validated here for water molecules in bulk liquid water. We consider this result as a first step towards more complex systems, anisotropic or charged ones, where electrostatic effects become even more important. In these new environments, the second hyperpolarizability of water might differ from the one of the bulk phase, and its fluctuations may become larger. With these next steps in mind, we show here that our explicit long-range correction can also be added to the system implicitly, during the QM calculation, *via* the addition of the homogeneous field  $\mathbf{e}(R_c)$  on the top of the heterogeneous MM embedding field. We note this approach PE + LI for Long-range Implicit correction, than can be noted

$$\beta^{\text{PE+LI}}(R_{\text{c}}) = \beta^{\text{PE}}(R_{\text{c}}; \mathbf{e}(R_{\text{c}})).$$
(9)

Fig. 7 illustrates that the PE + LI values of the first hyperpolarizability are very close to the reference values. The PE + LI approach is even more precise that the PE + L because (1) the second hyperpolarizability fluctuations are neglected in PE + L – not in PE + LI, and (2) the second hyperpolarizability is calculated with finite field approach in PE + L – not in PE + LI. The PE + LI correction does not require the knowledge of the second hyperpolarizability. It can be seen as a simplification of more elaborated QM/MM developments taking into account long-range environment in non-periodic QM/MM methods.<sup>60–63</sup> Indeed, long-range effects can be approximated by boundary potential models, by cutoff method coupled with switching and shifting functions, by the so-called Long-Range Electrostatic Correction, or by additional charges on the edge of the MM embedding, or in the MM embedding. Our simpler approach with a homogeneous electrostatic field is intended to provide the impact of the remote MM environment on the nonlinear optical property of the QM system. It is limited to sequential QM/MM approaches where the trajectory is obtained independently. The advantages of the PE + LI are (1) its insight in the physical origin of the correction and its link with usual equations used in the experimentalist community, and (2) its simplicity and low computational cost. Finally, the value of  $\mathbf{e}(R_c)$  could include periodic boundary conditions, but it is beyond the present work.

# 4 Conclusions

In this work, we have computed the first and second hyperpolarizability of water in the liquid phase using state-of-the-art QM/MM calculations. With this set of values, several approaches were tested to take into account the electrostatic environment in the first hyperpolarizability.

The second hyperpolarizability is high in the bulk phase. The neighborhood induces a dispersion on the second hyperpolarizability  $\gamma(2\omega, \omega, \omega, 0)$  values, but less than for the first hyperpolarizability.

Concerning the first hyperpolarizability  $\beta$ , linear expansions are valid for a weak and spatially-homogeneous electrostatic field, while the embedding electrostatic field created by the liquid surroundings is not. To compute individual molecular hyperpolarizability in the liquid phase, the molecular structure of the nearest neighbors have to be described explicitly, and QM/MM models are very relevant.

But long-range effects, up to several nanometers, can become heavy to take into account using QM/MM approaches, since a very large number of neighbors have to be included. This problem is expected to become all the more pregnant when the environment is anisotropic, with charges or polarization. We have proposed to consider the effect of long-range electrostatic environment on the liquid first hyperpolarizability using the average second hyperpolarizability  $\gamma$ . This long-term correction increases the precision of the hyperpolarizability calculation, and speeds up the convergence of the average value relative to the QM/MM environment size  $R_c$ . For pure bulk water, our approximation makes sense for a long-range part defined beyond 5 to 10 Å, depending on the quantity required. From a numerical point of view, the correction can be powerful for charged environments because the size of the explicit embedding region can be reduced drastically.

This work has demonstrated the promising potentialities of using the second hyperpolarizability, explicitly or implicitly to predict first hyperpolarizabilities. The present study is applied to pure water bulk phase at 300 K. However, the correction developed here should be also relevant for many other system geometries or compositions, where the long-range electrostatic effects can modify optical responses. If necessary, an extension including electric field gradients would also be straightforward using the Dalton package. Precision tests remain to be performed for more complex systems such as salted aqueous solutions, or charged interfaces.

# Author contributions

GLB: conceptualization, methodology, validation, writing. OB: supervision, validation, writing. EB: supervision, writing, review, editing. CL: supervision, validation, writing.

# Conflicts of interest

There are no conflicts to declare.

# Acknowledgements

We thank P.-F. Brevet for insightful discussions. We gratefully acknowledge support from the PSMN (Pôle Scientifique de Modélisation Numérique) of the ENS de Lyon for the computing resources.

# Notes and references

- 1 P.-F. Brevet, Surface Second Harmonic Generation, 1997.
- 2 M. C. Goh, J. M. Hicks, K. Kemnitz, G. R. Pinto, K. Bhattacharyya, K. B. Eisenthal and T. F. Heinz, *J. Phys. Chem.*, 1988, **92**, 5074–5075.
- 3 Q. Wei, D. Zhou and H. Bian, *Phys. Chem. Chem. Phys.*, 2018, 20, 11758–11767.
- 4 M. N. Nasir, E. Benichou, C. Loison, I. Russier-Antoine,
  F. Besson and P.-F. Brevet, *Phys. Chem. Chem. Phys.*, 2013, 15, 19919–19924.
- 5 G. Licari, J. S. Beckwith, S. Soleimanpour, S. Matile and E. Vauthey, *Phys. Chem. Chem. Phys.*, 2018, 20, 9328–9336.
- 6 E. Donohue, S. Khorsand, G. Mercado, K. M. Varney, P. T. Wilder, W. Yu, A. D. MacKerell, P. Alexander, Q. N. Van, B. Moree, A. G. Stephen, D. J. Weber, J. Salafsky and F. McCormick, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 17290–17297.
- 7 M. E. Didier, O. B. Tarun, P. Jourdain, P. Magistretti and S. Roke, *Nat. Commun.*, 2018, **9**, 1–7.
- 8 S. Bancelin, C. Aimé, I. Gusachenko, L. Kowalczuk, G. Latour, T. Coradin and M. C. Schanne-Klein, *Nat. Commun.*, 2014, 5, 1–8.
- 9 C. Liang, G. Tocci, D. M. Wilkins, A. Grisafi, S. Roke and M. Ceriotti, *Phys. Rev. B*, 2017, 96, 1–6.
- P. Kaatz, E. A. Donley and D. P. Shelton, *J. Chem. Phys.*, 1998, 108, 849–856.
- 11 J. Yang, E. V. Barnat, S.-k Im and D. B. Go, *J. Phys. D: Appl. Phys.*, 2022, 55(22), 225203.
- 12 B. Levine and C. Bethea, J. Chem. Phys., 1976, 65, 2429-2438.
- 13 T. N. Ramos, S. Canuto and B. Champagne, J. Chem. Inf. Model., 2020, 60, 4817–4826.

- 14 B. F. Levine and C. G. Bethea, J. Chem. Phys., 1975, 63, 2666–2682.
- E. C. Yan, Y. Liu and K. B. Eisenthal, J. Phys. Chem. B, 1998, 102, 6331–6336.
- 16 T. Joutsuka, T. Hirano, M. Sprik and A. Morita, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3040–3053.
- 17 S. Pezzotti, D. R. Galimberti, Y. R. Shen and M.-P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2018, **20**, 5190–5199.
- 18 L. Dalstein, K.-Y. Chiang and Y.-C. Wen, *J. Phys. Chem. Lett.*, 2019, **10**, 5200–5205.
- A. Marchioro, M. Bischoff, C. Lütgebaucks, D. Biriukov, M. Předota and S. Roke, *J. Phys. Chem. C*, 2019, 123, 20393–20404.
- 20 R. Hartkamp, A. L. Biance, L. Fu, J. F. Dufrêche, O. Bonhomme and L. Joly, *Curr. Opin. Colloid Interface Sci.*, 2018, 37, 101–114.
- 21 L. B. Dreier, C. Bernhard, G. Gonella, E. H. Backus and M. Bonn, J. Phys. Chem. Lett., 2018, 9, 5685–5691.
- 22 E. Ma, P. E. Ohno, J. Kim, Y. Liu, E. H. Lozier, T. F. Miller, H.-F. Wang and F. M. Geiger, *J. Phys. Chem. Lett.*, 2021, **12**, 5649–5659.
- 23 H. Chang, P. E. Ohno, Y. Liu, E. H. Lozier, N. Dalchand and F. M. Geiger, *J. Phys. Chem. B*, 2020, **124**, 641–649.
- 24 Y. Foucaud, B. Siboulet, M. Duvail, A. Jonchere, O. Diat, R. Vuilleumier and J.-F. Dufrêche, *Chem. Sci.*, 2021, 12, 15134–15142.
- 25 G. Maroulis, Chem. Phys. Lett., 1998, 289, 403-411.
- 26 J. Kongsted, A. Osted, K. V. Mikkelsen and O. Christiansen, J. Chem. Phys., 2004, 120, 3787–3798.
- 27 M. Hidalgo Cardenuto and B. Champagne, *Phys. Chem. Chem. Phys.*, 2015, **17**, 23634–23642.
- 28 T. Giovannini, M. Ambrosetti and C. Cappelli, *Theor. Chem. Acc.*, 2018, **137**(6), 74.
- 29 P. Beaujean and B. Champagne, J. Chem. Phys., 2019, 151, 064303.
- 30 T. N. Ramos, F. Castet and B. Champagne, *J. Phys. Chem. B*, 2021, **125**, 3386–3397.
- 31 L. Jensen and P. T. Van Duijnen, J. Chem. Phys., 2005, 123(7), 074307.
- 32 G. Le Breton, O. Bonhomme, P.-F. Brevet, E. Benichou and C. Loison, *Phys. Chem. Chem. Phys.*, 2021, 23, 24932–24941.
- 33 A. Willetts, J. E. Rice, D. M. Burland and D. P. Shelton, J. Chem. Phys., 1992, 97, 7590–7599.
- 34 O. Bonhomme, L. Sanchez, E. Benichou and P. Brevet, *J. Phys. Chem. B*, 2021, **125**, 10876–10881.
- 35 A. Pardon, O. Bonhomme, C. Gaillard, P.-F. Brevet and E. Benichou, *J. Mol. Liq.*, 2021, 322, 114976.
- 36 H. Kjaer, S. P. A. Sauer and J. Kongsted, J. Chem. Phys., 2011, 134, 44514.
- 37 Y. Luo, P. Norman, I. Introduction and H. Agren, *J. Chem. Phys.*, 1998, **109**, 3589–3595.
- 38 K. O. Sylvester-Hvid, K. V. Mikkelsen, P. Norman, D. Jonsson and H. Ågren, *J. Phys. Chem. A*, 2004, **108**, 8961–8965.
- 39 S. Plimpton, J. Comput. Phys., 1995, 117, 1-19.
- 40 J. L. Abascal and C. Vega, J. Chem. Phys., 2005, 123, 234505.
- 41 G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177–4189.

- 42 R. E. Isele-Holder, W. Mitchell and A. E. Ismail, J. Chem. Phys., 2012, 137.
- 43 R. E. Isele-Holder, W. Mitchell, J. R. Hammond, A. Kohlmeyer and A. E. Ismail, *J. Chem. Theory Comput.*, 2013, **9**, 5412–5420.
- 44 J. M. H. Olsen and J. Kongsted, *Advances in Quantum Chemistry*, Academic Press, 2011, vol. 61, pp. 107–143.
- 45 E. R. Kjellgren, J. M. Haugaard Olsen and J. Kongsted, J. Chem. Theory Comput., 2018, 14, 4309–4319.
- 46 K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani and P. Dahle, *et al.*, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, 4, 269–284.
- 47 H. C. Georg and S. Canuto, J. Phys. Chem. B, 2012, 116, 11247-11254.
- 48 N. H. List, H. J. A. Jensen and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2016, **18**, 10070–10080.
- 49 M. T. Beerepoot, A. H. Steindal, N. H. List, J. Kongsted and J. M. H. Olsen, *J. Chem. Theory Comput.*, 2016, 12, 1684–1695.
- 50 C. Steinmann, P. Reinholdt, M. S. Nørby, J. Kongsted and J. M. H. Olsen, *Int. J. Quantum Chem.*, 2019, **119**, e25717.
- 51 A. Marefat Khah, P. Reinholdt, J. M. H. Olsen, J. Kongsted and C. Hattig, *J. Chem. Theory Comput.*, 2020, **16**, 1373–1381.

- 52 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2008, 393, 51–57.
- 53 R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.
- 54 P. Sałek, O. Vahtras, T. Helgaker and H. Ågren, J. Chem. Phys., 2002, 117, 9630–9645.
- 55 P. Besalú-Sala, S. P. Sitkiewicz, P. Salvador, E. Matito and J. M. Luis, *Phys. Chem. Chem. Phys.*, 2020, 22, 11871–11880.
- 56 P. Beaujean and B. Champagne, *Theor. Chem. Acc.*, 2018, 137, 50.
- 57 C. Liang, G. Tocci, D. M. Wilkins, A. Grisafi, S. Roke and M. Ceriotti, *Phys. Rev. B*, 2017, **96**, 1–6.
- 58 P. Beaujean and B. Champagne, J. Chem. Phys., 2019, 151, 064303.
- 59 A. Osted, J. Kongsted, K. V. Mikkelsen, P. O. Åstrand and O. Christiansen, J. Chem. Phys., 2006, 124, 124503.
- 60 G. A. Cisneros, M. Karttunen, P. Ren and C. Sagui, *Chem. Rev.*, 2014, **114**, 779–814.
- 61 T. Vasilevskaya and W. Thiel, *J. Chem. Theory Comput.*, 2016, 12, 3561–3570.
- 62 X. Pan, K. Nam, E. Epifanovsky, A. C. Simmonett, E. Rosta and Y. Shao, *J. Chem. Phys.*, 2021, **154**, 024115.
- 63 J. Nochebuena, S. Naseem-Khan and G. A. Cisneros, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**, e1515.

# First Hyperpolarizability of Water in Bulk Liquid Phase: Long-Range Electrostatic Effects Included via the Second Hyperpolarizability

Guillaume Le Breton, Oriane Bonhomme, Emmanuel Benichou, and Claire Loison\*

University of Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-69622, Villeurbanne, France

E-mail: claire.loison@univ-lyon1.fr

# Contents

<b>S1</b>	Methodological details	3		
	S1.1 Second hyperpolarizability calculation	3		
	S1.2 Convergence of $\beta$ and $\gamma$ relative to the configuration number $\ldots \ldots \ldots \ldots$	9		
<b>S2</b>	Second hyperpolarizability $\gamma$ in the bulk phase	11		
	S2.1 Symmetry, averages and standard deviations of $\gamma$	11		
	S2.2 Relationship between bulk and vacuum values of water $\gamma$	14		
	S2.3 $\gamma$ dispersity	15		
<b>S</b> 3	Space heterogeneity of the Electrostatic field generated by the neighborhood	16		
S4	$\gamma$ -based correction: impact of $\gamma$ fluctuations	17		
Re	References 1			

# S1 Methodological details

## S1.1 Second hyperpolarizability calculation

#### **Numerical Implementation:**

To compute the second hyperpolarizability, we use the Finite Field (FF) method. The second hyperpolarizability is defined as the linear coefficient of the first hyperpolarizability evolution relative to an applied electrostatic field. For a vacuum calculation :

$$\gamma_{abcd}^{\text{vac}}(2\omega,\omega,\omega,0) = \frac{\delta\beta_{abc}^{\text{vac}}(2\omega,\omega,\omega)[e_d]}{\delta e_d},\tag{S1}$$

where the dependence relative to the exciting frequency is noted with parentheses (), and the dependence relative to an external static field is noted with hooks []. For the PE calculations, the embedding creates an electric field  $e^{PE}$ , and we define again  $\gamma$  in the PE embedding as the same derivative, but the field derivative is done around the field  $e^{PE}$ .

$$\gamma_{abcd}^{PE}(2\omega,\omega,\omega,0) = \frac{\delta\beta_{abc}^{PE}(2\omega,\omega,\omega)[e_d]}{\delta e_d}$$
(S2)

The Equations S1 and S2 are written in the molecular frame: the first and second hyperpolarizability tensors  $\beta$  and  $\gamma$  are expressed in the molecular frame as well as the electrostatic field **e**. Note that this electrostatic field **e** is **spatially-homogeneous** within the QM box, and is added to the one created by the PE environment. As a reminder, the later electrostatic field created by the environment  $\mathbf{e}^{PE}$  can be **spatially-heterogeneous**. In the following, we omit the *PE* notation and the explicit adding of  $\mathbf{e}^{PE}$  in the molecular frame, or  $\mathbf{E}^{PE}$  in the laboratory frame. Practically, the components second hyperpolarizability components are obtained using linear fit:

$$\beta_{abc}(2\omega,\omega,\omega)[e_d^0] = \beta_{abc}(2\omega,\omega,\omega)[e_d^0 = 0] + \gamma_{abcd}(2\omega,\omega,\omega,0)e_d^0$$
(S3)

$$\mathscr{B}ijk(2\omega,\omega,\omega)[E_l^0] = \mathscr{B}_{ijk}(2\omega,\omega,\omega)[E_l^0=0] + \Gamma_{ijkl}(2\omega,\omega,\omega,0)E_l^0,$$
(S4)

where  $\mathscr{B}$  and  $\Gamma$  are the first and second hyperpolarizability in the laboratory frame, **E** is also expressed in the laboratory frame. All the quantities (first and second hyperpolarizabilities, electrostatic fields, etc.) are in atomic units.

For every molecule, several QM simulations are performed:

- One without extra electrostatic field, to get  $\beta(2\omega, \omega, \omega)[0]$
- For every direction *l*, 6 QM calculations are performed with increasing electrostatic field, from 0.1 to  $1.5 \times 10^{-3}$  a.u.

For each of these calculations, the hyperpolarizability tensor is computed: each simulation returns 27 components. To obtain the second hyperpolarizability component *ijkl*, the evolution of the hyperpolarizability component *ijk* is plotted with respect to the electric field along the *l* direction. For instance, in the Figure S1 is plotted the evolution of the  $\beta_{ccc}$  component in function of the electric field along the *c* direction. The slope of this linear evolution is the  $\gamma_{cccc}$  components. Globally, to obtain all the second hyperpolarizability components, we require 19 QM calculations per molecule.

We have verified that the linear dependency remains valid if the electrostatic finite field is directed in an arbitrary direction – data not shown. Within our range of electrostatic field applied  $(10^{-3} \text{ a.u. or } 10^{-2} \text{ V.Å}^{-1})$ :

$$\beta(2\omega,\omega,\omega)_{ijk}[\mathbf{e}] = \beta(2\omega,\omega,\omega)_{ijk}[0] + \sum_{l} \gamma_{ijkl}(2\omega,\omega,\omega,0) \ e_l$$
(S5)

Finally, to convert the second hyperpolarizability from the molecular to the laboratory frame, we use the rotational matrix R defined from the expression of the laboratory-frame unit vectors



Figure S1: Evolution of one hyperpolarizability component as a function of the applied static and homogeneous electric field. Both hyperpolarizability and the electric field are expressed within the molecular frame. The obtained second hyperpolarizability component is shown in the legend, in atomic units.

 $\mathbf{V}^{\text{lab}}$  in the molecular frame unit vectors  $\mathbf{V}_{a}^{\text{mol}}$ :

$$\mathbf{V}_{i}^{\text{lab}} = \sum_{a} R_{ia} \mathbf{V}_{a}^{\text{mol}},\tag{S6}$$

$$\beta_{abc}(2\omega,\omega,\omega) = \sum_{ijk} R_{ai} R_{bj} R_{ck} \chi_{ijk}(2\omega,\omega,\omega), \qquad (S7)$$

$$\gamma_{abcd}(2\omega,\omega,\omega,0) = \sum_{ijkl} R_{ai} R_{bj} R_{ck} R_{dl} \Gamma_{ijkl}(2\omega,\omega,\omega,0).$$
(S8)

(S9)

#### **Finite Field Versus Response Theory:**

Using DALTON, we can compute the second hyperpolarizability within a response scheme. We have compared the laboratory second hyperpolarizability obtained using the Finite Field method (using 19 QM calculations) to the one using the response theory in Table S1 for some components of a molecule in the vacuum. Very good agreement is found: an error about 20 a.u. can be expected

due the numerical calculation for the second hyperpolarizability. Same results have been found for second hyperpolarizability calculation within the PE environment – data not shown. Hence, throughout the article, the values are rounded to tens of a.u.

Table S1: Second hyperpolarizability components in the laboratory frame for a water molecule in vacuum, in atomic units. The results obtained by the Finite Field method presented in the main text and the one provided by DALTON using the response theory are compared. The calculations have been made at 2 fundamental wavelength: 800 nm and the infinity wave-length limit.

$\Gamma(2\omega,\omega,\omega,0)$	$\lambda = 800 \text{ nm}$		$\lambda  ightarrow +\infty  \mathrm{nm}$	
	Response	Finite Field	Response	Finite Field
Γ <sub>xxxx</sub>	2604	2632	2094	2111
Γ <sub>xxyy</sub>	744	745	583	584
Γ <sub>xyyx</sub>	725	731	583	587
Γ <sub>xyxy</sub>	744	745	583	584
$\Gamma_{xxzz}$	978	978	790	789
$\Gamma_{xzzx}$	965	969	790	784
$\Gamma_{xzxz}$	978	978	790	789
$\Gamma_{yyxx}$	693	698	583	587
$\Gamma_{yxxy}$	699	700	583	584
$\Gamma_{yxyx}$	693	698	583	587
Γ <sub>уууу</sub>	1234	1233	1054	1054
$\Gamma_{yyzz}$	818	816	680	678
$\Gamma_{yzzy}$	827	825	680	679
Γ <sub>yzyz</sub>	818	816	680	678
$\Gamma_{zzxx}$	1025	1032	790	794
$\Gamma_{zxxz}$	988	987	790	789
$\Gamma_{zxzx}$	1025	1032	790	794
Γ <sub>zzyy</sub>	916	916	680	679
Γ <sub>zyyz</sub>	876	873	680	677
$\Gamma_{zyzy}$	916	916	680	679
$\Gamma_{7777}$	3491	3476	2719	2710

#### **Comparison with Literature:**

To conclude methodological checks, we have compared our results with the ones obtained in the literature. First, with the gold standard, Couple Cluster Single Double (CCSD), in the vacuum phase. We compare our results with two recent works by Liang *et. al.*<sup>1</sup> and Beaujean *et. al.*<sup>2</sup> using CCSD with the d-aug-cc-pVTZ basis set at zero frequency. Table S2 compares components at zero frequency. Table S3 compares the experimentally relevant  $\gamma_{\parallel}$ ,  $\gamma_{THS}$  or DR<sub>THS</sub> that are defined as:

$$\gamma_{\parallel} = \frac{1}{15} \sum_{i,j} \gamma_{iijj} + \gamma_{ijij} + \gamma_{ijji}, \qquad (S10)$$

$$\gamma_{\text{THS}} = \sqrt{\Gamma_{zzzz}^2 + \Gamma_{zxxx}^2},\tag{S11}$$

$$DR_{THS} = \Gamma_{zzzz}^2 / \Gamma_{zxxx}^2.$$
 (S12)

Indeed, these reduced values appear in scattering experiments since the experimental results cannot provide directly individual components. For complete  $\gamma_{THS}$  and DR<sub>THS</sub> expressions, see Ref<sup>3</sup> for instance.

The difference observed characterize the error made due to DFT/CAM-B3LYP instead of CCSD and is comparable for the first and second hyperpolarizability: around 10%. An important point is the large frequency dependence of the second hyperpolarizability, see Table S3, which is quite well reproduced by DFT. We can also note the recent work of Besalú-Sala<sup>4</sup> which provides advice for tuning DFT functional to get  $\gamma$  closer to CCSD ones. They mention that DFT/CAMB3LYP is the best standard functional to use for water in vacuum, and that it overestimates the  $\gamma$ .

Regarding the liquid phase, the work by Osted *et.al.*<sup>5</sup> provides a prediction for the water second hyperpolarizability. We are using the same methodology but with more approximation: (1) they used CCSD with the same basis ; (2) Their MD simulation has been made using a polarizable force field ; (3) Their electrostatic embedding also include a polarizable part while we are not. They propose an pure electric  $\gamma_{\parallel}(2\omega, \omega, \omega, 0) = 2070$  a.u. at  $\lambda \approx 1080$  nm. They compare this result by the one proposed by Levine and Bethea:<sup>6</sup>  $\gamma_{\parallel}(2\omega, \omega, \omega, 0) = 2134$  a.u. at the same frequency. We find,  $\gamma_{\parallel}(2\omega, \omega, \omega, 0)$  equal to 1730 a.u at 800nm and 1440 a.u at infinite wavelength. Our result is close to the experimental one, but is too small. However, in the liquid phase decreasing the wavelength seems to increase  $\gamma_{\parallel}$ . At 1080 nm we may hope our liquid  $\gamma_{\parallel}$  to be closer than the

Table S2: Water second hyperpolarizability in vacuum phase, in atomic unit, in the static limit. Our results (DFT/CAMB3LYP) are compared to the ones obtained by Liang *et. al.*<sup>1</sup> (CCSD).

γ	Zero frequency		
	$CCSD^1$	DFT/CAMB3LYP	
Yaaaa	863	965	
Yaabb	654	763	
Ybbbb	2711	3156	
Ycaac	450	507	
$\gamma_{cbbc}$	724	828	
$\gamma_{cccc}$	1506	1678	
$\gamma_{  }$	1747	1998	
ŶTHS	1794	2090	
DR <sub>THS</sub>	73	70	

Table S3: Water second hyperpolarizability in vacuum, in atomic unit. Our results (DFT/CAMB3LYP) are compared to the ones obtained by Beaujean *et. al.*<sup>2</sup> (CCSD). Several wave-length of calculation are shown, in nanometer. We also observe that  $\gamma_{THS}$  increases with decreasing frequency while the depolarization ratio decreases.

	$CCSD^2$		DFT/CAM	B3LYP
Frequency [nm]	<i>γ<sub>THS</sub></i> [a.u.]	DR <sub>THS</sub>	<i>γ<sub>THS</sub></i> [a.u.]	DR <sub>THS</sub>
+ inf	1821	76	2090	70
1064	2288	49	NA	NA
800	NA	NA	2640	47
694.3	3401	24	NA	NA

experimental value.

# S1.2 Convergence of $\beta$ and $\gamma$ relative to the configuration number

In Figure S2 and S3 the evolution of the mean values of the first and second hyperpolarizabilities at 800 nm are plotted as a function of the number of configurations used for the averaging. For every component, the evolution is compared to the last value (using 2400 configurations). Respectives errors of about 0.05 a.u. and 10 a.u. can be expected for the first and second hyperpolarizabilities due to the configuration averaging – which is far less than the error made using DFT. Finally, we



Figure S2: Convergence of  $\langle \beta^{env} \rangle$  relative to the number of configurations used. The reference value is given for N=2400 configurations.

have also checked that the sample obtained using these 2400 configurations is isotropic. Indeed, in top of the symmetry verification of the second hyperpolarizability, see the next section, the first hyperpolarizability in the laboratory frame has been computed. This tensor should be strictly null in average, because of the inversion symmetry inherent to the bulk phase. We obtained indeed a maximal average of 0.2 a.u. for all components.



Figure S3: Convergence of  $\langle \gamma^{env} \rangle$  relative to the number of configurations used. The reference value is given for N=2400 configurations.

# S2 Second hyperpolarizability $\gamma$ in the bulk phase

# S2.1 Symmetry, averages and standard deviations of $\gamma$

Table S4 and S5 report the second hyperpolarizability values in the molecular and laboratory frames, respectively. Only the components larger than 15 a.u. in average are shown.

Due to the  $C_{2\nu}$  molecular symmetry of water, only few  $\gamma$  components should be different of zero. The components  $\gamma_{iiii}$  and the  $\gamma_{iijj}$  with all permutations, where *i* and *j* can be *a*, *b*, or *c*. Regarding the second hyperpolarizability in the laboratory frame  $\Gamma$ , the averages should represent a centro-symmetric system. Hence, all  $\Gamma_{pppp}$  should be similar to 3 times  $\Gamma_{ppqq}$  (with all permutations), where *p* and *q* can be the laboratory axis *x*, *y*, or *z*.<sup>7</sup>

In the molecular frame, the  $C_{2\nu}$  molecular symmetry is respected by our results. The symmetry forbidden components have averages that are smaller than 15 a.u., our typical statistical error ; they are not strictly zero, due to statistical inaccuracies. In the laboratory frame, the centro-symmetry is also fulfilled regarding the  $\Gamma$  components. It confirms that our sample of 2400 molecules represents a centro-symmetric medium.

Table S4:  $\gamma(2\omega, \omega, \omega, 0)$  value in the vacuum and in bulk at 800 nm. For the bulk phase, the average,  $\langle \gamma^{\text{env}} \rangle$ , and standard deviation,  $\sigma[\gamma^{\text{env}}]$ , are presented in atomic unit. Absent components, which correspond to the  $C_{2\nu}$  symmetry forbidden ones, are below 15 a.u. in averaged for the bulk phase. The  $\gamma_{\parallel}$  is defined in Equation S10.

γ	$\gamma^{ m vac}$	$\langle \gamma^{ m env}  angle$	$\sigma[\gamma^{ m env}]$
Yaaaa	1110	740	140
Yaabb	930	640	200
Yaacc	590	370	60
Yabab	930	640	200
Yabba	940	660	140
Yacac	590	370	60
Yacca	590	370	80
Ybaab	1000	680	220
Ybaba	1070	710	150
Ybbaa	1060	710	150
Ybbbb	4090	2940	650
$\gamma_{bbcc}$	1120	770	170
Ybcbc	1120	770	170
$\gamma_{bccb}$	1060	750	200
Ycaac	610	380	60
Ycaca	610	380	80
$\gamma_{cbbc}$	1010	730	150
Ycbcb	1010	720	180
Yccaa	610	380	80
$\gamma_{ccbb}$	1000	720	180
$\gamma_{cccc}$	2000	1400	230
$\gamma_{  }$	2500	1730	

Table S5: Second hyperpolarizability in the laboratory frame  $\Gamma(2\omega, \omega, \omega, 0)$  in the liquid phase at 800 nm. The average and standard deviation are presented in atomic unit,  $\langle \Gamma^{env} \rangle$  and  $\sigma[\Gamma^{env}]$  respectively. Absent components, which correspond to the inversion symmetry forbidden ones, are bellows 15 a.u.

Γ	$\langle \Gamma^{\rm env} \rangle$	$\sigma[\Gamma^{env}]$
$\Gamma_{xxxx}$	1740	830
$\Gamma_{xxyy}$	580	150
$\Gamma_{xxzz}$	580	160
$\Gamma_{xyxy}$	580	150
$\Gamma_{xyyx}$	580	170
$\Gamma_{xzxz}$	580	160
$\Gamma_{xzzx}$	570	150
$\Gamma_{yxxy}$	580	150
$\Gamma_{yxyx}$	580	170
$\Gamma_{yyxx}$	580	170
$\Gamma_{yyyy}$	1740	690
$\Gamma_{yyzz}$	580	170
$\Gamma_{yzyz}$	580	170
$\Gamma_{yzzy}$	570	150
$\Gamma_{zxxz}$	570	150
$\Gamma_{zxzx}$	580	150
$\Gamma_{zyyz}$	570	170
$\Gamma_{zyzy}$	580	150
$\Gamma_{zzxx}$	580	150
$\Gamma_{zzyy}$	580	150
$\Gamma_{zzzz}$	1710	660

# S2.2 Relationship between bulk and vacuum values of water $\gamma$

Figure S4 displays the mean values of the  $\gamma$  components in the bulk, with respect to their values in the vacuum phase. For all components, a ratio of about 0.7 is found: the electrostatic embedding seems to reduce the second hyperpolarizability in the same way for all components.



Figure S4: Mean value of the bulk non vanishing  $\gamma$  components as a function of their vacuum values. The dashed line represents a linear fit. The second hyperpolarizabilities are taken in the molecular frame.

# S2.3 $\gamma$ dispersity

Figure S5 presents the individual values of two  $\beta$  and  $\gamma$  components in the liquid phase:  $\beta_{yyy}$  in function of  $\beta_{zyy}$  and  $\gamma_{yyyz}$  in function of  $\gamma_{zzxx}$ . Due to the  $C_{2\nu}$  symmetry,  $\beta_{yyy}$  and  $\gamma_{yyyz}$  are null on average, while  $\beta_{zyy}$  and  $\gamma_{zzxx}$  have a non-zero average value. However, due to the wide dispersion of all these components, some molecules have larger  $\beta_{yyy}$  than  $\beta_{zyy}$ , or larger  $\gamma_{yyyz}$  than  $\gamma_{zzxx}$ .



Figure S5: Distributions of some first and second hyperpolarizability components for liquid water. Each point represent one of the 2400 configurations of a water molecule in liquid water. Left :  $\beta_{yyyy}$  in function of  $\beta_{zyy}$ . Right :  $\gamma_{yyyz}$  in function of  $\gamma_{zzxx}$ . Calculation for an excitation wavelength of 800 nm.

# S3 Space heterogeneity of the Electrostatic field generated by the neighborhood

Here, we report the evolution of the spatial gradient of the electrostatic field generated by the PE embedding, depending on the environment size  $R_c$ . We have calculated

$$\langle |\Delta \delta e_i / \delta_j| \rangle(R_c) = \frac{1}{N} \sum_n \left| \left( \frac{\delta e_i}{\delta x_j} \right)^n [R_c] - \left( \frac{\delta e_i}{\delta x_j} \right)^n [R_f] \right|,$$
(S13)

where  $\left(\frac{\delta e_i}{\delta x_j}\right)^n [R_c]$  is the spatial gradient along the molecular direction *j* of the electrostatic field along the direction *i* generated by an environment up to  $R_c$  around the molecule *n*. Figure S6 displays  $\langle |\Delta \delta e_i / \delta_j| \rangle (R_c)$  with a logarithmic scale, and a reference gradient at  $R_c = R_f = 40$  Å. According to the Figure S6, the total spatial gradient is quite large at small distance: about  $10^{-2}$  a.u.. However, after few Angstroms, the neighborhood contribution to the spatial gradient drops.



Figure S6: Evolution of the electrostatic gradient with respect to the environment size  $R_c$ . The averaged difference with respect to each molecule value at  $R_c = R_f = 40$ Åis plotted in logarithmic scale. The electrostatic field direction are represented by the 3 curves: along *a* (blue circle), *b* (green diamond) and *c* (orange square). The spatial direction of the derivative is presented from left to right: along the molecular axis *a*, *b* and *c*.

# S4 $\gamma$ -based correction: impact of $\gamma$ fluctuations

In the main text, the same value of  $\gamma$  is attributed to all the water molecules to calculate our hyperpolarizability correction. To estimate the relevance of this approximation, 3 different ways of including the environment are compared here:

$$\beta^{PE}(R_c) = \beta^{PE}(R_c), \qquad (S14)$$

$$\boldsymbol{\beta}^{PE+L}(\boldsymbol{R}_c) = \boldsymbol{\beta}^{PE}(\boldsymbol{R}_c) + \langle \boldsymbol{\gamma}^{\text{env}} \rangle \cdot \Delta \mathbf{e}(\boldsymbol{R}_c), \qquad (S15)$$

$$\boldsymbol{\beta}^{PE+\gamma L}(R_c) = \boldsymbol{\beta}^{PE}(R_c) + \boldsymbol{\gamma}^{\text{env}} \cdot \Delta \mathbf{e}(R_c).$$
(S16)

The QM/MM calculation up to a distance  $R_c$ , is noted  $\beta^{PE}(R_c)$ . The same  $\beta^{PE}(R_c)$  but where the distant neighbors, between  $R_c$  and  $R_f$  are taken into account using the electric field they produced,  $\Delta \mathbf{e}(R_c)$ , using the averaged second hyperpolarizability  $\langle \gamma^{\text{env}} \rangle$  is noted  $\beta^{PE+L}(R_c)$  and is presented in the main text. Finally, we also present here the results for  $\beta^{PE+\gamma L}(R_c)$  that uses the individual values of the molecular second hyperpolarizability, instead of the averaged one.

Figure S7 presents the evolution of the average with respect to  $R_c$  and Figure S8 the  $\Delta\beta_T$ :

$$\Delta \beta_T^X = \frac{1}{27N} \sum_{ijk} \sum_{n=1}^N |\beta_{ijk}^{\text{env}(n)} - \beta_{ijk}^{X(n)}|, \qquad (S17)$$

also used in the main text. For the average, the result of  $\beta^{PE+L}$  and  $\beta^{PE+\gamma L}$  are almost in top of each others. For  $R_c$  larger than about 5 Å, the results are converged. For the  $\Delta\beta_T$ , both corrections using  $\Delta \mathbf{e}(R_c)$  are efficient. However, the  $\beta^{PE+\gamma L}$  error decreases more rapidly: if the effect of the  $\gamma^{\text{env}}$  dispersion cannot be seen in the average predicted  $\beta$ , it is more important for individual value quantity such as  $\Delta\beta_T$ . However, increasing  $R_c$  using  $\beta^{PE+L}$  seems to achieve the same result as  $\beta^{PE+\gamma L}$ . Hence, using the averaged  $\gamma^{\text{env}}$  is justified for water in the liquid phase.


Figure S7: Evolution of the  $\beta_{ccc}^{X}(R_c)$  averaged for the three corrections. The dashed line is the reference value obtained for  $R_c = R_f = 4$  nm.



Figure S8: Evolution of the  $\Delta \beta_T^X(R_c)$  for the three corrections. The dashed line represents the values at which convergence is considered to be obtained, and acts as a guide for the eyes.

# References

- Liang, C.; Tocci, G.; Wilkins, D. M.; Grisafi, A.; Roke, S.; Ceriotti, M. Solvent fluctuations and nuclear quantum effects modulate the molecular hyperpolarizability of water. <u>Phys. Rev. B</u> 2017, 96, 1–6.
- (2) Beaujean, P.; Champagne, B. Coupled cluster investigation of the vibrational and electronic second and third harmonic scattering hyperpolarizabilities of the water molecule.
   J. Chem. Phys. 2019, 151, 064303.
- (3) Beaujean, P.; Champagne, B. Coupled cluster evaluation of the second and third harmonic scattering responses of small molecules. Theor. Chem. Acc. 2018, 137, 50.
- (4) Besalú-Sala, P.; Sitkiewicz, S. P.; Salvador, P.; Matito, E.; Luis, J. M. A new tuned range-separated density functional for the accurate calculation of second hyperpolarizabilities. Phys. Chem. Chem. Phys. 2020, 22, 11871–11880.
- (5) Osted, A.; Kongsted, J.; Mikkelsen, K. V.; Åstrand, P. O.; Christiansen, O. Statistical mechanically averaged molecular properties of liquid water calculated using the combined coupled cluster/molecular dynamics method. J. Chem. Phys. 2006, <u>124</u>, 124503.
- (6) Levine, B.; Bethea, C. Effects on hyperpolarizabilities of molecular interactions in associating liquid mixtures. J. Chem. Phys. **1976**, 65, 2429–2438.
- (7) Brevet, P.-F. In Surface Second Harmonic Generation; Romandes, P. P., Ed.; 1997.

# IV. D | Conclusion and Perspectives

### **Conclusions:**

In this Chapter, we have presented our home-made code FROG developed during the thesis. This tool has already been used to compute the first hyperpolarizability of the water molecule, in the bulk phase, leading to the results of the article presented above and the data needed for the Chapter V, and at the interface, for the Chapter VI. The code has required an important amount of work to match the different goals that we had fixed: this code should be useful "not only" for this thesis.

Indeed, it works with several MD types or liquid compositions. Many structural analyses are available as different electrostatic embedding schemes. The native spatial analysis is useful to analyze the data and an important work has been done on the wiki and tutorials to help new users. I really hope that this code with also help others in the community in the future.

Our work on the electrostatic effects on the first hyperpolarizability, Section IV. C, presents several results:

- The values of the second hyperpolarizability at optical frequency in the liquid phase for both the average and fluctuations values.
- We have proven that the average second hyperpolarizability can be used to correct the  $\beta$  of water molecules with good accuracy if the electrostatic field is homogeneous enough.
- A physical explanation of the validity range of this correction, leading to the decomposition of the electrostatic environment into a short and long range.
- The fluctuations of the second hyperpolarizability are not very important.
- We propose an update for the QM/MM electrostatic embedding: adding a homogeneous electrostatic field directly in the Quantum Hamiltonian in order to represented the distant neighborhood. This last method has been implemented in FROG so that we can be more efficient numerically speaking for the inclusion of large electrostatic environment.

### Perspectives for FROG as a code:

We have already shared FROG on the platform Zenodo. We also consider publishing on a specific journal the code goal and specificities to gain more visibility. However we want first to finish the wiki and the tutorials and made the code even more user-friendly.

Numerically speaking, we have many features we would like to add:

• Add a robust automatized method to test the code when updates are made. For instance, inside the GitLab environment we are currently using.

- Use more complex QM/MM environments: with polarizable environmental (PE1) or with several molecules in the QM box (beyond the monomer approximation). We have already implemented these methods in FROG but not yet applied them.
- Take more advantages of previous calculations to compute the Ground States so that we do not have to start each QM calculations from scratch for several molecules.
- Add a Ewald-like scheme to compute efficiently the electrostatic long-range contribution. This
  way, we could take full advantage of the implicit QM/MM scheme presented above and talked
  systems with a net macroscopic electrostatic field.
- Implement more analysis for FROG. Today we have focalized our approach on the first hyperpolarizabilities. We can go beyond and compute any properties that can be computed by DALTON, for instance the second hyperpolarizabilities, the Two-Photon-Absorption (TPA) cross sections...

# Scientific perspectives: electrostatic potential of surfaces

Solid-liquid interfaces are of prime interest in the community [23–27, 110]. In my opinion, SHG is extremely interesting for the measurement of the electrostatic potential at these interfaces. Indeed, in order to understand the response of the solid-liquid interface, the induced mesoscopic polarization is often decomposed into two terms:

$$\mathbf{P}_{tot}^{2\omega} = \mathbf{P}_{s}^{2\omega} + \mathbf{P}_{liq}^{2\omega}.$$
 (IV.5)

The surface term is considered to be the one created by the solid surface and the first liquid layers, see Figure IV.6. To describe this SHG response, a susceptibility tensor  $\chi^{Stern}(2\omega, \omega, \omega)$  is defined so that:

$$\mathbf{P}_{\rm s}^{2\omega} = \chi^{Stern}(2\omega, \omega, \omega) : \mathbf{E}^{\omega} \mathbf{E}^{\omega} \qquad (\text{IV.6})$$



Figure IV.6: Scheme of a charged interface with three parts along the z-axis: the solid, the Stern-layer and the liquid-like part. The solid interface present negative charges which will create a negative potential compared to the one of the liquid. This will induce in the liquid-like phase an electrostatic potential which will decay to zero. This field induces an SHG response of the bulk-like phase according to a  $\chi^3$  term. The SHG response of the solid surface and the Stern-layer are expressed using a  $\chi^{Stern}$  tensor.

The second term is the one of the liquid

which is close enough to the liquid surface to be impacted by it, but far enough to consider that this effect is a small perturbation. If the solid surface presents net charges, it creates an electrostatic field  $\mathbf{E}^{s}[z]$  in the liquid phase. This field breaks the centro-symmetry of this "liquid-like" phase and thus induces a SHG response – we remind that in the liquid phase, if there is no electrostatic field

the "susceptibility tensor" is null in average. Hence,  $\mathbf{P}_{liq}^{2\omega} = 0$  for  $\mathbf{E}^{s}[z] = 0$ . Using the laboratory second hyperpolarizability,  $\Gamma^{(3)}(2\omega, \omega, \omega, 0)_{liq}$ , we can describe the effect of this electrostatic field to a slab at a position z:

$$\mathbf{P}_{\text{liq}}^{2\omega}(z) = \Gamma^{(3)}(2\omega, \omega, \omega, 0)_{\text{liq}}[z] : \mathbf{E}^{\omega}[z]\mathbf{E}^{\omega}[z]\mathbf{E}^{s}[z]$$
(IV.7)

By integration over the whole liquid phase along z we get the total response. In the literature [111], it is assumed that the second hyperpolarizability in the liquid phase is constant with respect to the altitude so that:

$$\mathbf{P}_{\text{liq}}^{2\omega} = iAV \ \Gamma^{(3)}(2\omega,\omega,\omega,0)_{\text{liq}} : \mathbf{E}^{\omega} \mathbf{E}^{\omega} \hat{z}$$
(IV.8)

With V the electrostatic potential of the solid-liquid surface and A a constant related to experimental condition. Hence, using such experimental approach, one could in principle measure the electrostatic potential of such solid surfaces. Nowadays, using phase-resolved SHG experiment [19, 58, 112, 113], we can even get an absolute value with less "fitting parameters": the Stern term should be real and the one from the liquid-phase complex.

Considering our results, we can really bring some new understanding on these systems:

- Values for  $\Gamma^{(3)}$  for neat water are proposed in the SI of our paper [70].
- We have shown that the fluctuation of  $\gamma$  is small in the liquid phase but what matters is the shape of the electrostatic field applied:  $\mathbf{E}^{s}[z]$  have to be spatially homogeneous at the scale of a molecule to use this approximation. Hence, we could perform more or less the same approach as in the previous paper but in the laboratory frame and decompose the liquid phase into two parts. One part close to the liquid where the inclusion of the electrostatic field created by the solid surface should be included explicit (direct space). The rest of the liquid phase, distant enough from the solid surface, where the effect of the surface can be described using  $\Gamma$  (long-range).
- Experimentally, one widespread method to extract the electrostatic potential using this formalism is to measure the SHG signal of the interface for several ionic concentration [25, 35, 111]. Indeed, using the Gouy-Chapman approximation, the evolution of the electrostatic potential with respect to the ionic concentration is known. However, in order to extract the electrostatic concentration, one has to know  $\Gamma^{(3)}$  for all the concentration, which is always assumed to the same. We could numerically check this hypothesis with FROG.

For course, we can also tackle liquid-gas interfaces where the electrostatic potential is also questioned using SHG in the experimental communality [14, 17–19], with similar method.

# V | Second Harmonic Scattering

I am always fair: from physicists to politicians I treat you alike. I do not mind if you like me or not when your approximations I strike. Often forgotten or inadequately evoked, I am not a detail, and never to be mocked.

What am I?

spinsippis



Figure V.1: When you realized that the simple and friendly approximations are not correct for your system and then you glimpse the theoretical nightmare that it trul y entails.

# V. A | Introduction: Experimental results & Context

In this Chapter, we will study how to use the molecular first hyperpolarizability  $\beta$  computed at the QM/MM level to predict the SHG response of the bulk liquid phase. To do so, we will compare to measurements realized on an experiment called polarization-resolved Hyper Rayleigh Scattering (HRS) or Second-Harmonic-Scattering (SHS). To start with, let us present the experimental setup and the usual framework used to understand the experimental signal. Then, we will focus on the questions we have addressed with the knowledge of the molecular  $\beta$ s.

### V. A.1 SHS experiment for neat water

During my thesis I have performed some SHS measurements on water and on ionic liquid salt dissolved in water. In this part I will briefly present the experimental setup, and the post-production Python code, named Alpaga, I have developed to analysis the raw data easily.

#### SHS setup:

In this part, the SHS typical scheme is presented in Figure V.2, we will only emphasize few elements, more details can be found in [28, 78].

For SHS, as for all Second Harmonic Generation based experiment in the group, we use a femtosecond laser. This means that the laser does not illuminate a continuous electromagnetic field, but pulses of typical duration 200 fs in our case. The frequency rate of these pulses is about 80 MHz, and the time-averaged power about 500 mW. To further enhance the intensity in the liquid phase, an x10 objective is used to condense spatially the light in the liquid sample. For collecting the second harmonic generated by the liquid, a lens (f=10cm) is placed at 90 degrees, so that its focalization point is the same as the one of the fundamental light.

For SHS at 90 degrees, controlling the in-



Figure V.2: Scheme of the experimental setup for SHG in the liquid phase. The first half-wave plate controls the light polarization injected by the x10 objective into the cell. The collection is made at 90 degrees by a lens, and an analyzer made of a half-wave plate and a polarizer cube control which polarization at the second harmonic frequency is sent to the detection. From [28].

coming and outcoming polarization is essential. The laser is naturally linearly polarized. Before the objective, we place a half-wave plate which will rotate the fundamental beam polarization. This plate is motorized to acquire the signal with respect to the incoming angle  $\gamma$  automatically. The outcoming light contains all the polarizations. To select one polarization (V or H), we use a half-wave plate and a polarizer cube. Once the SHG polarization selected, the light goes through a spectrometer. The image for the spectra is made on a cooled CCD camera. The camera will measure the intensity  $I(\lambda)$  as a function of the wavelength  $\lambda$  nearby the second harmonic frequency for the V and H outcoming polarization depending on the incoming polarization  $\gamma$ , see Figure V.3.

#### SHS acquisition:

For measuring the SHS signal from neat water, we first have to ensure that the water we use is as pure as possible. Indeed, the SHS signal of water is impacted by even very small pollution,



Figure V.3: Experimental intensity obtained from the SHS of neat water during 20 sec. The polarization is V-V ( $\gamma = 0$ ).

for instance ions [4]. We use ultra-pure water from a MilliQ device: the resistivity is measured to be 18.2 M $\Omega$ .cm<sup>-1</sup>. Once the cell filled of neat water, we can start the acquisition. As SHS is considered as an incoherent process, the obtained signal is way smaller than in Surface-SHG. Figure V.3 shows the typical signal we obtained on the camera from an acquisition. We can better understand the importance of the spectrometer and the camera compared to measuring only one specific wavelength. Indeed, there are two problems:

- Background: the water molecules are also generating other light using other processes than SHG. All this background due to the laser is a noise to the detection of the second harmonic signal. We can also detect parasite light coming from the experimental room despite our optical shielding. In Figure V.3 this noise is the slow varying signal: it increases from 325 to 460 nm, and then decreases.
- Spikes: These peaks are some extremely sharp bursts of intensity detected by the camera. In Figure V.3 we can see one at about 435 nm. Their amplitudes can be very large, and if it happens in the SHG signal it can be very difficult to disentangle as it is a stochastic noise.

We will see in the next paragraph how to use the spectra to remove these 2 kinds of noise in order to extract the SHS intensity.

For a given incoming and outcoming polarization angles, the typical acquisition time on our setup is about 200 s. This is done by saving for instance 12 acquisitions for 20 seconds each. This acquisition is repeated for all the  $\gamma$  angles (typically 90 angles are performed) and for the V and H outcoming polarization. In total, an SHS experiment of pure water lasts about half a day.

#### Post-treatment of the RAW data: Alpaga:

In order to remove the noise of the raw data, I have coded a python package called Alpaga. The procedure has been inspired by different already existing codes in the team.

In top of being automatic, robust and with an extensive wiki, this code aims to deal with the different setups of the team (SHS in several different rooms and also Surface-SHG). With the help other group members, Fabien Rondepierre (PhD), Maxime Fery (PhD) and also Oriane Bonhomme we now have an efficient and accessible tool. As Alpaga is provided with a wiki and tutorials <sup>1</sup>, we will only present briefly how to get ride of the peaks and the noise from the raw data provided by the detection.



For the spikes, we go through the spectra for one acquisition and com-

pute the local average on typically 2-3 nm range. If a point of the spectra is larger than a threshold compared to the local average, it is detected as 'spike'. This point is removed from the spectra. We repeat this operation for several local averaging sizes and thresholds. The goal is to avoid false-positive spike detection. We do this cleaning procedure for all the spectra of a given incoming/outcoming polarization (typically between 5 and 15 spectra) and then compute the average over all the acquisitions.

For the background noise, we work with the previously averaged spectra, *i.e.* without the spikes. To extract the SHG intensity, we first remove the wavelength range related to the second harmonic process (between 395 and 415 nm for instance) and fit the remaining spectra using a polynomial function. Then, difference between the total spectra and the noise-fit function is computed, for instance see Figure V.4. The SHG intensity is then obtained by fitting with a Gaussian-shape function along with 2 other control parameters: the central wavelength of the SHG signal and its width. If the SHG intensity may evolve with respect to the incoming and outcoming polarization angles, the central wavelength and the Gaussian width should not as there are related to the fundamental laser properties. Hence, if these 2 quantities evolve too much during the experiment, we can assume



Figure V.4: Averaged spectra for one incomingoutcoming polarization angles with the peaks and fluorescence noise removed. The line corresponds to the Gaussian fit from which we extract the intensity, 710.3, but also control parameters such as the central wavelength, 407.6 nm, and the width, 2.7 nm.

that there was a problem with the fundamental laser or the analysis process. Such control helps us avoid using wrong sets of data. Regarding the raw data analysis, the most critic part is the choice of the "wavelength range related to the SHG process". Hence, an automatic procedure has been

<sup>&</sup>lt;sup>1</sup>See the Zenodo page, https://doi.org/10.5281/zenodo.5639393



Figure V.5: a) Scheme of 90 degree polarization-resolved Second Harmonic Generation experiment in the bulk phase. The fundamental is polarized in the xy plane and propagate throughout +z. The detection is made along the +x axis with respect to the V = y or H = z polarization for the second harmonic frequency. b) Experimental results for SHS of neat water at room temperature for the V (dots) and H (stars) outcoming polarization. The intensity has been normalized using the maximal value of the V polarization. The depolarization ratio D is equal to  $0.14 \pm 0.01$ .

developed:

- Treat all the spectra for each set of polarizations, find the average central SHG wavelength and the associated width.
- Define a new SHG wavelength range based on these quantities (for instance away from 2 times the width with respect to the central wavelength).
- Treat again all the data with this new SHG wavelength range.

This procedure avoids operator-dependent choices. Moreover, it deals with the fact that the central wavelength and width can vary from one day to another.

#### SHS polarization-resolved intensity for neat water:

In Figure V.5 a) is reminded the definition of the incoming fundamental polarization definition, using the angle  $\gamma$ , and the outcoming polarization definition, either 'V' (for vertical) or 'H' (for horizontal). Figure V.5 b) presents the evolution of the second harmonic intensity for the outcoming V and H polarizations with respect to the fundamental beam polarization angle  $\gamma$ .

The V polarization curve presents oscillations with the incoming polarization  $\gamma$  while the H is almost flat. We usually fit the evolution of the intensity using three parameters (a, b and c) by outcoming polarization such that [29, 45, 78] :

$$I_{\Gamma} = a_{\Gamma} \cos^4 \gamma + b_{\Gamma} \cos^2 \gamma \sin^2 \gamma + c_{\Gamma} \sin^4 \gamma \tag{V.1}$$

Where the subscript  $_{\Gamma}$  stands for either the V or H outcoming polarization – the next Section V. A.2 will present the reason why we used this kind of equations to fit the experimental intensities. Due to the centro-symmetry of the liquid phase, we expect for the H polarization that  $a_H = c_H = b_H/2$  and thus that the H polarization is a plateau [42].

There are two main indicators that we determine experimentally. First, the **depolarization ratio** D which is the ratio between the minimal and maximal values of the V outcoming polarization:  $D = 0.14 \pm 0.01$  for neat water. Second, for pure water, the V intensity is always larger than the H one. We use the quantity  $\min(I_V) - \max(I_H) \approx 0.04$ , using normalized intensities, to characterize this behavior. The depolarization ratio is usually associated with the local symmetry in the literature while the difference in minimal and maximal intensity is related to long-range correlations within the molecule in the liquid phase. We will thus calculate and discuss these quantities through our numerical approach.

#### V. A.2 How to understand SHS results: scattering formalism

In order to understand the SHS intensity obtained in Figure V.5 b) we will use our molecular-based approach.

#### Building the experimental signal from an individual molecular perspective:

First of all, let us define the molecular response depending on the incoming polarization. The incoming fundamental field  $\mathbf{E}^{\omega}$  propagates along the +z-laboratory axis and has a wave vector:  $\mathbf{k}^{\omega} = \frac{2\pi n(\omega)}{\lambda} \mathbf{u}_z$ . The incoming field is linearly polarized in the xy plane. During the experiment, the incoming polarization is turned in this xy plane according to an angle  $\gamma$ , see Figure V.5 a). The electromagnetic field felt by a molecule n at a position  $\mathbf{r}_n$  is then:

$$\mathbf{E}^{\omega}[\mathbf{r}_n] = E\left[\sin\gamma \ \mathbf{u}_x + \cos\gamma \ \mathbf{u}_y\right] \exp\left[i\mathbf{k}^{\omega} \cdot \mathbf{r}_n\right] \tag{V.2}$$

At the dipolar approximation, the Second Harmonic response of an individual molecule n is given by its molecular hyperpolarizability in the laboratory frame  $(x, y, z) \mathscr{B}(2\omega, \omega, \omega)$  noted  $\mathscr{B}^2$ :

$$\mathbf{P}^{n} = \mathscr{B} : \mathbf{E}^{\omega}[\mathbf{r}_{n}]\mathbf{E}^{\omega}[\mathbf{r}_{n}]$$
(V.3)

Note that in this Chapter we will use  $\mathbf{P}^n$  for the induced dipole moment of a molecule n at the second harmonic frequency expressed in the laboratory frame (and not for the mesoscopic polarization).

The collection of the SHG intensity is made along the x-laboratory axis: at 90 degrees compared to the fundamental light direction. The electromagnetic field at the  $2\omega$  frequency generated in the direction of the detector, noted  $\mathbf{u}_D$ , is  $\mathcal{E}^n$ . This field is the one created by the dipole n oscillating

<sup>&</sup>lt;sup>2</sup>Here we have assumed that the excitation field is the electromagnetic one. This is not true since we have to take into account the response of the molecule at the fundamental frequency as presented in the theory chapter. However, this will not impact at all the comparison with the experiment because this we look at normalized intensity. Indeed, the ratio between the excitation and electromagnetic field do not depend on the fundamental polarization because we are in the bulk phase.

at the frequency  $2\omega$ ,  $\mathbf{P}^n$ , and is given by: <sup>3</sup>

$$\mathcal{E}^n \propto (\mathbf{u}_D \wedge \mathbf{P}^n) \wedge \mathbf{u}_D \exp\left[i\mathbf{k}^{2\omega} \cdot (\mathbf{u}_D - \mathbf{r}_n)\right].$$
 (V.4)

Where  $\mathbf{k}^{2\omega}$  is the wave vector at the second harmonic frequency:  $\mathbf{k}^{2\omega} = \frac{4\pi n(2\omega)}{\lambda} \mathbf{u}_x$ .

We collect either the Second Harmonic field polarized along the y-axis, 'V', or along the z-axis, 'H'. Hence, the electromagnetic field generated by a single molecule n that arrives at the detector is:

$$\mathcal{E}_{V}^{n} = \mathcal{E}^{n} \cdot \mathbf{u}_{y} = \mathcal{E}_{y}^{n} \propto (\mathbf{P}^{n} \cdot \mathbf{u}_{y}) \exp\left[i\mathbf{k}^{2\omega} \cdot (\mathbf{u}_{D} - \mathbf{r}_{n})\right]$$
(V.5)

$$\mathcal{E}_{H}^{n} = \mathcal{E}^{n} \cdot \mathbf{u}_{z} = \mathcal{E}_{z}^{n} \propto (\mathbf{P}^{n} \cdot \mathbf{u}_{z}) \exp\left[i\mathbf{k}^{2\omega} \cdot (\mathbf{u}_{D} - \mathbf{r}_{n})\right]$$
(V.6)

The detector measures the intensity produced by the response of all the molecules:

$$I_V = \left(\sum_n \mathcal{E}_V^n\right) \left(\sum_m \mathcal{E}_V^m\right)^*,\tag{V.7}$$

$$I_H = \left(\sum_n \mathcal{E}_H^n\right) \left(\sum_m \mathcal{E}_H^m\right)^*.$$
 (V.8)

#### Incoherent and Coherent contributions:

To discuss the intensity evolution, we derive a bit more the intensity collected by the detector for any outcoming polarization V or H:

$$I_{\Gamma} = \left(\sum_{n} \mathcal{E}_{\Gamma}^{n}\right) \left(\sum_{m} \mathcal{E}_{\Gamma}^{m}\right)^{*} = \sum_{n} |\mathcal{E}_{\Gamma}^{n}|^{2} + \sum_{n \neq m} (\mathcal{E}_{\Gamma}^{n}) (\mathcal{E}_{\Gamma}^{m})^{*}$$
(V.9)

$$I_{\Gamma} \propto \sum_{n} |\mathbf{P}_{\Gamma}^{n}|^{2} + \sum_{n \neq m} (\mathbf{P}_{\Gamma}^{n}) (\mathbf{P}_{\Gamma}^{m})^{*} \exp\left[i\Delta \mathbf{q} \cdot (\mathbf{r}_{n} - \mathbf{r}_{m})\right]$$
(V.10)

Where we have defined  $\Delta \mathbf{q} = 2\mathbf{k}^{\omega} - \mathbf{k}^{2\omega} = \frac{4\pi}{\lambda} (n(\omega)\mathbf{u}_z - n(2\omega)\mathbf{u}_x)$ . Note that  $\sum_{n \neq m}$  is a notation for  $\sum_n \sum_{m \text{ with } n \neq m}$ .

Here appears two terms: the first one which sums over all the molecules n called **incoherent** contribution and the second which sums over all possible molecule couples (n, m) called **coherent** contribution. A microscopic interpretation is presented in Figure V.6 for these two terms. For the incoherent case, the detector measures the intensity of the electromagnetic field generated molecule by molecule. For the coherent case, the detector measures the interferences between the field created by a molecule n with the one created by a molecule m. Experimentally, we cannot distinguish the contribution of these terms. Let us take two limit cases to further understand what represents these terms: the infinite dilution limit and the fully ordered system one.

<sup>&</sup>lt;sup>3</sup>see Appendix IX. C.1 for more details about this formula



Figure V.6: Molecular interpretation of Equation V.10. Left: Incoherent contribution where each molecule's contribution is added to the total signal directly: this is a one-body term. Right: Coherent contribution where the electromagnetic fields generated by the molecules n and m interfere: this is a two-body term.

The first limit corresponds to the Hyper Rayleigh Scattering (HRS) experiment. In such experiment, molecules with a strong non-linear response are diluted into a solvent. In this case, the contribution of the solvent to the second harmonic intensity is negligible. One example is the DiA molecule in methanol, see for instance [78] p.38. For low concentrations, each emitting molecule will be far away from each other. Equation V.10 shows that the coherent term is the product of the induced dipole moment of all the molecule couples at the second harmonic frequency. Hence, if all the emitting molecules are very distant from each another, we can expect their relative orientations to be random. Therefore, the induced dipole moment orientation is random and thus the product  $\sum_{n \neq m} (\mathbf{P}_{\Gamma}^n) (\mathbf{P}_{\Gamma}^m)$  is zero in average. The experimental signal depends only on individual molecular response, and is fully incoherent.

Now let us imagine a system perfectly arranged, for instance confined liquid, gel or liquid crystal: all molecules are arranged within a crystal-like structure. In general, the coherent term does not cancel anymore in the general case <sup>4</sup>. Moreover, if we assume that there are N molecules per unit volume, the incoherent term intensity scales with N while the coherent one scales as  $N^2$ . Hence, this coherent term can quickly overwhelm the incoherent one if there is orientational correlation within the liquid.

In the case of a liquid, the coherent term arises from orientational correlations between molecules. Such correlations have been evidenced in water [3] or in ionic liquids [45] for instance.

<sup>&</sup>lt;sup>4</sup>if we also assume that this ordered structure is not centro-symmetric

Hence, we define that a liquid presents "correlation" in the sense of SHS if  $\sum_{n \neq m} (\mathbf{P}_{\Gamma}^n) (\mathbf{P}_{\Gamma}^m)$  does not average to zero. In this case, the coherent terms is not zero without being the dominant contribution to the total SHS intensity.

#### Express the intensity using the molecular hyperpolarizabilities:

Using FROG, we have access to each molecule's first hyperpolarizability expressed in the laboratory frame  $\mathscr{B}$ . Hence, let us rewrite the intensity collected experimentally using this quantity instead of the induced dipole moment **P** for any given incoming fundamental polarization defined by the angle  $\gamma$ . To do so, we have to combine Equations V.2, V.3 which leads to:

$$\mathbf{P}_{V}^{n}(\gamma) = E^{2} \left( \mathscr{B}_{yyy}^{n} \cos^{2} \gamma + 2 \mathscr{B}_{yxy}^{n} \cos \gamma \sin \gamma + \mathscr{B}_{yxx}^{n} \sin^{2} \gamma \right) \exp\left[i 2\mathbf{k}^{\omega} \cdot \mathbf{r}_{n}\right]$$
(V.11)

$$\mathbf{P}_{H}^{n}(\gamma) = E^{2} \left( \mathscr{B}_{zyy}^{n} \cos^{2} \gamma + 2\mathscr{B}_{zxy}^{n} \cos \gamma \sin \gamma + \mathscr{B}_{zxx}^{n} \sin^{2} \gamma \right) \exp\left[i2\mathbf{k}^{\omega} \cdot \mathbf{r}_{n}\right]$$
(V.12)

Combining with Equation V.10 we obtain for the V and H intensities function:

$$I_{V} \propto E^{4} \sum_{n,m} (\mathscr{B}_{yyy}^{n} \cos^{2} \gamma + 2\mathscr{B}_{yxy}^{n} \cos \gamma \sin \gamma + \mathscr{B}_{yxx}^{n} \sin^{2} \gamma)$$

$$\times (\mathscr{B}_{yyy}^{m} \cos^{2} \gamma + 2\mathscr{B}_{yxy}^{m} \cos \gamma \sin \gamma + \mathscr{B}_{yxx}^{m} \sin^{2} \gamma)^{*} \times \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right]$$

$$I_{H} \propto E^{4} \sum_{n,m} (\mathscr{B}_{zyy}^{n} \cos^{2} \gamma + 2\mathscr{B}_{zxy}^{n} \cos \gamma \sin \gamma + \mathscr{B}_{zxx}^{n} \sin^{2} \gamma)$$

$$\times (\mathscr{B}_{zyy}^{m} \cos^{2} \gamma + 2\mathscr{B}_{zxy}^{m} \cos \gamma \sin \gamma + \mathscr{B}_{zxx}^{m} \sin^{2} \gamma)^{*} \times \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right]$$

$$(V.13)$$

Where  $\Delta \mathbf{q} = 2\mathbf{k}^{\omega} - \mathbf{k}^{2\omega} \approx 2.09 \times 10^{-2} (\mathbf{u}_z - \mathbf{u}_x) \,\mathrm{nm}^{-1}$  and  $\mathbf{r}_{nm} = \mathbf{r}_n - \mathbf{r}_m$ 

#### Definition of the total correlation function for $\mathscr{B}$ : $Y_{ijk,abc}$

In  $I_V$  and  $I_H$  appear terms of the form:  $\sum_{n,m} \mathscr{B}^n_{ijk} \mathscr{B}^m_{abc} \exp [i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}]$  Thus, thank to our QM/MM calculations, we can work directly with these quantities,  $\mathscr{B}$ , as an output from FROG to compare with the experimental intensity. Following the decomposition into an incoherent and coherent terms, we can define a new tensor with 6 components, Y, as follow:

$$Y_{ijk,abc} = \sum_{n,m} \mathscr{B}^n_{ijk} \mathscr{B}^m_{abc} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right]$$
(V.15)

$$=\sum_{n} \mathscr{B}_{ijk}^{n} \mathscr{B}_{abc}^{n} + \sum_{n \neq m} \mathscr{B}_{ijk}^{n} \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right]$$
(V.16)

$$Y_{ijk,abc} = Y_{ijk,abc}^{\text{incoh}} + Y_{ijk,abc}^{\text{coh}}$$
(V.17)

In Appendix IX. C.4 are given more details on these objects and which couple (ijk, abc) are expected to be null. Using these objects and rewriting the previous equation for the intensity, the SHS

#### Page 153

intensity is given by:

$$I_V \propto Y_{yyy,yyy} \cos^4 \gamma + (2\text{Re}Y_{yyy,yxx} + 4Y_{yxy,yxy}) \cos^2 \gamma \sin^2 \gamma + Y_{yxx,yxx} \sin^4 \gamma, \qquad (V.18)$$

$$I_H \propto Y_{zyy,zyy} \cos^4 \gamma + (2\text{Re}Y_{zyy,zxx} + 4Y_{zxy,zxy}) \cos^2 \gamma \sin^2 \gamma + Y_{zxx,zxx} \sin^4 \gamma.$$
(V.19)

These equations have a similar form than the ones used to fit the experimental data, see Equation V.1. In the following, the depolarization ratio D is given by:  $Y_{yxx,yxx}/Y_{yyy,yyy}$ , whereas the gap between the V and H is given by  $Y_{yxx,yxx} - Y_{zxx,zxx}$ .

#### Incoherent limit and molecular symmetry:

In the usual HRS framework, for instance where NLO dyes are highly diluted in a solvent, only the incoherent term is conserved. In our formalism, if we want to reproduce the depolarization ratio, it means that we should only compute:  $Y_{yxx,yxx}^{\text{incoh}}$  and  $Y_{yyy,yyy}^{\text{incoh}}$ :  $\sum_{n} \left(\mathscr{B}_{yxx}^{n}\right)^{2}$  and  $\sum_{n} \left(\mathscr{B}_{yyy}^{n}\right)^{2}$ .

For a given molecule, we obtain the following relation between the molecular and laboratory hyperpolarizability:

$$\mathscr{B}^n_{ijk}\mathscr{B}^n_{abc} = \sum_{IJKABC} R^n_{iI} R^n_{jJ} R^n_{kK} R^n_{aA} R^n_{bB} R^n_{cC} \beta^n_{IJK} \beta^n_{ABC}.$$
 (V.20)

Where  $\mathbb{R}^n$  is the rotational matrix which permits the conversion from the molecular frame of the molecule n to the laboratory one, see Appendix IX. C.2.

These averages of the square hyperpolarizability expressed in the laboratory frame, noted  $\langle \mathscr{B}^2 \rangle$ , can be computed easily if we assume that all molecules share the same first hyperpolarizability expressed in their own frame,  $\beta$ . In this hypothesis and in the case of an isotropic liquid,  $\langle \mathscr{B}^2 \rangle$  represents the average over all the possible orientations of the molecule in the laboratory frame. There are formula (see Appendix IX. C.2) which links  $\beta$  to  $\langle \mathscr{B}^2 \rangle$ , and thus to the depolarization ratio D in this pure incoherent case [42]. As a matter of fact, SHG in the liquid phase is in that case a good tool to probe the symmetry of  $\beta$ , and thus the molecular symmetry [42]. Indeed, within these hypotheses, the measurement of D can be related to ratios between  $\beta$  components. And since we know that the  $\beta$  should fulfills the molecule's symmetry, it can give us some insights on the molecule structure.

#### SHS of pure solvent:

However, in the case of pure solvent where there are no reasons to discard the coherent term, the analysis of the experimental results is more difficult. Indeed, the collected signal contains the two contributions. Hence, in this case, we will rather use the term Second-Harmonic-Scattering (SHS). This is the case for pure water. We know that this system is a centrosymetric liquid at the meso-scopic scale, but at the molecular scale there are strong orientational correlations between molecules (for instance through Hydrogen bonds). Recently, more and more SHS experiments [3, 44, 45] are performed on pure solvents in order to study the structure of the liquid phase thanks to this coher-

ent contribution to the Second Harmonic signal.

If the experimental signal may bear structural information, analyzing this signal for pure solvent is much more difficult than the usual HRS experiment. To emphasis this, let us decompose the coherent term further using the intermolecular distance  $\mathbf{r}_{nm} = \mathbf{r}_n - \mathbf{r}_m$ . Indeed, in liquid water the propagation difference vector  $\Delta \mathbf{q}$  is about  $2.09 \times 10^{-2} \text{ nm}^{-1}$ : the exponential may be developed using Taylor series. At the linear order, it gives:

$$I^{coh} = \sum_{n \neq m} \left( \mathbf{P}_{\Gamma}^{n} \right) \left( \mathbf{P}_{\Gamma}^{m} \right)^{*} \exp \left[ i \Delta \mathbf{q} \cdot \mathbf{r}_{nm} \right] \approx \sum_{n \neq m} \left( \mathbf{P}_{\Gamma}^{n} \right) \left( \mathbf{P}_{\Gamma}^{m} \right)^{*} + \frac{i \Delta \mathbf{q}}{2} \cdot \sum_{n \neq m} \mathbf{r}_{nm} \left( \mathbf{P}_{\Gamma}^{n} \right) \left( \mathbf{P}_{\Gamma}^{m} \right)^{*}$$
(V.21)

The second term emphasizes the difficulty to disentangle the different contributions in the coherent signal:

- On one hand, the term  $\sum_{n \neq m} (\mathbf{P}_{\Gamma}^n) (\mathbf{P}_{\Gamma}^m)^*$  should average to zero as the distance between molecules n and m increases as liquid water is a homogeneous and centro-symmetric liquid at the mesoscopic scale.
- On the other hand, the second sum is scaled by the distance between the n and m: hence even if the orientational correlation between the molecules decreases with the distance, its weight in the coherent signal increases due to the  $\mathbf{r}_{nm}$  factor! This can be even more dramatic when considering the full exponential form and not only the first order.

Hence, the coherent contribution contains both correlation between molecules at the very short distances (where  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} \approx 0$ ) and at larger distance (where  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} > 1$ ) even if they are very small!

#### **Problematics:**

In general, an important property used to disentangle the different contribution in the signal is the molecular symmetry. Indeed, if we assume that  $\beta$  fulfills strictly the molecular symmetry, it reduces the number of unknown quantities because many components are thus equal or vanish. Moreover, if one also assumes that the molecular first hyperpolarizability is the same for each molecule it allows using the ratio of few hyperpolarizability components to fit the experimental signal [42]. At the numerical point of view, this hypothesis is also made: the quantity computed to compare with experimental data is one hyperpolarizability tensor.

However, as we have seen in the Chapter IV, the first hyperpolarizability of the water in the liquid phase fluctuates a lot due to its electrostatic environment. Hence, in top of this difficulty regarding the molecule structure and orientational correlation, we should also include the  $\beta$ s fluctuations. We aim to explain the experimental SHS intensity of neat liquid water using individual molecular response calculated at the QM/MM level:

$$I \propto \sum_{n} \left(\mathscr{B}^{n}\right)^{2} + \sum_{n \neq m} \mathscr{B}^{n} \mathscr{B}^{m} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right].$$
(V.22)

Where the first term is the incoherent, and the second the coherent. First, we will see how to adapt the classical equation used in the fully incoherent case for liquid water where the molecular hyperpolarizability fluctuates. Then, we will compute the impact of the short-range coherent terms to the SHS intensity. In conclusion, we will suggest some perspectives to adapt the formalism developed here to the total coherent signal with short-range and long-range terms.

# V. B | Incoherent response: fluctuations matter

In this section, we first study the incoherent intensity which is a one-body average  $\langle \mathscr{B}^2 \rangle$ . In Chapter IV we have seen how to compute the first hyperpolarizability,  $\beta$ , of a molecule in its molecular frame in the liquid phase. Figure V.7 reminds the **distribution** of the  $\beta$  according to QM/MM calculations,  $\mathcal{D}[\beta]$ , for water in the liquid phase. The averaged value of each component respects the molecular symmetry, but very large fluctuations are observed. The largest distribution has been observed for the  $\beta_{yyy}$  components with a standard deviation about ten a.u. However, the experimental measurement is made in the laboratory frame. We then have to understand how to use these distributions, obtained in the molecular frame, to calculate the hyperpolarizabilities in the laboratory one.

Appendix IX. C.2 presents in detail how to compute  $\langle \mathscr{B}^2 \rangle$  when we assume that each molecule has the same  $\beta$ . We describe how to compute the **distribution** of the first hyperpolarizability in the laboratory frame,  $\mathscr{B}$ , using Euler matrices and orientational averaging, and compare it with our results using FROG.

In this section, we will compute  $\langle \mathscr{B}^2 \rangle$  for three possible cases: (i) by assuming that each molecule has the same  $\beta$  (ii) by accounting for  $\beta$  fluctuations using an independent distribution description for each  $\beta$  component and (iii) by using the data directly from FROG. Note that for simplicity's sake, we have used the results calculated in the zero frequency limit, because the Kleinman symmetry is strictly respected in this case. The final comparison with the experimental results discussed in the next Section V. C, including also the incoherent and coherent terms, has been made for a fundamental frequency of 800 nm.



Figure V.7: First hyperpolarizability distribution in the molecular frame for the water in the liquid phase in the zero-frequency limit. The distribution is not normalized: a total of 156 250 molecules have been used.

Table V.1: Mean and standard deviation of the first hyperpolazirability in the molecular frame for water in the liquid phase at zero frequency in atomic units. The standard deviation  $\sigma$  is defined in Equation V.27.

$\beta$	$<\beta>$	$\sigma$
aaa	0.0	2.0
bbb	0.0	7.3
ccc	3.9	3.9
abb	0.0	2.2
acc	0.0	1.6
baa	0.0	1.9
bcc	0.0	2.2
caa	-1.9	1.4
cbb	2.4	2.4
abc	0.0	1.2

#### V. B.1 Hypothesis 1: disregarding hyperpolarizability fluctuations

We start by the calculation of  $\langle \mathscr{B}^2 \rangle$  assuming that each molecule has the same first hyperpolarizability. To do so, we first have to choose a value for each component: one choice is to take the mean value of the  $\beta$  distribution presented in Figure V.7 and summarized in Table V.1. Then, as presented in the Appendix IX. C.2 we compute  $\langle \mathscr{B}^2 \rangle$  by sampling over all the possible orientations of the molecules in the laboratory frame. Using an Euler-based formalism, the orientational averaged of  $\mathscr{B}$  for a given  $\beta$  tensor  $\langle \mathscr{B}^2[\beta] \rangle$  is given by:

$$<\mathscr{B}^{2}[\beta]>_{ijk} = \frac{1}{8\pi^{2}} \int_{0}^{\pi} d\theta \sin(\theta) \int_{0}^{2\pi} d\psi \int_{0}^{2\pi} d\phi \left(\sum_{abc} R(\theta,\psi,\phi)_{ia} R(\theta,\psi,\phi)_{jb} R(\theta,\psi,\phi)_{kc} \beta_{abc}\right)^{2}$$
(V.23)

Where the rotational matrices  $R(\theta, \psi, \phi)$  are defined in Appendix IX. C.2. In the following we will focus on two components:  $\langle \mathscr{B}^2 \rangle_{zzz}$  and  $\langle \mathscr{B}^2 \rangle_{zxx}$ . Indeed, these components are directly related to the depolarization ratio. Since in this section we assume in the pure incoherent limit:  $D^{\text{incoh}} = \langle \mathscr{B}^2 \rangle_{zxx} / \langle \mathscr{B}^2 \rangle_{zzz}$ . According to Ref [42] we obtain:

$$<\mathscr{B}^{2}[\beta]>_{zzz} = \frac{1}{105} \sum_{abc} 2\beta_{abc}^{2} + \beta_{abb}\beta_{acc} + 4\beta_{aab}\beta_{bcc} + 4\beta_{aab}\beta_{ccb} + 4\beta_{abc}\beta_{bac} \tag{V.24}$$

$$<\mathscr{B}^{2}[\beta]>_{zxx} = \frac{1}{105} \sum_{abc} 6\beta_{abc}^{2} + 3\beta_{abb}\beta_{acc} - 2\beta_{aab}\beta_{bcc} - 2\beta_{aab}\beta_{ccb} - 2\beta_{abc}\beta_{bac} \tag{V.25}$$

In this analytical solution, we retrieve that the influence of the  $\beta$  components to the final result is not trivial: each new component increases the total value by the  $\beta_{abc}^2$  term, but can also decreases it because of the many other terms. This is especially true when mixing positive and negative  $\beta$ component, see Appendix IX. C.2. Using Equation V.24 with the  $\beta$  average values, we find that  $\langle \mathscr{B}^2 \rangle_{zzz} = 4.2$  and  $\langle \mathscr{B}^2 \rangle_{zxx} = 1.4$  a.u., leading to a depolarization ratio of 0.35. Using this hypothesis, we found that the prediction of the depolarization ratio in the incoherent limit using solely the mean  $\beta$  values do not match the experimental one  $(D = 0.14 \pm 0.01)$ .

#### V. B.2 Hypothesis 2: Taking into account independent fluctuations

Now, we take into account the fluctuation of the molecular  $\beta$  in the  $\langle \mathscr{B}^2 \rangle$  calculation using the  $\beta$  distributions obtained in Figure V.7. In a first method, we will sample explicitly the  $\beta$  fluctuations, and in a second one, we will use an analytical formula. Note that the two methods lead to the same results, but we will start by the first one to gain some physical insights.

#### Sampling 27 Gaussians: reducing the possible $\beta$ values

We have seen that we can easily compute for a given tensor  $\beta$  the associated averaging in the laboratory frame noted  $\langle \mathscr{B}^2[\beta] \rangle$  using Equation V.23 – see the solutions for the particular components of interest in Equation V.24. Averaging over all possible  $\beta$  values, noted  $\langle \mathscr{B}^2 \rangle \rangle$ , we arrive at this simple mathematical expression:

$$\langle \mathcal{B}^2 \rangle \rangle = \int \langle \mathcal{B}^2[\beta] \rangle P(\beta).$$
 (V.26)

Where the integral goes over all the possible  $\beta$  values with its associated probability  $P(\beta)$ . This formula is due to the fact that we are in the fully incoherent limit: the contribution of each molecule to the SHG intensity is accumulated via their square modulus.

 $P(\beta)$  is a **joint** probability:  $\beta$  has 27 components, and thus  $P(\beta)$  is the probability of finding each  $\beta$  component to a particular value at the same time. We could compute this joint probability using the results obtained by FROG. However, performing such analysis is challenging because of the large number of degrees of freedom (the 27  $\beta$  components) and we first tried to lower the complexity. Thanks to the SHG process, we know that the intrinsic symmetry should be respected: for any molecule we have that  $\beta_{ijk} = \beta_{ikj}$ . Hence, finding a molecule with a given value y for its component ijk,  $\beta_{ijk} = y$ , while its component ikj is not equal to y is not possible:  $P(\beta_{ijk} = Y \text{ and } \beta_{ikj} \neq Y) =$ 0. Therefore, all the components linked by the intrinsic symmetry are fully correlated. This reduces the number of meaningful components to 18, the other being obtained using this intrinsic symmetry.



Figure V.8: Correlation distribution of  $\beta_{cbb}$  and  $\beta_{bcb}$ . On the right the distribution of  $\beta_{bcb}$ . On the top the distribution of  $\beta_{cbb}$ . The total number of molecules used is 156 250. This plot shows a very strong correlation between  $\beta_{cbb}$  and  $\beta_{bcb}$ , which is very close to the Kleinman symmetry.

If we assume the Kleinman symmetry ( $\beta_{ijk} = \beta_{kij}$ ), then this number further reduces to 10. In Figure V.8 we check the validity of this assumption by plotting the correlation between such components (obtained by the QM/MM approach) related by the Kleinman symmetry at a frequency calculation of 800 nm. Even if the correlation is not perfect, we observe an extreme correlation: in a good approximation we can also assume that  $P(\beta_{ijk} = Y \text{ and } \beta_{jik} \neq Y) = 0$  at the optical frequency. However, we will stay in the static limit to avoid any mistake coming from this part: the components related by the Kleinman symmetry are thus strictly equal. In order to compute Equation V.26, we will assume only two possibilities for simplicity:

- If  $\beta_{ijk}$  is related to  $\beta_{abc}$  by Kleinman symmetry, their probabilities are fully correlated. For instance  $P(\beta_{abc} = Y \text{ and } \beta_{acb} \neq Y) = 0$  or  $P(\beta_{abc} = Y \text{ and } \beta_{bac} \neq Y) = 0$  or  $P(\beta_{abc} = Y \text{ and } \beta_{bac} \neq Y) = 0$  or  $P(\beta_{abc} = Y \text{ and } \beta_{acb} \neq Y) = P(\beta_{abc} = Y)$ .
- If  $\beta_{ijk}$  is not related to  $\beta_{abc}$  by Kleinman symmetry, their probabilities are independent. For instance:  $P(\beta_{caa} = Y \text{ and } \beta_{cbb} = Z) = P(\beta_{caa} = Y)P(\beta_{cbb} = Z)$

Doing so, we can find a simple analytical solution for the  $\langle \mathcal{B}^2 \rangle \rangle$  calculation while taking into account the  $\beta$  fluctuations – in the next subsection V. B.3 we will see the difference obtained without these approximations.

#### Numerical sampling of $\beta$ values:

Using this set of approximations, only 10 components represent the  $\beta$  tensor: *aaa*, *bbb*, *ccc*, *abb*, *acc*, *baa*, *bcc*, *caa*, *cbb* and *abc*.

We fit these different distributions with Gaussian functions:

$$P(\beta_{ijk} = Y) = A_{ijk} \exp\left[-\frac{1}{2} \left(\frac{Y - \langle \beta_{ijk} \rangle}{\sigma_{ijk}}\right)^2\right]$$
(V.27)

which describes quite well the  $\beta$  distributions. The obtained means and standard deviations are presented in Table V.1.

First, we assume that only the  $C_{2v}$ authorized components are relevant: *caa*, *cbb* and *ccc*. To compute the obtained distribution of  $\langle \mathscr{B}^2 \rangle$  we can perform a discretization of the  $\beta$  distributions and explore it iteratively to obtain  $\langle \mathscr{B}^2 \rangle$ . In Figure V.9 is presented the discretization of the 3 independents  $\beta$  which are  $C_{2v}$ -authorized. 151 points have been used, this means that  $151^3 = 3.4$  millions  $\beta$  tensor have been explored with for each possible  $\beta$  the associated probability:  $P(\beta) =$  $P(\beta_{caa} = X)P(\beta_{cbb} = Y)P(\beta_{ccc} = Z).$ 



Figure V.9: Independent probability for the three  $C_{2v}$ authorized  $\beta$  components. This represents the numerical sampling chosen to represent the distribution of Figure V.7. For each component, 151 points have been used to sample the distribution.



Figure V.10: Distributions of  $\langle \mathscr{B}^2 \rangle_{zxx}$  and  $\langle \mathscr{B}^2 \rangle_{zzz}$  when the  $\beta$  tensor presents fluctuations for the  $C_{2v}$ -authorized components, see Figure V.9. The averaged value for the zzz and zxx components are 8.3 and 2.6 atomic unit respectively.

For all of these 3.4 million  $\beta$ , we have computed the associated  $\langle \mathscr{B}^2[\beta] \rangle$  using Equations V.24. Figure V.10 presents the obtained distributions for  $\langle \mathscr{B}^2 \rangle_{zxx}$  and  $\langle \mathscr{B}^2 \rangle_{zzz}$ . These distributions show that the fluctuation of  $\beta$  also induces a large fluctuation for  $\langle \mathscr{B}^2 \rangle_{zzz}$ .

Using this approach, we find that  $\langle \mathscr{B}^2 \rangle_{zzz} = 8.3$  and  $\langle \mathscr{B}^2 \rangle_{zxx} = 2.6$  a.u., leading to a depolarization ratio of 0.32. Moreover, including only the  $C_{2v}$ -authorized component, we have doubled the total intensity generated by the incoherent term compared to the case neglecting  $\beta$ fluctuations. The fluctuations of  $\beta$  seem thus important: let us see what happens if we include all the possible components within this independent distribution hypothesis.

This explicit sampling method has the advantage to give the distributions of  $\langle \mathscr{B}^2 \rangle_{zxx}$  and  $\langle \mathscr{B}^2 \rangle_{zzz}$  when  $\beta$  fluctuates. However, it requires a lot of calculation which are not necessary to obtain the averaged value of  $\langle \mathscr{B}^2 \rangle$ ,  $\langle \langle \mathscr{B}^2 \rangle \rangle$ , as presented next. Moreover, to compute the distributions of  $\langle \mathscr{B}^2 \rangle$  for the 10 Kleinman independent components, the complexity of the calculation is of  $N^{10}$  – with N the numbers of points used to discretized each  $\beta$  component distribution.

#### Analytical solution:

Combining Equation V.26 and Equations V.24, we have in the general case:

$$<<\mathscr{B}^{2}>>_{zzz} = \frac{1}{105} \sum_{abc} \left[ 2 \int \beta_{abc}^{2} P(\beta) + \int \beta_{abb} \beta_{acc} P(\beta) + 4 \int \beta_{aab} \beta_{bcc} P(\beta) + 4 \int \beta_{aab} \beta_{bcc} P(\beta) + 4 \int \beta_{aab} \beta_{bcc} P(\beta) + 4 \int \beta_{abc} \beta_{bac} P(\beta) \right]$$
(V.28)

Note that in these expressions, the orientational averaging has already been performed. We can get a similar expression for  $\langle \mathcal{B}^2 \rangle \rangle_{zxx}$ . We are left with the averaging over the possible  $\beta$  values. To do so, we shall compute integrals like:  $\int \beta_{abc} \beta_{ijk} P(\beta)$ . Using the independent distribution hypothesis for the  $\beta$  components, there are two possibilities:

•  $\beta_{abc}$  and  $\beta_{ijk}$  are 2 components related by the Kleinman symmetry.

For instance (*aab*) and (*baa*) or (*bca*) and (*cab*). In this case,  $\beta_{abc}$  and  $\beta_{ijk}$  are always equal. Hence, this integral should be understood as:  $\int \beta_{abc}^2 P(\beta)$ . Since we have assumed that  $\beta_{abc}$  follows a Gaussian distribution, this integral is equal to  $\langle \beta_{abc} \rangle^2 + \sigma_{abc}^2$ .

•  $\beta_{abc}$  and  $\beta_{ijk}$  are 2 components **not** related by the Kleinman symmetry. For instance (*aaa*) and (*bbb*) or (*aab*) and (*abc*). In this case,  $\int \beta_{abc}\beta_{ijk}P(\beta)$  is the average over 2 (assumed) independent quantities. Hence,  $\int \beta_{abc}\beta_{ijk}P(\beta) = \langle \beta_{abc} \rangle \times \langle \beta_{ijk} \rangle$ .

Using these rules, we can compute  $\langle \mathscr{B}^2 \rangle_{zzz}$  and  $\langle \mathscr{B}^2 \rangle_{zxx}$  analytically knowing the independent distributions of the 10  $\beta$  components, presented in Table V.1. Including all the possible  $\beta$  components within this independent-distribution formalism, we find that  $\langle \mathscr{B}^2 \rangle_{zzz} = 21.1$  and  $\langle \mathscr{B}^2 \rangle_{zxx} = 6.3$  a.u., leading to a depolarization ratio of 0.30. This confirms that the fluctuations of **all** the  $\beta$  components are important and, in fact, dominant in the calculation of

 $<<\mathscr{B}^2>>$ . However, concerning the value of the depolarization ratio, we are still far from the experimental result.

# V. B.3 Hypothesis 3: Taking into account $\beta$ fluctuations and intramolecular correlations

To conclude this study of the incoherent contribution, we have also computed  $\langle \mathscr{B}^2 \rangle >$ without any assumption regarding the  $\beta$  distribution. The previous development aimed to gain some physical insights about what should or not should be included in the  $<<\mathscr{B}^2>>$ calculations, but are not needed for our data. Indeed, we rely on an important number of QM/MM calculations, which allows us to compute  $\langle \mathscr{B}^2 \rangle > brute$  force: by sampling the  $\beta$ and the orientation of the molecule at once. In Figure V.11 is plotted the independent distribution obtained by FROG for  $\mathscr{B}$ . These distributions have been obtained by computing directly  $\mathscr{B}$  for each molecule using its  $\beta$  and its orientation.



Figure V.11: First hyperpolarizability distribution in the laboratory frame for the water in the liquid phase at the QM/MM level in the zero-frequency limit. The distribution is not normalized: a total of 156 250 molecules have been used. This distribution takes into account correlation between  $\beta$  components if any.

Therefore, we can compute  $\langle \mathscr{B}^2 \rangle \rangle$  by taking the average of the square of the distribution shown in Figure V.11. The difference with respect to the previous case, where we consider all the  $\beta$  components within an independent approximation, is that we take into account the potential correlation between  $\beta$  components. Using this method, we obtain  $\langle \mathscr{B}^2 \rangle \rangle_{zzz} = 28.5$  and  $\langle \mathscr{B}^2 \rangle \rangle_{zxx} = 6.0$  a.u., leading to a depolarization ratio of 0.21, closer to the experimental value. Hence, there is clearly an impact of the correlations between the  $\beta$  components on the  $\langle \mathscr{B}^2 \rangle \rangle$ values.

## V. B.4 Comparison between all $\langle\langle \mathscr{B}^2 \rangle\rangle$ obtained with different hypotheses

Let us now compare the results obtained for the different methods. Table V.2 presents the obtained  $\langle \mathscr{B}^2 \rangle \rangle_{zzz}$ ,  $\langle \mathscr{B}^2 \rangle \rangle_{zxx}$  and depolarization ratio  $D = \langle \mathscr{B}^2 \rangle \rangle_{zxx} / \langle \mathscr{B}^2 \rangle \rangle_{zzz}$  for several hypotheses:

Table V.2: Averaged of the first hyperpolarizability squared in the laboratory frame for different hypotheses, see the main text. The ratio  $\langle \mathscr{B}^2 \rangle \rangle_{zxx} / \langle \mathscr{B}^2 \rangle_{zzz}$  is the depolarization ratio D in the fully-incoherent limit.

Case	$<<\mathscr{B}^2>>_{zzz}$	$<<\mathscr{B}^2>>_{zxx}$	D
Constant $\beta$	4.2	1.4	0.35
$\sigma$ only $C_{2v}$	8.4	2.7	0.32
$\sigma$ all	21.1	6.3	0.30
Correlated	28.5	6.0	0.21
Experimental	NA	NA	$0.14\pm0.01$

- The first case, "Constant  $\beta$ " corresponds to a Dirac distribution: the  $\beta$  components are represented by their average values.
- The second case, " $\sigma$  only  $C_{2v}$ " corresponds to the case where only the  $C_{2v}$ -authorized components are used (*ccc*, *caa* and *cbb*) with their standard deviations as presented in Table V.1.
- The third result, " $\sigma$  all", corresponds to the case where all the components are used with their relative standard deviations. Note that the  $C_{2v}$ -forbidden components participate only through their standard deviations since there are null in averaged.
- The fourth result, "Correlated", represents our best representation of the  $\beta$  distributions using directly the results from FROG, see Figure V.11.
- The experimental result obtained in the SHS experiment is recalled for the depolarization ratio. Note that we cannot have a value for  $\langle \mathscr{B}^2 \rangle >_{zzz}$  or  $\langle \mathscr{B}^2 \rangle >_{zxx}$  experimentally.

First of all, the  $\langle \mathscr{B}^2 \rangle \rangle$  increase with the number of  $\beta$  components included. The fluctuations of  $\beta$  tend to increase the total response thanks to the  $\beta^2$  term in Equation V.24. Similarly, when adding the  $C_{2v}$ -forbidden components, the  $\langle \mathscr{B}^2 \rangle \rangle$  increase dramatically. Thus, the fluctuations of all the  $\beta$  components have an important impact on the SHS intensity: assuming that water molecule has a constant hyperpolarizability, which perfectly respects the  $C_{2v}$  symmetry, is not valid for interpreting the incoherent term. Moreover, neglecting the fluctuation reduces the total Second Harmonic intensity: this may lead to the wrong ratio between the coherent and incoherent term later on. Finally, the results obtained with the assumption of independent distributions (' $\sigma$  all') and without ('Correlated') are different. This means that the correlation between the  $\beta$  components matters for the incoherent SHS intensity. We indeed observe some correlations between  $\beta$  components, for instance between  $\beta_{ccc}$  and  $\beta_{caa}$  as shown in Figure V.12.



Figure V.12: Correlation distribution of  $\beta_{ccc}$  and  $\beta_{caa}$ . On the right the distribution of  $\beta_{caa}$ . On the top the distribution of  $\beta_{ccc}$ . The total number of molecule used is 156 250. This plot shows a correlation between  $\beta_{ccc}$  and  $\beta_{caa}$ .

#### **Conclusions:**

To conclude this part regarding the incoherent contribution in the SHS signal for pure water, let us review what we have obtained:

- The  $\beta$  for water molecules in the liquid phase is a complex quantity. For a given molecule, we can expect that each component is non-zero and quite equivalent in amplitude. The  $C_{2v}$ is respected only on averaged. Moreover, the different components are correlated.
- Building the hyperpolarizability in the laboratory frame  $\mathscr{B}$  for a given  $\beta$  (with many components) is non-trivial. We rather work with the orientational averaged of  $\mathscr{B}^2$ ,  $\langle \mathscr{B}^2 \rangle$ , given by an analytical solution.
- By assuming independent Gaussian distribution for the 10 Kleinman-independent  $\beta$  components, we are able to compute analytically  $\langle \mathscr{B}^2 \rangle \geq \int \langle \mathscr{B}^2[\beta] \rangle P(\beta)$ . We observe

that including only the  $C_{2v}$ -authorized components underestimates greatly the  $\langle \mathscr{B}^2 \rangle \rangle$  compared to the results with all the components. Moreover, the role of the standard deviation of  $\beta$  seems preponderant.

• The value obtained for  $\beta$  with and without the independent approximation are close but still different: the correlations between the  $\beta$  components affect the  $\langle \mathcal{B}^2 \rangle \rangle$ .

Hence, when working with molecules with large  $\beta$  fluctuations, it is difficult to relate the experimental results to the molecular  $\langle \beta \rangle$ : we have to also include the fluctuations. Indeed, for pure water,  $\langle \mathcal{B}^2 \rangle \rangle$  is dominated by the fluctuations of **all** the  $\beta$  components, not only the ones authorized by the molecular symmetry. Moreover, it emphasizes the need to extract from numerical calculation not only the 'averaged'  $\beta$  components, but also their standard deviations. The best quantity to compute would be the **joint** probability for the  $\beta$  components since the correlations between components affect the results.

However, the calculation of the incoherent part does not allow to obtain the experimental depolarization ratio. We thus now question the effect of the coherent term on D.

# V. C | Short-Range Coherent Calculation

In this section we will present how we take into account the coherent contribution to the SHS intensity using our individual molecular response. We propose a new framework based on the first hyperpolarizability in the **laboratory** frame  $\mathscr{B}$  instead of the molecular one  $\beta$ . We are able to do so because we have the SHG-response of each molecule thanks to FROG.

Using this formalism, we will compute the impact of local structuration of liquid water up to the first and second solvation layers on the SHS signal. With our numerical approach, we are not able to study the correlation after the third solvation shell because of statistical noises. Finally, we will see the impact of this short-range coherent term on the SHS signal and we will compare to the experimental results.

### V. C.1 Molecular based Coherent calculation

In this section, we will develop the formalism used later on to predict the experimental signal from the molecular hyperpolarizabilities and molecular positions.

#### **Reminders:**

We have seen in the Section V. A.2 and particularly on Equation V.13 that we can express the experimental intensity using the laboratory first hyperpolarizability  $\mathscr{B}$  of each molecule. In  $I_V$  and  $I_H$  appear terms like:  $\sum_{n,m} \mathscr{B}^n_{ijk} \mathscr{B}^m_{abc} \exp[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}]$ . These terms can be reorganized using matrix with 6 components,  $Y_{ijk,abc}$ , so that:

$$Y_{ijk,abc} = \sum_{n,m} \mathscr{B}^n_{ijk} \mathscr{B}^m_{abc} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right]$$
(V.29)

$$=\sum_{n} \mathscr{B}_{ijk}^{n} \mathscr{B}_{abc}^{n} + \sum_{n \neq m} \mathscr{B}_{ijk}^{n} \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right]$$
(V.30)

$$=Y_{ijk,abc}^{\text{incoh}} + Y_{ijk,abc}^{\text{coh}} \tag{V.31}$$

Which leads to the following expression to describe the experimental signals:

$$I_V \propto Y_{yyy,yyy} \cos^4 \gamma + (2\text{Re}Y_{yyy,yxx} + 4Y_{yxy,yxy}) \cos^2 \gamma \sin^2 \gamma + Y_{yxx,yxx} \sin^4 \gamma$$
(V.32)

$$I_H \propto Y_{zyy,zyy} \cos^4 \gamma + (2\text{Re}Y_{zyy,zxx} + 4Y_{zxy,zxy}) \cos^2 \gamma \sin^2 \gamma + Y_{zxx,zxx} \sin^4 \gamma$$
(V.33)

In Section V. B we have seen how to compute  $Y^{\text{incoh}}$  and highlighted that the  $\beta$  fluctuations are important. Now let focus on the  $Y^{\text{coh}}$  calculations.

We can compute directly from our numerical calculations the Y tensor. However, an objective of this study is to disentangle the contributions of the different kinds of structural organizations in the liquid phase in the coherent SHS intensity. Therefore, let us work with the coherent part of Y,

#### Page 166

first by fixing a molecule n so that:

$$Y_{ijk,abc}^{coh} = \sum_{n} \mathscr{B}_{ijk}^{n} \left\{ \sum_{m \text{ with } n \neq m} \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right] \right\}$$
(V.34)

The vector between the molecule n and m is:  $\mathbf{r}_{nm} = \mathbf{r}_n - \mathbf{r}_m$ , and we note  $r = |\mathbf{r}_{nm}|$ . All the neighbors of a molecule n can be decomposed using the distance separating the molecule mfrom the molecule n, see Figure V.13. For instance, in Figure V.14 is plotted the average number of neighbors between r and r + dr. In this document, we noted this quantity N(r), and we use a width (dr) of 0.1 Å. Note that the point at N(r = 0) = 1 is the molecule nitself.



Figure V.13: Left: sketch of the liquid phase: here we highlight a particular molecule n and one neighbor m. The intermolecular distance is  $r_{nm} = R$ . Right: Scheme of all the molecules m within a distance R and R + dr of a particular molecule n.



Figure V.14: LEFT: Averaged number of neighbors as a function of the distance R, for a shell thickness of 0.1 Å. RIGHT: Radial Distribution Function (RDF), see Equation V.35. The red dots represent the first, second and third solvation shell at 2.7, 4.5 and 6.7 Å approximately. The vertical dashed lines are a guide for the eye: from the first to the second it represents the first solvation shell, from the second to the third the second shell. After the last line, there is the third solvation shell and the rest of the neighbors.

#### Decomposition with respect to neighbors distance:

Such decomposition with respect to the intermolecular distance is classic. One obvious example is the calculation of the Radial Distribution Function (RDF), see Figure V.14. The Radial Distribution Function (RDF) is defined to be:

$$RDF(r) = \frac{N(r)}{4\pi r^2 dr} \times \frac{1}{\rho_l}$$
(V.35)

With N(r) the number of neighbors found in a shell of radius r and width dr,  $\rho_l$  is the density of the liquid phase. The result is presented in Figure V.14 for bulk water at 300 K. There is no neighbor before 2.5 Å as expected (exclusion shell of the water molecule). The first, second and third solvation shells can be distinguished at approximately 2.7, 4.5 and 6.7 Å. As a guide for the eye, we have put red dots to represent the first, second and third solvatation shell. After 8 Å, we can no longer observe oscillations in the RDF – and thus defined a 4<sup>th</sup> solvation layer for instance. In other words, after this distance, the number of neighbors grows in  $r^2$ .

### Definition of the distance-correlation function of $\mathscr{B}$ : $C_{ijk,abc}(r)$

Therefore, similarly to the RDF, we define inter-molecular distance-based laboratory hyperpolarizability correlation function  $C_{ijk,abc}^n(r)$  so that:

$$C_{ijk,abc}^{n}(r) = \sum_{m \in S_{r}^{n}} \mathscr{B}_{ijk}^{n} \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right]$$
(V.36)

Where  $\sum_{m \in S_r^n}$  goes over all the neighbors m which are within a distance of r and r + dr from molecule n, see Figure V.13. The number of molecules m involved in this sum is given by N(r), see Figure V.14. Using this notation, we can show that: <sup>5</sup>

$$Y_{ijk,abc} = \sum_{n} \int_{r=0}^{+\infty} C^{n}_{ijk,abc}(r)dr = N \int_{r=0}^{+\infty} \langle C^{n} \rangle_{ijk,abc}(r)dr$$
(V.37)

and then:

$$Y_{ijk,abc}^{\text{incoh}} = \sum_{n} C_{ijk,abc}^{n}(r=0) \qquad \qquad Y_{ijk,abc}^{\text{coh}} = \sum_{n} \int_{0+}^{+\infty} C_{ijk,abc}^{n}(r) \qquad (V.38)$$

Let us study two simple limits for  $C^n$  when r goes to  $+\infty$ : if the molecule properties are completely "correlated" or "uncorrelated".

In the case where the molecules n and m are not correlated regarding both their relative orientation and their  $\beta$ , the  $C^n$  values become:

$$C_{ijk,abc}^{n}(r) = \sum_{m \in S_{r}^{n}} \mathscr{B}_{ijk}^{n} \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right] \approx N(r) \mathscr{B}_{ijk}^{n} < \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right] > \qquad (V.39)$$

And because the liquid is centro-symmetric,  $\langle \mathscr{B}^m_{abc} \rangle = 0$  and so  $C^n_{ijk,abc}(r) = 0$ .

On the contrary, if there are a strong correlations between the molecules n and m, for instance in a crystalline structure,  $C^n$  for a large  $r = R_{corr}$  becomes:

$$C_{ijk,abc}^{n}(R_{corr}) \approx N(R_{corr}) < \mathscr{B}_{ijk}^{n} \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right] >= N(R_{corr})A_{ijk,abc}[R_{corr}]$$
(V.40)

<sup>&</sup>lt;sup>5</sup>Note that in the following Equation V.37 there is a dr involved. This is needed for mathematically accuracy in the limit of infinitesimal shell width dr. In practice we present results obtained with a fixed width dr = 0.1 Å and the integral of Equation V.37 is performed by summing all the values obtained for the different dr.

Where  $A_{ijk,abc}$  is relatively constant for a range of inter molecular distance  $R_{corr}$ . Hence, if there is very strong correlation, the function  $C^n$  should grow as  $N(R_{corr})$ , while if there is no correlation it should go to zero.

To study the correlation at large distance, it can be useful to get rid of the N(r) evolution. For instance, we often define the "normalized" RDF for the density, see Equation V.35 and Table V.3. This quantity goes to 1 if there is no density correlations between molecules n and m instead of growing in a pure  $r^2$  shape. Thus, this renormalization shows better the deviation of the number of neighbors from the expected "uncorrelated" case.

In the same idea, we have defined a "normalized" correlation function, noted  $C^0$ :

$$C^{0}_{ijk,abc}(r) = \frac{1}{N(r)} < C^{n} >_{ijk,abc} (r)$$
(V.41)

This function will be used later to discriminate in the results the "statistical noise" from the "correlation".

To conclude this part, we present in Table V.3 the analogy between the C functions and the radial distribution functions. In order to link the microscopic quantities, the  $\mathscr{B}$ s, to the experimental intensity we have to compute the Y tensor, see Equation V.32. Using the C functions, we can compute the Y using Equations V.37. Several definitions of C will be used. In practice we first compute  $C^n$  for each molecule n. Then we compute the average,  $\langle C^n \rangle$ , over all molecules n. Then, we filter the noise thanks to  $C^0$  – see the following subsection. Finally, we compute Y using the filtered  $\langle C^n \rangle$  according to Equation V.37 in order to compare with the experimental results.

Table V.3: Definition of the different C functions used along with their density "equivalent". See Figure V.14 for N(r) and the RDF.

Averaging type	C definition	RDF equivalent
Single molecule	$C_{ijk,abc}^{n}(r) = \sum_{m \in S_{r}^{n}} \mathscr{B}_{ijk}^{n} \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right]$	For a given $n$ , $\sum_{m \in S_r^n}$
Averaged	$< C^n >_{ijk,abc} (r) = \frac{1}{N} \sum_n C^n_{ijk,abc}(r)$	$N(r) = \frac{1}{N} \sum_{n \in S_r^n} \sum_{m \in S_r^n}$
"Normalized"	$C^0_{ijk,abc}(r) = \frac{1}{N(r)} < C^n >_{ijk,abc} (r)$	$\operatorname{RDF}(r) = \frac{N(r)}{4\pi r^2 dr} \times \frac{1}{\rho_l}$

#### V. C.2 Laboratory first hyperpolarizability spatial correlation function C(r)

In this part we will discuss the obtained value for  $N < C^n >_{ijk,abc} (r)$ , which is directly linked to Y, for a sample composed of 10 frames of N = 15 625 molecules. The first hyperpolarizability of each molecule has been computed at a QM/MM level by including the neighborhood up to a distance of 20 Å– the first 10 Å explicitly and the last 10 Å implicitly, see Chapter IV. The calculation wavelength is 800 nm for the fundamental field. For each molecule of the 10 MD frames,  $C_{ijk,abc}^n(r)$ have been computed for a distance r up to 30 Å, the binarization is made every 0.1 Å.

#### Raw data:

The total value  $N < C^n >_{ijk,abc} (r) = \sum_n C_{ijk,abc}^n(r)$  for one time step is plotted in Figure V.15. The integral of  $N < C^n >$  over all the r, and over all the time steps, should be exactly Y, see Equation V.37. The point at r = 0 is the incoherent contribution <sup>6</sup>, and all the rest is the coherent contribution. We note that the first and second shell correspond to a strong signal: until approximately 5 Å  $N < C^n >$  is clearly none null. Between 5 and 15 Å,  $N < C^n >$  is very small, and after 15 Å it seems to increase again, especially for the yyy, yyy component.



Figure V.15: LEFT: Real part of  $C^T = N < C^n >$  for several (ijk, abc) couples for one time step. RIGHT: Effect of  $\Delta \mathbf{q}$  terms for the real part of  $C_{yyy,yyy}^T = N < C^n >_{yyy,yyy}$  for one time step. The difference is plotted with and without the  $\Delta \mathbf{q} = 2.09 \times 10^{-3} (\hat{z} - \hat{x}) \text{ Å}^{-1}$  inside the exponential terms. Comparing the amplitude of the two Figures, one can clearly see that the phase effect due to  $\Delta \mathbf{q}$  is negligible for r about the Å.

We have seen that if  $\langle C^n \rangle (r)$  is non-vanishing, it means that the first hyperpolarizabilities in the laboratory frame of the molecule distance for r are correlated – See Appendex IX. C.4. For small distances it makes sense: the first solvation shell is correlated to the core molecule. For instance, the neighbors orientation is so that they have H-bonds with the core molecule. Hence, the non-zero value obtained for C(r) over the first and second solvation would mean that the SHG response in the laboratory frame of molecules within 2 solvation shells are correlated. The zero value between 5 and 15 Å would mean that the SHG response is no longer correlated. We expect the correlation to

 $<sup>^{6}</sup>$  under proper renormalization, and with the same frequency, we find exactly the same result for r=0 as for the last line of Table V.2

be smaller and smaller as the distance between molecule increases: hence the increase of  $N < C^n >$  after 15 Å is counterintuitive and requires more attention.

### About $\Delta q$ :

One explanation of the behavior could be due to the phase factor  $\exp[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}]$  of  $N < C^n >$ , – see Equation V.36. Hence, the increase of  $N < C^n >$  after 15 Å may be due to the fact that  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm}$  starts to have a impact. To verify this hypothesis, right part of Figure V.15 plots the difference in the obtained  $N < C^n >_{yyyyyy}$  with the physical  $\Delta \mathbf{q} = 2.09 \times 10^{-3} (\hat{z} - \hat{x})$  Å<sup>-1</sup> and without ( $\Delta \mathbf{q} = \mathbf{0}$ ).

There is a difference, but it is extremely small compared to the value obtained for  $N < C^n >$ , presented in Figure V.15. Hence, assuming that  $\Delta \mathbf{q} = 0$  is a good approximation in our case. Therefore, this dephazing cannot explain the increase of  $N < C^n >$  after 15 Å. Note that in the following we will drop the exp $[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}]$  for shorter writing.

#### How to get rid of the statistical noise:

In order to explain the increase of  $N < C^n >$ , we compare the results obtained and the one due solely to a statistical noise. In a purely decorrelated case, the expected value of  $N < C^n >$  is:

$$N < C^n >_{ijk,abc} (r) = \sum_n \sum_{m \in S_r} \mathscr{B}^n_{ijk} \mathscr{B}^m_{abc} = \sum_n \mathscr{B}^n_{ijk} \sum_{m \in S_r} \mathscr{B}^m_{abc}$$
(V.42)

$$=\sum_{n} \mathscr{B}_{ijk}^{n} \left( N(r) < \mathscr{B}_{abc}^{m} > \right) = \left( N < \mathscr{B}_{ijk} > \right) \times \left( N(r) < \mathscr{B}_{abc} > \right)$$
(V.43)

$$= NN(r) < \mathscr{B}_{ijk} > < \mathscr{B}_{abc} > \tag{V.44}$$

Here we have used the fact that the value of  $\sum_{m \in S_r} \mathscr{B}^m_{abc}$  do not depend on the value of  $\mathscr{B}^n_{ijk}$ .

For large enough samples, we expect that  $\langle \mathscr{B}_{ijk} \rangle = \langle \mathscr{B}_{abc} \rangle = 0$ , due to the centro symmetry. This is indeed true of our total sample composed of 10 frames:  $|\langle \mathscr{B}_{ijk} \rangle|$  is less than 0.1 a.u. for any component. For a single frame, this value can be larger, see Figure IX.15 of Appendix IX. C.5. Therefore, for a single frame,  $\langle \mathscr{B}_{ijk} \rangle \langle \mathscr{B}_{abc} \rangle \neq 0$  and this has a dramatic impact: if we assume that the estimation of  $\langle \mathscr{B}_{ijk} \rangle$  and  $\langle \mathscr{B}_{abc} \rangle$  does not depend on r, the  $N \langle C^n \rangle_{ijk,abc}$  (r)increases with r because N(r) goes as  $r^2$ .

Hence, we have shown that  $N < C^n >_{ijk,abc} (r)$  can be very impacted by the noise due to the lack of sampling. A better quantity would be:

$$C^{0}_{ijk,abc}(r) = \frac{1}{NN(r)} N < C^{n} >_{ijk,abc} (r)$$
(V.45)

In the fully uncorrelated case,  $C_{ijk,abc}^{0}(r)$  should tend to  $\langle \mathscr{B}_{ijk} \rangle \langle \mathscr{B}_{abc} \rangle$ , which can be nonzero in real calculation due to the lack of sampling. Moreover, we can quantify the fluctuation by assuming that  $\mathscr{B}^{n}$  and  $\mathscr{B}^{m}$  are 2 independent random variables, following a Gaussian distribution. Hence, we can fit the associated average and standard deviation for both  $\mathscr{B}^{n}$  and  $\mathscr{B}^{m}$ . Within this independent hypothesis, we can predict the average and the standard deviation of this new random variable  $\mathscr{B}^n \times \mathscr{B}^m$  according to:

$$\langle \mathscr{B}^{n}_{ijk} \times \mathscr{B}^{m}_{abc} \rangle = \langle \mathscr{B}^{n}_{ijk} \rangle \langle \mathscr{B}^{m}_{abc} \rangle = \langle \mathscr{B}_{ijk} \rangle \langle \mathscr{B}_{abc} \rangle$$
(V.46)

$$\sigma\left[\mathscr{B}^{n}\times\mathscr{B}^{m}\right] = \left(\sigma\left[\mathscr{B}^{m}\right]^{2}\sigma\left[\mathscr{B}^{m}\right]^{2} + \langle\mathscr{B}^{n}\rangle^{2}\sigma\left[\mathscr{B}^{m}\right]^{2} + \langle\mathscr{B}^{m}\rangle^{2}\sigma\left[\mathscr{B}^{n}\right]^{2}\right)^{\frac{1}{2}}$$
(V.47)

Note that for the second equation we have dropped the (ijk, abc) for readability. Hence,  $C^0_{ijk,abc}(r)$  is an estimation of the mean value of the random variable  $\mathscr{B}^n \times \mathscr{B}^m$  with a number of configurations NN(r):

$$C^{0}(r) = \langle \mathscr{B}^{n} \times \mathscr{B}^{m} \rangle_{|NN(r)} \qquad (V.48)$$

Thanks to the central theorem, we know that  $C_{ijk,abc}^{0}(r)$  should be within  $\left[\mu - \frac{2\sigma}{\sqrt{n}}, \mu + \frac{2\sigma}{\sqrt{n}}\right]^{7}$  with an uncertainty of 95 %, with  $\mu$  and  $\sigma$  the average and standard deviation of the product defined in Equation V.46.

In Figure V.16 is plotted  $C^0(r) - \langle \mathscr{B}^n \times \mathscr{B}^m \rangle$ as a function of r for the component yyy, yyy(one of the largest). If we assume that we are



Figure V.16: Real part of  $C^0$  as a function of the distance R minus the expected noise due to the sampling  $<\mathscr{B}\times\mathscr{B}>$  for the yyy, yyy component and one time step. The red lines correspond to a confidence interface at 95 %: $\pm \frac{2\sigma[\mathscr{B}^n \times \mathscr{B}^m]}{\sqrt{NN(r)}}$ , and is the expected noise due to the number of configurations used NN(r).

in the uncorrelated case, we can expect  $|C^0(r) - \langle \mathscr{B} \times \mathscr{B} \rangle |$  to be smaller than  $\frac{2\sigma[\mathscr{B}^n \times \mathscr{B}^m]}{\sqrt{NN(r)}}$ <sup>8</sup>. In Figure V.16 is plotted in red this expected noise that comes from the lack of statistical sampling. Note that this noise is not relevant for the first and second solvatation shells were the RDF shows correlation  $^{9}$ .

For the time being, we will assume that the value obtained at a given r is "noise" if the amplitude of  $C^0(r) - \langle \mathscr{B}^n \times \mathscr{B}^m \rangle$  is below  $\frac{2\sigma[\mathscr{B}^n \times \mathscr{B}^m]}{\sqrt{NN(r)}}$ , and "signal" above. Figure V.17 shows for each distance r the number of "signal" and "noise" found over the 10 frames. Before 4 Å, the value of  $|C^0(r) - \langle \mathscr{B} \times \mathscr{B} \rangle|$  is way above the one expected by this sampling error. It is safe to assume that the intensity in this short-range area corresponds to a correlation and not to statistical noise. At approximately 5 Å, there is almost as many points in error bars than above. Note that if we take  $2 \times \frac{2\sigma[\mathscr{B}^n \times \mathscr{B}^m]}{\sqrt{NN(r)}}$ , corresponding to 97.5 % for uncertainty, for the noise almost all the points after

<sup>&</sup>lt;sup>7</sup>In our case we have at least NN(r) > 15000.

<sup>&</sup>lt;sup>8</sup>For simplicity we have assumed that the estimation of the average and standard deviation of  $\mathscr{B}$  do not depend on N(r). Assuming otherwise would increase this noise at small r compared to large r.

<sup>&</sup>lt;sup>9</sup>The small variation close to 5 Å is due to the variation of N(r) since N(r) is not monotonous in this area, see Figure V.14.

5 Å are considerate to be noise. This is make sense with the result obtained by the RDF: almost no density correlation can be seen after 7 Å.

we can rationalize the variation of Hence. N<  $C^n$ >(r) at large r: it may be due to statistical errors. Moreover, if we assume that the  $C^0(r)$  intensity after 7 Å is due to correlation, we would not be able to prove it because we are within the statistical noise. Therefore, we will assume that  $C^0(r > 7 \text{ Å}) = 0$  for sim-In other words, we will consider plicity. only the effect of the first, second and third solvation layer because we are almost sure that the obtained C is above the statisti-Studying the possible correlacal noise. tion for larger distance require more accuracy.



Figure V.17: For every bin used to discretized the distance between molecule n and m, declare if the resulting  $C_{yyyyyy}^0(r)$  could be considered as a "signal" or "noise" for all the frame. 10 frames have been used so that the sum over the noise and signal is equal to 10

In top of that, we will still apply a filtering: if for a given time step and a given r,  $C^0(r) - \langle \mathscr{B}^n \times \mathscr{B}^m \rangle$  is smaller than  $\frac{2\sigma[\mathscr{B}^n \times \mathscr{B}^m]}{\sqrt{NN(r)}}$ , we will set this value to zero. Hence, we will still analyze the result obtained from 5 to 7 Å, but keeping in mind that we are starting to reach the limit of our sampling.

Hence, we have found an explanation for the increase of  $N < C^n >$  at large r: it is due to the fact that the estimated value of  $\langle \mathscr{B} \rangle$  over a given MD frame is not perfectly zero. Using an independent assumption, we have shown that our results are within the statistical error for r > 7Å, and thus will not be discussed.

### Short-range C(r) and Y:

In Figure V.18 is plotted the "cleaned"  $< C^n >$  for the QM/MM results. Several components related to the V outcoming polarization are presented.

As expected, the component yxx, yxx and yxy, yxy are equal due to centro-symmetry and in the limit where  $\Delta \mathbf{q} \cdot \mathbf{r} = 0$  – which is respected for r about the Å. Similarly, other components that should be zero, for instance yyy, zyy or yxx, yxy or zyy, zxy or zxx, zxy, are indeed null throughout all r – data not shown.

The value obtained at r = 0 is the incoherent contribution, which is the dominant value. However, the first shell also contributes a lot to the signal: even if its value is smaller, it implies more couples, N(R), than the incoherent one, only 1 because n = m. As the r increases, this contribution becomes smaller and smaller. The value of  $\langle C^n \rangle$  is positive for all the r in this short-range domain. We do not have any explanation for this.



Figure V.18: Evolution of  $\langle C^n \rangle$  for the QM/MM  $\beta$  once the statistical noise removed. On the left the  $\langle C^n \rangle$  and on the right the integral. Several components are plotted corresponding to V-related outcoming polarization contributions. The vertical dashed lines are a guide for the eye: from the first to the second it represents the first solvation shell, from the second to the third the second shell and after the last line the third solvation shell.

In Figure V.18 is also plotted the integral of  $\langle C^n \rangle$  over r. This quantity is even more related to the experimental one:  $Y = Y^{\text{incoh}} + Y^{\text{coh}} \propto \langle C^n(r=0) \rangle + \int_{0+}^{+\infty} \langle C^n(r) \rangle dr$ . In this plot we can observe that the impact of the first solvation shell (from 2.3 Å to approx 3.3 Å) is of the same order of magnitude as the incoherent terms on the total SHS response. The second solvation shell also contributes for the yyy, yyy and less for the other components. The impact of the third shell, after 5.5 Å approximately, is almost negligible in the total amplitude. However we remind that the data obtained for this third shell starts to be affected by the statistical noise.
# V. C.3 Impact of intermolecular hyperpolarizability correlations on the depolarization ratio

Now, let us discuss more about the impact of the short-range correlations on the polarized resolved SHS intensity in order to compare with the experimental results. First, we briefly show the results obtained when setting the same first hyperpolarizability to all molecules, then using our QM/MM results.

# Results using $\beta^n = <\beta>$ :

As for the for the section dealing with the incoherent contribution to the signal, we have compared the results obtained by the QM/MM results, where each molecule can have a different  $\beta$ , with the one obtained by setting the molecular hyperpolarizability of all the molecules to the average one:  $\beta^n = \langle \beta \rangle$ . The results are presented extensively in the Appendix IX. C.6. Briefly, we found that the  $\langle C^n \rangle$  does not match at all the one obtained using the QM/MM results, Figure IX.18. Moreover, the predicted signal, presented in Figure IX.19, does not match the experimental one. The error from the  $\beta^n = \langle \beta \rangle$  hypothesis is first due to the incoherent term, see Table V.2, but also from the coherent one, as emphasized by Figure IX.18.

Therefore, the  $\beta^n = \langle \beta \rangle$  is not a good approximation to compute the SHS response of pure water regarding the incoherent nor the short-range coherent contributions.



Figure V.19: Evolution of the a, b, c parameters for the V, on the left, and H, on the right, outcoming polarization for increasing coherent contributions. Due to centro-symmetry, we expect that  $a^H = c^H = b^H/2$ . The vertical dashed lines are a guide for the eye: from the first to the second it represents the first solvation shell, from the second to the third the second shell and after the last line the third solvation shell.

# Results using $\beta$ QM/MM:

Now, let us present the results obtained with QM/MM values. First, in Figures V.19 are plotted the evolution of the a, b, c components (see Equation V.1) for the V or H outcoming polarizations depending on the amount of correlation taken into account. These parameters link the Y tensor to the SHS intensity – see Equations V.18 and V.1 or Appendex IX. C.4. We recover that  $a^H = c^H = b^H/2$  in Figure V.19 as expected by the centro-symmetry. Note that the small deviation from this expected behavior after  $r \approx 5$ Å is a sign that we start lacking statistical sampling for these large r: this confirms the trend obtained from the previous part regarding the statistical noise. As these parameters are linear combination of Y, the evolution is extremely similar to the one of  $\int_0^R \langle C^n(r) \rangle$  in Figure V.18. The values obtained only for the incoherent term (r = 0)provides the correct amplitude and order among the different parameters. However, the first and second solvation layers have an important impact on the amplitude of each component and most importantly on the relative amplitude between them.

Indeed, in Figure V.20 is plotted the ratio between  $a^V$  and  $c^V$ , which is exactly the depolarization ratio:  $D = \min(I_V)/\max(I_V) = c^V/a^V$ . This ratio evolves dramatically within the first solvation layer, but also the second and until the third one. This important result recalls that if intermolecular contributions are significant, the depolarization ratio cannot be linked to the molecular symmetry in a simple manner. We obtained that the majority of the ratio is indeed fixed by the incoherent part, but still the coherent one plays an important role for the depolarization ratio of water.

This effect is also visible on the SHS signal depending on the incoming polarization shown in Figure V.21. Figure V.21 compares the SHS intensities for the V and H outcoming polarization considering the incoherent contribution only or the incoherent & the coherent (until 7 Å)



Figure V.20: Evolution of the depolarization ratio when increasing coherent contribution. The horizontal line is the experimental value:  $D = 0.14 \pm 0.01$  for water at room temperature. The vertical dashed lines are a guide for the eye: from the first to the second it represents the first solvatation shell, from the second to the third the second shell and after the last line the third solvatation shell.

contributions and compares with the experimental signal <sup>10</sup>.

The V outcoming polarization predicted by the incoherent + coherent response matches the experimental signal with adjustable parameters. This is in agreement with the predicted depolarization ratio: experimentally we have  $D = 0.14 \pm 0.01$  as for the value obtained for the incoherent plus coherent part, see Figure V.20. If we use only the incoherent part, the agreement is worst.

However, the H outcoming polarization is not retrieved by the numerical approach: in all the cases we have a flat intensity with respect to  $\gamma$ , but experimentally the H intensity is lower than the V one. Numerically, we have always  $\min(I_V) = I_H$ . This due to the fact that our calculations respect the centro-symmetry. Indeed, as we are computing the coherent part in an area where  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} \approx 0$ . Hence, it is as if the whole system (structure and electromagnetic field) is centro-symmetric: which

<sup>&</sup>lt;sup>10</sup>For a given set of data, the V and H intensities have been renormalized by  $\max(I_V)$  as we have only access to relative intensity experimentally.



Figure V.21: SHS intensity depending on the incoming polarization angle  $\gamma$  for the V and H outcoming polarization, in blue and red respectively. The blue dots are the V experimental results and the red stars the H. The dashed lines the predicted signals using only the incoherent part using QM/MM  $\beta$ s. The full lines are the predicted signals of the incoherent and coherent part using QM/MM  $\beta$ s.

imposes  $\min(I_V) = I_H$ . We will discuss this last part as a perspective in the following Section V. C.4.

# V. C.4 First steps toward long-range correlations

In the previous part, we have seen that using the QM/MM results and including the coherent term up to the third solvation layer, we were able to obtain a good agreement with the experimental data. However, we still have a discrepancy regarding the gap between the V and H outcoming polarization plots, see Figure V.21. A possible explanation is that our coherent calculation is limited to the 'short-range" case. Indeed, the maximal distance between molecules involved in our calculation is 7 Å: within this range, we have almost no effect of the dephazing term, *i.e.*  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} \approx 0$ . In the following, I will first present why in this limit,  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} \approx 0$ , we cannot have the gap observed experimentally. Then, I will propose how to handle this issue with the help of existing formalism in the recent literature [43, 44].

# Consequences of $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} = 0$ :

This case represents the infinite wavelength limit or the case where the structural correlation of the liquid is largely smaller than the optical wavelength. In this limit, the coherent term loses its phase factor compared to its initial definition in Equation V.22 and is now equal to:

$$Y_{ijk,abc}^{coh} = \sum_{n \neq m} \mathscr{B}_{ijk}^n \mathscr{B}_{abc}^m \tag{V.49}$$

In this case, the incoming fundamental and the second harmonic emitted light are no longer described by their polarization **and** their propagation direction and phase velocity, but only by their polarization, see Figure V.22. Hence, if we assume that the liquid is centro-symmetric at the mescosopic scale, the only meaningful parameter is the angle between the incoming field polarization and the outcoming one. For instance, for a fundamental field along the x direction and an outcoming polarization in the x direction, we should obtain the same results as if the fundamental is in the y and the outcoming polarization along the y. This is not the case in the SHS experiment because the 3 axis are defined by the incoming field propagation direction, along +z, and by the outcoming propagation direction, along +x.

We are interested in the intensity obtained for a fundamental polarization at  $\gamma = \pi/2$  for the V and H outcoming polarization, given by:

$$I_V(\pi/2) = Y_{yxx,yxx}$$
  $I_H(\pi/2) = Y_{zxx,zxx}$  (V.50)

The incoming field polarization is along x, and the outcoming polarization along the y or z: the solid angle between the incoming and outcoming polarization is the same,  $\pi/2$ . If  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} = 0$ , this angle is the only parameter: thus the two intensities should be strictly equal. Hence:  $Y_{yxx,yxx} = Y_{zxx,zxx}$ . Note that using a similar argument, we can also find that  $I_H$  is expected to be constant for any  $\gamma$ .

Therefore, we obtain that:

If 
$$\Delta \mathbf{q} = 0 \rightarrow I_V(\pi/2) = I_H(\pi/2) = I_H(\gamma)$$
 (V.51)



Figure V.22: SHS intensity collected either along the V, a) and c), or H, b) and d), outcoming polarization for a fundamental polarization along the x axis (corresponds to  $\gamma = \pi/2$  in Figure V.21). We consider two cases: Either  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} \neq 0$ , a) and b), or  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} = 0$ , c) and d). If  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} \neq 0$ , the laboratory frame x, y, z is defined by the propagation of the fundamental and second harmonic. Hence, the a) and b) situation are different. If  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} = 0$ , the laboratory frame is less constrained: one can rotate the laboratory frame along the x axis. Therefore, the c) and d) polarization are in fact the same situation, and thus will be the SHS intensity for the V and H polarization for this fundamental polarization.

Since we are including the coherent term only up to few Angstrom, we are in this limit. Hence, using the dipolar approximation together with a short correlation range, we cannot reproduce the experimental gap between the V and H polarizations.

## How to include the long-range correlation?

Therefore, one way to solve this issue could be to include "long-range" correlations. Let us write again the coherent calculation using continuous equations. For simplicity, we study the  $\gamma = \pi/2$  case for the V outcoming polarization:

$$I_V^{\text{coh}}(\gamma = \pi/2) = \sum_{n \neq m} \mathscr{B}_{yxx}^n \mathscr{B}_{yxx}^m \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}_{nm}\right]$$
(V.52)

$$\propto \int_{r=0^+}^{+\infty} dr \int d\mathscr{B}^n \int d\mathscr{B}^m P[r](\mathscr{B}^n_{yxx}, \mathscr{B}^m_{yxx}) \mathscr{B}^n_{yxx} \mathscr{B}^m_{yxx} \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}\right]$$
(V.53)

# Page 179

Where we have defined  $P[r](\mathscr{B}_{yxx}^n, \mathscr{B}_{yxx}^m)$  the **joint** probability of finding 2 specific  $\mathscr{B}_{yxx}$  values for the *n* and *m* molecules distant from  $r = |\mathbf{r}| = |\mathbf{r}_n - \mathbf{r}_m|^{-11}$ . For writing simplicity, we will remove the index  $_{yxx}$  in the following. The integrals  $\int d\mathscr{B}^n$  and  $\int d\mathscr{B}^m$  go over the possible values of the first hyperpolarizabilities in the laboratory frame for the molecules *n* and *m* respectively (which is an exploration over the molecule's orientations and molecular first hyperpolarizability values). In Section V. B we have seen that for a given  $\beta$  tensor, the  $\mathscr{B}$  is given by the molecular orientation defined by the Euler angle:

$$\mathscr{B}[\beta]_{ijk}(\theta,\psi,\phi) = \sum_{ijk} R_{ia}(\theta,\psi,\phi) R_{jb}(\theta,\psi,\phi) R_{kc}(\theta,\psi,\phi) \beta_{abc} \qquad \mathscr{B}[\beta](\Omega) \equiv R^3(\Omega) \cdot \beta \qquad (V.54)$$

Where we have defined the solid angle  $\Omega$  to represent the molecular orientation –  $\Omega$  has 3 components. We have also defined the compressed notation  $R^3(\Omega) \cdot \beta$  as a shorter notation. Using these notations:

$$\mathscr{B}^{n}\mathscr{B}^{m} = \left(R^{3}(\Omega^{n})\cdot\beta^{n}\right)\left(R^{3}(\Omega^{m})\cdot\beta^{m}\right) \tag{V.55}$$

$$\int d\mathscr{B}^n \int d\mathscr{B}^m P[r](\mathscr{B}^n, \mathscr{B}^m) = \int d\Omega^n \int d\Omega^m \int d\beta^n \int d\beta^m P[r](\Omega^n, \Omega^m, \beta^n, \beta^m)$$
(V.56)

Which leads to  $^{12}$ :

$$I^{\text{coh}} \propto \int_{r=0^{+}}^{+\infty} dr \int d\Omega^{n} \int d\Omega^{m} \int d\beta^{n} \int d\beta^{m} P[r](\Omega^{n}, \Omega^{m}, \beta^{n}, \beta^{m}) \left(R^{3}(\Omega^{n}) \cdot \beta^{n}\right) \left(R^{3}(\Omega^{n}) \cdot \beta^{n}\right) \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}\right]$$
(V.57)

This Equation V.57 represents exactly the same thing as the one used previously, see for instance Equation V.37. The coherent intensity is related to the correlation between the first hyperpolarizability,  $\beta$ , and the orientation,  $\Omega$ , of molecules n and m over all the possible  $r_{nm}$  distances. Hence, we have to perform the sampling over all the possible orientations and  $\beta$  values for the n and mmolecules, with the associated probability  $P[r](\Omega^n, \Omega^m, \beta^n, \beta^m)$ .

In order to perform this sampling, one useful hypothesis is to assume that the first molecular hyperpolarizabilities of molecule n and m are not correlated. In this case:  $P[r](\Omega^n, \Omega^m, \beta^n, \beta^m) =$ 

<sup>&</sup>lt;sup>11</sup>This probability depends on  $r = |\mathbf{r}|$  and not  $\mathbf{r}$  because of the isotropicity of the fluid

 $<sup>{}^{12}\</sup>mathbf{r}$  is not affected by the orientation of the molecules n and m

 $P[r](\Omega^n, \Omega^m)P(\beta^n)P(\beta^m)$ , which simplifies the calculations dramatically:

$$I^{\text{coh, LR}} \propto \int_{r=0^+}^{+\infty} dr \int d\Omega^n \int d\Omega^m \int d\beta^n \int d\beta^m P[r](\Omega^n, \Omega^m) P(\beta^n) P(\beta^m) \left(R^3(\Omega^n) \cdot \beta^n\right)$$
(V.58)

$$\times \left( R^3(\Omega^n) \cdot \beta^n \right) \exp\left[ i \Delta \mathbf{q} \cdot \mathbf{r} \right] \tag{V.59}$$

$$\propto \int_{r=0^+}^{+\infty} dr \int d\Omega^n \cdot \int d\Omega^m P[r](\Omega^n, \Omega^m) \left( R^3(\Omega^n) \cdot \left[ \int d\beta^n \beta^n P(\beta^n) \right] \right)$$
(V.60)

$$\times \left( R^3(\Omega^m) \left[ \int d\beta^m \beta^m P(\beta^m) \right] \right) \exp\left[ i \Delta \mathbf{q} \cdot \mathbf{r} \right]$$
(V.61)

$$I^{\text{coh, LR}} \propto \int_{r=0^+}^{+\infty} dr \int d\Omega^n \cdot \int d\Omega^m P[r](\Omega^n, \Omega^m) \left(R^3(\Omega^n) \cdot <\beta>\right) \left(R^3(\Omega^m) \cdot <\beta>\right) \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}\right]$$
(V.62)

Using this approximation, the sampling over the  $\beta$  is replaced by the averaged value. This is convenient because it reduces the number of  $\beta$  components involved and paves the way for an analytical calculation.

Recent works [43, 114] have used these equations to link the SHS intensity to the molecular orientation using this approximation through the orientational correlation functions  $P[r](\Omega^n, \Omega^m)$ . More precisely, the above equation can be rewritten in the reciprocal space:

$$\int d\mathbf{r} P[r](\Omega^n, \Omega^m) \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}\right] = \rho^{\Delta \mathbf{q}}(\Omega^n, \Omega^m), \qquad (V.63)$$

so that:

$$I^{\text{coh, LR}} \propto \int d\Omega^n \cdot \int d\Omega^m \rho^{\Delta \mathbf{q}}(\Omega^n, \Omega^m) \left( R^3(\Omega^n) \cdot <\beta > \right) \left( R^3(\Omega^m) \cdot <\beta > \right).$$
(V.64)

Calculations based the Ornstein-Zernike Equations, see for instance [43], can provide the  $\rho^{\Delta \mathbf{q}}(\Omega^n, \Omega^m)$  values with great accuracy once provided the intermolecular interactions, for instance dipole-dipole interactions – whereas it is extremely difficult using classical MD. Moreover, recent works are based on this formalism to extract from the SHS intensity a trace of the long-range correlation in liquids [4, 44, 45].

# Short-range QM/MM accuracy and long-range accessibility: separation of the coherent term

Our approach, based on the direct space and QM/MM calculations, is able to describe the depolarization ratio of water. However we are limited in the correlation distance we can handle. The formalism described above can be a good path to take the long-range correlations into account, but it would be difficult to include the hyperpolarizability correlations observed at short range, and for the incoherent term. Hence, we propose to decompose the coherent terms into three parts:

$$I^{\text{coh}} \propto I^{\text{coh}, QM/MM}[l_c] + \left(I^{\text{coh}, LR} - I^{\text{coh}, SR < \beta >}[l_c]\right)$$
(V.65)

 $I^{\text{coh, }QM/MM}[l_c]$  is the coherent term obtained from the QM/MM results based until a distance  $l_c$ , typically 6 Å. In this range, there are correlations between the orientation of molecules and their SHS response and thus QM/MM calculations are required.  $I^{\text{coh, }QM/MM}$  can be computed in the direct space, for instance using trajectories obtained by MD. In this range, the phase factor  $\Delta \mathbf{q} \cdot \mathbf{r}$  can be neglected.

 $I^{\text{coh, }LR}$  is the coherent term computed for  $r \to +\infty$  assuming that each molecule has the same  $\beta = <\beta >$ . This calculation can be done in the reciprocal space using analytical tools already established [4, 43, 114].

 $I^{\text{coh, }SR<\beta>}[l_c]$  compensates from the error due to the second term in the short-range area where the first hyperpolarizability are correlated. Let us find the expression for  $I^{\text{coh, }SR<\beta>}[l_c]$ .

$$I^{\text{coh}} \propto \int_{r=0^+}^{+\infty} dr \int d\mathscr{B}^n \int d\mathscr{B}^m P[r](\mathscr{B}^n, \mathscr{B}^m) \mathscr{B}^n \mathscr{B}^m \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}\right]$$
(V.66)

$$\propto \left[ \int_{r=0^+}^{l_c} + \int_{r=l_c}^{+\infty} \right] dr \int d\mathscr{B}^n \int d\mathscr{B}^m P[r](\mathscr{B}^n, \mathscr{B}^m) \mathscr{B}^n \mathscr{B}^m \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}\right]$$
(V.67)

$$\propto \int_{r=0^+}^{l_c} dr \int d\mathscr{B}^n \int d\mathscr{B}^m P[r](\mathscr{B}^n, \mathscr{B}^m) \mathscr{B}^n \mathscr{B}^m \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}\right]$$
(V.68)

$$+\int_{r=l_{c}}^{+\infty} dr \int d\Omega^{n} \int d\Omega^{m} P[r](\Omega^{n}, \Omega^{m}) \left(R^{3}(\Omega^{n}) \cdot <\beta>\right) \times \left(R^{3}(\Omega^{m}) \cdot <\beta>\right) \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}\right].$$
(V.69)

Here we have assumed that  $l_c$  is the distance upon which  $P[r](\Omega^n, \Omega^m, \beta^n, \beta^m) = P[r](\Omega^n, \Omega^m)P(\beta^n)P(\beta^m)$ starts to be correct. The first term is by definition  $I^{\operatorname{coh}, QM/MM}[l_c]$  Now we remove and add  $\int_{r=0}^{+l_c}$  to the second term:

$$\left[\int_{r=l_c}^{+\infty} + \int_{r=0}^{+l_c} - \int_{r=0}^{l_c}\right] dr \int d\Omega^n \int d\Omega^m P[r](\Omega^n, \Omega^m) \left(R^3(\Omega^n) \cdot <\beta>\right) \left(R^3(\Omega^m) \cdot <\beta>\right) \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}\right]$$
(V.70)

$$= I^{\operatorname{coh}, LR} - \int_{r=0}^{l_c} dr \int d\Omega^n \int d\Omega^m P[r](\Omega^n, \Omega^m) \left( R^3(\Omega^n) \cdot \langle \beta \rangle \right) \left( R^3(\Omega^m) \cdot \langle \beta \rangle \right) \exp\left[i\Delta \mathbf{q} \cdot \mathbf{r}\right]$$
(V.71)

The last term is by definition  $I^{\text{coh}, SR < \beta >}[l_c]$  and corresponds to the correlation term from 0 to  $l_c$  when assuming that all molecules have the same  $\beta = <\beta >$ . This term has also been calculated in the present chapter, see for instance Figure IX.19.

Therefore, we have proved that we can decompose the coherent term into three parts in the Equation V.65. One term is the most accurate description of the short-range correlations where both the structure and the optical properties correlation are included in  $I^{\text{coh}, QM/MM}[l_c]$ . However, computing  $I^{\text{coh}, QM/MM}[l_c]$  for  $l_c$  greater than 5-6 Å seems very difficult. Hence, to include the rest of the interaction, we can set for each molecule the same molecular first hyperpolariability **to the average value** and compute the intensity in the reciprocal space, this term is  $I^{\text{coh}, LR}$ . However, this term also adds the short-range correlation by assuming that each molecule has the same first hyperpolarizability: we thus have to remove it using  $I^{\text{coh}, SR < \beta >}[l_c]$ .

# V. D | Conclusions

In this chapter, I have presented my numerical study on water in the bulk phase to predict the Second Harmonic Scattering signal. This was done thanks to the structure provided by a classical MD simulation and the individual first hyperpolarizability of every molecule calculated by FROG at the QM/MM level.

Liquid water is a (very) difficult system. First of all, experimentally speaking: obtaining its polarization-resolved SHS intensity for both the V and H outcoming polarization require a highly sensitive setup, knowledge and patience. Second, water is a pathological case for which the first hyperpolarizability is heavily impacted by its environment, see Chapter IV and VI. And finally, the SHS results cannot be interpreted solely by assuming an incoherent part because there are correlations within at least the first 3 solvation layers which affect the intensity.

First we study the incoherent response and especially questioned a widespread approximation consisting of fixing for each molecule the molecular first hyperpolarizability by a molecular symmetry based argument. The molecular first hyperpolarizability of water in the bulk phase shows large fluctuations due to its electrostatic environment and some of these components are correlated. The incoherent response being the average of the **square** of the first hyperpolarizability in the laboratory frame, the fluctuation matters more than the averaged value. We have shown that:

- Including the first hyperpolarizability fluctuations for the molecular symmetry authorized components impact drastically the incoherent response.
- Including all the molecular-symmetry-forbidden components impact even more the incoherent response.
- Including the correlations between the first hyperpolarizability components within the same molecule also impacts the incoherent response.

Considering only the incoherent contribution do not allow to recover the experimental signal.

Second, we have presented our formalism based on the first hyperpolarizability in the laboratory frame to compute the coherent part of the SHS intensity. As the statistical noise becomes too important for distance about 6-7 Å, we have focused on a "short-range" area composed for the first, second and third solvation layers. At this scale, there is no dephazing factor between molecules:  $\Delta \mathbf{q} \cdot \mathbf{r}_{nm} \approx 0$ . We found that the SHS responses of molecules within the first layer are correlated, and less throughout the second layer. This correlation is weaker within the third layer, and exists only for some components. The polarized-resolved SHS intensity for the V outcoming polarization is retrieved quantitatively when accounting for both the incoherent and coherent contributions. We have also found that the depolarization factor is impacted by the coherent part.

We have also tried to recover the SHS intensity with the coherent part using the approximation that each molecule share the same first hyperpolarizability. This approach failed both for the incoherent and this short-range coherent part. This shows that within the first solvation layers, there are both orientational correlation and correlation between the molecular first hyperpolarizability, and that we should take both into account to interpret SHS results.

I would like to propose two perspectives for this work on SHS. First I suggest a formalism including the long-range correlations. It would be interesting to see if this mixed formalism allows us to explain the remaining discrepancy with respect to the experimental result for pure water. A second perspective is the study of another chemical system for instance mixture of methanol-water. In the group, we have already performed SHS measurement of these mixtures at room temperature, and performed the MD simulations, and we have started the QM/MM calculations.

# VI | Surface-SHG

A negative result is not sexy, but it is a result anyway! From the PhD advisor of one of my PhD advisors.



Figure VI.1: Why research take so long

# VI. A | Introduction



Figure VI.2: LEFT: The S-SHG experimental setup. The fundamental beam arrives from the right and the second harmonic is collected to the left. RIGHT: Second harmonic intensity of the water liquid-gas interface with respect to the fundamental polarization,  $\gamma$ , for the two possible outcoming polarization direction of the second harmonic, namely P or S.

In this chapter, I present the results obtained on the study of the Surface Second Harmonic Generation of liquid interfaces. The experimental setup was presented in Section II. A. A picture of the setup is shown in Figure VI.2. Briefly, the fundamental beam arrives with an angle of 70 degrees with respect to the surface normal and the second harmonic generated signal is collected in the reflection direction. The studied system is composed of pure water at the liquid-gas interface, the intensity collected with respect to the fundamental polarization is reminded in Figure VI.2. This system is of almost importance as we use it as a day-to-day calibration in the lab: the intensity collected of a given liquid-gas interface is generally renormalized by the one of the air-water interface.

However, the understanding of these results at a microscopic point of view is still controversial. Even nowadays, many experimental and numerical work tackle this system: understanding the response of the water molecule is essential since many systems of interest are in the aqueous phase.

As presented in the Chapter III, many phenomena should be taken into account to predict the Surface SHG intensity: at the mesoscopic scale and microscopic one. Before presenting our numerical results and our current understanding, let us remind the different steps of our molecular-based approached.

# Fundamental excitation:

Experimentally, we apply the fundamental with an angle of 70 degrees with respect to the interface normal vector, see left part of Figure VI.2. Due to the difference in the optical indices of the gas and the liquid phase, the fundamental field propagation direction is modified. This phenomenon will impact the fundamental polarization direction differently. For instance, at this large angle, the polarization perpendicular to the incidence plane, named "S", will be far less transmitted to the liquid phase compared to the one in the plane, named "P". Hence, the electromagnetic field controlled by the operator, in the gas phase, will be different from the one at the liquid surface.

The other major problem is the difference between the electromagnetic field inside a media, here the "interface", and the exciting field noted  $e^{\omega}$ . The exciting field is the electromagnetic field felt by the molecule before its own reaction: this field can be expressed easily in the bulk phase using the optical index for homogeneous liquids (see Chapter III). However, at interfaces where the optical index is evolving with the altitude, the definition of the exciting field is less obvious.

Finally, since the electromagnetic field evolves, and so will the exciting field, at the interface there should be a non-zero spatial gradient:  $\nabla \cdot \mathbf{e}^{\omega}$ . This field gradient will induce another second harmonic response due to the quadrupolar order.

# Microscopic second harmonic generation process:

If we assume that we know the exciting field and its spatial gradient at any point of the interface then we have to compute the response tensors for each molecule. To do so, we use FROG to compute the first hyperpolarizability of all the molecules in their electrostatic environment. Up to the quadrupolar order, we have to compute the three first hyperpolarizabilities tensors involved, so that the induced dipole and quadrupole moments are:

$$\boldsymbol{\mu}^{SHG} = \frac{1}{2!} \beta^{dd}(2\omega, \omega, \omega) : \mathbf{e}^{\omega} \mathbf{e}^{\omega} + \frac{1}{2!} \beta^{dq}(2\omega, \omega, \omega) : \mathbf{e}^{\omega} \nabla \mathbf{e}^{\omega}$$
(VI.1)

$$Q_{ij}^{SHG} = \frac{1}{2!} \beta^{qd} (2\omega, \omega, \omega) : \mathbf{e}^{\omega} \mathbf{e}^{\omega}$$
(VI.2)

Where the definition of the  $\beta$ s using the response framework is reminded in Section III. B.1.5. Note that for readability we will drop the explicit frequency dependency in this chapter.

# The mesoscopic response at the second harmonic frequency:

As presented in Chapter III, the microscopic equations above describe the individual second harmonic sources, not the total dipolar and quadrupolar moments. Indeed, the molecules generating the second harmonic electromagnetic field have still a polarizability at the second harmonic frequency: they may induce to one another a new dipole moment! Hence, we may have to compute the difference between the second harmonic source terms, given by  $\mu^{SHG}$  and  $Q_{ij}^{SHG}$ , and the total dipolar moment with the intermolecular interaction, called dressed dipole.

We will have to deal with this part later on. For the moment, we will focus on the induced averaged mesoscopic polarization  $\mathbf{P}^{SHG}$ , the source term:

$$\mathbf{P}^{SHG} = \langle \boldsymbol{\mu} \rangle^{SHG} - \frac{1}{2} \nabla \cdot \langle Q \rangle^{SHG}$$
(VI.3)

Which is given by the spatial average of the microscopic induced dipole moment  $\langle \mu \rangle^{SHG}$  and quadrupole moment  $\langle Q \rangle^{SHG}$  at the second harmonic frequency in the laboratory frame. To compute these quantities, we will compute the first hyperpolarizability averaged in the laboratory frame.

# Second harmonic electromagnetic field collected experimentally:

Finally, we have to solve the Maxwell equations in a layered system. At the bottom there is the liquid phase with no second harmonic sources but with an optical index  $n \approx 1.33$ . At the interface, there is a sheet of non-linear sources given by  $\mathbf{P}^{2\omega}$ , and finally on the top the gas phase with no source and an optical index of 1. To solve this problem, we have used the well known three layer models (or polarization sheet) [76]. The main issue of this approach is the definition of the "interface" media optical index.

# Overview of this chapter:

This system was chronologically the first studied tried in my thesis, which leads to our first paper [71].

This paper shows a first use of our QM/MM approach in this new kind of heterogeneous system and is presented in the first part of this chapter. The main objective was to compute how evolves the dipolar first hyperpolarizability across the liquid-gas interface. Moreover, we questioned a very standard approximation made to understand the SHG response within the dipolar approximation: can we attribute to each molecule the same **molecular** first hyperpolarizability in order to compute the response of the interface?

Using these results, we were then able to compute the induced dipole moment at the second harmonic frequency in the laboratory frame, first term of Equation VI.3, and thus define the surface susceptibility tensor at the dipolar order. However, the obtained results did not reproduce the experimental measurements well. Hence, we tried to deepen our understanding of the microscopic properties and we found that the problem may not be the value of the dipolar first hyperpolarizability at the liquid-gas interface: it seems that it was necessary to go beyond the dipolar approximation.

At the beginning of my last year of PhD, I had the opportunity to participate to the Molecular Response Property Summer School 2021 at Stockholm which helped me to understand the nature of the molecular response at the quadrupolar order. Simultaneously, Foucaud and co-worker [41] published an important numerical article which shows the important contribution of the quadrupolar terms for the neat water liquid-gas interface. Hence, we included in FROG the first hyperpolarizability calculation at the quadrupolar order. The second section of this chapter presents the value obtained for neat water in the bulk and at the liquid-gas interface.

Moreover, we had to adapt and built a complete multi-scale approach which includes the terms needed for the quadrupole using our molecular-based formalism. We ended up with the formalism presented in Chapter III and summarized previously. However, we have still some work to do regarding the link between the microscopic properties and the induced microscopic ones: the end of this chapter should be seen as a snapshot of our current understanding and results. We hope to conclude this approached soon and to be able to compare with the experimental signal.

# VI. B | Neat water at the liquid-gas interface: how to compute $\chi$ at the dipolar level?

Cet article a été publié en 2021 dans le journal PCCP.

# Résumé et mise en contexte:

La Génération de Second Harmonique en surface (S-SHG) est un outil expérimental unique en son genre pour sonder la structure des interfaces. En effet, ce phénomène est intrinsèquement sensible à la brisure de centro symétrie, et donc à l'interface. Néanmoins, un problème important de la S-SHG est l'interprétation des résultats expérimentaux, et son lien à la structure moléculaire.

Il est commun de supposer que la réponse individuelle des molécules dans leur référentiel, le tenseur d'hyperpolarizabilité moléculaire  $\beta(2\omega, \omega, \omega)$ , est le même quelque soit la position de la molécule ou son orientation, voir Section 4 du SI. Dans le cas des molécules d'eau, nous savons que ce tenseur est très sensible à son environnement électrostatique: nous pouvons donc nous demander à quel point cette hypothèse est valable à l'interface liquide-gaz.

Pour ce faire, nous avons calculé  $\beta(2\omega, \omega, \omega)$  en fonction de la distance à l'interface en utilisant une approche mêlant l'environnement électrostatique autour de chaque molécule à une résolution quantique de leur degré électronique en utilisant FROG. Cette approche multiéchelle est schématisée dans la Figure 1. Cette description de l'environnement de façon explicite nous permet de calculer les variations de ce tenseur moléculaire au travers de l'interface. Ces évolutions sont représentées dans les Figures 6 et 7. De façon surprenante, nous observons que dans la phase liquide comme à l'interface la symétrie fréquentielle de Kleinman est respectée, alors qu'elle ne l'est pas en phase gaz.

Grâce à cette approche, nous avons aussi pu mettre en évidence que le signal S-SHG (à l'ordre dipolaire) provient uniquement du dernier nanomètre formant l'interface: nous avons bien une réponse spécifique à la surface. De plus, nous avons testé l'hypothèse qui consiste à attribuer à toutes les molécules la même hyperpolarizabilitée moléculaire pour calculer la réponse dans le référentiel du laboratoire. Nous avons comparé les résultats obtenus avec et sans cette approximation dans la Figure 9. La différence observée est importante et remet largement en question le formalisme habituellement utilisé pour comprendre le signal SHG de l'eau aux interfaces: la réponse de chaque molécule ne peut pas être expliquée uniquement par son orientation puisque sa réponse à la seconde harmonique dépend fortement de son environnement.

# PCCP

# PAPER

Check for updates

Cite this: Phys. Chem. Chem. Phys., 2021, 23, 24932

Received 21st May 2021, Accepted 30th August 2021

DOI: 10.1039/d1cp02258j

rsc.li/pccp

# 1 Introduction

Liquid water is ubiquitous on Earth; understanding the properties of its interfaces with other media is essential for many fields of physics, chemistry, and biology. Surface analytical tools have been developed to obtain information on the molecular organization at interfaces; in particular, non-linear optical techniques such as Surface Second Harmonic Generation (S-SHG) and Surface Sum Frequency Generation (S-SFG) are increasingly used to probe biological materials at interfaces,<sup>1-3</sup> and to investigate the molecular structure of liquid–gas,<sup>4-9</sup> liquid–liquid<sup>10,11</sup> or liquid– solid<sup>12-16</sup> interfaces. The SHG is an optical process whereby two photons at a fundamental frequency are converted into one photon at twice the fundamental frequency, *i.e.* the harmonic frequency. Coherent SHG techniques are surface-sensitive because of the cancellation of the SHG process within

University of Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-69622, Villeurbanne, France. E-mail: claire.loison@univ-lyon1.fr

# centrosymmetric media like bulk liquids<sup>17</sup>. Moreover, the SHG response is strongly dependent on the molecular arrangement and the molecular electrostatic environments, 18,19 so that S-SHG has become a useful tool for probing molecular structure<sup>1,20,21</sup> or surface electrostatics.<sup>2,8,9,12,14,22-24</sup> However, linking the experimental S-SHG intensity to interface molecular structure often remains a challenge, for which different experimental characterizations,<sup>9,11</sup> theoretical developments,<sup>17,25</sup> and molecular modeling<sup>26-28</sup> can benefit from each others.<sup>10,29-32</sup>

The interpretation of the S-SHG signal of the water/air interface has been a subject of debate for decades.<sup>9,10,33</sup> Efforts have been made to assess the contribution of the interface (local) contribution of "the Bonded Interface Layer" (BIL) and the one of bulk electric quadrupole and magnetic dipole contributions to the SHG signal appearing in the diffuse layer (DL).<sup>9,27</sup>

For a neat water/air interface, Dalstein et al.9 and Zhang et al.<sup>33</sup> attribute the signal to the BIL, *i.e.* the layer where the inversion symmetry is broken by the interface-specific structure. Typically, under the electric dipole approximation, this contribution to the quadratic surface susceptibility tensor  $\chi^{(2)}$  is modeled as a sum of the molecular hyperpolarizability tensors  $\beta(-2\omega,\omega,\omega)^n$ , where *n* is not a power notation, but an index running over the molecules.<sup>10,33</sup>



interface: a QM/MM study guestions standard

Surface Second-Harmonic Generation (S-SHG) experiments provide a unique approach to probe interfaces. One important issue for S-SHG is how to interpret the S-SHG intensities at the molecular level. Established frameworks commonly assume that each molecule emits light according to an average molecular hyperpolarizability tensor  $\beta(-2\omega,\omega,\omega)$ . However, for water molecules, this first hyperpolarizability is known to be extremely sensitive to their environment. We have investigated the molecular first hyperpolarizability of water molecules within the liquid-vapor interface, using a quantum description with explicit, inhomogeneous electrostatic embedding. The resulting average molecular first hyperpolarizability tensor depends on the distance relative to the interface, and it practically respects the Kleinman symmetry everywhere in the liquid. Within this numerical approach, based on the dipolar approximation, the water layer contributing to the Surface Second Harmonic Generation (S-SHG) intensity is less than a nanometer. The results reported here question standard interpretations based on a single, averaged hyperpolarizability for all molecules at the interface. Not only the molecular first hyperpolarizability tensor significantly depends on the distance relative to the interface, but it is also correlated to the molecular orientation. Such hyperpolarizability fluctuations may impact the S-SHG intensity emitted by an aqueous interface.

**View Article Online** 

View Journal | View Issue

<sup>†</sup> Electronic supplementary information (ESI) available: Details of the QM/MD method and comparison with literature, MD structural analysis, molecular water hyperpolarizability analysis in the bulk and at the liquid/vapor interface, susceptibility expressions for the study of molecular orientation/ $\beta$  correlations. See DOI: 10.1039/d1cp02258j

#### Paper

To express  $\chi^{(2)}$  in the laboratory frame, one takes into account the orientation of the *n*th-molecule *via* the matrix  $\mathbf{R}^n$  of its molecular coordinate system  $\{a,b,c\}$  in the laboratory frame  $\{X, Y, Z\}$ :  $\chi^{(2)} \propto \sum \mathbf{R}^n \mathbf{R}^n \beta^n$ .<sup>10,33</sup> To interpret the S-SHG signal of the water/air interface in terms of interface structure using this formalism, one therefore needs to know the hyperpolarizabilities of individual water molecules.33 Moreover, one considers that the signal can be interpreted in terms of molecular hyperpolarizability. The decomposition of the non linear optical response of a cluster into a sum of molecular properties is not obvious,<sup>36</sup> since intermolecular interactions may strongly impact non linear optical properties (for a recent example see ref. 37). In the special case of water, Maroulis<sup>38</sup> or Liang et al.<sup>39</sup> compared the polarizabilities and/or hyperpolarizabilities of dimers to the sum of the properties of the individual molecules. Their results indicate that the single-molecule summation is a reasonable first step.

The molecular hyperpolarizability is commonly assumed to be the same for all the water molecules of a given phase, and respecting the molecular symmetry group  $C_{2v}$ . These hypotheses permit to extract from the susceptibilities the mean molecular orientation – *via* the orientation matrices.<sup>4,20</sup> However, it is known that solvation has a substantial effect on the measured hyperpolarizability of water molecules, and the liquid environment breaks the  $C_{2v}$  symmetry.<sup>36,39–43</sup> To reach quantitative interpretation of S-SHG signal of aqueous interfaces, the water hyperpolarizability must be known, precisely at these interfaces.

In particular, the origin of the failure of Kleinman symmetry observed experimentally for the neat air/water interface ( $\chi^{(2)}_{ZXX} \neq \chi^{(2)}_{XZX}$ , for example) is debated.<sup>33,44,45</sup> It was attributed by Zhang *et al.* to the asymmetry of the molecular hyperpolarizability, for instance to  $\beta_{aca} \neq \beta_{caa}$  or  $\beta_{bcb} \neq \beta_{cbb}$ , where the *c*-axis is along the water dipole moment.<sup>33</sup> A symmetry breaking of the hyperpolarizability tensor  $\beta$  indeed exists for water in the gas phase,<sup>10,33</sup> but to our knowledge, it has been reported neither in the bulk liquid, nor at the interface.

In this article, we report theoretical chemistry calculations of the molecular hyperpolarizability of water molecules nearby the interface, and discuss the results in terms of the S-SHG measurements. Liquid water is a difficult system for computational modeling in general, and for the prediction of electric properties in particular because of the extended H-bond network.<sup>46,47</sup> Modeling the molecular environment as homogeneous dielectric material is not sufficient to reproduce the sign change of water hyperpolarizability from the gas phase to the liquid phase.<sup>48</sup> Here, we propose a sequential Quantum Mechanical/Molecular Mechanics (QM/MM) approach to provide new insights on the contribution of water molecular hyperpolarizability to the S-SHG signal of the liquid–vapor interface, see Fig. 1.

Such methods have been successful to describe the impact of solvation on the optical properties of NLO-dyes,<sup>28,49-54</sup> and on other non-resonant molecules.<sup>39,41-43,55-57</sup> First, a classical molecular description of the structure of the liquid–vapor water interface is obtained using Molecular Dynamics (MD) simulations. Then, the frequency dependent  $\beta(-2\omega,\omega,\omega)$  of each water molecule at the interface is computed using a QM description within an electrostatic embedding framework<sup>58</sup> to mimic the condensed phase environment.

In Section 2, we present the MD and QM methods and justify some important calculation parameters. In Section 3, we report our results concerning the interface structure, and the evolution of the hyperpolarizability components when the water molecule gets close to the liquid-vapor interface.

# 2 Method description and validation

# 2.1 Molecular mechanics simulations and analysis

LAMMPS<sup>59</sup> V.11.08.2017 is used to perform the MD simulation along with the rigid TIP4P/2005 water force field.<sup>60</sup> Using equilibrated bulk configurations, 9000 rigid TIP4P/2005 water molecules are placed in a simulation box ( $5 \times 5 \times 40 \text{ nm}^3$ ) to form a water film, about 10 nm thick, see Fig. 1(a). Thanks to the sufficient film thickness, the middle part of the fluid film well mimics bulk water. Periodic boundary conditions were applied in all three directions, so that the simulation is actually that of a multi-film system, but the vacuum layer of 30 nm is thick enough to neglect interactions between the water interfaces. The target temperature is 300 K using a Nose–Hoover thermostat with the period of the temperature fluctuations set to  $\tau = 0.4$  ps. 3 ns of *NVT* equilibration is performed before the 20 ns *NVT* production run with a time step of 2 fs. The center of



**Fig. 1** Scheme of the numerical procedure (performed by the home-made software FROG). (a) Molecular dynamics snapshot of the water liquid–vapor simulation within a box of  $5 \times 5 \times 40$  nm<sup>3</sup>. (b) Zoom on the interface and on one water molecule at a given time step. (c) A single water molecule with its electrostatic environment, within a radius  $R_c$ , used for quantum mechanics calculation of individual hyperpolarizabilities. Average hyperpolarizability profiles are obtained as a function of the coordinate normal to the interface (*Z*) by selecting and analyzing molecules within 0.1 nm-thick slices. Pictures were prepared using VMD<sup>34</sup> and Inkscape<sup>35</sup> softwares.

mass of the whole system is fixed every timestep. Both electrostatic and Lennard-Jones intermolecular interactions are computed using the long-range PPPM formalism.<sup>61,62</sup> Neighbor lists were updated every timesteps with a radius of 1 nm.

To compute the structural properties of the simulated system, such as the molecular orientations or the hyperpolarizabilities, a home-made Python software FROG is used, using the module MDAnalysis.<sup>63</sup> Local composition and molecular orientation are computed along with the hydrogen bond (H-bond) network. The H-bond analysis uses the geometric criteria from Pezzotti *et al.*, *i.e.* an H-bond is present if the distance  $O(-H)\cdots O$  is smaller than 3.2 Å and the angle  $O-H\cdots O$  is in the range of  $[140-220]^{\circ}$ .<sup>27</sup> For the bulk phase, we obtain 3.37 H-bonds per water molecule, to be compared to the value obtained by Pezzotti *et al.* of 3.4.<sup>27</sup> The structural properties such as densities and H-bond numbers are averaged over time and space to produce profiles, as precised below.

### 2.2 Space averaging

In the system, two liquid–vapor interfaces are present, but we shall focus on a single interface here, see Fig. 1(b). The average position of the interface is supposed to be planar, parallel to the {*X*,*Y*} plane. To calculate profiles along the *Z*-direction, the various properties are averaged over the molecules belonging to 0.1 nm-thick slices normal to the *Z* direction. In the following, the profiles are plotted as function of the "altitude"  $\Delta Z = Z - Z_0$  relative to the average position of the interface  $Z_0$ . To calculate the error-bars, we have estimated space-correlations within each slice, see ESI,† for details (Section S1.3).

### 2.3 Response properties within density functional theory

The QM/MM workflow was handled by our software FROG that reads the MD trajectories and writes input files for the QM calculations, to compute NLO properties within the PE framework.<sup>51</sup> After the hyperpolarizability calculations have been performed, FROG analyses the individual QM outputs and computes the altitude-dependent distribution of the hyperpolarizability (see Section S1.1 of ESI<sup>†</sup>).

Hyperpolarizability calculations. Calculations of the static and dynamic NLO responses of water at the interface were performed at the density functional theory (DFT) CAM-B3LYP<sup>64</sup>/ d-aug-cc-pVTZ<sup>65</sup> level using the DALTON software,<sup>66</sup> release 2018.2 package. DFT presents inaccuracies that we discuss in the following, but its computational efficiency, necessary for a proper QM/MM sampling of several thousands of calculations, was the reason of our choice.

Concerning the basis, large sets including polarization and diffuse functions are necessary to obtain reliable hyperpolarizabilities.<sup>67,68</sup> The Dunning's correlation consistent (cc) basis sets,<sup>65,69,70</sup> in particular d-aug-cc-VTZ basis set is commonly used for water.<sup>10,39,42,71</sup> It can be considered as a very good compromise between efficiency and accuracy for the present QM/MM applications that has been used for water molecules in the gas and liquid bulk phases. For an isolated water molecule, Sections S1.2.1 and S1.2.2 of ESI,† show the convergence of DFT approaches relative to basis set.

Concerning the functional, CAM-B3LYP is a range-separated hybrid functional which adds long-range correction using the Coulomb-attenuating method and includes 19% of HF exchange at short-range and 65% at long-range.<sup>64</sup> It is expected to provide reliable results for the non linear optical properties of medium-size molecules.<sup>72,73</sup> While it is true that DFT/CAM-B3LYP is not very accurate to compute cross section at resonant frequency,<sup>74</sup> the present work investigates SHG response for the fundamental frequency of 800 nm, i.e. in the non-resonant case, where no excited state is populated. For water, the excitation energies of the first states are around 7-10 eV, corresponding to a fundamental wavelength about 300 nm. Even if the excitation wavelength of interest in our article is 800 nm, these states should still contribute to the SHG response. CAM-B3LYP seems a reasonable choice, since it performs best to predict the cross section and the frequency compared to other DFT functionals.74 For single water molecules, the use of long-range-corrected functionals permits to obtain DFT hyperpolarizabilities in reasonable accordance with CCSD/d-aug-cc-pVTZ results: the typical absolute difference is 1.3 a.u., averaged over non-zero hyperpolarizability components (see Table S1, ESI<sup>†</sup> and ref. therein<sup>67,75</sup>). Finally, since the QM/ MM approach involves an electrostatic environment, the second hyperpolarizability y plays an important role. Besalú-Sala et al. have showed recently that CAM-B3LYP is among the best choices regarding the second hyperpolarizability of water.<sup>73</sup>

The first hyperpolarizability is extracted using the frequencydependent quadratic response scheme.<sup>76</sup> The hyperpolarizability values are given within the T convention,<sup>77</sup> in atomic units. Frequency-dependent calculations were carried out using an incident wavelength  $\lambda$  of 800 nm, and compared to the results for a static field ( $\lambda \rightarrow \infty$ ). The vibrational contribution to the first hyperpolarizability is neglected. This contribution is typically of 10% for water in the gas phase.<sup>71</sup>

QM/MM embedding. To model the liquid phase effect, an electrostatic environment is considered. The QM system is always composed of a single water molecule, surrounded by point charges representing the neighboring water molecules. The typical size of the electrostatic environment is defined by  $R_{\rm c}$ . To preserve the electroneutrality of the whole system, only complete molecules are included in the electrostatic environment. If a neighbor molecule posses at least one atom in radius  $R_c$  around the center of mass of the target water molecule, the whole molecule is included in the electrostatic environment. This heterogeneous and discrete solvation model is still in active development.51,58,78-80 Even if potential improvement can be gained using more accurate electrostatic description of the neighbors78 or the calculation of the local effective field,<sup>51</sup> we have used a point charge description (PE0 scheme in DALTON) of the electrostatic environment fully coherent with the MD simulation protocol. This ensures consistency within the QM embedding and the MD structures and provides a first robust and simple calculation.

To characterize the effect of the MM embedding, we have calculated the electrostatic field generated by the point charges on the QM water molecule. More precisely, this electrostatic

**Table 1** Molecular hyperpolarizability  $\beta(-2\omega,\omega,\omega)$  in the gas and bulk liquid water phases. Hyperpolarizabilities are in a.u. and the excitation wavelength  $\lambda$  is in nm. For the liquid phase, the mean values are presented along with the distribution width in brackets. Table S1 of ESI reports the other components, which average to 0

	Gas		Liquid							
λ	$\infty^{a}$	800 <sup><i>a</i></sup>	$\infty_p$	$\infty^{a}$	800 <sup><i>a</i></sup>					
$\beta_{caa}$	-10.6	-12.5	-2.1[1.1]	-1.9 [1.0]	-1.9 [1.1]					
$\beta_{aca}$	-10.6	-12.4	-2.1[1.1]	-1.9[1.0]	-2.0[1.1]					
$\beta_{cbb}$	-4.2	-5.0	1.6 [1.8]	2.6[1.7]	2.7 [2.0]					
$\beta_{bcb}$	-4.2	-7.4	1.6 [1.8]	2.6[1.7]	2.3 [2.1]					
$\beta_{ccc}$	-12.1	-15.3	3.0 [3.0]	3.9 [2.8]	4.0 [3.1]					

<sup>*a*</sup> This work, using DFT-CAM-B3LYP/d-aug-pVTZ level within the PE-0 response scheme using the charges of the TIP4P-2005 force field. The values for the liquid are averages over 48 000 configurations obtained from 32 MD frames separated from 20 ps, with an electrostatic environment built up to  $R_c = 2$  nm around the target molecule. <sup>*b*</sup> Combined coupled cluster/molecular mechanics (CC/MM) method by Liang *et al.*<sup>39</sup>

field is calculated at the position of the negative charge in the TIP4P/2005 force field -0.15 Å away from the oxygen atom, along the dipole vector.

Validation of the QM/MM method for bulk liquid water. The averages and standard deviations of the  $\beta$  components with CAM-B3LYP/d-aug-pVTZ for water in liquid water are compared to the values of Liang et al.<sup>39</sup> in Table 1 (see Table S2 of ESI,† for all components). Both calculation schemes use the TIP4P/2005 rigid force field and the basis set d-aug-pVTZ. Liang et al. have used couple-cluster calculation, while we have used DFT. The bulk hyperpolarizability values differ from the ones of Liang et al. by 0.2 to 1 a.u., but the main features of the hyperpolarizability of solvated water are well reproduced by our scheme: (i) the hyperpolarizability of water in bulk liquid is very different from the one of water in the gas phase, with a change of sign of  $\beta_{ccc}$ ,  $\beta_{bcc}$  and  $\beta_{cbc}$ ; (ii) the component  $\beta_{ccc}$ remains the most important in absolute value; (iii) the standard deviations of the hyperpolarizability distributions are very large, often similar to the average values, (iv) the tensor respects the  $C_{2v}$  symmetry on average, but it is generally not the case for the hyperpolarizability of individual water molecules within their anisotropic environments. Moreover, our results for the first hyperpolarizability of water in bulk match well the ones obtained with the a polarizable environment (PE1 level) by Osted et al.<sup>42</sup> This is in qualitative accordance with the results by Liang et al.,<sup>39</sup> who have shown the small impact of the water flexibility on  $\beta$  values for water in bulk phase.

Based on these results, the rest of the work concerning the interface was done using the CAM-B3LYP functional and the d-aug-cc-VTZ basis set.

Impact of electrostatic environment radius ( $R_c$ ). To choose the value for  $R_c$ , the convergence of the hyperpolarizability relative to it has been performed for the water molecules belonging to the 0.1 nm-thick slice at the altitude  $\Delta Z =$ -0.06 nm. This altitude has been chosen because it is typical of the interface: for this altitude, the structural properties differ strongly from the bulk ones while the density remains large



**Fig. 2** Convergence of the averaged molecular first hyperpolarizability component  $\langle \beta_{ccc} \rangle$  relative to the environment cutoff radius  $R_c$ , for molecules at the altitude  $\Delta Z = -0.06 \pm 0.05$  nm. The percentage of the relative absolute error  $|\langle \beta_{ccc} \rangle - \langle \beta_{ccc} \rangle^0 | / \langle \beta_{ccc} \rangle^0$  is plotted with a log. scale, where the reference value  $\langle \beta_{ccc} \rangle^0$  is obtained for  $R_c = 5$  nm. The dashed line represents an error of 1%. See ESI,† Fig. S7 for other  $\beta$  components.

enough to obtain precise averages. Moreover, due to the large number of molecules with a net orientation, this slice is expected to contribute strongly to the second harmonic response of the interface. The convergence of  $\langle \beta_{ccc} \rangle$  relative to  $R_c$  is illustrated by Fig. 2, where  $\langle \cdot \rangle$  denotes here an average over the molecules at the chosen altitude (in this case, 4294 molecules extracted from 80 MD frames separated from 20 ps).

A radius of about 2 nm is needed to reach a relative error below 1%, *i.e.* 0.1 atomic unit (dashed line on Fig. 2). The convergence is similar for molecules selected in the bulk phase (see Fig. S6 of ESI<sup>†</sup>). Therefore, in the following, the neighborhood is built up to a distance  $R_c = 2.3$  nm. This result highlights that the hyperpolarizability is sensitive to the electrostatic environment up to several nanometers: for aqueous interfaces in general, the S-SHG signal of the BIL might be influenced indirectly by the structure and composition of a thicker interfacial layer. Indeed, special care should be granted to the choice of the electrostatic embedding description for QM/MM SHG studies.<sup>51,78</sup>

#### 2.4 Time averaging of hyperpolarizability

The hyperpolarizability profiles are obtained from the analysis of  $N_{\rm f}$  selected frames of the MD trajectory, separated from a time span  $\Delta t$ . This Section presents our methods to choose  $\Delta t$  and  $N_{\rm f}$ .

Hyperpolarizability time autocorrelation. To avoid timecorrelation in the hyperpolarizabilities, one waits for a given time span  $\Delta t$  between treated MD frames. To choose this time span, we have calculated the autocorrelation function of the hyperpolarizability tensors of two arbitrary molecules chosen either in the bulk phase, or at the interface. For this particular test, 1000 configurations of these two molecules are treated, every  $\delta t = 0.2$  ps, with the same functional and basis as for the rest of the study (CAM-B3LYP/d-aug-pVTZ), with an electrostatic embedding up to  $R_c = 2$  nm. The autocorrelation function (AC) is defined as:

$$AC[\beta_{ijk}](\Delta t) = \frac{\delta t}{T} \sum_{t=\delta t}^{T} \tilde{\beta}_{ijk}(t + \Delta t) \tilde{\beta}_{ijk}(t)$$
(1)

for  $\Delta t$  in range 0 to 200 $\delta t$ ,  $T = 800\delta t$  and  $\beta_{ijk}(t) = \beta_{ijk}(t) - \langle \beta_{ijk} \rangle$ . In order to represent the contribution of all the components, a "total autocorrelation" (TAC) function is defined:

$$TAC[\beta](\Delta t) = \left(\sum_{ijk} AC[\beta_{ijk}]^2(\Delta t)\right)^{\frac{1}{2}}$$
(2)

The normalized TAC (TAC( $\Delta t$ )/TAC(0)) is plotted in Fig. 3. In the interface region, as in the bulk region, the autocorrelation time of the hyperpolarizability of a single water molecule nearby the interface is of the order of a few picoseconds. This is coherent with the value used for the  $\beta$  calculation in bulk phase in ref. 42 (1 ps) and for properties of larger molecules in ref. 79 (10 ps). Finally, we have exploited the MD configurations obtained every  $\Delta t = 20$  ps of simulation, which is also larger than the surface tension autocorrelation time.<sup>81</sup>

Hyperpolarizability time convergence. To choose the total number of treated frames  $N_{\rm f}$ , we have calculated the average hyperpolarizability obtained as a function of  $N_{\rm f}$  for the typical altitude of  $\Delta Z = -0.06$  nm. Fig. 4 depicts the non-zero components, relatively to our reference (80 frames, *i.e.* 4294 configurations).

Approximately 4200 configurations are needed to obtain relative deviation of the order of 2% at this typical altitude. These correspond to the  $N_{\rm f}$  = 80 selected MD frames, which was chosen for the rest of the study, for all altitudes. Fig. S5 of ESI<sup>†</sup>



**Fig. 3** Time auto-correlation (TAC) of the fluctuations of the beta tensor of a single molecule, either in the bulk phase (blue solid line), or at the interface (orange solid line), averaged over the components as defined by eqn (2). The ordinate axis follows a logarithmic scale. The dashed line represents a TAC of 5%.



Fig. 4 Relative deviation of the average  $\beta$  components as a function of the number of QM calculations, for the typical altitude of  $\Delta Z - Z_0 = -0.06 \pm 0.05$  nm. The relative deviation (in %) is calculated with respect to the reference value obtained using the 4294 molecular configurations obtained for  $N_{\rm f}$  = 80 selected MD frames. The dashed lines represent a deviation of 5%.

reports the evolution for the other components, that converge towards zero, and the conclusions are the same.

For a fixed number of frames  $N_{\rm f}$ , the number of QM calculations included in the hyperpolarizability average of a given slice is proportional to the density in this slice. Therefore, the hyperpolarizability averages are less precise for altitudes where the density is lower.

# 3 Results and discussion

# 3.1 Interface structure from classical molecular dynamics simulations

In a first step, the structure of the liquid–vapor interface has been obtained from the analysis of one interface in an MD simulation of a water film surrounded by vapor, see Fig. 1(a and b).

We discuss here the density profile and the water dipole orientation profile displayed in Fig. 5(b). The density decreases progressively from the bulk value to zero with a profile close to a hyperbolic tangent shape.<sup>82</sup> Its inflection point permits to define an absolute position  $Z_0$  for the interface, and the distance relative to it ( $\Delta Z = Z - Z_0$ ) that is called "altitude" here.

The projection of the water dipole *c*-axis on the interface normal (*Z*), illustrated by Fig. 5(a), is particularly relevant for the interpretation of the S-SHG intensities. Indeed, the  $C_{\infty}$ symmetry of the interface, associated to an expected  $C_{2v}$ symmetry of the water hyperpolarizability tensor, yields surface susceptibility components expressed as linear combinations of  $\langle \beta_{ijk}(\cos \theta)^I \rangle$  where  $\theta$  is the (*c*,*Z*) angle<sup>33,83</sup> and *l* is 1 or 3. For example, if one focuses on the  $\chi^{(2)}_{ZZZ}$  element (the dominant one, according to experimental observations on the water-air interface<sup>10</sup>):

$$\chi_{ZZZ}^{(2)} \propto \langle \beta_{\perp} \cos \theta \rangle + \langle (\beta_{ccc} - \beta_{\perp}) \cos^3 \theta \rangle, \qquad (3)$$

where  $\langle \cdot \rangle$  defines the average over molecules and

PCCP

Paper



**Fig. 5** (a) Definition of the molecular frame  $\{a,b,c\}$  and of the angle  $\theta = (c,Z)$  characterizing the orientation of the water dipole moment with respect to the interface normal Z-axis. (b) Water density profile (black circles) and  $\langle \cos \theta \rangle$  profile (gray squares) as a function of the altitude. The solid black line is proportional to a hyperbolic tangent function fitted onto the density profile. The error bars are calculated using eqn (S3) of ESI.†

$$\beta_{\perp} = (\beta_{caa} + \beta_{cbb} + 2\beta_{aca} + 2\beta_{bcb})/2. \tag{4}$$

Section S4 of ESI<sup>†</sup> recalls the complete expressions.

The profile of  $\langle \cos \theta \rangle$  is therefore plotted in Fig. 5(b) as a function of the altitude. The region where a net orientation appears, *i.e.*  $\langle \cos \theta \rangle \neq 0$ , spotlights the BIL and the thickness of the layer which should contribute to the S-SHG signal. Noticeably, the net water orientation appears few Angstroms further away from the interface than the density drop. The deviation for the H-bond network relative to the bulk occurs at a similar altitude (see Fig. S3 of ESI†).

Moreover,  $\langle \cos \theta \rangle$  provides structural information: the negative values indicate that the dipole moment of the water molecule points on average with the hydrogens towards the bulk phase. But this propensity is weak, and the dipole moment of the water molecule in the most external water layers is preferentially close to parallel to the interface plane (see Fig. S4 of ESI<sup>†</sup>). These results agree with previous studies.<sup>10,47,84</sup>

To guide the interpretation, we have divided the system into 3 regions shown in Fig. 5: bulk (in purple), interface (in cyan) and vapor (in white). The limit between the bulk and interface is fixed at the altitude where  $\langle \cos \theta \rangle$  significantly deviates from its bulk value. The edge of the vapor area is defined where the density reaches 5% of the bulk value. In practice, we have limited the vapor region at the point where accurate statistical averaging over the number of water molecules became difficult because of too small densities.

# 3.2 Water hyperpolarizabilities from quantum mechanics/ molecular mechanics (QM/MM) calculations

The second step consists in computing, for each MD snapshot, the molecular hyperpolarizability  $\beta(2\omega,\omega,\omega)$  of all water molecules. After justifying several calculation parameters (see Section 2.3), we discuss here the molecular hyperpolarizability components as a function of the altitude.

First, the average hyperpolarizability tensor respects the  $C_{2v}$  symmetry at the interface. Note that this was already true in the bulk (see Table S2 of ESI†). Indeed, the  $C_{2v}$ -forbidden  $\beta$  components remain negligible (less than 0.3 a.u. in average) throughout the whole interface (see Fig. 6). All these values average to zero in the region where the density is large enough to obtain precise results, but their fluctuations are strong, as already noted by Liang *et al.*<sup>39</sup> for the bulk phase. Concerning the non-vanishing hyperpolarizability components, Fig. 7 shows that they start to deviate from their bulk values roughly at the same altitude as the density, molecular orientation or H-bond network.

Concerning the frequency dispersion, characterized for example by the difference between  $\langle \beta_{cbb} \rangle$  and  $\langle \beta_{bcb} \rangle$ , it is much smaller in the water bulk phase ( $\simeq 0.5$  a.u.) than in the vacuum phase ( $\simeq 2.5$  a.u., see Table 1). In the interface region, the frequency dispersion remains weak, and the Kleinman symmetry is practically respected, *i.e.*  $\langle \beta_{caa} \rangle = \langle \beta_{aca} \rangle$  and  $\langle \beta_{cbb} \rangle \simeq \langle \beta_{bcb} \rangle$ . Therefore, within our methodology, the Kleinman symmetry breaking observed experimentally seems difficult to be explained solely by the individual dipolar molecular responses (see Section S4.3 of ESI†). Our calculations do not take into account the effective field (cavity field) effect, which is beyond the scope of this work. However, since the dispersion of the water polarizability between 800 and 400 nm is very small,<sup>10</sup> it should not affect much our conclusion. More precise QM/MD calculations, local effective field effects, collective effects, or



**Fig. 6** Averages of the hyperpolarizability  $\beta(-2\omega,\omega,\omega)$  as a function of the altitude, in the molecular frame {*a,b,c*}, for the  $C_{2v}$ -forbidden components. The exciting wavelength is 800 nm. Errors bars are calculated using eqn (S3) of ESI.†



**Fig. 7** Averages of the components of the hyperpolarizability  $\beta(-2\omega,\omega,\omega)$  as a function of the altitude, in the molecular frame  $\{a,b,c\}$ , for an exciting wavelength of 800 nm.  $\beta_{caa}$  and  $\beta_{aca}$  are indistinguishable.  $\beta_{\perp}$  is defined by eqn (4). The errors bars, calculated using eqn (S3) of ESI,† are smaller than the symbols.

quadrupolar contributions should be further explored to explain this behavior.

The evolutions of individual components are relatively modest, and the hyperpolarizability tensor remains closer to the bulk one than to the gas one. Indeed, similarly to the structural properties depicted in Fig. 5, the electric field generated by the point-charge environment differs from the bulk one in the interface area. However, its projection along the molecular *c*-axis loses only about 30% of its bulk value. This can be seen on Fig. 8 that shows the electrostatic field generated by the point-charged environment on the central water molecule, projected along the *c*-molecular axis. We attribute this strong remaining field to the presence of close neighbors, and H-bonds, even for the most external layer of water molecules,<sup>47</sup> see Fig. S3 of ESI.<sup>†</sup>

#### 3.3 Link with surface-second harmonic generation (S-SHG)

Even if the variations of individual molecular hyperpolarizability components could be considered as weak, the evolution of  $\beta_{ccc}$ and  $\beta_{\perp}$  is significant for the S-SHG interpretation. Whereas  $\beta_{ccc}$ is dominant in the bulk (4.0 vs. 0.5 a.u.), at the altitude of 0.04 nm,  $\beta_{ccc}$  and  $\beta_{\perp}$  have opposite values so that ( $\beta_{ccc} - \beta_{\perp}$ ) vanishes. The ratio of the two components of eqn (3) is strongly modulated by the altitude. This questions some frameworks used to analyze S-SHG signals, in which the ratio between  $\beta_{ccc}$ and  $\beta_{\perp}$  is typically considered as constant throughout the interface (see for example Section S4.4 of ESI<sup>†</sup>).

Further, we tackle an even more basic question: can the  $\beta$  tensor distribution be replaced by a single, average value at a given altitude? In the dipolar approximation, one may write the susceptibility  $\chi^{(2)}$  as proportional to the average over molecular configurations of the product of the rotational matrix *R* with the  $\beta$  tensor. In practice, it is often assumed that all molecules can be described by a single hyperpolarizability tensor  $\langle \beta_{ijk} \rangle$ , and that the fluctuations of the hyperpolarizability tensor and the molecular orientation are uncorrelated,



**Fig. 8** Electrostatic field generated by the environment along the molecular *c*-axis throughout the interface. In insert the electrostatic field distribution. The environment is built up to  $R_c = 2$  nm. The average values of the electric field along the *a* and *b*-axis remain null across the interface – data not shown.

*i.e.*  $\langle \beta_{ijk} \cos^l \theta \rangle = \langle \beta_{ijk} \rangle \langle \cos^l \theta \rangle$ . In our data, such equalities do not hold: Fig. 9 illustrates it for eqn (3), where the contributions of  $\beta_{\perp}$  and  $\beta_{ccc}$  to  $\chi^{(2)}_{ZZZ}$  are calculated either taking into account the fluctuations, or using the average  $\beta$  tensor and average orientation at each altitude (see Section S4.2 of ESI,† for details). The difference between the two curves emphasizes correlations between the molecular orientation and molecular hyperpolarizabilities at the interface.

In Fig. 9,  $\chi^{(2)}$  is expected to vanish in the bulk region. The deviations relative to zero in this area indicate that the convergence relative to sampling is not completely achieved. Indeed, the configuration-convergence study presented in Fig. 4 is focused on the hyperpolarizability components  $\beta$ . Since the  $\chi^{(2)}$  elements also include molecular orientations, their convergence is more demanding. However, at the interface, the differences between the two ways of calculating  $\chi^{(2)}$  on our data exceed these errors.



**Fig. 9**  $\chi^{(2)}_{ZZZ}$  as a function of altitude calculated using the right hand side of eqn (3) in 0.1 nm-thick slices. Either the correlations between hyperpolarizability and orientation are considered (blue squares, using  $\langle \beta_{ijk} \cos^l \theta \rangle$ ), or they are neglected (red circles, using  $\langle \cos^l \theta \rangle \langle \beta_{ijk} \rangle$ ).

## Paper

We conclude that orientation/hyperpolarizability correlations may be relevant to interpret quantitatively S-SHG intensities. Such correlations have been mentioned by Champagne and co-workers in a more complex system, composed of an organic dye embedded in a lipid bilayer,<sup>85</sup> and we show here that these may appear even at the neat liquid-vapor interface of pure water.

# 4 Conclusions

To summarize, a QM/MD approach was applied at the water liquid-vapor interface, to compute first hyperpolarizabilities of individual water molecules within the very specific environment where the S-SHG signal is generated. There is still a lot of room for improving the accuracy of the QM/MM approach (QM framework, electronic delocalization on several molecules, environment polarizability, long range electrostatic effects, local field factors,...), but qualitative results emerge. Within our approach, the dipolar contribution to the NLO response appears in a molecular layer of about half a nanometer thickness, where both the molecular orientation and the hyperpolarizability significantly differ from the bulk ones. The molecular first hyperpolarizabilities depend on the molecular environment up to radii larger than 2 nm, so that the S-SHG is very sensitive to the electrostatic properties of the interface. The hyperpolarizability tensor elements calculated in our PE approach almost respect the Kleinman symmetry, both in the bulk water phase and at the liquid-vapor interface, so that the Kleinman symmetry breaking observed experimentally may not originate in individual dipolar molecular response. The variations of the hyperpolarizability within the interface have an impact on the relative weights of its different components. Finally, we spotlight that the calculated molecular hyperpolarizabilities and molecular orientations are correlated: single average values for the molecular hyperpolarizability tensor and orientation are not sufficient to describe the entire distribution of molecular responses. The hyperpolarizability cannot be considered as constant for all water molecules at the liquid-vapor interface. More generally, this work indicates that one may need to consider different water populations at the interface to interpret the dipolar contributions to the surface second harmonic generation of aqueous interfaces. These hyperpolarizability fluctuations revealed at the neat air/water interface may also be relevant in more complex systems and could be more generally considered in quantitative interpretations of S-SHG experiments.

# Conflicts of interest

There are no conflicts to declare.

# Acknowledgements

We thanks Zacharie Behel for his help during the writing of FROG. We gratefully acknowledge support from the PSMN (Pôle Scientifique de Modélisation Numérique) of the Ecole Normale Supérieure de Lyon for the computing resources.

# Notes and references

- 1 E. Donohue, S. Khorsand, G. Mercado, K. M. Varney, P. T. Wilder, W. Yu, A. D. MacKerell, P. Alexander, Q. N. Van, B. Moree, A. G. Stephen, D. J. Weber, J. Salafsky and F. McCormick, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 17290–17297.
- 2 A. C. McGeachy, E. R. Caudill, D. Liang, Q. Cui, J. A. Pedersen and F. Geiger, *Chem. Sci.*, 2018, **9**, 4285–4298.
- 3 G. Licari, J. S. Beckwith, S. Soleimanpour, S. Matile and E. Vauthey, *Phys. Chem. Chem. Phys.*, 2018, **20**, 9328–9336.
- 4 H.-T. Bian, R.-R. Feng, Y. Guo and H.-F. Wang, J. Chem. Phys., 2009, 130, 134709.
- 5 H. C. Allen, N. N. Casillas-Ituarte, M. R. Sierra-Hernandez, X. Chen and C. Y. Tang, *Phys. Chem. Chem. Phys.*, 2009, **11**, 5538–5549.
- 6 M. Sulpizi, M. Salanne, M. Sprik and M.-P. Gaigeot, *J. Phys. Chem. Lett.*, 2012, 4, 83–87.
- 7 Y. Chen, H. I. Okur, N. Gomopoulos, C. Macias-Romero,
  P. S. Cremer, P. B. Petersen, G. Tocci, D. M. Wilkins,
  C. Liang, M. Ceriotti and S. Roke, *Sci. Adv.*, 2016, 2, 1–9.
- 8 P. E. Ohno, H. Chang, A. P. Spencer, Y. Liu, M. D. Boamah, H.-F. Wang and F. M. Geiger, *J. Phys. Chem. Lett.*, 2019, 10, 2328–2334.
- 9 L. Dalstein, K.-Y. Chiang and Y.-C. Wen, *J. Phys. Chem. Lett.*, 2019, **10**, 5200–5205.
- 10 T. T. Pham, A. Jonchère, J. F. Dufrêche, P. F. Brevet and O. Diat, *J. Chem. Phys.*, 2017, **146**, 144701.
- L. B. Dreier, C. Bernhard, G. Gonella, E. H. Backus and M. Bonn, *J. Phys. Chem. Lett.*, 2018, 9, 5685–5691.
- 12 R. K. Campen, A. K. Pymer, S. Nihonyanagi and E. Borguet, J. Phys. Chem. C, 2010, 114, 18465–18473.
- 13 S. Xu, S. Xing, S.-S. Pei, V. Ivanišev, R. Lynden-Bell and S. Baldelli, *J. Phys. Chem. C*, 2015, **119**, 26009–26019.
- 14 A. Marchioro, M. Bischoff, C. Lütgebaucks, D. Biriukov, M. Předota and S. Roke, *J. Phys. Chem. C*, 2019, 123, 20393–20404.
- J. H. Raberg, J. Vatamanu, S. J. Harris, C. H. van Oversteeg, A. Ramos, O. Borodin and T. Cuk, *J. Phys. Chem. Lett.*, 2019, 10, 3381–3389.
- 16 S. Pezzotti, D. R. Galimberti and M.-P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2019, **21**, 22188–22202.
- 17 Y. Shen, Annu. Rev. Phys. Chem., 1989, 40, 327-350.
- 18 N. A. Murugan, R. Apostolov, Z. Rinkevicius, J. Kongsted, E. Lindahl and H. Ågren, J. Am. Chem. Soc., 2013, 135, 13590–13597.
- 19 G. C. Gschwend, M. Kazmierczak, A. J. Olaya, P.-F. Brevet and H. H. Girault, *Chem. Sci.*, 2019, **10**, 7633–7640.
- 20 H.-T. Bian, R.-R. Feng, Y.-Y. Xu, Y. Guo and H.-F. Wang, *Phys. Chem. Chem. Phys.*, 2008, **10**, 4920–4931.
- 21 Q. Wei, D. Zhou and H. Bian, *Phys. Chem. Chem. Phys.*, 2018, 20, 11758–11767.
- 22 S. Ong, X. Zhao and K. B. Eisenthal, *Chem. Phys. Lett.*, 1992, 191, 327–335.
- 23 Y. Liu, E. C. Yan, X. Zhao and K. B. Eisenthal, *Langmuir*, 2001, **17**, 2063–2066.

- 24 G. C. Gschwend, A. Olaya and H. H. Girault, *Chem. Sci.*, 2020, **11**, 10807–10813.
- 25 A. G. F. de Beer and S. Roke, J. Chem. Phys., 2010, 132, 234702.
- 26 G. Tocci, C. Liang, D. M. Wilkins, S. Roke and M. Ceriotti, J. Phys. Chem. Lett., 2016, 7, 4311–4316.
- 27 S. Pezzotti, D. R. Galimberti, Y. R. Shen and M.-P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2018, **20**, 5190–5199.
- 28 C. Bouquiaux, C. Tonnelé, F. Castet and B. Champagne, J. Phys. Chem. B, 2020, 124, 2101–2109.
- 29 S. Roke, ChemPhysChem, 2009, 10, 1380-1388.
- 30 M. N. Nasir, E. Benichou, C. Loison, I. Russier-Antoine, F. Besson and P.-F. Brevet, *Phys. Chem. Chem. Phys.*, 2013, 15, 19919–19924.
- 31 C. Loison, M. N. Nasir, E. Benichou, F. Besson and P.-F. Brevet, *Phys. Chem. Chem. Phys.*, 2014, 16, 2136–2148.
- 32 R. Hartkamp, A.-L. Biance, L. Fu, J.-F. Dufrêche, O. Bonhomme and L. Joly, *Curr. Opin. Colloid Interface Sci.*, 2018, 37, 101–114.
- 33 W.-K. Zhang, D.-S. Zheng, Y.-Y. Xu, H.-T. Bian, Y. Guo and H.-F. Wang, *J. Chem. Phys.*, 2005, **123**, 224713.
- 34 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.
- 35 http://www.inkscape.org/.
- 36 I. Harczuk, O. Vahtras and H. Ågren, *Phys. Chem. Chem. Phys.*, 2016, **18**, 8710–8722.
- 37 T. N. Ramos, F. Castet and B. Champagne, J. Phys. Chem. B, 2021, 125, 3386–3397.
- 38 G. Maroulis, J. Chem. Phys., 2000, 113, 1813-1820.
- 39 C. Liang, G. Tocci, D. M. Wilkins, A. Grisafi, S. Roke and M. Ceriotti, *Phys. Rev. B*, 2017, 96, 1–6.
- 40 P. Kaatz, E. A. Donley and D. P. Shelton, *J. Chem. Phys.*, 1998, 108, 849–856.
- 41 T. D. Poulsen, P. R. Ogilby and K. V. Mikkelsen, J. Chem. Phys., 2001, 115, 7843–7851.
- 42 A. Osted, J. Kongsted, K. V. Mikkelsen, P. O. Åstrand and O. Christiansen, *J. Chem. Phys.*, 2006, **124**, 124503.
- 43 C. B. Nielsen, O. Christiansen, K. V. Mikkelsen and J. Kongsted, *J. Chem. Phys.*, 2007, **126**, 154112.
- 44 P. Guyot-Sionnest and Y. Shen, Phys. Rev. B: Condens. Matter Mater. Phys., 1988, 38, 7985.
- 45 C. A. Dailey, B. J. Burke and G. J. Simpson, *Chem. Phys. Lett.*, 2004, **390**, 8–13.
- 46 J. Kongsted, A. Osted, K. V. Mikkelsen and O. Christiansen, J. Chem. Phys., 2004, 120, 3787–3798.
- 47 S. Pezzotti, D. R. Galimberti and M. P. Gaigeot, *J. Phys. Chem. Lett.*, 2017, 8, 3133–3141.
- 48 K. V. Mikkelsen, Y. Luo, H. Ågren and P. Jørgensen, *J. Chem. Phys.*, 1995, **102**, 9362–9367.
- 49 K. Garrett, X. Sosa Vazquez, S. B. Egri, J. Wilmer, L. E. Johnson, B. H. Robinson and C. M. Isborn, *J. Chem. Theory Comput.*, 2014, **10**, 3821–3831.
- 50 S. Osella, N. A. Murugan, N. K. Jena and S. Knippenberg, J. Chem. Theory Comput., 2016, 12, 6169–6181.
- 51 N. H. List, H. J. A. Jensen and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2016, **18**, 10070–10080.
- 52 G. Licari, L. Cwiklik, P. Jungwirth and E. Vauthey, *Langmuir*, 2017, **33**, 3373–3383.

- 53 M. De Wergifosse, E. Botek, E. De Meulenaere, K. Clays and
  B. Champagne, *J. Phys. Chem. B*, 2018, **122**, 4993–5005.
- 54 C. Tonnelé, B. Champagne, L. Muccioli and F. Castet, *Chem. Mater.*, 2019, **31**, 6759–6769.
- 55 M. H. Cardenuto and B. Champagne, *Phys. Chem. Chem. Phys.*, 2015, **17**, 23634–23642.
- 56 M. de Wergifosse, F. Castet and B. Champagne, *J. Chem. Phys.*, 2015, **142**, 194102.
- 57 P. Beaujean and B. Champagne, *Theor. Chem. Acc.*, 2018, 137, 50.
- 58 C. Steinmann, P. Reinholdt, M. S. Nørby, J. Kongsted and J. M. H. Olsen, Int. J. Quantum Chem., 2019, 119, e25717.
- 59 S. Plimpton, J. Comput. Phys., 1995, 117, 1-19.
- 60 J. L. Abascal and C. Vega, J. Chem. Phys., 2005, 123, 234505.
- 61 R. E. Isele-Holder, W. Mitchell and A. E. Ismail, *J. Chem. Phys.*, 2012, **137**, 174107.
- 62 R. E. Isele-Holder, W. Mitchell, J. R. Hammond,
  A. Kohlmeyer and A. E. Ismail, *J. Chem. Theory Comput.*, 2013, 9, 5412–5420.
- 63 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *J. Comput. Chem.*, 2011, 32, 2319–2327.
- 64 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2008, 393, 51–57.
- 65 R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.
- 66 K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani and P. Dahle, *et al.*, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, 4, 269–284.
- 67 J. R. Hammond and K. Kowalski, J. Chem. Phys., 2009, 130, 194108.
- 68 F. Castet, E. Bogdan, A. Plaquet, L. Ducasse, B. Champagne and V. Rodriguez, *J. Chem. Phys.*, 2012, **136**, 024506.
- 69 D. E. Woon and T. H. Dunning, J. Chem. Phys., 1993, 98, 1358–1371.
- 70 D. E. Woon and T. H. Dunning, J. Chem. Phys., 1994, 100, 2975–2988.
- 71 P. Beaujean and B. Champagne, *J. Chem. Phys.*, 2019, **151**, 064303.
- 72 F. Castet and B. Champagne, J. Chem. Theory Comput., 2012, 8, 2044–2052.
- 73 P. Besalú-Sala, S. P. Sitkiewicz, P. Salvador, E. Matito and J. M. Luis, *Phys. Chem. Chem. Phys.*, 2020, 22, 11871–11880.
- 74 M. J. Paterson, O. Christiansen, F. Pawłowski, P. Jørgensen, C. Hättig, T. Helgaker and P. Sałek, *J. Chem. Phys.*, 2006, 124, 054322.
- 75 P. Beaujean and B. Champagne, *J. Chem. Phys.*, 2016, **145**, 044311.
- 76 P. Sałek, O. Vahtras, T. Helgaker and H. Ågren, J. Chem. Phys., 2002, 117, 9630–9645.
- 77 A. Willetts, J. E. Rice, D. M. Burland and D. P. Shelton, J. Chem. Phys., 1992, 97, 7590–7599.
- 78 M. T. Beerepoot, A. H. Steindal, N. H. List, J. Kongsted and J. M. H. Olsen, *J. Chem. Theory Comput.*, 2016, 12, 1684–1695.
- 79 E. R. Kjellgren, J. M. Haugaard Olsen and J. Kongsted, J. Chem. Theory Comput., 2018, 14, 4309–4319.

- 80 A. Marefat Khah, P. Reinholdt, J. M. H. Olsen, J. Kongsted and C. Hattig, J. Chem. Theory Comput., 2020, 16, 1373-1381.
- 81 G. Le Breton and L. Joly, J. Chem. Phys., 2020, 152, 241102.
- 82 W. Bu, D. Kim and D. Vaknin, J. Phys. Chem. C, 2014, 118, 12405-12409.
- 83 P.-F. Brevet, Surface Second Harmonic Generation, PPUR Presses Polytechniques, 1997.
- 84 W. Gan, D. Wu, Z. Zhang, R. R. Feng and H. F. Wang, J. Chem. Phys., 2006, 124, 114705.
- 85 C. Bouquiaux, C. Tonnelé, F. Castet and B. Champagne, J. Phys. Chem. B, 2020, 124, 2101-2109.

Paper

# Journal Name

# ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxx	ELECTRONIC SUPPLEMENTARY INFORMATION First Hyperpolarizability of Water at the Air-Vapor Interfa QM/MM Study Questions Standard Experimental Approxima	CTRONIC SUPPLEMENTARY INFORMATION Hyperpolarizability of Water at the Air-Vapor Interface: A MM Study Questions Standard Experimental Approximations				
	Guillaume Le Breton*, Oriane Bonhomme*, Pierre-François Brevet*, Emmanuel B and Claire Loison* <sup>‡</sup>	enichou*				
Contents						
S1 Methodological details		<b>S</b> 1				
S2 Interface structure		<b>S</b> 4				
S3 Hyperpolarizability		<b>S</b> 5				
S4 Susceptibility expressions		<b>S</b> 7				

# S1 Methodological details

S1.1 QM/MM workflows using the FROG tool.



Fig. S1 Scheme of the QM/MM procedure, handled by the FROG script. See the text for more details.

To compute the structural properties of the simulated system, such as the molecular orientations and the hyperpolarizabilities, a home-made software FROG (FROm molecular dynamics to second harmonic Generation) is used. This Python software opens trajectories from MD codes using the module MDAnalysis<sup>1</sup>. First, common structural analyses are performed: molecular orientation (see

<sup>\*</sup> University of Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-69622, Villeurbanne, France

<sup>‡</sup> Corresponding author. E-mail : claire.loison@univ-lyon1.fr

Figure 5(b) or S4), or H-bond network (see Figure S3). Then, for each molecule located nearby the interface, the neighborhood at a given time step is built up to a distance  $R_c$ , see Figure S1 (a). From the structural environment, an electrostatic one is built: each of the neighbor molecule is described using the same point charges as in the MD (TIP4P/2005), Figure S1 (b). The electrostatic potential generated by these neighbors is taken into account within the Polarizable Embedding scheme<sup>2,3</sup> of the DALTON software<sup>4</sup>, release 2018.2. In practice, the Hamiltonians of individual molecules are different since they depend on the neighborhood. For every configuration, FROG writes the QM input. Individual frequency-dependent hyperpolarizabilities are calculated for every molecule, Figure S1 (c). The QM calculations of the Non Linear Opical (NLO) properties are performed on a cluster. They are still more expensive that the MD run, but distributing them over many cores makes this approach numerically affordable. Finally, FROG analyses the individual QM outputs and compute the altitude-dependent distribution of the hyperpolarizability, Figure S1 (d).

FROG code is available on demand to the corresponding author.

#### S1.2 Quantum Chemistry Calculations

### S1.2.1 Basis set and functional for gas-phase calculations

The impact of basis set and DFT functional on the hyperpolarizability components for water molecule in gas phase are summarized in Table S1. We use the quantity  $\beta_{\parallel}$  to compare with the values of literature <sup>5–7</sup>. The definition of  $\beta_{\parallel}$  is :

$$\beta_{\parallel} = \frac{\sum_{i} \beta_{i} \mu_{i}}{|\mu|}, \quad \text{with} \quad \beta_{i} = \frac{1}{5} \sum_{j} \beta_{ijj} + \beta_{jij} + \beta_{jji}, \tag{S1}$$

and  $\mu_i$  are the dipole moment components (the permament dipole has been calculated at the DFT level in the gas phase).

The Dunning's correlation consistent (cc) basis sets<sup>8–10</sup> studied here (cc-VTZ, cc-VQZ, aug-cc-VDZ, aug-cc-VTZ, aug-cc-PVTZ and d-aug-cc-pVTZ) include polarization functions. The notations VDZ, VTZ, VQZ mean respectively double, triple and quadruple gaussian functions for each valence atoms. The augmented basis set – noted with aug and d-aug – respectively includes one or two diffuse functions of each angular momentum to represent the electron delocalization. We have added our own results for CCSD/d-aug-cc-pVTZ (performed using Dalton and its default settings).

Table S1 Comparison with literature of hyperpolarizability components for water molecule in gas phase (in atomic unit). Each components are computed for a static field and for a wavelength of 694 nm. The  $\beta_{\parallel}$  is defined in the Eq. S1. Reference notations : <sup>*a*</sup> is Ref. 6. <sup>*b*</sup> is Ref. 11. <sup>*c*</sup> is Ref. 5. <sup>*d*</sup> is the experimental value of Ref. 7 with its uncertainty. Results without references are our calculations. The absolute error (MAE) averaged of all non-zero  $\beta$  components is given for DFT/d-aug-cc-pVTZ relative to CCSD/d-aug-cc-pVTZ.

	$\beta_c$	aa	$\beta_c$	іса	$\beta_{cbb}$ $\beta_{bc}$		$\beta_{bcb}$ $\beta_{ccc}$		$oldsymbol{eta}_{\parallel}$		MAE			
Wavelength	8	694	8	694	8	694	8	694	8	694	8	694	8	694
					B3L	ΥP								
aug-cc-pVTZ	-14.9	-17.4	-14.9	-17.4	-5.0	-4.2	-5.0	-10.2	-11.75	-14.5	-19.0	-24.1		
aug-cc-pVTZ <sup>a</sup>	-14.4	-17.7	-14.4	-17.6	-5.0	-4.3	-5.0	-10.0	-11.2	-14.4	-18.3	-24.1		
d-aug-cc-pVTZ	-11.8	-14.2	-11.8	-14.0	-5.3	-6.5	-5.3	-9.5	-14.7	-18.8	19.1	-24.8	1.1	1.7
CAM-B3LYP														
aug-cc-pVTZ	-13.6	-15.7	-13.6	-15.7	-4.2	-3.5	-4.2	-8.4	-10.3	-12.5	-16.1	-21.0		
aug-cc-pVTZ <sup>a</sup>	-13.0	-15.5	-13.0	-15.5	-4.2	-3.5	-4.2	-8.2	-9.6	-12.0	-16.1	-20.5		
d-aug-cc-pVTZ	-10.6	-12.5	-10.6	-12.4	-4.2	-5.0	-4.2	-7.4	-12.1	-15.3	-16.1	-20.6	1.3	1.3
					CCS	D								
d-aug-cc-pVDZ <sup>b</sup>											-15.0			
d-aug-cc-pVTZ <sup>b</sup>											-17.9			
d-aug-cc-pVQZ <sup>b</sup>											-18.3			
d-aug-cc-pV5Z <sup>b</sup>											-18.2			
d-aug-cc-pVTZ <sup>c</sup>	-9.9		-9.9		-5.7		-5.7		-14.0					
d-aug-cc-pVTZ	-9.9	-11.5	-9.9	-11.4	-5.7	-6.7	-5.7	-8.5	-14.0	-16.9	-17.7	-21.7	0	0
Exp. $ESHG^d$	cp. ESHG <sup>d</sup> $-22.0 \pm 0.9$													

The impact of basis set and DFT functional on the hyperpolarizability components for a water molecule in **gas phase** are summarized in ESI Table S1, and compared to calculated or experimental values of literature <sup>5–7</sup>. The results obtained with the d-aug-cc-pVTZ basis set are considered as satisfactory.

## S1.2.2 Bulk liquid phase calculations

The impact of the basis set on the  $\beta$  components and on  $|\beta|$  for water in the bulk phase is depicted in Fig. S2, with

$$\boldsymbol{\beta}| = \left(\sum_{i,j,k} \beta_{ijk}^2\right)^{\frac{1}{2}}.$$
(S2)

The value of  $\beta$  is calculated using DFT (CAM-B3LYP) and an embedding up to  $R_c = 2$  nm for about 55 000 configurations. As for the



Fig. S2 Evolution of  $\beta$  components for several basis set, using DFT with the CAM-B3LYP functional,  $R_c = 20$  Å, at 800 nm.  $|\beta|$  is defined by Eq. S2.

gas phase, polarization and diffuse function basis set are necessary, and we have chosen to use the d-aug-pVTZ in the our work. The relative smooth convergence of the first hyperpolarizability as a function of the basis set permits to gain confidence in the absence of the "spill-out" effect which can appear using QM/MM approaches with large basis sets<sup>12,13</sup>.

The averages and standard deviations of the  $\beta$  components within our response scheme in liquid water phase are compared to literature results in Table S2. Both calculation schemes use the TIP4P/2005 rigid force field,  $R_c = 15$  Å and the basis is d-aug-pVTZ. Liang et al. have used Couple-Cluster calculation (CCSD), while we have used DFT.

Table S2 Molecular SHG first hyperpolarizability  $\beta(-2\omega, \omega, \omega)$  for  $\omega = 0$  of Liang et al.<sup>5</sup> using CCSD(left column), compared to our CAM-B3LYP results CAM-B3LYP (right column), both using the d-aug-pVTZ basis set. Hyperpolarizabilities are in atomic units. Standard deviations are provided in brackets.

	Liquid						
λ	∞5	×					
$\beta_{caa}$	-2.06 [1.06]	-1.87 [1.01]					
$\beta_{cbb}$	1.59 [1.83]	2.43 [1.74]					
$\beta_{ccc}$	3.00 [3.04]	3.80 [2.82]					
$\beta_{bbb}$	0.00 [5.74]	0.00 [5.21]					
$\beta_{bcc}$	0.00 [1.64]	0.01 [1.2]					
$\beta_{baa}$	0.00 [1.38]	-0.01 [1.31]					
$\beta_{aaa}$	0.00 [1.56]	0.01 [1.47]					
$\beta_{abb}$	0.00 [1.64]	0.00 [1.56]					
$\beta_{acc}$	0.00 [1.13]	0.01 [1.11]					
$\beta_{abc}$	0.00 [0.78]	0.00 [0.85]					

### S1.3 Error Bars

Error bars  $\Delta$  for any quantity *X* were computed as follow:

$$V[X] = \frac{1}{N} \sum_{n=1}^{N} (X_n)^2 - \left(\frac{1}{N} \sum_{n=1}^{N} X_n\right)^2, \text{ and } \Delta[X] = \sqrt{\frac{V[X]}{N_f \times S}} ,$$
 (S3)

where *X* can be the orientation of a water molecule or components of  $\beta$  tensor. *N* stands for the total number of configurations (individual calculations) considered for the average value, i.e. for all molecules of all the frames treated – typically several tens of thousands. For the altitude profiles, the *N* configurations describe molecules within *XY*-slices of thickness of a few angstrom. *N*<sub>f</sub> stands for the number of different time steps treated (typically 32 and 80 for bulk and interface respectively). These are considered as independent with respect to the SHG response (see main text Section 2.3). *S* stands for a geometrical factor taking into account possible spatial correlations between water molecules which are close to each other.

We have considered that molecules distant from more than 0.8 nm are independent with respect to their NLO properties. Therefore, the system can be divided into small sub-box of 1.6 nm per side. For the bulk phase, these sub-boxes are  $1.6 \times 1.6 \times 1.6 \times 1.6$  mm<sup>3</sup> cubes. At the interface they are  $1.6 \times 1.6 \times 0.1$  nm<sup>3</sup> parallelepipeds (*X*, *Y*, *Z* directions) since the response is discretized along the *Z* direction. Thus, for the bulk phase, using box size in nm, *S* = (size of the MD box/1.6)<sup>3</sup> and for the interface *S* = (size of the MD box along *xy*/1.6)<sup>2</sup>. For the bulk phase  $\sqrt{S} = 6$  and at the interface  $\sqrt{S} = 3$ .

# S2 Interface structure

Figure S3 reports the number of given and accepted H-bonds per water molecule, as a function of the altitude, i.e. the position relative to the liquid-vapor interface along the Z-direction.



Fig. S3 Number of given and accepted H-bonds per water molecule, as a function of the altitude  $\Delta Z = Z - Z_0$ , where  $Z_0$  is the average position of the interface. Squares and circles are used for the average accepted and given H-bonds, respectively. The errors bars are calculated using Eq S3. Hyperbolic tangent water density fit in dark solid line.

Figure S4 reports the orientational order parameter  $P_2$  of the water dipole moment relative to the normal of the liquid-vapor interface:  $P_2 = \langle \frac{1}{2}(3\cos^2\theta - 1) \rangle$ , where  $\theta = (c, Z)$ .  $P_2 \simeq 0$  in the bulk indicates an isotropic distribution.  $P_2 \simeq 1$  indicates a propensity to align along the surface normal  $(\pm Z)$ .  $P_2 \simeq -1/2$  indicates a propensity to align parallel to the surface. We measure negative values and conclude that the dipole moment of the water molecule show a propensity to align parallel to the surface.



Fig. S4 Orientational order parameter of the water dipole moment as a function of the altitude  $\Delta Z = Z - Z_0$ , where  $Z_0$  is the average position of the interface. Hyperbolic tangent water density fit in dark solid line.

## S3 Hyperpolarizability

In this section, we complete the methodological tests presented in the main text on the number of configurations needed form averaging  $N_f$ , and the radius of the electrostatic environment  $R_c$ .

### S3.1 Convergence relative to the number of configurations

This section provide supplementary information concerning the convergence of the  $\beta$  components relative to the value obtained with largest frame number (80 frames, 4294 water molecules in total) for a chosen altitude ( $Z - Z_0 = -0.6 \pm 0.5$  Å) for which the density is about half the bulk density.

The Figure S5 depicts the  $C_{2\nu}$  forbidden components, which converges within  $\pm 0.05$  atomic unit after 2000 configurations.



Fig. S5 Evolution of the  $\beta$  C<sub>2 $\nu$ </sub> forbidden components mean value in atomic unit with respect to the reference value obtained using 4294 molecular configurations. The dashed line represent 0.05 a.u. deviation.

Using 4000 configurations (80 frames for our systems size) lead to converged results. Overall, the statistical error made by this configuration number should be smaller than the one made by our choice of QM method – DFT instead of CCSD for instance. Note that because of the density evolution with the altitude, the number of molecular configurations available decreases for increasing altitudes. At altitudes higher than ( $Z - Z_0 = -0.6$  Å), *i.e.* closer to the vapor phase, our hyperpolarizability values are less precise.

In total, 48 650 hyperpolarizability calculations were necessary for the Main-Text-Figure 7. They have cost approximately 11 k-hours for a single 2.6 Ghz CPU.

### **S3.2** Electrostatic embedding : *R<sub>c</sub>* convergence

This section provides supplementary information concerning the convergence of the  $\beta$  components relative to the size of the embedding environment  $R_c$ . Beyond a given distance, the electrostatic field generated by the surrounding no longer affects the mean  $\beta$  value.

#### S3.2.1 Bulk phase

Figure S6 depicts the deviation of the  $\beta$  component mean values for several  $R_c$ , with respect to the results obtain with  $R_c = 2$  nm. 48 000 configurations have been used per  $R_c$ . The  $C_{2\nu}$  forbidden components are not shown since their mean value are zero for all  $R_c$ . From 1.0 to 1.5 nm, an important evolution is observed for all the non-zero components while between 1.5 and 2.0 nm, there is almost no evolution. Therefore, an  $R_c$  of 2.0 nm seems sufficient to converge the hyperpolarizability components in the water bulk phase for our PE0-TIP4P/2005 scheme.



Fig. S6 Relative evolution of the  $\beta$  components in the bulk phase for several  $R_c$  with respect to the results obtain with  $R_c = 2.0$  nm.

#### S3.2.2 At the interface:

As for the time-convergence, the evolution of the mean  $\beta$  as a function of the  $R_c$  has not been computed for all the altitudes but only at one:  $Z - Z_0 = -0.6 \pm 0.5$  Å. 80 frames have been used for each  $R_c$  to ensure time-convergence. The Figure S7 shows the evolution of the relative mean  $\beta$  with respect to  $R_c$  for the non-zero components. The components forbidden by the  $C_{2\nu}$  water symmetry remains null in averaged for all  $R_c$  (data not shown). As for the bulk phase, an  $R_c$  about 2.0 nmm is sufficient to converge the averaged  $\beta$ . We have studied the convergences of  $\beta$  for slices at other altitudes, and the results are similar – data not shown.



Fig. S7 Evolution of the non-zero  $\beta(-2\omega, \omega, \omega)$  components with respect to the  $R_c$  using the relative deviation compared to the results obtained for  $R_c = 5.0$  nm at the altitude  $\Delta Z = -0.06$  nm. The exciting wavelength is 800 nm. The dashed black lines represent a deviation of  $\pm 1$  %. The norms of the  $C_{2\nu}$  forbidden components are all bellow 0.3 atomic units for all  $R_c$ .

# S4 Susceptibility expressions

To discuss relevant orientation/hyperpolarizability correlations, we recall here common expressions of the surface susceptibility components for the S-SHG from neat air/water interface considering only the dipolar contribution, as can be found in Refs. 6,14–16.

The molecular frame has the *c*-axis along the  $C_{2\nu}$  axis of the water molecule,  $\{a, c\}$  contains the three atoms. The water molecule belongs to  $C_{2\nu}$  symmetry, with seven non-zero polarizability tensor elements in SHG :  $\beta_{ccc}$ ,  $\beta_{caa}$ ,  $\beta_{cbb}$ ,  $\beta_{aac} = \beta_{aca}$ ,  $\beta_{bbc} = \beta_{bcb}$ . For the  $C_{\infty}$  neat air/water interface, there are seven non-zero elements, for the surface susceptibility tensor  $\chi^{(2)}$ :  $\chi_{ZZZ}$ ,  $\chi_{ZXX} = \chi_{ZYY}$ ,  $\chi_{XZX} = \chi_{YZY} = \chi_{XXZ} = \chi_{YYZ}$ .

Considering the twisted angle  $\psi$  and the azimuthal angle  $\phi$  as isotropic and uncorrelated with the  $\beta$  tensor elements yields expressions found in Ref. 6 :

$$\chi_{ZXX}^{(2)} = \frac{N}{4} \left\{ 2 \left\langle \left(\beta_{caa} + \beta_{cbb}\right) \cos \theta \right\rangle + \left\langle \left(2\beta_{ccc} - \beta_{caa} - \beta_{cbb} - 2\beta_{aca} - 2\beta_{bcb}\right) \cos \theta \sin^2 \theta \right\rangle \right\}, \\ \chi_{XZX}^{(2)} = \frac{N}{4} \left\{ 2 \left\langle \left(\beta_{aca} + \beta_{bcb}\right) \cos \theta \right\rangle + \left\langle \left(2\beta_{ccc} - \beta_{caa} - \beta_{cbb} - 2\beta_{aca} - 2\beta_{bcb}\right) \cos \theta \sin^2 \theta \right\rangle \right\}, \\ \chi_{ZZZ}^{(2)} = \frac{N}{4} \left\{ 2 \left\langle \beta_{ccc} \cos^3 \theta \right\rangle + \left\langle \left(\beta_{caa} + \beta_{cbb} + 2\beta_{aca} + 2\beta_{bcb}\right) \cos \theta \sin^2 \theta \right\rangle \right\}.$$
(S4)

where N is the number of molecules. If one defines intermediate values

$$\beta'_{\perp} = \beta_{caa} + \beta_{cbb}, \tag{S5}$$

$$\beta_{\perp}^{\prime\prime} = \beta_{aca} + \beta_{bcb}, \tag{S6}$$

$$\beta_{\perp} = \frac{1}{2} \left( \beta_{\perp}' + 2\beta_{\perp}'' \right). \tag{S7}$$

It simplifies to :

$$\chi_{ZXX}^{(2)} = \frac{N}{2} \left\{ \langle \beta_{\perp}' \cos \theta \rangle + \left\langle (\beta_{ccc} - \beta_{\perp}) \cos \theta \sin^2 \theta \rangle \right\}, \\ \chi_{XZX}^{(2)} = \frac{N}{2} \left\{ \langle \beta_{\perp}'' \cos \theta \rangle + \left\langle (\beta_{ccc} - \beta_{\perp}) \cos \theta \sin^2 \theta \rangle \right\rangle \right\}, \\ \chi_{ZZZ}^{(2)} = \frac{N}{2} \left\{ \left\langle \beta_{ccc} \cos^3 \theta \right\rangle + \left\langle \beta_{\perp} \cos \theta \sin^2 \theta \right\rangle \right\}.$$
(S8)

Noticeably, these expressions neglect  $C_{2\nu}$ -forbidden elements. In our calculations, the averages of these elements are null at all altitudes (see Main Text Figure 6).

#### S4.1 Neglecting orientation/hyperpolarizability correlations

Now, the common assumption  $\langle \beta_{ijk} \cos^l \theta \rangle = \langle \beta_{ijk} \rangle \langle \cos^l \theta \rangle$  is considered. This assumption allows to extract the molecule's orientation from the S-SHG signal in many studies. Equation S8 simplifies to:

$$\chi_{ZXX}^{(2)} = \frac{N}{2} \left\{ \langle \beta_{\perp}' \rangle \langle \cos \theta \rangle + \langle (\beta_{ccc} - \beta_{\perp}) \rangle \left\langle \cos \theta \sin^2 \theta \right\rangle \right\},$$
  

$$\chi_{XZX}^{(2)} = \frac{N}{2} \left\{ \langle \beta_{\perp}'' \rangle \langle \cos \theta \rangle + \langle (\beta_{ccc} - \beta_{\perp}) \rangle \left\langle \cos \theta \sin^2 \theta \right\rangle \right\},$$
  

$$\chi_{ZZZ}^{(2)} = \frac{N}{2} \left\{ \langle \beta_{ccc} \rangle \left\langle \cos^3 \theta \right\rangle + \langle \beta_{\perp} \rangle \left\langle \cos \theta \sin^2 \theta \right\rangle \right\}.$$
(S9)

### S4.2 Taking into account altitude variations

Equations S8 can be rewritten to take into account variations of  $\langle \beta_{ijk} \cos^l \theta \rangle$  or density  $\rho$  as a function of the altitude  $Z - Z_0$  (see Figure 5). The susceptibility is the sum over the whole system, which can be decomposed into slices of different positions  $Z^6$ . Equations S8

become:

$$\chi_{ZXX}^{(2)} = \frac{A}{2} \int_{Z} \left\{ \left\langle \beta_{\perp}' \cos \theta \right\rangle_{Z} + \left\langle \left( \beta_{ccc} - \beta_{\perp} \right) \cos \theta \sin^{2} \theta \right\rangle_{Z} \right\} \rho(Z) dZ, \\ \chi_{XZX}^{(2)} = \frac{A}{2} \int_{Z} \left\{ \left\langle \beta_{\perp}'' \cos \theta \right\rangle_{Z} + \left\langle \left( \beta_{ccc} - \beta_{\perp} \right) \cos \theta \sin^{2} \theta \right) \right\rangle_{Z} \right\} \rho(Z) dZ, \\ \chi_{ZZZ}^{(2)} = \frac{A}{2} \int_{Z} \left\{ \left\langle \beta_{ccc} \cos^{3} \theta \right\rangle_{Z} + \left\langle \beta_{\perp} \cos \theta \sin^{2} \theta \right\rangle_{Z} \right\} \rho(Z) dZ.$$
(S10)

Where the notation  $\langle \cdot \rangle_Z$  corresponds to the averaging over molecules situated in a slab parallel to the interface, characterized by its position *Z* and its thickness *dZ*. In our work, *dZ* is 1 Å. The part between curly brackets of Equation S10,  $\{\langle \beta_{ccc} \cos^3 \theta \rangle_Z + \langle \beta_{\perp} \cos \theta \sin^2 \theta \rangle_Z\}$ , is plotted in Main-Text-Figure 9 (blue squares).

Similarly, Equations S9 neglecting orientation/hyperpolarizability correlations become:

$$\chi_{ZXX}^{(2)} = \frac{A}{2} \int_{Z} \left\{ \langle \beta_{\perp}' \rangle_{Z} \langle \cos \theta \rangle_{Z} + \langle \beta_{ccc} - \beta_{\perp} \rangle_{Z} \left\langle \cos \theta \sin^{2} \theta \right\rangle_{Z} \right\} \rho(Z) dZ,$$
  

$$\chi_{XZX}^{(2)} = \frac{A}{2} \int_{Z} \left\{ \langle \beta_{\perp}'' \rangle_{Z} \langle \cos \theta \rangle_{Z} + \langle \beta_{ccc} - \beta_{\perp} \rangle_{Z} \left\langle \cos \theta \sin^{2} \theta \right\rangle_{Z} \right\} \rho(Z) dZ,$$
  

$$\chi_{ZZZ}^{(2)} = \frac{A}{2} \int_{Z} \left\{ \langle \beta_{ccc} \rangle_{Z} \left\langle \cos^{3} \theta \right\rangle_{Z} + \langle \beta_{\perp} \rangle_{Z} \left\langle \cos \theta \sin^{2} \theta \right\rangle_{Z} \right\} \rho(Z) dZ.$$
(S11)

The part between curly brackets of Equation S11 has been plotted in Main-Text-Figure 9 (red circles).

The experimental S-SHG signal is proportional to the integral. Whereas the *Z*-dependency of the density can be guessed, the evolution of the NLO-part is not known. To extract molecular information from experimental data, it is often assumed that the S-SHG signals emerge from an 'one-layer' system, i.e. the integrand is independent of *Z*.

#### S4.3 Molecular Kleinman symmetry

In the numerical calculations, we observe that the average hyperpolarizability tensor almost respects the Kleinman symmetry throughout the interface, see Figure 7. Therefore, we shall suppose in the following that for all altitudes,

$$\langle \beta_{\perp}^{\prime\prime} \rangle_{Z} \simeq \langle \beta_{\perp}^{\prime} \rangle_{Z},$$
 (S12)

$$\langle \beta'_{\perp} \rangle_Z \simeq \frac{2}{3} \langle \beta_{\perp} \rangle_Z.$$
 (S13)

Which remains true if one neglects the Z dependence.

For our system, using (Eq. S12), neglecting both Z-dependance and orientation/hyperpolarizability correlations, Equations S11 become

$$\chi^{(2)}_{XZX} \simeq \chi^{(2)}_{ZXX}, \tag{S14}$$

$$\chi_{ZXX}^{(2)} = \frac{N}{2} \left\{ \beta_{\perp}' \left\langle \cos \theta \right\rangle + \left( \beta_{ccc} - \beta_{\perp} \right) \left\langle \cos \theta \sin^2 \theta \right\rangle \right\},$$
  
$$\chi_{ZZZ}^{(2)} = \frac{N}{2} \left\{ \beta_{ccc} \left\langle \cos^3 \theta \right\rangle + \beta_{\perp} \left\langle \cos \theta \sin^2 \theta \right\rangle \right\}.$$
 (S15)

Where  $\langle \beta_{ijk} \rangle$  is noted  $\beta_{ijk}$  for simplicity. The difference between  $\chi^{(2)}_{XZX}$  and  $\chi^{(2)}_{ZXX}$  observed experimentally is not explained using this formalism.

### S4.4 Link with the molecular orientation parameter

Equations S15 are straightforwardly equivalent to :

$$\chi_{ZXX}^{(2)} \simeq \frac{N}{2} \left\{ \left( \beta_{ccc} - \frac{1}{3} \beta_{\perp} \right) \left\langle \cos \theta \right\rangle + \left( \beta_{\perp} - \beta_{ccc} \right) \left\langle \cos^{3} \theta \right\rangle \right\},$$
  
$$\chi_{ZZZ}^{(2)} = \frac{N}{2} \left\{ \beta_{\perp} \left\langle \cos \theta \right\rangle + \left( \beta_{ccc} - \beta_{\perp} \right) \left\langle \cos^{3} \theta \right\rangle \right\}.$$
 (S16)
Noting  $\alpha$  the ratio between  $\beta_{ccc}$  and  $\beta_{\perp}$  with  $\beta_{\perp} = \alpha \beta_{ccc}$ , then

$$\chi_{ZXX}^{(2)} = \frac{N}{2} \beta_{ccc} \left\{ \frac{3-\alpha}{3} \left\langle \cos \theta \right\rangle + (\alpha - 1) \left\langle \cos^3 \theta \right\rangle \right\},$$
  
$$\chi_{ZZZ}^{(2)} = \frac{N}{2} \beta_{ccc} \left\{ \alpha \left\langle \cos \theta \right\rangle + (1-\alpha) \left\langle \cos^3 \theta \right\rangle \right\}.$$
 (S17)

Within this framework, neglecting the altitude dependence, knowing the ratio  $\alpha$ , and measuring the experimental observable  $\chi$ , one typically studies the molecular orientation parameter *D* (e.g. Refs. 14,17):

$$D = \frac{\left\langle \cos^3 \theta \right\rangle}{\left\langle \cos \theta \right\rangle}.$$

Note that some authors use the inverse as a definition of molecular orientation parameter<sup>15</sup>. If one supposes here  $\alpha$  different from 1, then :

$$D = \frac{\alpha \chi_{ZXX} - (1 - \alpha/3) \chi_{ZZZ}}{(\alpha - 1)(\chi_{ZXX} + \chi_{ZZZ})}.$$
(S18)

Results of our calculations question such a framework:

- $\alpha = \beta_{\perp} / \beta_{ccc}$  varies through the interface, see Figure 7.
- At several altitudes, the common approximation (β<sub>ijk</sub> cos<sup>l</sup> θ) = (β<sub>ijk</sub>)(cos<sup>l</sup> θ) does not hold. Figure 9 shows the difference for the NLO part of the integrand in χ<sup>(2)</sup><sub>ZZZ</sub>, with and without this assumption.

# Notes and references

- 1 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, J. Comput. Chem., 2011, 32, 2319–2327.
- 2 J. M. Olsen, K. Aidas and J. Kongsted, J. Chem. Theory Comput., 2010, 6, 3721-3734.
- 3 C. Steinmann, P. Reinholdt, M. S. Nørby, J. Kongsted and J. M. H. Olsen, Int. J. Quantum Chem., 2019, 119, e25717.
- 4 K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle et al., Wiley Interdisciplinary Reviews: Computational Molecular Science, 2014, 4, 269–284.
- 5 C. Liang, G. Tocci, D. M. Wilkins, A. Grisafi, S. Roke and M. Ceriotti, Phys. Rev. B, 2017, 96, 1-6.
- 6 T. T. Pham, A. Jonchère, J. F. Dufrêche, P. F. Brevet and O. Diat, J. Chem. Phys., 2017, 146, 144701.
- 7 D. P. Shelton and J. E. Rice, Chemical Reviews, 1994, 94, 3–29.
- 8 R. A. Kendall, T. H. Dunning and R. J. Harrison, J. Chem. Phys., 1992, 96, 6796-6806.
- 9 D. E. Woon and T. H. Dunning, J. Chem. Phys., 1993, 98, 1358-1371.
- 10 D. E. Woon and T. H. Dunning, J. Chem. Phys., 1994, 100, 2975-2988.
- 11 P. Beaujean and B. Champagne, The Journal of Chemical Physics, 2016, 145, 044311.
- 12 P. Reinholdt, J. Kongsted and J. M. H. Olsen, J. Phys. Chem. Lett., 2017, 8, 5949-5958.
- 13 C. Steinmann, P. Reinholdt, M. S. Nørby, J. Kongsted and J. M. H. Olsen, Int. J. Quantum Chem., 2019, 119, e25717.
- 14 P.-F. Brevet, Surface Second Harmonic Generation, PPUR presses polytechniques, 1997.
- 15 W.-k. Zhang, D.-s. Zheng, Y.-y. Xu, H.-t. Bian, Y. Guo and H.-f. Wang, J. Chem. Phys., 2005, 123, 224713.
- 16 M. N. Nasir, E. Benichou, C. Loison, I. Russier-Antoine, F. Besson and P.-F. Brevet, Phys. Chem. Chem. Phys., 2013, 15, 19919–19924.
- 17 A. G. F. de Beer and S. Roke, J. Chem. Phys., 2010, 132, 234702.

# VI. C | Molecular Response at the Quadrupolar order

The relation between the first hyperpolarizabilities and the induced dipole and quadrupole moment at the second harmonic frequency,  $\mu^{SHG}$  and  $Q^{SHG}$  respectively, is given by:

$$\boldsymbol{\mu}^{SHG} = \frac{1}{2!} \beta^{dd} (2\omega, \omega, \omega) : \mathbf{e}^{\omega} \mathbf{e}^{\omega} + \frac{1}{2!} \beta^{dq} (2\omega, \omega, \omega) : \mathbf{e}^{\omega} \nabla \mathbf{e}^{\omega}$$
$$Q^{SHG} = \frac{1}{2!} \beta^{qd} (2\omega, \omega, \omega) : \mathbf{e}^{\omega} \mathbf{e}^{\omega}$$
(VI.4)

Where  $\beta^{dd}$  is the dipolar (presented in the previous article),  $\beta^{dq}$  the dipole-quadrupole and  $\beta^{qd}$  the quadrupole-dipole hyperpolarizabilities. See Section III. B.1.5 for more details. However, as evoked early on, we have to include the quadrupole contributions to correctly describes the Surface SHG signal of pure water [41]. Hence, this section presents the computed quadrupolar first hyperpolarizability of water using our QM/MM scheme. The study was performed in the vacuum, in the liquid phase and at the liquid-gas interface using FROG.

# VI. C.1 Vacuum and liquid phase

First of all, we compute the molecular first hyperpolarizability of the water molecules in the vacuum and in the water liquid phase. The details of the calculations for the dipolar term were presented in the previous section and summarized in Chapter III. B.2.3. For the quadrupolar terms, the comparison with vacuum phase literature results [89], obtained using the Finite Field scheme and CCSD, was made in Section III. B.2.3. In short, we get a good agreement between DFT/CAM-B3LYP and CCSD. For simplicity we did not try to change the basis set for the quadrupolar terms. We have found for the dipolar ones that using d-aug-cc-pVTZ led to meaningful results, and we assumed that this would also be the case for the quadrupolar first hyperpolarizability in the vacuum and bulk phase.

## **Reminders on symmetries:**

In general, the response tensors shall respect two types of symmetries: geometrical and related to the perturbative operator frequencies.

The geometrical one depends on the structure of the molecule if the tensor is expressed in the molecular frame, and on the system if the tensor is expressed in the laboratory frame. For instance, water molecule has a  $C_{2v}$  symmetry: it has 2 inversion plane, see Figure VI.3 and a rotational axis. It implies that the response components with an odd number of a or b should be zero – for instance  $\beta_{abc}^{dd}$  or  $\beta_{abcc}^{dq}$  have to be null.

The frequency symmetry is related to the nature of the response tensor. The dipolar first hyperpolarizability is the perturbation of the molecule's dipole moment upon a quadratic dipolar perturbation,  $\beta^{dd}(2\omega, \omega, \omega)_{ijk} = +\langle \langle \hat{\mu}_i; \hat{\mu}_j^{\omega}, \hat{\mu}_k^{\omega} \rangle \rangle$ . The nature of perturbations,  $\hat{\mu}_j^{\omega}$ and  $\hat{\mu}_k^{\omega}$ , are the same: there are dipole operators associated with the  $\omega$  frequency. In the particular case of second harmonic generation, the *intrinsic symmetry* shall be respected, meaning that both perturbations can be permuted. Therefore, this imposes that  $\beta_{ijk}^{dd} = \beta_{ikj}^{dd}$ . If the fundamental frequency goes to zero, the observable and the permutations can be permuted because they are all dipole operators. This case is often called *Kleinman symmetry*, and this leads to these equalities  $\beta_{ijk}^{dd} = \beta_{ikj}^{dd} = \beta_{jik}^{dd}$  for the dipolar first hyperpolarizability.



Figure VI.3: Sketch of the water molecule in its frame, abc, along with its planes of symmetry: ac and bc planes.

The quadrupolar first hyperpolarizabilities respect similar frequencies related to the operators. However, in our formalism, they are slightly more complicated, see the last part of Chapter III, Section III. B.1.5. To sum up the equalities between the components due to these operator permutations, we have for optical frequencies:

$$\begin{aligned} \beta_{ijk}^{dd} &= \beta_{ikj}^{dd}, \text{ intrinsic symmetry} \\ \beta_{ijkl}^{dq} &= \beta_{ijlk}^{dq}, \text{ because } \hat{Q}_{lk}^{\omega} = \hat{Q}_{kl}^{\omega} \\ \beta_{ijkl}^{qd} &= \beta_{ijlk}^{qd}, \text{ intrinsic symmetry} \\ \beta_{ijkl}^{qd} &= \beta_{jikl}^{qd}, \text{ because } \hat{Q}_{ij} = \hat{Q}_{ji} \end{aligned}$$

and in the zero-frequency limit (Kleinman symmetry):

$$\beta_{ijk}^{dd} = \beta_{ikj}^{dd} = \beta_{jik}^{dd}, \text{ intrinsic + Kleinman symmetry} \beta_{ijkl}^{dq} = \beta_{jikl}^{dq} = \beta_{klij}^{qd} = \beta_{klji}^{qd}, \text{ intrinsic + Kleinman symmetry} \beta_{ijkl}^{dq} = \beta_{ijlk}^{dq}, \ \beta_{ijkl}^{qd} = \beta_{jikl}^{dq}, \text{ because } \hat{Q}_{lk} = \hat{Q}_{kl}$$

Note that we have equality between the  $\beta^{dq}$  and  $\beta^{qd}$  rather than between components of the same response tensor thanks to the intrinsic symmetry. Moreover, in our case we are supposed to be off resonances (the fundamental wavelength is 800 nm). This means that the denominator (involving  $\omega - \omega_{n0}$  where  $\omega$  is the optical frequency and  $\omega_{n0}$  transition energy in pulsation units, see for instance in Chapter III Equation III.32) of the response tensor should not be impacted by the order of the operators even for optical fundamental: in this limit we recover the Kleinman frequency. Therefore, we expect to be close to the zero-frequency limit behavior with respect to the components equalities presented above for the water molecule for calculations at 800 nm.

# Symmetry for the dipolar hyperpolarizability:

For the first hyperpolarizability at the dipolar level,  $\beta^{dd}$ , the C<sub>2v</sub> fixes 3 triplets: (*aac*, *bbc*, *ccc*). The combination within these triplets along with the intrinsic frequency equality for second harmonic process leads to only 5 non-zeros and independent components:  $\beta^{dd}_{aca}, \beta^{dd}_{caa}, \beta^{dd}_{bcb}, \beta^{dd}_{cbb}$  and  $\beta^{dd}_{ccc}$ . Note that more components can be non-zero: we have that  $\beta^{dd}_{aca} = \beta^{dd}_{aac}$  and  $\beta^{dd}_{bcb} = \beta^{dd}_{bbc}$  due to the intrinsic symmetry. In the zero-frequency limit, there are only 3 non-zeros and independent components:  $\beta^{dd}_{caa}, \beta^{dd}_{caa}, \beta^{dd}_{cbb}, \beta^{dd}_{ccc}$  because we have another set of equalities:  $\beta^{dd}_{caa} = \beta^{dd}_{aca} = \beta^{dd}_{aac}$  and  $\beta^{dd}_{cbb} = \beta^{dd}_{bcb}$ 

### Symmetry for the quadrupolar hyperpolarizability::

 $\beta^{dq}$  and  $\beta^{qd}$  are 3x3x3x3 tensors: applying the C<sub>2v</sub> symmetry of the water molecule leaves us to 6 quadruplets:(*aabb*, *aacc*, *bbcc*, *aaaa*, *bbbb*, *cccc*). To compute the number of independent components, we should also discuss the frequency symmetry for these tensors, but with more care. Indeed, the operator involved in the computation of these response tensors are not of the same nature. The dipolar operator and the quadrupolar ones cannot be "inverted" within the response tensor. That is why in our formalism, the Kleinman symmetry would impose equality between the  $\beta^{dq}$  and  $\beta^{qd}$  tensor.

In the general case, there are 15 non-zero and independent components for the dipole-quadrupole term:  $\beta_{aaaa}^{dq}$ ,  $\beta_{bbbb}^{dq}$ ,  $\beta_{aabb}^{dq}$ ,  $\beta_{bbaa}^{dq}$ ,  $\beta_{baab}^{dq}$ ,  $\beta_{baab}^{dq}$ ,  $\beta_{baab}^{dq}$ ,  $\beta_{baab}^{dq}$ ,  $\beta_{bcaa}^{dq}$ ,  $\beta_{ccaa}^{dq}$ ,  $\beta_{acac}^{dq}$ ,  $\beta_{caac}^{dq}$ ,  $\beta_{bbc}^{dq}$ ,  $\beta_{bbcb}^{dq}$ ,  $\beta_{bbcb}^{dq}$ ,  $\beta_{cbbc}^{dq}$ ,  $\beta_{bbc}^{dq}$ ,  $\beta_{bbcb}^{dq}$ ,  $\beta_{bbc$ 

In the case of the Kleinman symmetry, it drops to 12 because we can invert the two dipolar operator at the first and second component position:  $\beta_{ijkl}^{dq} = \beta_{jikl}^{dq}$ , for instance  $\beta_{abab}^{dq} = \beta_{baab}^{dq}$ . Hence, the independent components are:  $\beta_{aaaa}^{dq}$ ,  $\beta_{bbbb}^{dq}$ ,  $\beta_{cccc}^{dq}$ ,  $\beta_{aabb}^{dq}$ ,  $\beta_{bbaa}^{dq}$ ,  $\beta_{aacc}^{dq}$ ,  $\beta_{ccaa}^{dq}$ ,  $\beta_{acac}^{dq}$ ,  $\beta_{bbcc}^{dq}$ ,  $\beta_{bcbc}^{dq}$ ,  $\beta_{bcbc}^{dq}$ .

Similarly, for the quadrupolar-dipole first hyperpolarizability for a non-zero frequency, there are 12 independent non-zero components:  $\beta_{aaaa}^{qd}$ ,  $\beta_{bbb}^{qd}$ ,  $\beta_{cccc}^{qd}$ ,  $\beta_{aabb}^{qd}$ ,  $\beta_{abab}^{qd}$ ,  $\beta_{aacc}^{qd}$ ,  $\beta_{acac}^{qd}$ ,  $\beta_{bbcc}^{qd}$ ,  $\beta_{ccbb}^{qd}$ ,  $\beta_{bcbc}^{qd}$ . At the zero frequency limit, the number of independent components is not reduced for  $\beta^{qd}$  but it creates equality with  $\beta^{dq}$  as explained above.

# Molecular symmetry in the liquid phase:

In the liquid phase this picture is more complex. As presented extensively in the Chapter V devoted to liquid phase SHG, the molecular  $C_{2v}$  symmetry is broken due to the electrostatic environment. Hence, each molecule can have  $\beta$  tensors which do not fulfill the  $C_{2v}$  symmetry. This has important consequences for SHS as the square of the individual  $\beta$ s is involved directly in the collected intensity, see Chapter V. Yet, each molecule respect strictly the frequency symmetries presented above since they are related to the nature of the response components: they are independent of the wave function shape.

When computing the averaged values, we retrieve the  $C_{2v}$  molecular symmetry:  $\langle \beta^{dd} \rangle$  as indeed five independent components with a non-zero average at 800 nm and 3 in the zero frequency limit. In order to describe the fluctuations, we use the standard deviation defined as  $\sigma[X] = \sqrt{\langle X^2 \rangle - \langle X \rangle^2}$  for each component. This term represents how strongly the first hyperpolarizabilities are impacted by the fluctuation of the electrostatic environment. Hence there is no standard deviation for the molecular values in the vacuum phase.

Table VI.1: Dipolar first hyperpolarizability components in the vacuum and bulk phase at the DFT/CAM-B3LYP (d-aug-cc-pVTZ) level for zero frequency and at 800 nm. All values are expressed in atomic units and in the T convention, see Equation VI.4. For the bulk phase, the averages and the standard deviations (in bracket) are shown. The components shown are either equal to one of the presented component using the tensor intrinsic symmetry, here  $\beta_{ijk}^{dd} = \beta_{ikj}^{dd}$ , either with a value less than 0.1 atomic unit. Note that standard deviation in the bulk phase of the values with a zero average are presented in the table SI.2 of the previous paper [71].

$\beta^{dd}$	Vacuum		Liquid		
	$+\infty$	$800~\mathrm{nm}$	$+\infty$	$800~\mathrm{nm}$	
aac	-9.8	-11.0	-1.9 [1.4]	-2.0 [1.6]	
bbc	-4.4	-6.5	2.4 [2.4]	2.3 [2.9]	
caa	-9.8	-11.1	-1.9 [1.4]	-1.9 [1.6]	
cbb	-4.4	-5.0	2.4 [2.4]	2.6 [2.8]	
ccc	-11.7	-14.0	3.9 [ 3.9]	4.0 [4.4]	

# Molecular first hyperpolarizabilities:

The data for the dipole-dipole terms has been obtained using 156 250 configuration with PE-Long FROG-like embedding, see the end of Chapter IV. The direct inclusion is up to 10 Å, the implicit part up to 20 Å. The data for the quadrupole terms has been obtained using a direct inclusion up to 10 Å with 2392 configurations. All the calculations have been made using TIP4P/2005 bulk water simulation at 300 K.

The results for the dipolar term is presented in Table VI.1, for the dipolar-quadrupolar and quadrupolar-dipolar in Table VI.2. For all of these quantities, only the components with a calculated non-zero average value are reported. If components are linked with intrinsic symmetry, they also are not shown. The distribution obtained for some other components are given in Annex IX. D.1.

First of all, we indeed observe that the  $C_{2v}$  symmetry is respected for all the hyperpolarizabilities – for instance  $\beta_{abbc}^{dq} = 0$  (not shown in Table VI.2). As discussed in the article presented previously [71], the dipolar first hyperpolarizability in the liquid phase almost respected the Kleinman symmetry at 800 nm. For the quadrupolar terms, the frequency dispersion between the zero frequency and 800 nm is important in both the vacuum and gas phase: about 10%. We also observe at 800 nm that the Kleinman symmetry is almost respected but is not rigorously. At zero frequency, there is indeed the expected equality between some components of  $\beta^{dq}$  and  $\beta^{qd}$  For instance,  $\beta^{dq}(0,0,0)_{aacc} = \beta^{qd}(0,0,0)_{ccaa}$  in both the vacuum and bulk phases. However, this is no longer the case at 800 nm. As recalled by Table VI.1, the effect of the electrostatic environment

Table VI.2: Dipole-Quadrupole and Quadrupole-Dipole first hyperpolarizability components in the vacuum and bulk phase at the DFT/CAM-B3LYP (d-aug-cc-pVTZ) level for the zero frequency and at 800 nm. All values are expressed in atomic units and in the T convention, see Equation VI.4. For the bulk phase, the averages and the standard deviations (in bracket) are shown. The components not shown are either equal to one of the presented component using the tensor intrinsic symmetry, here  $\beta_{ijkl}^{dq} = \beta_{ijlk}^{qd}$  or  $\beta_{ijkl}^{qd} = \beta_{jikl}^{qd}$ , either with a value less than 1 atomic unit.

Component	$\beta^{dq} V$	acuum	$\beta^{dq}$ I	Liquid	$  \beta^{qd} V$	acuum	$\beta^{qd}$ L	liquid
	$+\infty$	800 nm	$+\infty$	800 nm	$+\infty$	800 nm	$+\infty$	800 nm
aaaa	-170.9	-184.9	-129.4 [8.8]	-138.2 [9.8]	-170.9	-180.8	-129.4 [8.8]	-135.5 [9.5]
aabb	-55.9	-62.4	-46.5[5.0]	-51.2 [5.9]	-60.0	-68.0	-46.1 [5.8]	-51.3 [6.9]
aacc	-45.4	-50.1	-33.3 [2.8]	-36.1 [3.2]	-47.7	-52.4	-32.4 [3.0]	-34.8 [3.3]
abab	-102.7	-114.0	-85.1 [7.3]	-92.9 [8.6]	-102.7	-114.1	-85.1 [7.3]	-92.5 [8.6]
acac	-88.5	-96.4	-68.1 [4.7]	-73.0 [5.3]	-88.5	-94.9	-68.1 [4.7]	-72.0 [5.2]
baab	-102.7	-121.7	-85.1 [7.3]	-97.4 [9.5]	-102.7	-114.1	-85.1 [7.3]	-92.5 [8.6]
bbaa	-60.0	-77.7	-46.1 [5.8]	-56.4 [7.9]	-55.9	-62.7	-46.5 [5.0]	-51.3 [6.0]
bbbb	-340.7	-400.8	-296.8 [32.8]	-339.5 [41.2]	-340.7	-383.0	-296.8 [32.8]	-327.0 [38.8]
bbcc	-70.8	-88.5	-57.5 [7.1]	-68.2 [9.4]	-67.1	-74.4	-56.2 [6.5]	-61.4 [7.5]
bcbc	-108.7	-127.1	-95.6 [8.9]	-108.4 [11.2]	-108.7	-120.6	-95.6 [8.9]	-103.9 [10.3]
caac	-88.5	-98.3	-68.1 [4.7]	-74.1 [5.5]	-88.5	-94.9	-68.1 [4.7]	-72.0 [5.2]
cbbc	-108.7	-121.6	-95.6 [8.9]	-105.1 [10.4]	-108.7	-120.6	-95.6 [8.9]	-103.9 [10.3]
ccaa	-47.7	-55.3	-32.4 [3.0]	-36.5 [3.5]	-45.4	-49.5	-33.3 [2.8]	-35.7 [3.1]
ccbb	-67.1	-76.0	-56.2 [6.5]	-62.5 [7.6]	-70.8	-79.6	-57.5 [7.1]	-63.2 [8.4]
cccc	-229.7	-255.6	-193.8 [17.7]	-211.6 [20.4]	-229.7	-248.0	-193.8 [17.7]	-206.4 [19.6]

on the dipolar first hyperpolarizability is dramatic: the averaged values evolve importantly and the induced standard deviations are of the same order of magnitude as the averaged values. In this case, we should consider that each molecule has its own  $\beta^{dd}$  tensor. This was rationalized in Chapter IV in which we explain this behavior using the dipolar second hyperpolarizability  $\gamma^{(3)}(2\omega, \omega, \omega, 0)$  to quantify the impact of the electrostatic environment on  $\beta^{dd}$ .

For the two quadrupolar first hyperpolarizabilities, the effect of the environment is weaker with no modification of the sign of the component. Indeed, all the components tend to be more positive in the liquid phase but their order in amplitude is not changed. This behavior is comparable with the one observed for the dipolar second hyperpolarizability between the vacuum values and the bulk ones, see Chapter IV. Hence, we plotted in Figure VI.4 all the  $\beta^{dq}$  averaged components for the liquid phase with respect to the vacuum ones at 800 nm. We observed a linear dependency where the values of the vacuum are 0.8 times smaller than the one obtained for the bulk. Similar results are found for the quadrupole-dipole hyperpolarizability – data not shown. No explanation has been found yet to explain this 0.8 factor.

Finally, the relative standard deviations of the quadrupolar hyperpolarizabilities in the liquid phase are smaller compared to the dipolar one. The electrostatic environment induces fluctuations, but only about 10 % of the average value. This behavior is also found for the component with zero averaged: the standard deviation is about 5 a.u. As for the dipolar terms, we found that the components involving many b directions are more impacted by the fluctuation: the *bbbb* component has the highest standard deviation. Therefore, in a first glance, we can say that each molecule has the same quadrupolar first hyperpolarizability.

#### Laboratory first hyperpolarizabilities:

Now let us discuss the final molecular quantity of interest: the first hyperpolarizability ex-



Figure VI.4: Averaged dipole-quadrupole first hyperpolarizability components of water in the bulk phase with respect to the vacuum value. The dots are the values obtained for all the components at the DFT/CAM-B3LYP (d-aug-cc-pVTZ) level for 800 nm. The dashed line is a guide to the eye and corresponds to a linear dependency of 0.8. The values are presented in atomic units.

pressed in the laboratory frame. As discuss before, SHG within the dipolar approximation in a centro-symmetric media imposes that the first hyperpolarizability average in the laboratory frame is zero:  $\langle \mathscr{B}^{dd} \rangle = 0$ . However, this is not the case for the quadrupolar terms,  $\mathscr{B}^{dq}$  and  $\mathscr{B}^{qd}$ . This tensor should respect the centro-symmetry: components with an even number of x, y and zcan be non-zero. Moreover, the permutation of the indexes, xyz, should provide the same values because of isotropy. Hence, there are only six quadruplets, (xxxx, yyyy, zzzz, xxzz, yyzz, xxyy), by applying the first rule, and the second one reduces to two independent quadruplets: (xxxx, xxzz). For the dipolar second hyperpolarizability tensor in the laboratory frame,  $\Gamma(2\omega, \omega, \omega, 0)$ , this leads to only two independent components because all the operators involved in the response are dipolar operators:  $\Gamma \propto \langle \langle \hat{\mu}; \hat{\mu}^{\omega}, \hat{\mu}^{\omega}, \hat{\mu}^{0} \rangle \rangle$ . But we have to be more careful for the quadrupolar first hyperpolarizabilities. Indeed, there are dipole and quadrupolar operators which cannot be exchanged,  $\mathscr{B}^{dq} \propto \langle \langle \hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\mu}}^{\omega}, \hat{Q}^{\omega} \rangle \rangle$ . For a physical insight:  $\mathscr{B}^{dq}_{xyxy}$  links the induced dipole moment along the x axis for an exciting field along the y axis and another exciting field along the y direction which has a gradient along the x axis.  $\mathscr{B}_{xxuu}^{dq}$  links the induced dipole moment along the x axis for an exciting field along the x axis and another exciting field along the y direction which has a gradient along the y axis. In the first case the gradient of the electric field is not in the direction of the field polarization (xy), and in the other the direction of the gradient and the field matches (yy). Hence, we may have  $\mathscr{B}_{xyxy}^{dq} \neq \mathscr{B}_{xxyy}^{dq}$  because they do not represent the same physical light conversion process!

Hence, for the dipolar-quadrupolar first hyperpolarizability we have three independent compo-

Table VI.3: Average Dipole-Quadrupole and Quadrupole-Dipole first hyperpolarizability components expressed in the laboratory frame in the liquid phase, DFT/CAM-B3LYP (d-aug-cc-pVTZ) for the zero frequency and at 800 nm. The standard deviations are presented in brackets. All values are expressed in atomic units and in the T convention, see Equation VI.4. Only the components larger than 1 atomic unit are shown.

Components	$\mathscr{B}^{dq}$ Liquid		$\mathscr{B}^{qd}$ Liquid	
	$+\infty$	800 nm	$+\infty$	800 nm
XXXX	-209.2 [48.6]	-232.9[58.8]	-209.2 [48.6]	-226.0 [55.9]
ххуу	-44.3 [8.3]	-50.6 [10.2]	-44.3 [8.3]	-48.4 [9.4]
XXZZ	-44.6 [8.4]	-50.9 [10.4]	-44.6 [8.4]	-48.7 [9.6]
xyxy	-81.9 [10.3]	-90.6 [12.4]	-81.9 [10.3]	-88.2 [11.8]
хуух	-81.9 [10.3]	-90.6 [12.4]	-81.9 [10.3]	-88.2 [11.8]
XZXZ	-82.2 [10.4]	-90.9 [12.6]	-82.2 [10.4]	-88.5 [12.0]
XZZX	-82.2 [10.4]	-90.9 [12.6]	-82.2 [10.4]	-88.5 [12.0]
yxxy	-81.9 [10.3]	-90.5 [12.4]	-81.9 [10.3]	-88.2 [11.8]
yxyx	-81.9 [10.3]	-90.5 [12.4]	-81.9 [10.3]	-88.2 [11.8]
yyxx	-44.3 [8.3]	-50.5 [10.2]	-44.3 [8.3]	-48.4 [9.5]
уууу	-207.2 [48.1]	-230.5[58.1]	-207.2 [48.1]	-223.8 [55.2]
yyzz	-44.3 [8.3]	-50.5 [10.2]	-44.3 [8.3]	-48.4 [9.5]
yzyz	-81.9 [10.2]	-90.5 [12.4]	-81.9 [10.2]	-88.2 [11.8]
yzzy	-81.9 [10.2]	-90.5 [12.4]	-81.9 [10.2]	-88.2 [11.8]
ZXXZ	-82.2 [10.4]	-90.9 [12.6]	-82.2 [10.4]	-88.5 [12.0]
ZXZX	-82.2 [10.4]	-90.9 [12.6]	-82.2 [10.4]	-88.5 [12.0]
zyyz	-81.9 [10.2]	-90.6 [12.4]	-81.9 [10.2]	-88.2 [11.8]
zyzy	-81.9 [10.2]	-90.6 [12.4]	-81.9 [10.2]	-88.2 [11.8]
ZZXX	-44.6 [8.4]	-50.9 [10.3]	-44.6 [8.4]	-48.7 [9.6]
zzyy	-44.3 [8.3]	-50.6 [10.2]	-44.3 [8.3]	-48.4 [9.4]
ZZZZ	-209.1 [47.4]	-232.8 [57.4]	-209.1 [47.4]	-225.9[54.5]

nents which can be non-zero:  $\mathscr{B}_{xxxx}^{dq}$ ,  $\mathscr{B}_{xxyy}^{dq}$ ,  $\mathscr{B}_{xyxy}^{dq}$ . The rest is given by permutation between x, y and z. For the quadrupolar-dipole term, there are also three independent components:  $\mathscr{B}_{xxxx}^{qd}$ ,  $\mathscr{B}_{xxyy}^{qd}$ ,  $\mathscr{B}_{xyxy}^{qd}$ .

The results obtained with our sampling made on 2392 configurations are presented in Table VI.3 for the quadrupolar first hyperpolarizabilities in the laboratory frame. We indeed retrieve this strong centro-symmetry: there are only three non-zero independent components for both quadrupole terms. The fact that we have a good agreement between the permutation of the xyz axis, as expect for a centro-symmetric liquid, shows that the sampling made is sufficient to sample the orientational degrees of freedom. With respect to the quadrupolar first hyperpolarizability, a sample of about 2500 molecules is centro-symmetric for the bulk phase.

What about the frequency symmetries? We may think that the Kleinman symmetry is always respected for a given quadrupolar tensor (for instance  $\mathscr{B}_{xyxy}^{qd} = \mathscr{B}_{yxxy}^{qd}$ ) but this is due to the

centro-symmetry. At zero frequency, the Kleinman symmetry acts on the dipole-quadrupole and quadrupole-dipole tensor, for instance,  $\mathscr{B}^{dq}(0,0,0)_{xxxx} = \mathscr{B}^{qd}(0,0,0)_{xxxx}$ . As for the molecular values, this symmetry is no longer strictly respected at 800 nm, but the values are extremely close, for example  $\mathscr{B}^{dq}_{xxyy}$  and  $\mathscr{B}^{qd}_{yyxx}$ .

The standard deviations obtained with the quadrupolar tensors are much larger in the laboratory frame compared to the molecular one. This is expected because the fluctuations evaluated in the laboratory frame are the combination of the molecular orientations and response tensor fluctuations. If the standard deviation of the laboratory response tensor is zero, it would mean that each molecule shares the same molecular response tensor and that they present exactly the same orientation in the laboratory frame, which is not possible in the liquid phase.

# VI. C.2 At the liquid-gas interface

In this last section, we present the evolution of these tensors at the liquid-gas interface. We focus on the calculations at 800 nm since it is the most relevant for the comparison with the experiment. Note that in the zero-frequency limit shows similar evolution.

For bulk SHG, for instance SHS, the standard deviation plays an important role, but less for the surface SHG. Indeed, for both the dipolar and quadrupolar terms, there is a net average at the interface for the laboratory first hyperpolarizabilities: hence the "incoherent" term can be fully neglected. Hence, we will use the Equations VI.4 as the source term for the second harmonic field: this equation involves the averaged value of the laboratory hyperpolarizabilities and not their fluctuations. Hence, even if standard deviations hold an important physical meaning, we will rather focus on the average value for all the hyperpolarizabilities in the laboratory frame.



Figure VI.5: Left: Density in kg/L at the interface. The dots are the numerical results and the full line the fit using a tanh function. RIGHT: The number of molecules involved in each altitude. The dashed vertical line is the altitude 2.5 Å and the horizontal ones to a density of 0.05 kg/L or 700 QM/MM calculations.

### The system and the sampling:

As for the previous article, we have computed a neat water slab in a huge box:  $[50\text{\AA} \times 50 \text{\AA} \times 400 \text{\AA}]$ . The liquid phase spread about 90 Å in the z direction. We have computed the first hyperpolarizabilities tensor for only one interface, the upper one. To select which molecule should be computed at the QM/MM level, we have used the floating interface layer selection FROG: the first 4 molecular layers are treated – using the Pytim package [109]. In this case, we are sure to have the interface and a part of the bulk phase to describe the system efficiently. Moreover, adding too much bulk phase would be a waste of numerical computational power.

Today, we have computed the results for 160 MD frames separated by 10 ns to ensure timeuncorrelation. This sample represents 97625 QM/MM calculations. On the left part of Figure VI.5, the density obtained at the upper interface with respect to the altitude is presented. The density is also fitted using the usual tanh function: the altitude is renormalized with respect to the Gibbs dividing altitude: at z = 0 there is about half of the liquid density.

### Page 220

The space discretization done along the z axis is 1 Å. On the right side of Figure VI.5 the number of molecules involved in every z-slab used later on to define the averaged value of the hyperpolarizabilities along the z axis is plotted. Every slab is a [50 Å  $\times$  50 Å  $\times$  1 Å] rectangle. For the bulk-like altitude (large positive z), the number of QM/MM calculations on these slabs for the 160 time steps used is about 12 000. For slabs corresponding to the interface, this number decreases and consequently the statistical averaging is less precise. Hence, we have decided to present the results only for slabs with meaningful sampling number. The last altitude is about 2.5 Å: after this altitude the density is less than 0.05 kg/L, and the number of molecules involved in the QM/MM calculation less than 700.

# Evolution of the quadrupolar molecular hyperpolarizabilities:

The dipolar evolution in the molecular frame was already described in the previous section, see Figure 7 of the article. The evolution the average and standard deviation of some component of the dipole-quadrupole  $\beta^{dq}$  for the water molecule are presented in Figure VI.6. The quadrupole-dipole  $\beta^{qd}$  follows the same trends. Several components were not plotted since many of them are null in average. These components correspond to the ones forbidden by the  $C_{2v}$  symmetry. Hence, as for the dipolar molecular first hyperpolarizability, the quadrupolar ones respect in average the  $C_{2v}$  symmetry.

The evolution of the average value are small across the interfaces. For all the components, the average evolves monotonously from their bulk value, the big dot, to their vacuum one, the dashed lines. As for the dipolar molecular  $\beta$ , the averaged values are closer to the bulk value rather than the vacuum ones. However, contrary to the dipolar term, the quadrupolar first hyperpolarizability in the bulk phase is close to their vacuum value. Hence, the quadrupolar  $\beta$ s evolution is smaller compared to the dipolar term throughout the interface.

The evolution of the standard deviation is not very insightful: there is almost no changes through the interface. The slight evolution of the highest component, *bbbb*, may be attributed to a lack of statistical averaging. Similar results are found for the dipolar  $\beta$  – data not shown.

Hence, as for the bulk phase, it would make sense to attribute for each molecule the same quadrupolar molecular first hyperpolarizability at the interface: there are few fluctuations of the individual values and the evolution of the average value throughout the interface is also small.

# Laboratory hyperpolarizabilities: dipolar term

Yet, we are more interested in the laboratory first hyperpolarizabilities,  $\mathscr{B}$ , since there are directly involved in the second harmonic process. Contrarily to the bulk case, the laboratory axis definition matters at the surface. Hence, we have defined the +z-axis to be normal to the liquid-gas interface, pointing toward the gas phase. The x and y are defined in the surface plane in order to construct



Figure VI.6: Dipole-Quadrupole first hyperpolarizability in the molecular frame at the interface. On the left is plotted its averaged value, on the right its standard deviation. The full lines are the computed value at the interface, the big dots the bulk values and the dashed lines the vacuum ones.



Figure VI.7: Average dipolar first hyperpolarizability in the laboratory frame. LEFT: Some components expected to be null with respect to the centro-symmetry or  $C_{\infty}$  symmetry. The dashed line represents the maximal fluctuation observed. RIGHT: Two components that should be equal according to the  $C_{\infty}$  symmetry. The green curve is the difference between these two components.

a direct orthonormal frame.

First of all, let us check that  $\mathscr{B}$  at the interface respects the  $C_{\infty}$  symmetry, as we have checked that its respect the centro-symmetry in the bulk phase. The  $C_{\infty}$  symmetry imposes that the non-zero tensor component can have only even number of x and y, and that permuting x and y leaves the tensor unchanged. Of course there can have fluctuations at the molecular scale: we expect the  $C_{\infty}$  symmetry to be respected in averaged. For the dipolar terms, we expect to have only three independent and non-zero terms:  $\mathscr{B}_{zxx}^{dd}$ ,  $\mathscr{B}_{xzx}^{dd}$  and  $\mathscr{B}_{zzz}^{dd}$ .

The left part of Figure VI.7 presents components of  $\mathscr{B}^{dd}$  that should be null in the bulk and at the interface due to the  $C_{\infty}$  symmetry. We found for the  $C_{\infty}$  forbidden components some fluctuation of the average value within a maximal amplitude about 0.2 a.u., dashed lines in Figure VI.7.

The right part shows the  $\mathscr{B}_{zxx}^{dd}$  and  $\mathscr{B}_{zyy}^{dd}$  components. They should be null in the bulk phase and can be non-zero at the interface. Yet, these components should always be equal, hence is also plotted the difference:  $\mathscr{B}_{zxx}^{dd} - \mathscr{B}_{zyy}^{dd}$ . The zxx and zyy components presented in Figure VI.7 are always very close to each other, within 0.1 a.u. Note that the great evolution at altitude about 2 Å may be attributed to the very small sampling.

The aim of this study is to use the average laboratory first hyperpolarizability as the basic element to build the SHG response of the interface. As we are limited by the numerical cost of the QM/MM calculation, we have to deal with numerical errors due to a small statistical sampling. This is especially true at the interface for the laboratory tensors: we have to sample both the molecular orientation and its electrostatic environment. Therefore, we should characterize the statistical "noise". This permits to decide which averaged components can be considerate to have a net average, and which should be disregarded and put to zero. To solve this problem, we can use the  $C_{\infty}$  symmetry: we know that many components have to be zero in or-



Figure VI.8: Average dipolar first hyperpolarizability in the laboratory frame for the components involved in the S-SHG signal. The dashed lines correspond to the statistical noise estimated from symmetry consideration, see the main text.

der to describe this liquid-gas interface. Using this approach, one would say that we have an error of 0.2 a.u. for the dipolar laboratory first hyperpolarizability. Using this, we can consider that the components of the left part of Figure VI.7 are null as expected. Within this error, we verify that  $\mathscr{B}_{zxx}^{dd} = \mathscr{B}_{zyy}^{dd}$  as imposed by the  $C_{\infty}$  symmetry.

In Figure VI.8 is presented the relevant component involved in the S-SHG:  $\mathscr{B}_{zxx}^{dd}$ ,  $\mathscr{B}_{xzx}^{dd}$  and  $\mathscr{B}_{zzz}^{dd}$  at 800 nm. All the averages are above the "noise" threshold at 0.2 a.u. after -1.5 Å approximately. However, we have a problem for the  $\mathscr{B}_{zzz}^{dd}$  value at the beginning of the "interface", between -4 Å and -1.5 Å, because the average value is within the "noise". Hence, we will have to refine more our noise analysis to be sure that the value used later on can be differentiated from the noise. For instance, we can perform a sliding average by sampling over an increasing number of MD frame. It should also be noted that the quantity involved in the SHG process is not the averaged  $\mathscr{B}$ , but the averaged  $\mathscr{B}$  weighted by the density.

Apart from this point, we have obtained results for  $\mathscr{B}$  which are acceptable: the  $C_{\infty}$  symmetry is respected and we are able to determine the evolution of the three independent  $\mathscr{B}^{dd}$  components involved in the Surface-SHG process.

# Laboratory hyperpolarizabilities: quadrupolar terms

Similarly to the dipolar terms, Figure VI.9 presents some components that should be zero on the left part, and equal and non-zero on the right part. Following the same arguments discussed for the dipolar terms, we would say that the quadrupolar terms respect the  $C_{\infty}$  symmetry up to 4 a.u. Using this idea, we show that odd number of x and y in a components lead to zero averages, and that permuting x and y leaves the component unchanged on average.



Figure VI.9: Average dipole-quadrupole first hyperpolarizability in the laboratory frame. LEFT: Some components expected to be null with respect to the centro-symmetry or  $C_{\infty}$  symmetry. The dashed line represents the maximal fluctuation observed. RIGHT: Difference between components that should be equal according to the  $C_{\infty}$  symmetry.



Figure VI.10: Quadrupolar average first hyperpolarizabilities in the laboratory frame at the interface. On the left is plotted the dipole-quadrupole, on the right the quadrupole-dipole. The full lines are the computed value at the interface, the dots the bulk values and the dashed line the vacuum ones. Note that  $\langle \mathscr{B}_{zzxx}^{dq} \rangle$  and  $\langle \mathscr{B}_{xzzx}^{dq} \rangle$  are almost on top of each other.

Some relevant components for the dipole-quadrupole and quadrupole-dipole are presented in Figure VI.10. All of these components are non-zero in the bulk, the big dots, in the vacuum one, the dashed lines, and also throughout the interface. We observe that each component evolves smoothly from the bulk value to the vacuum one, while staying closer to the bulk one. This evolution and the average are above the "statistical noise" of 4 a.u defined using the symmetries. Moreover,

the Kleinman symmetry is respected for the quadrupole terms, see for instance the values of the components  $\langle \mathscr{B}_{zxzx}^{dq} \rangle$  and  $\langle \mathscr{B}_{xzzx}^{dq} \rangle$  or the  $\langle \mathscr{B}_{xxzx}^{dq} \rangle$  and  $\langle \mathscr{B}_{zxxx}^{qd} \rangle$ .

# VI. D | Conclusion and Perspectives

# **Conclusions:**

In conclusion, we have successfully computed the molecular first hyperpolarizability up to the quadrupolar order for water in the liquid phase and at the liquid-gas interface within an electro-static embedding.

Contrarily to the dipole first hyperpolarizability, we have found that the quadrupolar first hyperpolarizabilities tensors,  $\beta^{dq}$  and  $\beta^{qd}$ , in the liquid phase are close to the vacuum ones. We still observe some fluctuations due to the electrostatic environment, but these variations are small relative to the averaged values. Hence, in a good approximation one can attribute to each molecule the same quadrupolar  $\beta$ s. The evolution of each  $\beta$  component is continuous at the interface, going from the bulk values to the vacuum ones. Throughout the interface, the  $\beta$  are closer to the liquid than the vacuum ones. No important evolution of the standard deviation has been observed. For each quantity, the Kleinman approximation is closely respected.

Computing these tensors in the laboratory frame, the  $\mathscr{B}s$ , require more statistical effort. We have verified that both the dipolar and quadrupolar hyperpolarizabilities respect in average the interface  $C_{\infty}$  symmetry within a certain error. This error based on the symmetry provides a first estimation of the statistical validity of our results. The average value computed for the quadrupolar terms and their evolution at the interface are larger than this error. The picture is less clear for the dipolar term as one component,  $\mathscr{B}_{zzz}^{dd}$ , is within the statistical error. Hence, we will have to refine our error estimation for this particular term.

Overall, we have obtained microscopic results which are almost usable for the SHG intensity prediction, and thus comparison with experimental results. We are still working on the link between the microscopic tensors,  $\mathscr{B}s$ , and the Surface-SHG intensity. As presented in the Chapter III, we still have to treat carefully the fundamental electromagnetic field intensity at the interface, but also the link between the induced dipole moment at the second harmonic frequency and the mesoscopic polarization at the second harmonic frequency.

# **Perspectives:**

An important perspective for us would be for example to compare the results obtained numerically with the experimental ones with another system: methanol-water mixtures at the liquid-gas interface. Indeed, such systems are more complex to understand since there are more effects occurring when we increase the concentration of methanol of the liquid: the methanol can populate more and more the interface. Hence, numerically speaking we should first performed the MD simulations at the interface for several concentration and then compute the optical response at each concentration. We have already performed the MD simulations, and we are currently working on the bulk optical response of these mixtures in parallel. We have already performed first experimental runs:



Figure VI.11: S-SHG maximal intensity measured experimentally on liquid-gas mixtures of methanol-water. The concentration of methanol in the mixture is expressed in % of weight. The acquisition time is 1sec and the signal was detected using a PM at the second harmonic frequency (404 nm).

the results are presented in Figure VI.11. Here are plotted the maximal SHG intensity for the outcoming P and S polarizations. We observed that adding a small amount of methanol impacts a lot the S-SHG signal.

We would like to study these systems to also understand more the role of the quadrupolar term. Indeed, the methanol and water almost have the same optical index: we expect the exciting field to be the same for all the mixtures. Hence, the modification of the SHG signal can be only due to the evolution of the individual tensor, at the dipolar and quadrupolar order, at the interface. Thus, we hope to be able to propose a link between the molecular structure of the mixture, and their S-SHG intensity for these mixtures.

# VII | Toward Second Harmonic Generation in Confined Liquid

How do you recognize a theoretician in an experimentalist lab?

He says in the morning " Today let's do some real experimental work" and you find him coding an overkill analysis program in the afternoon.

How do you recognize an experimentalist in a theoretical lab?

He says in the morning "Today let's do some real theoretical work" and you find him wondering how reproducible the coffee machine is on the afternoon.

# What is your PhD student doing?



# VII. A | Introduction

Understanding the property of liquids at interfaces is of prime interest. Indeed, many macroscopic properties can be impacted by microscopic behaviors at the microscopic scale. For instance, in lubrication, the slip length highly dependents on the interaction between the first liquid layer and the solid substrate. Depending on the size of the channel, this slip-length can have a dramatic impact: this is particularly true in nanofluidics where some dimensions of the channel are comparable to slip lengths.

In a more general sense, close to an interface, the behavior of the molecules in the "liquid" phase deviates from the one of the bulk. For instance, close to charged surfaces, the first layers are called "Stern layer" as they do not follow the expected Poisson-Boltzmann behavior [37, 115, 116]. Or, at the top of a water ice/air interface it has been shown that there are a remaining small number of "quasi-liquid layers" which are nor solid-like nor bulk-like [102]. When the molecule-solid interaction typical distance is comparable to the one where expends the liquid phase, this effect can become even more important. Especially if the liquid is trapped, the evolution of the behavior can affect the whole liquid phase. For example, "confined"



Figure VII.1: Scheme of the usual SFA experiment. The liquid in confined between two solids: a flat surface and a sphere. The distance between the solid is controlled,  $\Delta z$ . As this distance evolved with a small amount  $\delta z$ (dynamically or not), the force applied to one solid substrate is measured, F.

liquids are in fact quite common outside the labs. In clays where the different layers of solids can be very close to each other. Also in cellular membranes where there are nano-flows in and out the cell, or in foam films.

# Surface Force Apparatus:

One way to study the properties of confined liquids is to use an experiment where this confinement can be continuously controlled. One famous setup is the Surface-Force-Apparatus (SFA), schematized in Figure VII.1. During such experiment, a liquid is confined between two atomically smooth solid surfaces (usually cleaved-mica or floated borosilicate glass) where the inter-solid distance,  $\Delta z$ , is controlled up to the angstrom. In the first experiment of Israelachvili and coworkers, this distance is reduced slowly and the force applied to the bottom solid is recorded. They have shown that the force starts to oscillate when the inter-solid distance starts to be close to the liquid molecular size, see for instance [46, 47]. This technique has grown in popularity [117–120], and has also evolved as the dynamic properties have also been tackled.

To obtain dynamics information, a small oscillating displacement,  $\delta z(t)$ , is done at a given distance  $\Delta z$  and at a given frequency. During the oscillation, the time-dependent force is recorded: the frequency-dependent impedance is thus measured for different inter-solid distance. Such setup is called dynamic-Surface-Force-Apparatus (dSFA). Indeed, the characterization of the rheological property of the confined fluid [48–50, 121, 122] can provide more information than just the force needed to displace the fluid in the static limit. Yet, models have to be built in order to rationalize the observed properties: many questions are still not solved. For instance, recently it has been shown that for small enough confinement, the solid deformation starts to have an important impact on the measured force [123] and should be included in the model.

Hence, the combination of several probing methods on this technique can narrow down the unknown phenomena. Therefore, SFA setup as also been coupled with X-ray measurement [124, 125], fluorescence microscopy [126] or Raman microscopy [127]. In this work, we propose to couple SFA with Second Harmonic Generation measurement.

## Why use SHG?

As we have seen in the previous chapters, the SHG phenomenon is very sensitive to symmetries: at the molecular scale and at the mesoscopic one. The total signal of an isotropic non-resonant liquid is small compared to the one of an interface where the centro-symmetry is broken. Hence, for liquid between two solid surfaces with an organization similar to the bulk, we shall measure using SHG only the response of the two surfaces.

Under a strong confinement, where the measured force in classical SFA oscillates for instance, the structure of the liquid phase is generally considered as layered. This new liquid structure (with new rheological properties) might no longer be centro-symmetric, and thus generates strong SHG signals. Therefore, measuring the SHG response of a liquid in a SFA experiment might shed some light on the understanding of the liquid structure under confinement.

# **Requirements and bed stones:**

We built the experiment from scratch, but with knowledge inside the collaboration regarding the different parts involved: Liquid at Interfaces for the SFA setup and Non-Linear Optics at Interfaces for the SHG part. We have also benefited from the help of the Mechanical, and Interfacing, teams of the lab.

An overview of the setup is presented in Figure VII.2. A



Figure VII.2: Overview of the SFA-SHG experiment.

femtosecond laser produces the fundamental light which is injected in the confinement setup thanks to a microscope. We decided to use a plan-sphere geometry, as Figure VII.1 for the confinement: the fundamental and SHG signals will be applied and collected throughout the plan. The Second Harmonic generated by the system is then conveyed into an optical detection where the intensity is measured. For practical reason we have decided not to measure the forces. We will still call the confinement device an SFA because for the long term runs we hope to have also a force measurement in addition to the second harmonic one. The requirement of both worlds, non-linear optics and extreme confinement, should be met. We will go throughout them in this chapter, but here are the more demanding ones:

- 1. The plane-sphere distance should be known and tunable at the order of nanometers. At this scale, any mechanical vibration can perturb the measurement.
- 2. The acquisition procedure should be reproducible and preserve physical integrity of the different components.
- 3. SHG needs high electromagnetic fields: femtosecond laser shall be used along with a strong focalization.
- 4. To perform SHG and to acquire this signal, the fundamental and the second harmonic should pass through the plan. Moreover, the position of the plan-sphere with respect to the fundamental beam should be tunable.
- 5. The whole system, mechanical and optical, should be stable enough to allow the SHG acquisition to be converged with respect to acquisition time.

To answer points 1 and 5, usual SFA are built with rigid, and heavy materials with large crosssections. Typically, the confinement device can weigh several kilograms. To answer point 3, we use an x60 objective with large magnification and high Numerical Aperture. This usually induces that the microscope has a short working distance, therefore the plane should also be thick. That is why we chose to use a microscope lamella for the plane solid surface. To answer point 4, we can either modify the light injection (using galvanometer mirrors for instance), or displace the confinement setup. We chose to rather displace the confinement because the optical injection and collection for the non-linear optical part seem to be more difficult to modify. But this leads to one of the major challenges of this setup: the nanometric translation piezzo plate which displaces the confinement setup can bear only a limited weight, less than 500 g. Hence, the whole confinement device weight should be less than 500 g. To answer point 2, we have worked on the reflection on the plane and the sphere to measure the distance by interferometry.

# Goals and overlook:

This first setup is thus a home-made ultra-light SFA-like confinement setup couple with SHG measurement. This is a proof of concept: the goal is to probe by SHG a liquid between two solid interfaces, with tunable inter-solid distance. The current state of the experiment does not allow us to perform state-of-the-art measurement as we are still facing some challenges for the confinement and non-linear optics parts. However, we have yet promising results which show that we are in the right direction.

First, we will present today's setup: the different parts, how the confinement works and the optical lines. Second, we will present the results regarding the inter-solid distance control and the SHG measurement. In conclusion, we will present our newest results on confined systems probed by SHG.

# VII. B | Setup description

In this section, we will first describe the different parts of the setup. Then, we will present the confinement procedure using interferometry. Finally, we will present the SHG line and some calibrations we have made.

# VII. B.1 General presentation

Figure VII.3 presents the different parts of the setup in detail. First, let us present how the SHG measurement of a confined liquid proceeds.

# The SFA-SHG in a nutshell:

The "confinement device" is made of the sphere and a lamella which are approached up the micron meter, or lower. To measure this distance, an interferometric system is used. The fundamental beam (here at 1048 nm, in red in Figure VII.3) goes through an optical cage, system to guide it toward the confinement setup, and the intensity of the electromagnetic field is further increased using an objective. At the solid-liquid interface or within the liquid phase, second harmonic is generated (here green light at 524 nm in Figure VII.3 b)). This scarce second harmonic light is collected in the backward direction using the same objective. Then, the second harmonic signal is separated from the fundamental one using a dichroic mirror in the optical cage. Finally, a spectrometer and a Photo Multiplier Tube (PMT), acquires the second harmonic signal generated by the sample.

Now, let us go through the different parts of this experiment, starting with the confinement part. The confinement setup is made of two parts: the cell, presented in Figure VII.4, and the tower, presented in Figure VII.3 b). The cell contains the liquid we want to confine and the tower performs this confinement.



Figure VII.3: a) Scheme of the optical paths. On right part of the Figure, there is the fundamental laser, and the Helium-Neon laser (HeNe), which comes from the bottom of the optical cage and go through the dichroic mirror. The second harmonic signal, the light from the LED and from the white source are completely reflected by the dichroic mirror. On the left part of the Figure, continuous green lines correspond to light paths used for distance measurement and alignment whereas the dotted green line is the SHG signal path toward the detection. b) Scheme for the confinement setup. Left: from the optical table to the confinement device. Right: The confinement device. For both parts, a side view is plotted. Note that the macro screw is attached to the tower while the micro piezzo can move with respect to the tower since it is embedded inside the circular spring. The cell is represented by the lamella.



Figure VII.5: Sphere melting: Agnès Piednoir moving the blowtorch to melt the top of the glass steam. The steam is rotating between 100 RPM to 500 RPM.

# The cell and the sphere:

The cell is composed of several pieces that can be secure to the XYZ translation stage, see Figure VII.4 and VII.3 b). This cell can be filled with liquid from above once placed inside the confinement setup. Regarding our requirements, the most important part of the cell is the lamella: the liquid is confined between the lamella and the sphere while the lights (fundamental and second harmonic signal) go through the lamella.

The lamella and the sphere are made in Boro-Silicate (Pyrex). This material has a very small absorption in the green wavelength where the second harmonic is generated and chemically inert. Moreover, it is quite easy to obtain a roughness (pick-to-valley high) less than a parameter with such material. The lamella is



Figure VII.4: Side and top view of the cell. Three screws are used to maintain the cell together by compression. Two extra screws not presented here are used to secure the cell to the XYZ translation stage.

a nanometer with such material. The lamella is a classical microscopic slide of width 170  $\mu$ m.

The spheres are made at the ILM by Agnès Piednoir, by melting a rotating glass steam using a blowtorch, see Figure VII.5: a spherical apex is created with a typical roughness of 0.2 nm. The heating and rotating parameters were carefully chosen to ensure that the sphere apex is aligned with the steam and to avoid air bubbles during the formation.

# The confinement device:

The tower structure (see right part of Figure VII.3 b) is made of a first 3D-print plastic piece. In the tower is placed a module composed of an adaptative piece, the micro piezzo and the sphere. This module is not directly in contact within the tower: a circular spring makes the junction between the tower and the module. The top of the module can be pushed in the z-direction against the spring.



Figure VII.6: Displacement control in the z direction. The macro screw can achieve displacements of the scale of a millimeter with micrometric precision. The micro piezzo can perform movement smaller but very precise one: up to  $80 \ \mu m$  with a nanometric precision.

To achieve this, a macro screw is placed at the very top of the tower, see Figure VII.6. This macro screw is in contact with the module and fasten within the tower. Hence, when we screw the macro screw in the -z direction, it lowers the module and thus the sphere. The spring pushes the module in the +z direction when the screw is screwed back. The module is embedded in the spring within the XY direction: the module and the tower should not move with respect to one other in the XY direction. This is quite efficient when the spring is compressed.

## From the optical table to the cell:

The whole experiment is developed on to an active optical table which should reduce the mechanical noise from the ground. The confinement device starts with four legs set to the optical table. The leg purpose is to set the SFA device in top of the optical cage, but also to reduce the low-frequency mechanical noise. For that, these legs are filled with sand. The junction between the legs and the adaptative plate is made by a rubber joint, reducing the mechanical high frequency noise.

The SFA setup (tower and the cell) is mounted on a motorized XYZ translation stage. We use this translation stage to control the position of the confinement device with respect to the objective – and thus with respect to the fundamental laser. This allows numerically controlled displacement on the range of the hundreds of  $\mu$ m.

This XYZ translation stage as a nanometer resolution and quite a large displacement range (200  $\mu$ m in the XYZ directions). However, the maximum weight supported by the XYZ translation stage is very limited. Therefore, the **whole** tower+cell should weigh less than 500 g, while being rigid enough.

Hence, a lot of mechanical development has been made to find a suitable device. The actual confinement device is made of two independent parts: the cell and the tower attached to the XYZ translation stage.

# **Optical focus:** *XYZ* **control**

The displacement in all the directions of the confinement device (tower+cell) is made by the XYZ translation stage as presented in the Figure VII.7. If the XY alignment between the confinement



Figure VII.7: Left: Displacement control of the whole confinement setup using the XYZ translation stage. Right: Manual displacement of the microscope for the focus.

setup and the objective requires large displacement (>  $200\mu$ m), we can use the manual XY plate. Before any measurement, a manual focus of the objective is roughly made, as presented in the right part of the Figure VII.7, before adjusting it with the Z-axis translation.

# Interferometry line:

To measure the distance between the sphere and the lamella, we take advantage of the light reflection on both interface (Pyrex/air). These two interfaces act as an interferometric system: the phase of the reflected light depends on the inter-solid distance  $\Delta z$ , and on the optical index of the liquid between them. Hence, in Figure VII.3 a) is presented on the left part the interferometric setup, which coexists with the SHG one. Two light sources are used (the LED and a white source) which and are injected in the objective using the dichroic mirror in reflection. The collected light, with interferences, follows the same path in the opposite direction until a 50:50 beam splitter. To measure this signal, we use a spectrum analyzer (for the spectral interferences) and a monochromatic camera (for the spatial interferences).

# SHG line:

Finally, the SHG line is also presented in Figure VII.3 a). The source is a fiber-femtosecond laser (HighQ-2-IR from Spectra Physics) at 1048 nm, its output power is about 1.5 W. According to the constructor, the repetition rate is 63 MHz and the pulse duration is between 150 and 250 fs. To control the power, a half plate ( $\lambda/2$ ) and a polarizer cube are used after the laser source. The fundamental intensity just after this power control is plotted in Figure IX.21 as a function of the polarizer angle.

The fundamental beam is too small compared to the aperture of the objective. Hence, we use a telescope made of two lenses to match the size of the objective. At the focalization spot of this telescope, a pinhole of 50  $\mu$ m is added to filter the potential non-Gaussian mode of the fundamental beam.

The SHG line is then injected in the optical cage. Before going through the dichroic mirror, the

light is filtered using a low-frequency filter (red filter in Figure VII.3 a)). This is done in order to avoid the detection of any second harmonic signal generated by previous part of the line. Then, the dichroic mirror lets through the fundamental beam, and reflects the second harmonic signal (see Figure IX.25 in appendix).

The second harmonic is conducted to the detection made of a spectrometer linked to a Photo Multiplier Tube (PMT). In short, the light is focalized at the entrance of the spectrometer which spreads the light according to the wavelength. The PMT is basically a big pixel: it returns the intensity collected across it whole surface regardless of the wavelength. The PMT efficiency is wavelength dependent, see Figure IX.28 and is especially sensitive at the second harmonic frequency with a quantum efficiency about 20 %. We can either acquire the intensity collected by the PMT with time within a certain spectral range (about few nanometers) or scan around the second harmonic wavelength to acquire spectrum.

# VII. B.2 Confinement procedure

Now, let us go in more details concerning the experimental procedure developed to confine the liquid between the sphere and the lamella.

	Objective x10		Objective >	<b>k60</b>
Eye	Reflection	Spatial Interferometry	Spectral Interferomet	ry
mm	100	μm	10 µm	confinement

Figure VII.8: Confinement procedure. We start by naked eyes to confine up to several hundred of  $\mu m$ . Then we use the reflection of the light from the sphere and the lamella to approach up to approximately 200  $\mu m$ . Then, spatial interferences allow us to center with precision the sphere with respect to the microscope objective. Finally, we use spectral interferences to reach confinement distances about dozens of  $\mu m$ .

# **Requirements and goals:**

During the whole process, the lamella and the sphere shall never touch each other, otherwise the sphere can be damaged and the confinement being impacted. Indeed, at the nanometer scale, a stripe or a scratch on the lamella or the sphere can modify a lot the confinement.



Figure VII.9:

We have aligned the objective and the optical line along the laboratory Z-axis and the lamella along the horizontal XYplane. This alignment has been done, and checked from time to time, but does not require attention during a particular experiment.

On the contrary, between each measurement the sphere and <sup>a</sup>

the cell have to be cleaned. Hence, the position of the sphere should be adjusted before each experiment. The choice of a sphere ensures that the surface close to the apex (the area where the distance between the sphere and the lamella is minimal) is horizontal with respect to the lamella, see Figure. VII.9.

Therefore, before any new measurement, we have to:

- Align the fundamental beam and the sphere apex along XY.
- Control and measure the lamella-sphere minimal distance  $\Delta z$ .

### **Procedure:** rough alignment

First, the sphere is approached by naked eyes using the macro screw, up to less than a millimeter. Then we use optical feedback to approach closer using the light coming from the reflection of the lamella and the sphere. At this point, we know that the sphere and the lamella distance  $\Delta z$  is

is tilted with respect to the lamella. Right: The sphere is perfectly aligned. In both cases, the localization on the sphere that minimizes the lamellasphere distance, called the apex, is close to a XY plane (horizontal dashed line).

Left:

The sphere



Figure VII.10: Illustration of the sphere XY centering procedure a-c): Alignment of the sphere using its image on the camera with a white light source. In a) the sphere is the upper dot and the lamella the center one, in b) the sphere is getting closer and in c) they cannot be distinguished. d-e) Alignment of the sphere using spatial interferences. In d) the sphere apex is not centered, here on the left, and centered in e).

about 200  $\mu$ m.

We now have to align the apex with respect to the focalization point using the reflection on the camera. The left part of Figure VII.7 emphasis this XY movement of the translational plate while the top part of Figure VII.10 shows the evolution of the reflection pattern during the XY procedure on the camera. We can distinguish two spots if the apex is not aligned with the objective, Figure VII.10 a) and b), one spot being the reflection on the lamella, the other the reflection on the apex. When moving the XYZ translational plate in the XY direction, the reflection of the lamella does not change, the optical axis and the lamella being perpendicular. However, the apex reflection on the camera moves: when the two spots are indistinguishable, it means that the apex is roughly aligned in the optical axis.

# Limits of the reflection technic:

At this point,  $\Delta z$  is about 200  $\mu$ m, and the apex is close to the optical axis. We may approach the sphere closer using the reflection technic but it presents three problems.

First, the precision depends on how strongly the microscope focalizes the light in the z-direction – *i.e.* the waist along the propagation direction. Typically, the order of magnitude of this waste in the Z direction is tens of  $\mu$ m for the x10 objective.

Second, the focalized light is affected by the lamella z position. Indeed, the light first go through the air (or the microscope oil), then through the lamella, through the air and finally is reflected by

the sphere. In the simplest case of the microscope oil and the x60, there should be only one optical gap: when passing from the upper interface of the lamella to the air. The optical gap is about 1.5 to 1: this will move the focalization point. When moving the lamella by 50  $\mu$ m, the focalization point gets closer to the lamella by about 30  $\mu$ m instead of 50 $\mu$ m. Thus, measuring the position of the sphere using the reflection of the light requires to know perfectly the position of the focalization point which is not straightforward. However, it gives us a good order of magnitude.

Finally, looking only at the reflected intensity can lead to miss the apex. Indeed, we performed a rough xy alignment of the sphere using the image on the camera but this can lead to a misalignment of tens of  $\mu$ m. To get a better apex position, we use the spatial interferences.

# Centering the sphere: spatial interferences

When the  $\Delta z$  distance is about hundreds of  $\mu m$ , the coherence length of the HeNe allows the observation of spatial interferences between the sphere and the lamella, as seen in the bottom of Figure VII.10. These rings are due to the fact that the apex is curved: the distance between the lamella and the sphere depends on the XY position, see Figure VII.11. Therefore, we can perform a very good alignment in the xy-direction taking advantage of these interferences: the center of the rings indicates the apex! For closer  $\Delta z$  we use the LED source with smaller coherence length.

It should be noted that such images can only be obtained using the x10 objective. The x60 focuses too strongly: we can observe maybe one interfrange at maximum. Moreover, since the focalization is extremely strong in the x60, it is harder to collect the signal reflected by both the lamella and the sphere: it requires very short  $\Delta z$ . Hence, for the beginning of the procedure, we use the X10 objective.

When moving  $\Delta z$ , we can observe the inter-

ferences shifting, but we cannot measure an absolute value for the  $\Delta z$  distance. To do so, we use frequential interferometry.





Figure VII.11: Scheme of the spatial interferences using the HeNe or LED source. Top: lamella-sphere confinement where the green light from the LED is shined. Several points are selected depending on their distance along z with respect to the lamella. Bottom: scheme of the obtained image on the camera. All the points are on a constructive distance from the lamella, leading to high intensity. Each ring corresponds to an increase of  $\delta z = 2d/\lambda$ . Experimentally we observed this kind of interference on the camera, see Figure VII.10.

## The last microns: spectral interferences

Once the apex of the sphere and the objective are centered in the XY position using the HeNe or the LED, we switch to the white source. We record the spectrum of the light reflected by both the sphere and the lamella. When the  $\Delta z$ is close enough <sup>1</sup>, we start to observe oscillations in the spectrum due to interferences, see Figure VII.12.

From these interferences, we can measure  $\Delta z$  with precision about hundreds of nm, see Section VII. C.2. At this point, we can now safely approach the sphere to any given distance above the micron: we are measuring the apexlamella distance in real time. For instance, we



Figure VII.12: Spectrum of the reflected light by the lamella and the sphere using the white light source, x10 objective. The detection is made using the spectrum analyzer and the measured distance is 9360 nm, see Section VII. C.2.

set  $\Delta z$  to 50  $\mu$ m, and switch the objective: we go for the x60. Indeed, at the end of the day we need to use the x60 for the SHG acquisition. Note that since the change in the objective may create vibration, it is not safe to not set  $\Delta z$  too small. Once with the x60, we confirm the  $\Delta z$ , and can make the final approach up too few  $\mu$ m.

#### Sum up:

Now that we have presented the whole procedure, we can better understand why we do not fill the cell with liquid in the first place. All of these technic are based on the reflection of the light at the sphere and lamella interface. Therefore, the more reflection between the glass and the phase between the sphere and the lamella is, the better contrast we will have. Therefore, here is the typical procedure regarding the confinement:

- 1. Clean the surfaces, built the cell. Add the cell on the confinement device, and then finish building the tower.
- 2. Use the X10 objective
- 3. By naked eyes: approach the sphere up too few millimeters
- 4. Using the reflection on the sphere (HeNe or LED) approach to 200 microns.
- 5. Center the sphere and the objective using the reflection on the camera.
- 6. Use the spatial interferometry (HeNe or LED) to center the apex and the objective.
- 7. Use the frequency interferometry (white source) to approach around tens of microns.

<sup>&</sup>lt;sup>1</sup>About tens of microns: if  $\Delta z$  is too large, the interfringe is smaller than the spectrum minimal wavelength precision.

- 8. Switch to the X60 objective
- 9. Use the frequency interferometry (white source) to approach at the desired  $\Delta z$ .
- 10. Add liquid
- 11. Confirm the  $\Delta z$  by frequency interferometry (white source).
- 12. Starts SHG

# VII. B.3 SHG microscopy

In this part, we present the SHG line and some control measurements to confirm the Second Harmonic behavior.

### First measurement: quartz

We first studied the Second Harmonic generated by a piece of quartz, which is well known in the literature. To do so, we remove the cell and place this solid slab (about few centimeters thick) on the XYZ-plate while the x10 objective is used. We focalize the objective on the lower interface. The spectra obtained near the second harmonic wavelength is presented in Figure VII.13. The ratio signal over noise is excellent because the quartz is known to generate very strong second harmonic: we recover a Gaussian shape for the intensity with respect to the wavelength as expected.



Figure VII.13: Spectrum measured by the detection line in the backward direction of at the air-quartz interface using the x10 objective and the femtosecond laser.

We have also checked the SHG intensity generated by the quartz sample with respect to the fundamental fre-

quency with the x10 objective. A quadratic evolution of the second harmonic is observed as expected. The control of the fundamental power is presented in Appendix IX. E.1.

To conclude this test system, I have measured the collected intensity at the second harmonic wavelength depending on the position of the sample, we call it "z-scan". Using the XYZ translational stage, we move the sample from 0 to +200  $\mu$ m while the detection record the intensity at a given wavelength. Figure VII.14 shows the evolution of the SHG signal for an acquisition time of 5 sec. For small z, the light is focalized inside the quartz. At approximately 90  $\mu$ m the light is focalized at the interface: the generated intensity is maximum. For large z, the light is focalized before the quartz, on the air.

The smooth shape of the intensity depending on the beam position can be explained by the extension of the laser focalization along the z-direction. We are using a x10 objective which does not focalize the light very strongly: the Rayleigh length can be estimated to be about tens of microns, in agreement with the laser beam characterization, from Figure VII.14.

Using this system, we were able to confirm that we can record the SHG signal and perform meaningful z-scan. Then, we measure the second harmonic signal coming from the lamella-air interface using the x60 objective.

We should emphasize that the response of the oil, when placed between the objective and the lamella as it should be, has never been detected in our setup: if the objective is not focalized at the second lamella interface, we see no signal near the second harmonic wavelength. Hence, we will study the upper lamella interface with the air: air-lamella. The typical power used is about



Figure VII.14: Second harmonic intensity as a function of the quartz z-position. For small z the laser is focalized inside the quartz, for large z the laser is focalized in the air before the quartz.

100 mW together with the x60 objective for these air-glass systems.

#### Effect of the detection slit: Air-lamella interface

Here we would like to see the effect of the slits of the SHG detection. The detection is made of a spectrometer linked to a PM. At the entrance and at the end of the spectrometer, there are 2 slits. The spectrometer is made so that these slits are optically conjugate – hence the aperture of the two slits should always be the same. That is why before the spectrometer there is a lens to focalize the light on the first entrance slit, see Figure VII.3 a).

The first slit main purpose is to reduce the optical noise coming from the room. When closing this slit, the light coming into the spectrometer with not the proper angle is more likely to be lost within the spectrometer – and hence not detected by the PM. The second slit purpose is to get more restrictive with respect to the wavelength measured. When closed, the second slit let come through less light in the horizontal direction: this axis corresponds to wavelengths thanks to the grating.

Closing the slits lead to less noise, and to a better wavelength resolution. However, it also decreases the signal detected, and can make the SHG detection more difficult. To study the effect of the slit, we have recorded for a system composed of the x60 objective, the adaptative oil and the lamella the SHG signal with different slit aperture. The objective has been focalized to the air-lamella interface for a fundamental power about 100 mW before the objective.

In Figure VII.15 is presented the spectrum obtained depending on the slits aperture. For closed slits (0.25 for instance), the intensity collected is small because few light can enter the spectrometer,



Figure VII.15: Spectrum obtained for the air-lamella interface using the x60 objective. Different slit apertures are tested for the detection part. The entrance and detection slit are always at the same size. The vertical dashed line is a guide to the eye and represents the central second harmonic wavelength.

and the spectrum looks like a Gaussian. As the slits are opening, the intensity increases, and the spectrum becomes wider and wider. For the largest aperture, we can clearly see that the spectrum is less defined in wavelengths: there is a convolution of the expected Gaussian shape with a larger door.

What slit aperture should we use? On one hand, a full aperture is useful experimentally speaking: the acquisition time can be decreases once the noise of the room controlled. On the other hand, this broadening of the spectrum can be an obstacle if the system presents fluorescence. Indeed, in this case, there are "noises" inside the spectrum which came from the sample itself: they cannot be avoided by optical shielding and are participating to the intensity collected in the whole second harmonic wavelength region. To remove them, we shall measure the spectrum, fit the fluorescence background and remove it to isolate the second harmonic signal, see Section V. A.1. But, if the wavelengths are mixed due to the aperture of the exit slit, this disentanglement of the second harmonic to the fluorescence background can be much more complicated.

Hence, for systems with no background around the second harmonic frequency, we can use open slits. For systems where the extraction of the second harmonic intensity requires more work from the spectrum, we should use close slits. Note that this problem will be soon solved by the use of a camera: the spectrum will be acquired with the resolution of the camera pixel directly. Hence, we would have to deal with just one slit: the entrance one. This slit would be fully open for SHG acquisition to increase the intensity collected, without any loss in the spectrum resolution.
# VII. C $\mid$ Results

In this section we present the results achieved on our SFA-SHG setup. First, we present the SHG intensity obtained for non-resonant system without confinement: at glass-air and glass-liquid interfaces. Then, the control of the lamella-sphere distance,  $\Delta z$ , using the spectral interferometry is discussed. Finally, as a preliminary result, I will present our first measurement of second harmonic signal when the lamella-sphere distance is about dozens of microns.

# VII. C.1 SHG at the lamella interface

Before performing SHG on confined system, we should understand the response of the material composing the cell. Indeed, the cell response will always be part of the total signal, whatever the liquid composition. Hence, the first system to study is composed of the x60 objective, the adaptative oil and the lamella.

# Air-Lamella: Power law

First of all, we check that we are indeed measuring the second harmonic process. In Figure VII.16 is plotted the spectra obtained for different fundamental power for 30 sec of acquisition and the slit almost closed (0.5). For each spectrum, the noise has been removed using the home-made Alpaga code, see Chapter V. The full line is the Gaussian fit obtained with fixed central wavelength and the Gaussian width for all powers <sup>2</sup>. On the right part of Figure VII.16 is plotted the obtained SHG intensity with respect to the fundamental power in logarithm scale. If we do have a non-linear behavior of the SHG intensity with respect to the fundamental power, the power law seems to be not the quadratic expected one.

Here are two possible reasons:

- The objective itself. At 1048 nm, the objective absorbs about half of the fundamental, see Figure IX.27. For large intensity required to detect the SHG (around 100 mW), we may induce non-linear absorption effect such as saturation. This will be checked by measuring the outcoming power from the objective with such large fundamental power.
- Self-focusing Kerr-effect. With high enough electromagnetic field, the optical index of a material can increase or decrease. In the case of an increase, this would lead to a self-focusing effect: an already focusing light is getting more and more compressed as it goes through the material due to the increasing change of optical index. Such effect may happen either in the objective, the adaptative oil or the lamella.

On quartz samples with the X10 objective, we have not faced this issue: we will try to measure the SHG of the quartz with the X60 objective with and without oil at the same fundamental power. While the nature of the signal detected is questionable, we have still acquired the response in the

 $<sup>^2\</sup>mathrm{This}$  was done to avoid compensation of small power where the SHG intensity is small.



Figure VII.16: Second harmonic intensity generated by the air-lamella interface using the x60 objective. Left: The dots are the spectra obtained for several fundamental power in mW once the noise is removed. The fit in full lines are made using the Alpaga code with fixed central wavelengths and width for all spectra. The fundamental power has been measured just after the power control, see Figure VII.3 a). Right: Logarithm plot of the obtained SHG intensity with respect to the fundamental power. The dashed lines are a guide for the eye: they show an ideal quadratic power law second harmonic intensity with respect to the fundamental power.

SHG region of several positions of the beam with respect to the air-lamella interface and when adding water.



Figure VII.17: Normalized intensity collected with respect to the position of the XYZ plate along the Z axis for the air-lamella system. For small Z, the objective focalized on the air, for large Z inside the lamella. In red is plotted the collected intensity when illuminating with the Helium Neon laser (HeNe). In green the intensity at the central second harmonic frequency once the noise removed, see Figure VII.16, using the fundamental laser. The objective x60 is used at 350 mW.

# Lamella-air: z-scan

First, we have measured the intensity reflected by the system with the HeNe laser beam. The aim of this first characterization is to find the position of the lamella-air interface. The intensity collected is plotted in red in Figure VII.17 as a function of the z-position of the system. Note that such acquisition is extremely fast compared to a SHG one: 100 ms per altitude for the HeNe compared to 30 s for the SHG. The intensity at the HeNe frequency presents one main peak, at  $Z \approx 150 \ \mu m$ , and two shoulders at approximately 80  $\mu m$  and 180  $\mu m$ . These shoulders are reproducible but not explained yet. The position of the main peak corresponds to lamella-air interface <sup>3</sup>.

Figure VII.17 also presents the second harmonic intensity in green respect to the Z position of the sample generated using the femtosecond laser. Note that the noise has been removed: that is why for low z (when the fundamental beam is focalized above the interface in the air) the green curve of Figure VII.17 goes to zero. It should be noted that the amount of noise near the second harmonic frequency is constant for all z – data not shown.

The SHG intensity is null or not detected when the laser is focalized in the air (small z) as expected. When the laser is focalized in the lamella, we still get a remaining signal: this is the plateau after  $Z = 150 \ \mu\text{m}$ . This signal can be explained by a volume contribution of the amorphous glass composing the lamella (SHS). Note that both curves of Figure VII.17 has been normalized with respect to their maximal value to easier their comparison.

Similarly to the HeNe profile, the SHG intensity shows a peak corresponding to the air-lamella interface. However, this peak is shifted in altitude: this can be due to the frequency dispersion of the objective, oil and lamella. Surprisingly, the SHG signal starts to rise approximately at the same altitude then the HeNe first shoulder (without explanation for the moment).

# Lamella-liquid: z-scan

When adding water, the evolution of the SHG signal with respect to the focalization changes dramatically. The SHG intensity of the lamella-water system is plotted along with the one obtained on the lamella-air system in Figure VII.18. In that case, as for the lamella-air system, the noise measured around the central SH frequency is constant throughout the Z plate position. We conclude that the SHG signal of bulk water is not measured: if it were the case we should measure a larger noise for small Z than for high Z.

The asymptotic value of  $I_{SHG}$  are equal to the lamella-air configuration: for small Z no SHG is detected, and for large Z there is a plateau. This confirms that at large Z, the second harmonic signal comes from the bulk lamella response. Moreover, both curves start to increase at the same altitude (around  $Z = 70 \ \mu m$ ), but the water-lamella system does not present any peak with respect to the sample altitude. This confirms results obtained on a similar setup by Lucile Sanchez, see here PhD [29]. One explanation can be that the second harmonic generated by the lamella-water surface is too weak compared to the one generated by the bulk lamella.

<sup>&</sup>lt;sup>3</sup>We have cross-checked using the image on the camera.



Figure VII.18: SHG intensity collected with respect to the position of the XYZ plate along the Z axis. For the lamella-air in green and for the lamella-water in blue. The same acquisition time (30s) and the same fundamental power is used along with the objective x60. For both curves, the noise has been removed.

This can be due to several reasons. First, the surface susceptibility tensor of the solid interface is impacted by the fluid. For instance, adding water with low pH will protonate the borosilicate interface, and thus reducing the surface charge and decrease the second harmonic signal emitted [111]. Secondly, the optical indices gap has an important role for the electromagnetic field felt at the interface, and for the SHG collected. For instance, if the optical index of the solid and the liquid matches, with a similar composition for the liquid, the SHG signal decreases [29]. Therefore, a reduction of the SHG signal when adding water is not really surprising and confirms previous observations, even if it is not well understood yet.

#### **Conclusions:**

We have shown that we were able to collect the SHG intensity of both systems, lamella-air and lamella-liquid. It seems that adding neat water makes the second harmonic generation of the interface too small to be detected by our setup. However, a deep understanding of the Z-scan profile and the evolution of the SHG intensity with respect to the fundamental power are not reached yet. In the future we will perform complementary studies with other objectives to go further.

Moreover, we have bought a new detection setup made of a new spectrometer and a cooled CCD camera to get the spectrum near the second harmonic wavelength directly. It will allow us to get sure that we study only the second harmonic signal by removing any background. To finish, we plan to use electrolyte solutions at different pH to confirm that our setup can compare the second

harmonic intensity collected for different liquids– the evolution of the SHG intensity with respect to the pH is well known the literature [35, 111]

# VII. C.2 $\Delta z$ measurement and mechanical stability

In this section we present how we have measured the lamella-sphere distance using spectral interferometry. To do so, we use a white light source and collect the spectra reflected by the lamella and the sphere, see Figure VII.3. The typical spectra obtained present oscillations (Figure VII.12, IX.23 or VII.19). Using a classical interferometric approach, see Appendix IX. E.2, we can obtain the distances between the two solid interfaces by analyzing the spectra.

To obtain the inter-solid distance  $\Delta z$  from the measured spectra, I have coded a Python analysis named Guanaco. This procedure has been inspired by the code of Joachim Trosseille who did a Post-Doc at the ILM. It runs in continuous way to analyze the last spectrum directly acquired by the spectrometer. This way, we have an almost instantaneous feedback on the  $\Delta z$ . We intend to share it in open access once a small documentation written.

We have checked our setup using a  $SiO_2$  wafer, see Appendix IX. E.3, and find that we were able to measure inter-solid distance with an error of about 10 nm. Now, let us focus on the system of interest: the lamella-sphere probed by the x60 objective.

### XY spatial averaging in the field of view:

One important difference between the flat wafer and the sphere-lamella system is the spatial evolution of the  $\Delta z$  along the XY direction due to the curvature of the apex, see Figure VII.11. Hence, if the light illuminates a too wide area in the XY direction we expect to have less and less contrast as a spatial averaging of the interferences occur and that the measured distance increases.

To reduce this issue, we have two irises: one just before the objective in the optical cage, and one before the lens of the interferometry line, see Figure VII.3 a). When closing both irises, we expect the XY averaging to be reduced. In Figure VII.19 is plotted the interference spectrum obtained when the irises are open or closed. We observe that the contrast is better when the iris



Figure VII.19: Normalized spectrum when the irises are open or close using the x10 objective. The extrema positions are at different wavelengths: the distances extracted from these interference spectra using Equation IX.119 are different. For opened iris: 6900 nm, for closed iris: 6780 nm.

is open, which is counterintuitive. However, the measured distance is indeed larger: 6900 nm for open iris while close ones lead to 6780 nm.

# $\Delta z$ drift over time

To test the  $\Delta z$  control, we have performed a controlled approach of the sphere using piezzo steps (1 step = 100 nm) while measuring the  $\Delta z$  using spectral interferences. The results are presented in FigureVII.20 and shows very good correspon-We indeed retrieve steps about dence. 100 nm as requested. Note that the valwhich deviates largely from the mean ues curve are due to the fact that the spectra are averaged over 0.5 sec: these points correspond to a change in the distance while the spectra are collected. However, we observe a deviation of the expected displacement at the end of the approach (the dark dots are above the dashed blue line in Figure VII.20.



Figure VII.20: Measure of the sphere-lamella distance while the sphere is approached by step of 100 nm every 10 sec. The red dashed line corresponds to the initial distance, and the blue dashed line the expected final distance. The green dashed line should correspond to the  $10^{th}$  step. The dark dots are the measured distance in time.

To observe the origin of this shift, we have measured the  $\Delta z$  distance during more than 2 hours, using the same interferometric measurement with the x10 objective, and for a given position of the piezzo. The results reported in Figure VII.21 also shows a continuous increase, about 0.15 nm per second. Note that this drift rate is not reproducible.

Today we have no clear explanation for this non-reproducible drift but we have three hypotheses:

- The piezzo is shifting when we do not explicitly communicate with it.
- The tower (made in polymers by 3D impression) is relaxing through time when the spring is compressed. This can lead the platform containing the piezzo and the sphere to go up.
- The different parts of the system are compressing/dilating due to thermal drift. Indeed during these experiments the airconditioning was not working.

We will test these hypotheses in the future.



Figure VII.21:  $\Delta z$  measured over time using the x10 objective for a given position imposed by the piezzo. The observed drift is about 0.15 nm per second for this situation.

# Mechanical noise

Finally, we quantify the stability of our setup with respect to the surrounding mechanical vibration. Indeed, the mechanical vibration can propagate thought the confinement device, and affect the  $\Delta z$  by creating small oscillations. We can note mechanical noise sources at different scales.

On the optical table, electric devices such as the laser or the spectrometer can be some mechanical noise sources. However, we cannot put them off the optical table because it would be too difficult for the optical alignment and stability. For the long term, we can put them on anti-vibrational legs, as for the SFA device.

Within the room, the air-conditioning, the computer, or all the electric power sources, can also produce induce some vibrations. To prevent the air flow, the whole confined devices, as most of



Figure VII.22: Measured  $\Delta z$  of the lamella-air-sphere system with x10 objective. The optical table was activated or not. During the measurement, the operator has made his best to create vibration on the floor near the table.

the setup, is under an optical shielding – also required for the SHG detection. Recently, the optical table has been put on air-compressed activable legs. This active mechanical noise reduction of the optical table legs can be clearly seen in Figure VII.22.

#### Conclusions and Perspectives for the $\Delta z$ control:

We have shown that we can measure the  $\Delta z$  distance with an accuracy of tens of nanometers. Yet, several questions are still not answered:

- How to prevent the time drift of  $\Delta z$ ?
- Does the  $\Delta z$  remains the same with liquid?
- Are there more vibrations or drifts when using the x60 objective?

Moreover, as evoked previously, we can measure distances in the order of the micrometer using this interferometric measurement. Hence, once all these problems solved for the stability of the order of the micrometer, we shall also try to measure the fluctuations and  $\Delta z$  for smaller distances.

# VII. C.3 SHG of confined-like system

Finally, we present the SHG z-profile of the sphere-air-lamella.

The intensity at the second harmonic wavelength depending on the z position of the XYZ plate is presented in Figure VII.23. We have not recorded the initial position of the sphere, but it is less than 50  $\mu$ m from the lamella. The first profile in blue, when the sphere is far from the lamella, shows two peaks: one very sharp at 90  $\mu$ m (full dark vertical line) and one large at 170  $\mu$ m. The first peaks correspond to the sphere-air interface, and the second to the air-lamella interface.

The second curve has been obtained after moving the sphere by  $+20 \ \mu m$  toward the lamella. The air-lamella peak is retrieved (at large Z), but not the one corresponding to the sphere-air one. We may detect a shoulder at 130  $\mu$ m which may be the sphere-air trace on the SHG signal. But the position is not the one we have moved the sphere by expected:  $+20 \ \mu m$ , which in a first approximation would correspond to a change in  $20 \times 1.5/1$  $30 \mu m^{4}$ .

Hence, we indeed manage to observe the sphereair-lamella for small intersolid distance in SHG, but not with a good understanding of what is



Figure VII.23: Z-scan at the Second Harmonic wavelength for a system composed of a lamella and the sphere separated by air for two positions of the sphere. The full vertical line is a guide to the eye that should correspond to the sphere position at the initial stage and the dashed line is the expected position of the SHG intensity peak of the sphere when moved to  $+20 \ \mu m$ .

happening. This experiment has been done only once due to material issues. Yet, its shows that the main goal of the setup is possible: we have a first coupling of an SFA-like confinement setup probed by SHG. Once all the electronic-device functional, we will be able to tackle liquid under confinement by SHG.

<sup>&</sup>lt;sup>4</sup>because the focal point is affected by the lamella-air optical gap of approximately 1.5.

# VII. D | Conclusions

In this chapter I have presented the setup built during my thesis and the first results obtained. The experimental procedure to confine the liquid phase designed for the setup starts to be robust. The confinement is controlled by interferometric measurement, and an automatic software allows us to have an almost instantaneous measurement. We have still some work to do regarding the intersolid distance drift, and to better control the mechanical vibration.

The SHG line is functional, and we have found qualitative results for many reference systems. Yet, we have observed deviations from the expected behavior that we attribute to the objective used or a too high fundamental power. We have access to many other objectives, and a more detailed study is on the way. Today, we are also limited by the PMT used for the detection as the SHG intensity is impacted by other optical responses for liquids (fluorescence). We expect to have better results in less acquisition time with the upcoming camera and new spectrometer.

While the "results" part may seem small, I see it already as a success: after more than five years of development and many technical issues, we have proven that this device can perform a controlled confinement of a liquid and probe it by SHG. I would like to emphasize that this project has involved many people with complementary skills, and I thank them again for their help and support.

For the first test of confined liquid probed by SHG, we can think of two kinds of systems:

• Confinement of SHG-dyes.

For instance DiA in methanol. In this case we can expect two behaviors as the liquid get confined. First a decrease of the SHG response because there is less and less dye in the laser waist. Second an increase if the dye get structured for very small inter solid distance.

• Confinement of non-resonant liquids which are easy to get structured.

For this part we are talking about liquids that were already probed by classical SFA. The liquid should be transparent enough not to reabsorb the second harmonic generated, and the inter-solid distance required to see structural modification accessible with our setup. For instance, we can think of polymers, ionic liquids or mematic fluids.

To conclude, this setup is now in the hand of Maxime Fery during his thesis: I have no doubt that he will improve this setup and get astonishing results.

# VIII | Conclusions and Perspectives

You may strengthen your tools and prepare your spirit, Listen to the fools or the one with the highest merit. Learn the map and do the math, be the first of every tourney, But you cannot know the path until the end of the journey. Wise man maxim.

#### General conclusions:

In this manuscript I have presented our molecular-based approach used to describe the Second Harmonic Generation of non-resonant liquids. The optical response is governed by the electronic degrees of freedom which interact with the applied external electromagnetic field. We have seen how to use the perturbative theory along with response tensors to extract from ground-state electronic properties the SHG ones, both in the exact scheme, and in approximated Quantum Chemistry calculation – DFT in our case. Then, we have seen that depending on the geometry of the system, either in volume or at the interface, some ingredients describing the electromagnetic properties shall be added.

Then, we have presented the main numerical tool developed to perform this multi-scale approach: FROG. This software aims at computing the first hyperpolarizabilities,  $\beta$ , of molecules in condensed phase, in the volume or at the interface, from a Molecular Dynamics trajectory. We have also seen how we have brought new insights about the effect of the electrostatic environment on the molecular dipolar first hyperpolarizability using the second order one. For that, we have computed the second dipolar hyperpolarizability,  $\gamma$ , in the bulk phase. We highlight that the liquid phase has a smaller impact on the molecular values of  $\gamma$  compared to  $\beta$  since there are very few fluctuations from one molecule to another. Hence, the electrostatic environment may be described in two parts: the closest neighbors (within the first nanometer) which have to be included explicitly, and the other (after 1 nanometer) which effect on the first hyperpolarizability can be accounted for; using the averaged second hyperpolarizability.

Once the software developed, we used it to describe the Second Harmonic Scattering of bulk liquid water. This experiment is growing in interest in the community, and we have studied two contributions involved in SHS signal: the incoherent contribution and the short-range coherent one. We found that the fluctuations of the molecular first hyperpolarizability,  $\beta$ , have a dramatic impact on the incoherent signal. Indeed, this signal is a measure of the square of first hyperpolarizability in the laboratory frame, the deviations from the zero-value can be more important than some components that are authorized by the molecular symmetry (and thus with a net average in the liquid phase!). We also compute the coherent contribution for molecules within that are closer than 7 Å: correlation at larger distances could not be observed with our current statistical sampling. We found that the contribution of the short-range coherent part to the signal is quite important in amplitude, with the same order of magnitude as the incoherent term. Moreover, we have observed that this term impacts the depolarization ratio, D In this case, D cannot be described solely by simplified formula based on the symmetry of the molecular hyperpolarizability. Considering the two contributions, we have reproduced numerically and quantitatively the experimental results. We attributed the remaining mismatch to the remaining long-range correlations which are not included in our calculation.

The second study was performed on the neat water liquid-gas interface. We first have calculated the dipolar first hyperpolarizability,  $\beta$ , with respect to the position relative to the interface. We have found that this dipolar tensor evolves when approaching the surface and remains closer to the bulk value compared to the gas-phase one. Moreover, we have observed that the Kleinman symmetry was almost respected at 800 nm. Finally, we have demonstrated that the usual approximation consisting in setting the same molecular first hyperpolarizability for each molecule was not valid at this interface for the dipolar one: there are correlations between the molecular  $\beta$ s and the molecule orientation.

We also have calculated the hyperpolarizabilities at the quadrupolar order. The impact of the liquid phase is less important for the quadrupolar hyperpolarizability than the dipolar one. The values are rescaled by a factor about 0.8 and there are only few fluctuations. Moreover, at the interface, the molecular quadrupolar hyperpolarizabilities do not evolve much, as well as the one written in the laboratory frame. At this stage, we have all the numerical and analytical tools to compute the second harmonic intensity generated for these liquid-gas interface, and we will apply them in the future.

Experimentally, I have also worked with several experimental setups: SHS, Surface-SHG and confocal SHG. Few results are presented in this document: I rather focus on the most original contribution regarding the building of confined liquid setup probed by SHG. This project has been impacted by the sanitary crises and suffered from the inherent inertia of building a new original experiment <sup>1</sup>. Yet, the determination, knowledge and competences of this multidisciplinary team, have led over the years to a first prototype which shows very encouraging results.

I achieved a first reproducible confinement procedure to the micrometer. The optical measure of the distance between the two interfaces allows for an instantaneous feedback. First Second Harmonic

<sup>&</sup>lt;sup>1</sup>Wait for new pieces, test them, rebuilt them, adapt one more time the acquisition software, reshape half of the device, built and order new parts, change the spectrometer, fix a broken cable...

signals were collected, proving the feasibility of the experiment. I have shown that measuring the non linear signal at the second harmonic frequency of a confined system is possible.

# Perspectives on Confined liquid:

The next step for the confined liquid experiment would be to have a deeper understanding of the measured z-profiles. Even if we do not have a total understanding, we could try to check the basic idea: if the liquid is structured, the Second Harmonic signal should skyrocket. For that, we will try to confine liquids which are known to be impacted even if the liquid confinement thickness is as large as microns (liquid crystal for instance).

In the long term, and after more studies on our home-made SFA, I think that we should return to a well-established SFA, as the one of Cecile Cottin-Bizonne and Agnès Piednoir at ILM, and modify it so that we can measure the SHG signal. Indeed, I think that the weight limitation may be an impassable problem which prevents us for achieving confinement up to the Angstrom.

## Perspectives for the multiscale approach:

FROG has been developed to be shared and to treat many different systems. In this thesis, I have worked with water, but a study on methanol-water mixture is on the way. Documentation and tutorials will soon be available on Zenodo. Hence, for the short-range term, the most important perspectives would be to promote this code and approach: inside the ILM but also abroad. SHG of liquid is becoming more and more studied, and as the computational power increases, project involving experiments and numerical calculations should be more common.

At the longer term, I would see two main improvements for FROG: the implementation of longrange electrostatic environment and Machine Learning (ML). As we have seen, the electrostatic field of distant neighbors can be either included through the second hyperpolarizability or a constant electric field during the  $\beta$  calculation. But, numerically speaking, this it has limitations: the coulombic interactions need to be included using Periodic Boundary Conditions (PBC). To achieve this, Ewald formalism is more efficiently and more precisely and is often used in classical Mechanical Dynamics or Dynamical Quantum calculation to tackle condensed phases. Practically, we can either load the electrostatic field in FROG from the MD (so that it does not have to be computed again in FROG), or compute it in FROG by coding Ewald approach.

However, we would still have to compute thousands of configurations, and recalculate all the QM/MM computations even for a small change in the structure of the liquid. One way to tackle this problem would be to use Machine Learning. First, it would have to be trained on a system representative of the problem (for instance using QM/MM results of pure water, pure methanol and half-half methanol-water mixture), and then can be applied with almost no numerical cost (for instance for any water-methanol mixture concentrations).

# Perspectives for SHS:

Our numerical results on water for SHS are encouraging and we should extend this methodology to other systems. Indeed, methanol-water mixtures have just been studied by SHS experiments in the lab, and we are also halfway numerically speaking. Many other systems are studied in the lab and by the community: aqueous ionic solutions, ionic liquids, surfactants... These systems could be treated by our molecule-based formalism, provided a more or less important numerical cost. Importantly, our formalism, based on the molecular first hyperpolarizability, can tackle alike small and rigid molecules, such as water, than flexible and longer ones. However, for chemical systems where the definition of a "single molecule" is difficult (for instance ion pair or complexes) may require more development regarding the QM/MM scheme and the molecular-based formalism.

Moreover, we can also study the effect of temperature in our calculation. it could help to disentangle the long-range and short-range correlation effects, as initiated in Section V. C.4. This approach, if working, may be important for the community: we would have shown how to connect the local fluctuations, which have to be described at the QM/MM level, to the "long-range" orientations between molecules inside solvents.

# Perspectives for Surface SHG:

Our modelisation of the Surface Second Harmonic Generation is not complete in this manuscript. However, we have developed a multi-scale approach with combines several tools and approximations at very different levels. Weighting the impact of the different terms takes some time, but we expect to have soon some results for pure water.

On a more general perspective, our study is in line with the work of Foucaud and co-worker [41] and thus question the use of SHG as a tool to probe the molecular structure of aqueous liquid interfaces. Indeed, we have first found that each molecule has its own molecular hyperpolarizability at the dipolar level: its response in the laboratory frame is correlated to its orientation, but not only. Moreover, the quadrupolar terms have an important contribution to the signal do not strongly depend on the orientation of the molecules. Therefore, even if we can observe the evolution of the SHG signal with respect to the composition of the mixture experimentally, extracting from this signal the orientation of the molecule seems very difficult for non-resonant molecules without supplementary information (for example provided by numerical modeling).

However, Surface-SHG can provide the electrostatic environment of aqueous solutions at interfaces. Indeed, the electrostatic environment is almost more important than the molecule orientation and we can take advantage of this extreme sensitivity. Moreover, we have shown that the second hyperpolarizability tensor is almost the same for all molecules, as the quadrupolar first hyperpolarizability. Hence, a macroscopic field (for instance produced by a charged surface) would impact only the dipolar response in a known manner: using the laboratory second hyperpolarizability which in a very good approximation is constant! Surprisingly, I think that at charged interfaces it would be easier to draw a "simple picture" to understand the SHG signal, and to measure important properties such as surface potential, compared to pure water at the liquid-gas interface.

# Last words:

After 3 years, and the difficulties highlighted in this manuscript, I am more than convinced that SHG is a rich process that holds promises in probing the structure of the liquids. The framework is complicated due to the nature of this non-linear process and requires hypotheses to analyze the experimental results. However, thanks to our numerical approach, I think that we have gained knowledge on the SHG processes.

Does this thesis really helps the community to understand SHG of non-resonant liquids? Will FROG be another unknown tool which will be coded in few years by another PhD student? I am afraid I do not know: we will find out in the future. However, I am sure that this kind of multidisciplinary approach can deal with these difficult problems: one calculation and one experience after another.

If I may, I would have a bit of very personal advice for young (or less young?) scientists: do not consider numerical simulations as a tool that **must** be perfect. It is not exact and never will, as the experiment will never bypass any bias or limitations. I, for too much time, considered numerical simulations with "convergence glasses": keeping only the "golden-standard" methods or parameters and disregarding the rest because it was x % off the tabulated results (x may be really, really small...). Of course, comparing the results obtained with different numerical approaches or parameters is important (as the important numbers shown in the several Tables of this document emphasize), as providing the community references calculation with the highest possible accuracy. But, if your goal is to answer a question related to a specific system, you really should focus on whether the method you developed provides any new insight or not, even if the results do not match the tabulated or experimental results. Predicting the impact of a hypothesis is valuable even with slightly off numbers!

# IX | Appendix

# IX. A | Chapter III

IX. A.1 More on the Exact scheme

# IX. A.1.1 Electromagnetic interaction

# Taylor development of the fields: multipolar expansion

We have shown that the external electromagnetic field can be considerate as almost constant within the molecule vicinity. Therefore, to describe the fields and potential we use a Taylor expansion inside the molecule up to the quadratic order:

$$\phi(\mathbf{r},t) = \phi(\mathbf{0},t) + \mathbf{r}_i \frac{\delta\phi}{\delta\mathbf{r}_i}(\mathbf{0},t) + \frac{1}{2}\mathbf{r}_i\mathbf{r}_j \frac{\delta^2\phi}{\delta\mathbf{r}_i\mathbf{r}_j}(\mathbf{0},t)$$
(IX.1)

Where the position  $\mathbf{r} = \mathbf{0}$  is the "reference position" of the molecule. In the right gauge <sup>1</sup> where the field is  $\mathbf{e}_i(\mathbf{0}, t) = -\frac{\delta\phi}{\delta\mathbf{r}_i}(\mathbf{0}, t)$  and  $\frac{\delta e_i}{\delta\mathbf{r}_j}(\mathbf{0}, t) = -\frac{\delta^2\phi}{\delta\mathbf{r}_i\mathbf{r}_j}(\mathbf{0}, t)$  we have that:

$$\phi(\mathbf{r},t) = \phi(\mathbf{0},t) - \mathbf{r}_i \mathbf{e}_i(\mathbf{0},t) - \frac{1}{2} \mathbf{r}_i \mathbf{r}_j \frac{\delta \mathbf{e}_i}{\delta \mathbf{r}_j}(\mathbf{0},t)$$
(IX.2)

In the integral form of the interaction potential given by Equation III.8:

$$V(t) = \langle \Psi | \hat{V}(t) | \Psi \rangle = \int \left[ -\rho(\mathbf{r}, t)\phi(\mathbf{r}, t) - \mathbf{J}(\mathbf{r}, t) \cdot \mathbf{A}(\mathbf{r}, t) \right] d^3\mathbf{r}, \qquad (IX.3)$$

we can replace the explicit space dependency of the external fields and the by powers of  $\mathbf{r}$ . We use the electric and magnetic multipole moment,  $\mathbf{Q}^{[n]}$  and  $\mathbf{m}^{[n]}$  respectively, defined as <sup>2</sup>

$$\mathbf{Q}_{j_1\cdots j_n}^{[n]} = -\int \mathbf{r}_{j_1} \mathbf{r}_{j_2} \cdots \mathbf{r}_{j_n} \rho(\mathbf{r}, t) \mathrm{d}^3 \mathbf{r}, \qquad (\mathrm{IX.4})$$

$$\mathbf{m}_{i;j_1\cdots j_{n-1}}^{[n]} = \frac{n}{n+1} \int \mathbf{r}_{j_1} \mathbf{r}_{j_2} \cdots \mathbf{r}_{j_n} \left[ \mathbf{r} \wedge \mathbf{J}(\mathbf{r},t) \right]_i \mathrm{d}^3 \mathbf{r}, \qquad (\mathrm{IX.5})$$

<sup>&</sup>lt;sup>1</sup>See page 57 and next of Ref [1]

<sup>&</sup>lt;sup>2</sup>Note that  $\mathbf{r}_{j_i}$  are direction in space: either  $\hat{\mathbf{a}}$ ,  $\hat{\mathbf{b}}$  or  $\hat{\mathbf{c}}$ .

in order to express the light-matter interaction potential as:

$$V(t) = Q^{[0]}\phi(\mathbf{0},t) - \sum_{n=1}^{+\infty} \frac{1}{n!} \mathbf{Q}^{[n]}_{j_1\cdots j_{n-1}} \cdot \mathbf{e}^{[n-1]}_{j_1\cdots j_{n-1}}(\mathbf{0},t) - \sum_{n=1}^{+\infty} \frac{1}{n!} \mathbf{m}^{[n]}_{j_1\cdots j_{n-1}} \cdot \mathbf{b}^{[n-1]}_{j_1\cdots j_{n-1}}(\mathbf{0},t).$$
(IX.6)

Where  $\mathbf{e}_{j_1\cdots j_{n-1}}^{[n-1]}$  and  $\mathbf{b}_{j_1\cdots j_{n-1}}^{[n-1]}$  are the field derivatives measured at the molecule "center":

$$\mathbf{e}_{j_1\cdots j_{n-1}}^{[n-1]}(t) = \frac{\delta^{n-1}\mathbf{e}_{j_1}}{\delta\mathbf{r}_{j_2}\delta\mathbf{r}_{j_3}\cdots\delta\mathbf{r}_{j_{n-1}}}(\mathbf{0},t)$$
(IX.7)

$$\mathbf{b}_{j_1\cdots j_{n-1}}^{[n-1]}(t) = \frac{\delta^{n-1}\mathbf{b}_{j_1}}{\delta\mathbf{r}_{j_2}\delta\mathbf{r}_{j_3}\cdots\delta\mathbf{r}_{j_{n-1}}}(\mathbf{0},t)$$
(IX.8)

# Application to water:

For neutral molecules, the zeroth order  $Q^{[0]}$  is the total charge: thus zero in our case. Therefore, the first term  $Q^{[0]}\phi(\mathbf{0},t)$  is always null. The second terms involve the electric multipoles of the molecules, for instance the dipole moment  $\boldsymbol{\mu}$  and the quadrupolar one  $Q_{ij}$ :

$$\mathbf{Q}^{[1]} = -\int \mathbf{r}\rho(\mathbf{r})d^{3}\mathbf{r} = \langle \Psi | \,\hat{\boldsymbol{\mu}} \, | \Psi \rangle = \boldsymbol{\mu}, \qquad (IX.9)$$

$$\mathbf{Q}_{ij}^{[2]} = -\int \mathbf{r}_i \mathbf{r}_j \rho(\mathbf{r}) \mathrm{d}^3 \mathbf{r} = \langle \Psi | \, \hat{Q}_{ij} \, | \Psi \rangle = Q_{ij}. \tag{IX.10}$$

For instance, if the molecule is under an electric field along the molecular c axis,  $e^c \neq 0$ , and that this field present also a gradient along the c direction,  $\frac{\delta e^c}{\delta c} \neq 0$ , then the electric part of the interaction potential is:

$$\langle \Psi | V | \Psi \rangle (t) = -\boldsymbol{\mu}_c e_c(t) - \frac{1}{2} Q_{cc} \frac{\delta e^c}{\delta c}(t)$$
(IX.11)

In the following, we will assume that the electric term can be cut after the quadrupolar terms because the field spatial derivatives are too small and also that we can neglect the magnetic terms. Therefore, the interacting Hamiltonian is :

$$\hat{H}(t) = \hat{H}^0 - \hat{\boldsymbol{\mu}} \cdot \mathbf{e}(t) - \frac{1}{2}\hat{Q} : \boldsymbol{\nabla}\mathbf{e}(t)$$
(IX.12)

# IX. A.1.2 Response scheme

# Equation of motion of the time-dependent wave-function:

Let us decompose the time-dependent wave-function on the eigenstate of the isolated Hamiltonian  $\{|\Psi^n\rangle\}$  noted  $\{|n\rangle\}$  with associated eigenvalue  $\{E_n\}$  – or frequencies  $\{\omega_n\}$  with  $E_n = \hbar\omega_n$ . To do so, we will use the time dependent coefficients  $\{d_n(t)\}$ :

$$|\Psi\rangle(t) = \sum_{n} d_n(t) \exp[-i\omega_n t] |n\rangle$$
(IX.13)

# Page 264

We will assume that the interaction potential is small, and that the wave-function can be decomposed in order of magnitude:

$$|\Psi\rangle(t) = |\Psi^{(0)}\rangle(t) + |\Psi^{(1)}\rangle(t) + |\Psi^{(2)}\rangle(t) + \cdots$$
 (IX.14)

To follow this decomposition, several sets of  $\{d_n^{(0)}(t)\}, \{d_n^{(1)}(t)\}, \{d_n^{(2)}(t)\}, \cdots$  are also defined:

$$|\Psi^{(m)}\rangle(t) = \sum_{n} d_{n}^{(m)}(t) \exp[-i\omega_{n}t] |n\rangle$$
(IX.15)

the Schrödinger equation becomes:

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\left[|\Psi^{(0)}\rangle(t) + |\Psi^{(1)}\rangle(t) + |\Psi^{(2)}\rangle(t)\cdots\right] = \left[\hat{H}^{0} + \hat{V}(t)\right]\left[|\Psi^{(0)}\rangle(t) + |\Psi^{(1)}\rangle(t) + |\Psi^{(2)}\rangle(t)\cdots\right]$$
(IX.16)

The interacting potential is viewed as a perturbation. By regrouping the terms with the same order of magnitude, one obtains:

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \left| \Psi^{(0)} \right\rangle(t) = \hat{H}^{0} \left| \Psi^{(0)} \right\rangle(t) \tag{IX.17}$$

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\Psi^{(1)}\rangle(t) = \hat{H}^{0} |\Psi^{(1)}\rangle(t) + \hat{V}(t) |\Psi^{(0)}\rangle(t)$$
(IX.18)

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\Psi^{(2)}\rangle(t) = \hat{H}^{0} |\Psi^{(2)}\rangle(t) + \hat{V}(t) |\Psi^{(1)}\rangle(t)$$
(IX.19)

# Perturbation calculation:

The zeroth order gives:

$$\begin{split} &\hbar \sum_{n} \left[ i \frac{\mathrm{d}}{\mathrm{d}t} d_{n}^{(0)}(t) + \omega_{n} d_{n}^{(0)}(t) \right] \exp[-i\omega_{n}t] \left| n \right\rangle = \sum_{n} E_{n} d_{n}^{(0)}(t) \exp[-i\omega_{n}t] \left| n \right\rangle \\ &\sum_{n} \frac{\mathrm{d}}{\mathrm{d}t} d_{n}^{(0)}(t) \exp[-i\omega_{n}t] \left| n \right\rangle = 0 \\ &\forall n, \frac{\mathrm{d}}{\mathrm{d}t} d_{n}^{(0)}(t) = 0, \text{because } \left\{ \left| n \right\rangle \right\} \text{ orthogonal basis} \end{split}$$

The non-interacting wave-function, can be expressed as a stationary state. Hence, if we assume that before any light-matter interaction the system is in the electronic ground states  $^3$ :

$$|\Psi^{(0)}\rangle(t) = \exp[-i\omega_0 t] |0\rangle. \qquad (IX.20)$$

<sup>&</sup>lt;sup>3</sup>Note that this is defined up to a global phase factor, for instance  $e^{i\theta}$ .

The first order leads to:

$$\begin{split} &\hbar \sum_{n} \left[ i \frac{\mathrm{d}}{\mathrm{d}t} d_{n}^{(1)}(t) + \omega_{n} d_{n}^{(1)}(t) \right] \exp[-i\omega_{n}t] \left| n \right\rangle = \sum_{n} E_{n} d_{n}^{(1)}(t) \exp[-i\omega_{n}t] \left| n \right\rangle + \hat{V}(t) \exp[-i\omega_{0}t] \left| 0 \right\rangle \\ &i\hbar \sum_{n} \frac{\mathrm{d}}{\mathrm{d}t} d_{n}^{(1)}(t) \exp[-i\omega_{n}t] \left| n \right\rangle = \sum_{\alpha} \sum_{\omega} \hat{V}_{\alpha} \mathbf{F}_{\alpha}^{\omega} \exp[-i\omega t] \exp[-i\omega_{0}t] \left| 0 \right\rangle \\ &\forall n, \frac{\mathrm{d}}{\mathrm{d}t} d_{n}^{(1)}(t) = \frac{1}{i\hbar} \sum_{\alpha} \sum_{\omega} \langle n | \hat{V}_{\alpha} | 0 \rangle \mathbf{F}_{\alpha}^{\omega} \exp[-i(\omega + \omega_{0} - \omega_{n})t] \\ &\forall n, d_{n}^{(1)}(t) = \frac{1}{\hbar} \sum_{\alpha} \sum_{\omega} \frac{\langle n | \hat{V}_{\alpha} | 0 \rangle}{\omega - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega} \exp[-i(\omega - \omega_{n0})t] \end{split}$$

Where the difference in energy between the ground state and the excited n has been defined in the frequency domain:  $\omega_{n0} = \omega_n - \omega_0$ .

Hence, the first perturbation is:

$$|\Psi^{(1)}\rangle(t) = \frac{1}{\hbar} \sum_{n} \sum_{\alpha} \sum_{\omega_{1}} \frac{\langle n | \hat{V}_{\alpha} | 0 \rangle}{\omega_{1} - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega_{1}} \exp[-i(\omega_{1} - \omega_{n0} + \omega_{n})t] | n \rangle$$
(IX.21)

The second order gives:

$$\begin{split} &\hbar \sum_{m} \left[ i \frac{\mathrm{d}}{\mathrm{d}t} d_{m}^{(2)}(t) + \omega_{m} d_{m}^{(2)}(t) \right] \exp[-i\omega_{m}t] \left| m \right\rangle = \sum_{m} E_{m} d_{m}^{(2)}(t) \exp[-i\omega_{m}t] \left| m \right\rangle + \hat{V}(t) \left| \Psi^{(1)} \right\rangle(t) \\ &i\hbar \sum_{m} \frac{\mathrm{d}}{\mathrm{d}t} d_{m}^{(2)}(t) \exp[-i\omega_{m}t] \left| m \right\rangle = \sum_{\alpha',\omega_{2}} \mathbf{F}_{\alpha'}^{\omega_{2}} \exp[-i\omega_{2}t] \times \frac{1}{\hbar} \sum_{n,\alpha,\omega_{1}} \frac{\langle n | \hat{V}_{\alpha} | 0 \rangle}{\omega_{1} - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega_{1}} \exp[-i(\omega_{1} - \omega_{n0} + \omega_{n})t] \hat{V}_{\alpha'} \left| n \right\rangle \\ &\forall m, \frac{\mathrm{d}}{\mathrm{d}t} d_{m}^{(2)}(t) = \frac{1}{i\hbar^{2}} \sum_{\alpha,\alpha'} \sum_{\omega_{1},\omega_{2}} \sum_{n} \frac{\langle n | \hat{V}_{\alpha} | 0 \rangle}{\omega_{1} - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega_{1}} \langle m | \hat{V}_{\alpha'} \left| n \right\rangle \mathbf{F}_{\alpha'}^{\omega_{2}} \exp[-i(\omega_{1} + \omega_{2} - \omega_{n0} + \omega_{n} - \omega_{m})t] \\ &\forall m, \frac{\mathrm{d}}{\mathrm{d}t} d_{m}^{(2)}(t) = \frac{1}{i\hbar^{2}} \sum_{\alpha,\alpha'} \sum_{\omega_{1},\omega_{2}} \sum_{n} \frac{\langle n | \hat{V}_{\alpha} | 0 \rangle}{\omega_{1} - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega_{1}} \langle m | \hat{V}_{\alpha'} \left| n \right\rangle \mathbf{F}_{\alpha'}^{\omega_{2}} \exp[-i(\omega_{1} + \omega_{2} - \omega_{m0})t] \\ &\forall m, d_{m}^{(2)}(t) = \frac{1}{\hbar^{2}} \sum_{\alpha,\alpha'} \sum_{\omega_{1},\omega_{2}} \sum_{n} \frac{\langle n | \hat{V}_{\alpha} | 0 \rangle}{\omega_{1} - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega_{1}} \frac{\langle m | \hat{V}_{\alpha'} | n \rangle}{\omega_{1} + \omega_{2} - \omega_{m0}} \mathbf{F}_{\alpha'}^{\omega_{2}} \exp[-i(\omega_{1} + \omega_{2} - \omega_{m0})t] \end{split}$$

Hence, the second order perturbation is:

$$\left|\Psi^{(2)}\right\rangle(t) = \frac{1}{\hbar^2} \sum_{\alpha,\alpha'} \sum_{\omega_1,\omega_2} \sum_{n,m} \frac{\langle n | \hat{V}_{\alpha} | 0 \rangle}{\omega_1 - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega_1} \frac{\langle m | \hat{V}_{\alpha'} | n \rangle}{\omega_1 + \omega_2 - \omega_{m0}} \mathbf{F}_{\alpha'}^{\omega_2} \exp\left[-i(\omega_1 + \omega_2 - \omega_{m0} + \omega_m)t\right] | m \rangle$$
(IX.22)

# First order response function:

Let us compute this function at the first order.

$$\begin{split} \langle \Psi^{(1)} | \,\hat{\Omega} | \Psi^{(0)} \rangle \left( t \right) &= \frac{1}{\hbar} \sum_{\alpha} \sum_{\omega_{1}} \sum_{n} \frac{\langle 0 | \,\hat{\Omega} | n \rangle \langle n | \,\hat{V}_{\alpha} | 0 \rangle}{\omega_{1} - \omega_{n0}} \exp[-i(-\omega_{n0} + \omega_{n})t] \exp[+i(\omega_{0})t] \mathbf{F}_{\alpha}^{\omega_{1}} \exp[-i\omega_{1}t] \\ \langle \Psi^{(1)} | \,\hat{\Omega} | \Psi^{(0)} \rangle \left( t \right) &= \frac{1}{\hbar} \sum_{\alpha} \sum_{\omega_{1}} \sum_{n} \frac{\langle 0 | \,\hat{\Omega} | n \rangle \langle n | \,\hat{V}_{\alpha} | 0 \rangle}{\omega_{1} - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega_{1}} \exp[-i\omega_{1}t] \\ \langle \Psi^{(0)} | \,\hat{\Omega} | \Psi^{(1)} \rangle \left( t \right) &= \frac{1}{\hbar} \sum_{\alpha} \sum_{\omega_{1}} \sum_{n} \frac{\langle 0 | \,\hat{\Omega} | n \rangle^{*} \langle n | \,\hat{V}_{\alpha} | 0 \rangle^{*}}{\omega_{1} - \omega_{n0}} \left\{ \mathbf{F}_{\alpha}^{\omega_{1}} \exp[-i\omega_{1}t] \right\}^{*} \end{split}$$

Since the electromagnetic field is a real quantity, we have:  $\mathbf{F}_{\alpha}^{\omega_1} = {\{\mathbf{F}_{\alpha}^{-\omega_1}\}}^*$ . Therefore, using a variable change:  $\omega_1 \to -\omega_1$  for the second term:

$$\begin{split} \langle \Psi^{(1)} | \,\hat{\Omega} | \Psi^{(0)} \rangle \left( t \right) + \langle \Psi^{(0)} | \,\hat{\Omega} | \Psi^{(1)} \rangle \left( t \right) = &\frac{1}{\hbar} \sum_{\alpha} \sum_{\omega_1} \sum_{n} \frac{\langle 0 | \,\hat{\Omega} | n \rangle \langle n | \,\hat{V}_{\alpha} | 0 \rangle}{\omega_1 - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega_1} \exp[-i\omega_1 t] \\ &+ \frac{1}{\hbar} \sum_{\alpha} \sum_{\omega_1} \sum_{\alpha} \sum_{\omega_1} \sum_{n} \frac{\langle 0 | \,\hat{V}_{\alpha} | n \rangle \langle n | \,\hat{\Omega} | 0 \rangle}{-\omega_1 - \omega_{n0}} \mathbf{F}_{\alpha}^{\omega_1} \exp[-i\omega_1 t] \end{split}$$

For a given frequency  $\omega_1$  associated with only one interacting terms  $\hat{V}_{\alpha}$ , we can now define the response function:  $\langle \langle \hat{\Omega}; \hat{V}_{\alpha}^{\omega_1} \rangle \rangle$ . It describes how the observable  $\Omega$  is impacted by an external excitation  $\mathbf{F}_{\alpha}^{\omega_1}$  associated with an interacting potential  $\hat{V}_{\alpha}$ . To keep track of the external field frequency, the interacting potential is written  $\hat{V}_{\alpha}^{\omega_1}$  so that: <sup>4</sup>

$$\langle \langle \hat{\Omega}; \hat{V}^{\omega_1}_{\alpha} \rangle \rangle = \frac{\langle \Psi^{(1)} | \hat{\Omega} | \Psi^{(0)} \rangle (t) + \langle \Psi^{(0)} | \hat{\Omega} | \Psi^{(1)} \rangle (t)}{\mathbf{F}^{\omega_1}_{\alpha} \exp[-i\omega_1 t]}$$
(IX.23)

Which leads to the equation proposed in the main text:

$$\langle\langle\hat{\Omega};\hat{V}^{\omega_{1}}_{\alpha}\rangle\rangle = -\frac{1}{\hbar} \left[\sum_{n} \frac{\langle 0|\,\hat{\Omega}\,|n\rangle\,\langle n|\,\hat{V}_{\alpha}\,|0\rangle}{\omega_{n0} - \omega_{1}} + \frac{\langle 0|\,\hat{V}_{\alpha}\,|n\rangle\,\langle n|\,\hat{\Omega}\,|0\rangle}{\omega_{n0} + \omega_{1}}\right] \tag{IX.24}$$

<sup>&</sup>lt;sup>4</sup>You can also find the same equation where the summation over states exclude the ground state  $(n \neq 0)$ . The reason is that the case n = 0 cancels out using the 2 terms.

## Second order response function:

Similarly, for the second order response function:

$$\langle \Psi^{(2)} | \hat{\Omega} | \Psi^{(0)} \rangle (t) = \frac{1}{\hbar^2} \sum_{\alpha, \alpha'} \sum_{\omega_1, \omega_2} \sum_{n, m} \frac{\langle 0 | \hat{V}_{\alpha'} | m \rangle \langle m | \hat{V}_{\alpha} | n \rangle \langle n | \hat{\Omega} | 0 \rangle}{(\omega_{n0} + \omega_1 + \omega_2)(\omega_{m0} + \omega_2)} \mathbf{F}_{\alpha}^{\omega_1} \mathbf{F}_{\alpha'}^{\omega_2} \exp[-i(\omega_1 + \omega_2)t]$$
(IX.25)

$$\langle \Psi^{(1)} | \hat{\Omega} | \Psi^{(1)} \rangle (t) = \frac{1}{\hbar^2} \sum_{\alpha, \alpha'} \sum_{\omega_1, \omega_2} \sum_{n, m} \frac{\langle 0 | \hat{V}_{\alpha} | n \rangle \langle n | \hat{\Omega} | m \rangle \langle m | \hat{V}_{\alpha'} | 0 \rangle}{(\omega_{n0} + \omega_1)(\omega_{m0} - \omega_2)} \mathbf{F}_{\alpha}^{\omega_1} \mathbf{F}_{\alpha'}^{\omega_2} \exp[-i(\omega_1 + \omega_2)t]$$
(IX.26)

$$\langle \Psi^{(0)} | \hat{\Omega} | \Psi^{(2)} \rangle (t) = \frac{1}{\hbar^2} \sum_{\alpha, \alpha'} \sum_{\omega_1, \omega_2} \sum_{n, m} \frac{\langle 0 | \hat{\Omega} | n \rangle \langle n | \hat{V}_{\alpha} | m \rangle \langle m | \hat{V}_{\alpha'} | 0 \rangle}{(\omega_{n0} - \omega_1 - \omega_2)(\omega_{m0} - \omega_2)} \mathbf{F}_{\alpha}^{\omega_1} \mathbf{F}_{\alpha'}^{\omega_2} \exp[-i(\omega_1 + \omega_2)t]$$
(IX.27)

Here two interacting potential operators appear:  $\hat{V}_{\alpha}$  and  $\hat{V}_{\alpha'}$  associated with two external fields at frequencies  $\omega_1$  and  $\omega_2$ :  $\mathbf{F}_{\alpha}^{\omega_1}$  and  $\mathbf{F}_{\alpha'}^{\omega_2}$ . Even if these terms may look non-symmetric with respect to these two potential, they are because of the summation over  $\alpha, \alpha'$  goes over the same quantity so that  $\alpha'$  and  $\alpha$  are getting exchanged at some point. The same goes for the frequency summation.

But, the explicit symmetrization is often performed by adding explicitly an exchange operator noted  $\mathcal{P}_{1,2}$  which exchange  $(\alpha, \omega_1)$  with  $(\alpha', \omega_2)$ . To keep the same absolute value, a factor 1/2!need to be added:

$$\Omega^{(2)}(t) = \frac{1}{2!} \frac{1}{\hbar^2} \sum_{\alpha,\alpha'} \sum_{\omega_1,\omega_2} \mathcal{P}_{1,2} \sum_{n,m} \left[ \frac{\langle 0 | \hat{V}_{\alpha'} | m \rangle \langle m | \hat{V}_{\alpha} | n \rangle \langle n | \hat{\Omega} | 0 \rangle}{(\omega_{n0} + \omega_1 + \omega_2)(\omega_{m0} + \omega_2)} + \frac{\langle 0 | \hat{V}_{\alpha} | n \rangle \langle n | \hat{\Omega} | m \rangle \langle m | \hat{V}_{\alpha'} | 0 \rangle}{(\omega_{n0} + \omega_1)(\omega_{m0} - \omega_2)} + \frac{\langle 0 | \hat{\Omega} | n \rangle \langle n | \hat{V}_{\alpha} | m \rangle \langle m | \hat{V}_{\alpha'} | 0 \rangle}{(\omega_{n0} - \omega_1 - \omega_2)(\omega_{m0} - \omega_2)} \right] \mathbf{F}_{\alpha'}^{\omega_1} \mathbf{F}_{\alpha'}^{\omega_2} \exp[-i(\omega_1 + \omega_2)t]$$

For readability, we will keep this 1/2! factor written using the factorial notation. As for the first order response function, we can identify  $\langle \langle \hat{\Omega}; \hat{V}^{\omega_1}_{\alpha}, \hat{V}^{\omega_2}_{\alpha'} \rangle \rangle$  for two given interaction potential. Since the frequencies of interaction matters, we will keep track of them by incorporating them in the notation:  $\hat{V}^{\omega_1}_{\alpha}$  and  $\hat{V}^{\omega_2}_{\alpha'}$ <sup>5</sup>:

<sup>&</sup>lt;sup>5</sup>You can also find the same equation where the summation over states exclude the ground state  $(n \neq 0)$ . The reason is that the case n=0 cancels out: the first and last terms with each other. The second term cancels out when summing over all the frequencies: the second term for  $(\omega_1, \omega_2)$  is canceled by the one obtained at  $-(\omega_1, \omega_2)$ . The same demonstration also applies for n=m.

$$\langle \langle \hat{\Omega}; \hat{V}_{\alpha}^{\omega_{1}}, \hat{V}_{\alpha'}^{\omega_{2}} \rangle \rangle = \frac{1}{\hbar^{2}} \mathcal{P}_{1,2} \sum_{n,m} \left[ \frac{\langle 0 | \hat{V}_{\alpha'} | m \rangle \langle m | \hat{V}_{\alpha} | n \rangle \langle n | \hat{\Omega} | 0 \rangle}{(\omega_{n0} + \omega_{1} + \omega_{2})(\omega_{m0} + \omega_{2})} + \frac{\langle 0 | \hat{V}_{\alpha} | n \rangle \langle n | \hat{\Omega} | m \rangle \langle m | \hat{V}_{\alpha'} | 0 \rangle}{(\omega_{n0} + \omega_{1})(\omega_{m0} - \omega_{2})} + \frac{\langle 0 | \hat{\Omega} | n \rangle \langle n | \hat{V}_{\alpha} | m \rangle \langle m | \hat{V}_{\alpha} | m \rangle \langle m | \hat{V}_{\alpha'} | 0 \rangle}{(\omega_{n0} - \omega_{1} - \omega_{2})(\omega_{m0} - \omega_{2})} \right]$$
(IX.28)

# IX. A.1.3 Conventions: traceless quadrupole moment

In the literature [89], the quadrupolar term can also be defined using another tensor  $\Theta$  named traceless:

$$\Theta_{ij} = \frac{3}{2} \left( Q_{ij} - \frac{1}{3} \delta_{ij} Q_{ll} \right)$$

Where is the trace of the quadrupolar moment  $Q_{ll} = \sum_l Q_{ll}$  – using the Einstein summation convention. The idea of this formalism is to take advantage of the Gauss-Maxwell equation:

$$\boldsymbol{\nabla} \cdot \mathbf{e} = \boldsymbol{\nabla}_l \mathbf{e}_l = \frac{\rho}{\epsilon_0} \tag{IX.29}$$

Let us assume that the field is created by an external charge away from the point of interest. In the case of the undepleted approximation (the part of space where the field is measured does not impact the field), the equation fulfilled by the field is:  $\nabla_l \mathbf{e}_l = 0$ . Hence, one can define a new electric field gradient which is traceless by definition:

$$\mathbf{\nabla}_i \tilde{\mathbf{e}}_j = \mathbf{\nabla}_i \mathbf{e}_j + rac{1}{3} \delta_{ij} \mathbf{\nabla}_l \mathbf{e}_l$$

Using such definition, we make sure that the new electric field gradient is traceless – note that it is solely equal to the original gradient in the case  $\nabla_l \mathbf{e}_l = 0$ .

The two sets of definition are similar once summed over indexes, for instance let us define a value A so that:

$$\begin{split} A &= \frac{1}{3} \sum_{ij} \Theta_{ij} \nabla_i \tilde{\mathbf{e}}_j = \frac{1}{3} \sum_i \Theta_{ii} \nabla_i \tilde{\mathbf{e}}_i + \frac{1}{3} \sum_{i \neq j} \Theta_{ij} \nabla_i \tilde{\mathbf{e}}_j \\ &= \frac{1}{3} \sum_i \frac{3}{2} \left( Q_{ii} - \frac{1}{3} Q_{ll} \right) \left( \nabla_i \mathbf{e}_i + \frac{1}{3} \nabla_l \mathbf{e}_l \right) + \frac{1}{3} \sum_{i \neq j} \frac{3}{2} Q_{ij} \nabla_i \mathbf{e}_j \\ &= \frac{1}{2} \sum_i Q_{ii} \nabla_i \mathbf{e}_i + \frac{1}{6} \sum_i Q_{ii} \nabla_l \mathbf{e}_l - \frac{1}{6} \sum_i Q_{ll} \nabla_i \mathbf{e}_i + \frac{1}{2} \sum_{i \neq j} Q_{ij} \nabla_i \mathbf{e}_j = \frac{1}{2} \sum_i Q_{ii} \nabla_i \mathbf{e}_i + \frac{1}{2} \sum_{i \neq j} Q_{ij} \nabla_i \mathbf{e}_j \\ &= \frac{1}{2} \sum_{ij} Q_{ij} \nabla_i \mathbf{e}_j \end{split}$$

Hence, compared to some articles, the quadrupolar hyperpolarizability may have a 3/2 scaling factor for the off-diagonal elements. The diagonal elements are more complicated because it requires to use the trace of Q. However, the two formalisms described the same quantity and the same physics. We chose to stay in the Q convention rather than the traceless because it seems easier to use for SHG at the surface or in the bulk. Moreover,  $\nabla_l \mathbf{e}_l \neq 0$  at the interface.

# IX. A.1.4 Conventions: T and V

At least two conventions exist for the response tensor definition: the T and the V. For the sake of example, let us compute the induced dipole moment for an external electric field composed of:  $\mathbf{e}_a^{\omega}, \mathbf{e}_b^{\omega}, \nabla_c \mathbf{e}_b^{\omega}$ . The induced dipole moment at the frequency  $2\omega, \mu^{2\omega}$ , along the *c* direction is in the T and V conventions:

$$T: \boldsymbol{\mu}_{c}^{2\omega} = \frac{1}{2}\beta_{caa}^{dd}\mathbf{e}_{a}^{\omega}\mathbf{e}_{a}^{\omega} + \frac{1}{2}\beta_{cbb}^{dd}\mathbf{e}_{b}^{\omega}\mathbf{e}_{b}^{\omega} + \beta_{cba}^{dd}\mathbf{e}_{b}^{\omega}\mathbf{e}_{a}^{\omega} + \frac{1}{2}\beta_{cacb}^{dq}\mathbf{e}_{a}^{\omega}\boldsymbol{\nabla}_{c}\mathbf{e}_{b}^{\omega} + \frac{1}{2}\beta_{cbcb}^{dq}\mathbf{e}_{b}^{\omega}\boldsymbol{\nabla}_{c}\mathbf{e}_{b}^{\omega}$$
(IX.30)

$$V: \boldsymbol{\mu}_{c}^{2\omega} = \tilde{\beta}_{caa}^{dd} \mathbf{e}_{a}^{\omega} \mathbf{e}_{a}^{\omega} + \tilde{\beta}_{cbb}^{dd} \mathbf{e}_{b}^{\omega} \mathbf{e}_{b}^{\omega} + 2\tilde{\beta}_{cba}^{dd} \mathbf{e}_{b}^{\omega} \mathbf{e}_{a}^{\omega} + \tilde{\beta}_{cacb}^{dq} \mathbf{e}_{a}^{\omega} \boldsymbol{\nabla}_{c} \mathbf{e}_{b}^{\omega} + \tilde{\beta}_{cbcb}^{dq} \mathbf{e}_{b}^{\omega} \boldsymbol{\nabla}_{c} \mathbf{e}_{b}^{\omega}$$
(IX.31)

However,  $\tilde{\beta}_{ijkl}^{dq} = \tilde{\beta}_{ijk}^{dq}$  Hence, for second order response tensors the convention T is 2 times bigger than convention V. In the following, we will use the T convention. Please note that the V convention is very often used in the experimental literature, and the T in the numerical one.

# IX. A.1.5 Gauge (in-)variability

To compute the response function, one has to compute dipole or quadrupole transition moment:

$$\langle n | \, \hat{\boldsymbol{\mu}}_i \, | m \rangle = -\int \phi^n(\mathbf{r})^* \mathbf{r}_i \phi^m(\mathbf{r}) d\mathbf{r}, \langle n | \, \hat{Q}_{ij} \, | m \rangle = -\int \phi^n(\mathbf{r})^* \mathbf{r}_i \mathbf{r}_j \phi^m(\mathbf{r}) d\mathbf{r}$$

The origin of the integrant  $\mathbf{r}$  should not impact the results. For neutral molecules it is the case for the dipole: let's define a new frame where  $\mathbf{r}' = \mathbf{r} + \mathbf{C}$ 

$$\langle n | \, \hat{\boldsymbol{\mu}}_i \, | m \rangle = -\int \phi^n(\mathbf{r}')^* \mathbf{r}'_i \phi^m(\mathbf{r}') d\mathbf{r}' = -\int \phi^n(\mathbf{r})^* \, (\mathbf{r}_i + \mathbf{C}_i) \, \phi^m(\mathbf{r}) d\mathbf{r} = -\mathbf{C}_i \int \phi^n(\mathbf{r})^* \phi^m(\mathbf{r}) d\mathbf{r} + \langle n | \, \hat{\boldsymbol{\mu}}_i \, | m \rangle$$
$$= -\mathbf{C}_i \delta_{nm} + \langle n | \, \hat{\boldsymbol{\mu}}_i \, | m \rangle$$

Since the summation goes over  $n \neq m$ , the transition dipole moment is gauge invariant. For the quadrupolar transitions:

$$\langle n | \hat{Q}_{ij} | m \rangle = -\int \phi^n(\mathbf{r}')^* \mathbf{r}'_i \mathbf{r}'_j \phi^m(\mathbf{r}') d\mathbf{r}' = -\int \phi^n(\mathbf{r})^* (\mathbf{r}_i + \mathbf{C}_i) (\mathbf{r}_j + \mathbf{C}_j) \phi^m(\mathbf{r}) d\mathbf{r}$$
$$= \langle n | \hat{Q}_{ij} | m \rangle + \mathbf{C}_i \langle n | \hat{\boldsymbol{\mu}}_j | m \rangle + \mathbf{C}_j \langle n | \hat{\boldsymbol{\mu}}_i | m \rangle - \mathbf{C}_i \mathbf{C}_j \delta_{nm}$$

Hence, the frame is important for the calculation of the quadrupole moment. In DALTON by default, the gauge is fixed at the center of mass position. Hence, our quadrupole calculations are done using for referential the one centered at the barycentric position. Note that this problem is not solved using the traceless convention.

# IX. A.2 About Dunning basis set

In this short annex, we describe what is hidden behind the notation of the Dunning basis set. For instance, for water we use "d-aug-cc-pVTZ" basis set.

The smallest AO set would be 'cc-pVDZ'. 'cc', stands for 'correlation-consistent' because it was optimized for correlated systems. The 'pV' is for 'polarized Valence' and DZ is for 'Double-Zeta'. 'pVDZ' means that for both the Hydrogen and the Oxygen atoms, a full extra valence band is added. The minimal basis set would be only 1 AOs for each Hydrogen and 8 AOs for the Oxygen, but this would be insufficient to describe the 'polarization' of the electron cloud: the chemical bonds.

There will be 3 sub-basis centered at every nuclei position: 2 for the Hydrogen and 1 for the Oxygen atoms. For each of these sub-basis set, we will define AOs centered at the atom positions. For the Hydrogens, 5 functions are used: 1s, 2s and 2p. The 's' functions are described using only one function each while the 'p' requires 3 functions – for each momenta direction. For the Oxygen, 14 functions are used: 1s, 2s, 2p, 3s, 3p and 1d – the 1d requires 5 functions. For our calculation, we have estimated good convergence for a Triple zeta basis set (*i.e.* the "VTZ").

For SHG processes, it appears that delocalized electrons have a very important role – at the dipolar and even more at the quadrupolar level. Thus, if we want to describes efficiently the most delocalized electrons, we would have to increase more and more the *n* and thus go to QZ or 5Z. However, these AO are still very localized near the nuclei and a very slow convergence is expected, see Table III.1. Hence, "augmented" AO can be added to the basis set: they are an extra layer of valence basis set with a different set of parameters. Hence, in the following we are using the d-aug-cc-pVTZ for the water molecule. The non-diffuse AO for the Hydrogen goes up to 3s2p1d and the Oxygen up to 4s3p2d1f, and 2 full-valence AO are added (d-aug) optimized for the "diffuse" behavior. This makes a basis of 126 AO for the water molecule which will be used to build the MO.

# IX. A.3 About DFT framework

The DFT theory is based on two theorems demonstrated by Hohenberg and Kohn (HK) [128] and on the Kohn-Sham (KS) ansatz [129]. This formalism assumes that the ground state is non-degenerate. Another formulation made by Levy and Lieb [130, 131] can be used to treat degenerate ground states. Here, we will use the Bohr Oppenheimer approximation - which is not required to use a DFT framework – and neglect the spin for simplicity.

In this formalism, the Hamiltonian is decomposed into purely electronic terms ( $\hat{T}$  and  $\hat{V}_{ee}$  of Equation III.3) and the interaction between the electrons and the rest. Hence, the nuclei-electrons are included in an *external* potential  $\hat{V}_{ext}$ , which would also include the electromagnetic interaction if applied an electromagnetic field. The Hamiltonian is thus:  $\hat{H} = \hat{T} + \hat{V}_{ee} + \hat{V}_{ext}$ .

We should also emphasize that for common external potential (electron-nulclei or electric fieldelectron interaction):  $\langle \Psi | \hat{V}_{\text{ext}} | \Psi \rangle = \int V_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r}.$ 

#### IX. A.3.1 Exact DFT: Hohenberg-Kohn Theorems

Theorem I: For any system of interacting particles in an external potential  $\hat{V}_{\text{ext}}$ , the potential  $\hat{V}_{\text{ext}}$  is determined uniquely, except for a constant, by the ground state particle density  $\rho_0(\mathbf{r})$ .

Corollary I: Since the Hamiltonian is thus fully determined, except for a constant shift of the energy, it follows that the many-body wave-functions for all states (ground and excited) are determined. Therefore, all observable of the system is completely determined given only the ground state density  $\rho_0(\mathbf{r})$ .

Demonstration: Let us assume that there are two different external potentials  $\hat{V}_1$  and  $\hat{V}_2$  that are not equal up to a constant that leads to the same one-body ground state density  $\rho_0$ . These two different external potential lead to two different Hamiltonians  $\hat{H}_1$  and  $\hat{H}_2$  which have different ground state wavefunctions  $|\Psi_1\rangle$  and  $|\Psi_2\rangle$ .  $|\Psi_1\rangle$  and  $|\Psi_2\rangle$  must also lead to the same density  $\rho_0$  by construction. Since  $|\Psi_1\rangle$  is not the ground state of  $\hat{H}_2$  and that the ground state is assumed to be non-degenerate:

$$E_{2} = \langle \Psi_{2} | \hat{H}_{2} | \Psi_{2} \rangle < \langle \Psi_{1} | \hat{H}_{2} | \Psi_{1} \rangle$$
  
$$\langle \Psi_{1} | \hat{H}_{2} | \Psi_{1} \rangle = \langle \Psi_{1} | \hat{H}_{1} | \Psi_{1} \rangle + \langle \Psi_{1} | [\hat{H}_{2} - \hat{H}_{1}] | \Psi_{1} \rangle = E_{1} + \int [\hat{V}_{2} - \hat{V}_{1}](\mathbf{r})\rho_{0}(\mathbf{r})d\mathbf{r}$$
  
$$E_{2} < E_{1} + \int [\hat{V}_{2} - \hat{V}_{1}](\mathbf{r})\rho_{0}(\mathbf{r})d\mathbf{r}$$

One can do the same with  $|\Psi_2\rangle$  and  $\hat{H}_1$ , which leads to:

$$E_1 < E_2 + \int \mathrm{d}\mathbf{r} [\hat{V}_1 - \hat{V}_2](\mathbf{r}) 
ho_0(\mathbf{r})$$

By adding the last two equations, we obtained the absurd result:  $E_1 + E_2 < E_1 + E_2$ . Therefore,

there is not such external potentials  $\hat{V}_1$  and  $\hat{V}_2$  that lead to the same ground state density  $\rho_0$ . Hence the external potential is uniquely determined by the ground state density.

Theorem II: A universal functional of the electronic density for the energy  $E[\rho]$  in terms of the density can be defined, valid for any external potential  $\hat{V}_{\text{ext}}$ . For any particular  $\hat{V}_{\text{ext}}$ , the exact ground state energy of the system is the global minimum value of this functional, and the density  $\rho(\mathbf{r})$  that minimizes the functional is the exact ground state density  $\rho_0(\mathbf{r})$ .

Corollary II: The functional  $F_{\text{HK}}[n]$  alone is sufficient to determine the exact ground state energy and density. In general, excited states of the electrons must be determined by other means.

Demonstration: Let us find the wave-function  $\Psi$  so that it minimizes the energy:

$$E = \min_{\Psi} \left[ \left\langle \Psi \right| \hat{T} + \hat{V}_{\text{ee}} + \hat{V}_{\text{ext}} \left| \Psi \right\rangle \right]$$

Instead of comparing the energy obtained for all wave-function, we defined a subset of wave-function leading to a given density  $\Psi \rightarrow \rho$ . Then, we minimize with respect to the density and to the wavefunction:

$$E = \min_{\rho} \min_{\Psi \to \rho} \left[ \left\langle \Psi | \hat{T} + \hat{V}_{ee} + \hat{V}_{ext} | \Psi \right\rangle \right]$$
  
=  $\min_{\rho} \left[ \min_{\Psi \to \rho} \left\langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \right\rangle + \int V_{ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \right]$   
$$E = \min_{\rho} \left[ F_{HK}[\rho] + \int V_{ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \right]$$

Using for the definition of the universal functional  $F_{\rm HK}[\rho]$ :

$$F_{\rm HK}[\rho] = \min_{\Psi \to \rho} \left\langle \Psi \right| \hat{T} + \hat{V}_{\rm ee} \left| \Psi \right\rangle$$

This second theorem does not provide any information about the form of  $F_{\text{HK}}[\rho]$ : it only proves its existence. At this point, we have proven that i) all of the electronic properties of the system are given by the ground state density and ii) that the ground state energy can be written as a functional of the one-body ground state density. The initial problem written in terms of the many-body wavefunction (defined in  $C^{3N}$ ) has been reformulated exactly in terms of the total density (defined in  $\mathcal{R}^3$ ). There is an enormous gain in simplicity! One can also reformulate this in terms of density and current or in terms of Green functions.

# IX. A.3.2 Kohn-Sham Ansatz

Hence, the main object of the DFT framework is not the wave-function, but the one-body density. However, in practice, to construct the density we have to know the wave-function. Therefore, to work with a simple wave-function we will use the Kohn Sham ansatz: the exact ground state density can be calculated from an auxiliary system of non-interacting particles. This non-interaction wavefunction is written using a single Slater Determinant to enforce Fermi exclusion principle using one-body wave-functions { $\psi^i$ }. Its density  $n_{\rm KS}(\mathbf{r})$  is equal to the ground state one of the real interacting systems,  $\rho_0$ , by construction:

$$n_{\rm KS}(\mathbf{r}) = \sum_{n} |\psi^i(\mathbf{r})|^2 = \rho_0(\vec{r})$$
 (IX.32)

Note that the one-body density is way easier to compute in the HK wave-function than in the real correlated system.

According to the second Hohenberg-Kohn theorem:

$$E_0 = \min_{\rho} \left[ F_{\rm HK}[\rho] + \int V_{\rm ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \right]$$
(IX.33)

$$= F_{\rm HK}[\rho_0] + \int V_{\rm ext}(\mathbf{r})\rho_0(\mathbf{r})d\mathbf{r} = F_{\rm HK}[n_{\rm KS}] + \int V_{\rm ext}(\mathbf{r})n_{\rm KS}(\mathbf{r})d\mathbf{r}$$
(IX.34)

Thus, to find the ground state density, we can minimize  $E_0$  using the auxiliary system. The external potential acts in the same way for the real and auxiliary system. However, we have no information about  $F_{\text{HK}}[n_{\text{KS}}]$ : let us try out an expression that can be used for numerical calculation.

We know that:  $F_{\text{HK}}[\rho] = \min_{\Psi \to \rho} \langle \Psi | \hat{T} + \hat{V}_{\text{ee}} | \Psi \rangle$ . But we want to use the auxiliary wave-function which is un-correlated. For uncorrelated system (one-Slater Determinant) the kinetic energy is:

$$T_{\rm KS} = \langle \Psi | T_s | \Psi \rangle = -\frac{1}{2} \sum_i \langle \psi^i | \hat{\boldsymbol{\nabla}}^2 | \psi^i \rangle = \frac{1}{2} \sum_i | \hat{\boldsymbol{\nabla}} \psi^i |^2$$
(IX.35)

The electron-electron interaction is:

$$\begin{split} V_{\mathrm{KS}}^{ee}[n_{KS}] &= \langle \Psi | \, \hat{V}_{ee} \, | \Psi \rangle - \frac{1}{2} \sum_{i,j} \int \frac{|\psi^i|^2(\mathbf{r})|\psi^j|^2(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' - \frac{1}{2} \sum_{i,j} \int \frac{|\psi^i|^2(\mathbf{r})\psi^j(\mathbf{r})] \times \left[\psi^i(\mathbf{r}')\psi^j|(\mathbf{r}')\right]^*}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \\ &= \frac{1}{2} \sum_i \int \frac{|\psi^i|^2(\mathbf{r})n_{KS}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_x^{exact} \\ &= \frac{1}{2} \int \frac{n_{KS}(\mathbf{r})n_{KS}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_x^{exact} = V_{Hartree}[n_{KS}] + E_x^{exact} \end{split}$$

Hence, let us define an energy error  $E_{xc}$  between the real interacting electron-electron energy and the auxiliary one:

$$F_{\rm HK}[n_{\rm KS}] = F_{\rm KS}[n_{\rm KS}] + E_{xc}[n_{\rm KS}]$$
 (IX.36)

Where the universal functional  $F_{\rm KS}$  defined for the auxiliary system is:

$$F_{\rm KS}[n_{\rm KS}] = T_{\rm KS}[n_{KS}] + V_{Hartree}[n_{KS}]$$
(IX.37)

# Page 274

So that:

$$E_{xc}[n_{\rm KS}] = (T_{\rm HK}[n_{KS}] - T_{\rm KS}[n_{KS}]) + (V_{\rm HK}^{ee}[n_{KS}] - V_{Hartree}[n_{KS}])$$
(IX.38)

The exchange-correlation energy  $E_{xc}$  is the difference between the kinetic and the internal interaction energies of the true interacting many-body systems from these of the fictitious independentparticle system with electron-electron interactions replaced by the Hartree energy. We will define after a concrete expression for  $E_{xc}[n_{\text{KS}}]$ . Using these equations, we have defined univoquely a Hamiltonian for the auxiliary system to minimize:

$$H_{\rm KS}[n_{\rm KS}](\mathbf{r}) = -\frac{1}{2}\boldsymbol{\nabla}^2 + V_{\rm ext}(\mathbf{r}) + V_{\rm Hartree}[n_{\rm KS}](\mathbf{r}) + V_{\rm xc}[n_{\rm KS}](\mathbf{r}) \quad V_{\rm Hartree}[n_{\rm KS}](\mathbf{r}) = \frac{1}{2}\int \frac{n_{\rm KS}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}'$$
(IX.39)

One can then derive the variational equation with the constraint  $\langle \psi_i | \psi_j \rangle = \delta_{i,j}$ :

$$\frac{\delta}{\delta\psi_i^*(r)} \left[ E_{\rm KS}[n_{\rm KS}] + \sum_{m,n} L_{mn}(\langle\psi_m|\psi_n\rangle - \delta_{m,n}) \right] = 0 \tag{IX.40}$$

$$\frac{\delta}{\delta L_{ab}} \left[ E_{\rm KS}[n_{\rm KS}] + \sum_{m,n} L_{mn} (\langle \psi_m | \psi_n \rangle - \delta_{m,n}) \right] = 0 \tag{IX.41}$$

With  $L_{ij}$  Lagrange multipliers. Defining  $\epsilon_i$  the energy of every one-electron wave-function, this leads to the Kohn-Sham equation:

$$H_{\rm KS}[n_{\rm KS}]\psi^i = \epsilon_i \psi^i \tag{IX.42}$$

Compared to the original eigenstate equation of the full interacting system, Equation III.4, there is N equations that must be fulfied by every one-electronic wave-function – instead of 1 equation for an N-electronic one! The Kohn-Sham equations are self-consistent:  $H_{\rm KS}$  depends on all the  $\psi^i$ .

# IX. A.3.3 Functional for $E_{xc}$ : Hybrid

Now, let us discuss the form of the functional to define the exchange-correlation energy.

The exchange-correlation energy contains 2 terms:

$$E_{xc}[n_{KS}] = \underbrace{V_{\text{HK}}^{ee}[n_{KS}] - V_{Hartree}[n_{KS}]}_{\text{exchange}} + \underbrace{T_{\text{HK}}[n_{KS}] - T_{\text{KS}}[n_{KS}]}_{\text{correlation}}$$
(IX.43)

This separation of  $E_{xc}$  between an exchange  $E_x$  and correlated  $E_c$  energy is extremely widespread – even if no formal proof of this decomposition exist.

To parametrize these terms, a local approximation is made:  $E_{xc}[n_{KS}] = \int \tilde{E}_{xc}[n_{KS}(\mathbf{r}), \nabla n_{KS}(\mathbf{r})] d\mathbf{r}$ .  $\tilde{E}_{xc}$  is defined locally in different ways. For instance, the exchange part for a uniform electron gas has been defined by Dirac [132]:  $E_x^{LDA}[n_{KS}] = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} \int n_{KS}^{4/3}(\mathbf{r}) d\mathbf{r}$ . To complete the description of the uniform electron gas, Vosko, Wilk and Nusair [133] proposed a correlation energy fitted based on Monte Carlo simulations:  $E_c^{VWN}$ . Using these 2 functional depending only on the density value at each point  $\mathbf{r}$  is extremely efficient for describing systems with a high and slow varying density: for instance solids. However, for molecules with large spacial density evolution, one should include explicitly the  $\nabla n_{KS}$  in the  $E_{xc}$ . One among many examples is the exchange energy proposed by Becke [134]  $E_x^{Becke}$  and the correlation energy by Lee, Yang and Parr [135]  $E_c^{LYP}$ . The advantage of using local functional is that the Kohn-Sham Hamiltonian is local:  $\delta(\mathbf{r} - \mathbf{r}')\hat{H}_{KS}[n_{KS}] |\psi^i\rangle = H_{KS}[n_{KS}](\mathbf{r})\psi^i(\mathbf{r})$ .

We will not present here these functional in detail but rather highlight some key point relevant for our use with this key argument in mind: we aim to reproduce the NLO molecular properties of very few systems (water and alcohol) with the cheapest numerical cost possible. Hence, according to the literature, a very good functional is CAM-B3LYP [136], see Section III. B.2.3.

CAM-B3LYP is based on the B3LYP [137] functional, which is **hybrid**:  $E_{xc}$  is not given solely by functional defined locally using  $n_{KS}(\mathbf{r})$  and its gradient  $\nabla n_{KS}(\mathbf{r})$ . In a hybrid functional, a part of "exact" exchange energy is included:  $E_x^{exact}$ . We should remind that  $E_x^{exact}$  is not the true exact exchange energy because this one is calculated using a wave-function build using one Slater Determinant. The advantage of this energy is that its represents well the exchange energy for delocalized electrons and low density in general – because in this case the correlation in general plays a smaller role! The drawback of this expression is that the calculation of  $E_{xc}$  do not depend explicitly on the density but also on the MO themselves. Therefore,  $\hat{H}_{KS}[n_{KS}]$  is no longer local due to  $E_x^{exact}$ .

The functional B3LYP is thus composed of several other functional:

$$E_{xc}^{B3LYP} = 0.8E_x^{LDA} + 0.72E_x^{Becke} + 0.2E_x^{exact} + 0.19E_c^{VWN} + 0.81E_c^{LYP}$$
(IX.44)

The aim of using several functional is to describe the most efficiently possible the different aspect of the electronic behavior in the molecule. In top of this, the CAM-B3LYP adds a varying impact of the exact term depending on the distance of interaction  $(|\mathbf{r} - \mathbf{r}'|)$ . This is named Coulomb Attenuated Method (CAM) functional.

# IX. A.4 More on exciting field in the bulk phase

In this appendix we present in more detail some reminders and equations about the exciting field.

# What if $e^M = E$ ?

Let us assume that the exciting field is equal to the electromagnetic one for neat water. In this case:

$$\mathbf{P}^{\omega} = <\alpha(\omega,\omega) > \mathbf{E} = \epsilon_0(\epsilon_r - 1)\mathbf{E} \qquad \rightarrow \qquad \epsilon_r = n^2 = 1 + \frac{1}{\epsilon_0} < \alpha(\omega,\omega) > \qquad (IX.45)$$

For pure water, we have  $\langle \alpha(\omega, \omega) \rangle = 9.5$  a.u. and for the density 1kg/L (or 0.05 molecules per volume in atomic units). We remind that in atomic units:  $4\pi\epsilon_0 = 1$  Using the above equation we found  $n \approx 2.6$ . At optical frequencies we shall find  $n \approx 1.33$ . Hence, the electric field felt by the molecules is not **E**.

# Shape of $\mathcal{T}^M$ for the liquid phase:

We have assumed that there is a matrix  $\mathcal{T}^M$  so that:

$$\mathbf{e}^M = \mathcal{T}^M \cdot \mathbf{E}(\mathbf{r}_M) \tag{IX.46}$$

 $\mathcal{T}^M$  is a 3x3 matrix which links the exciting field to the electromagnetic one. At the scale of several molecules,  $\mathbf{E}(\mathbf{r}_M)$  is constant: the spatial wave length is  $\lambda/n \approx 600$  nm. Hence, let us consider a large subsystem of several nanometers of dimensions so that the electromagnetic field is constant and so that there are enough molecule so that the centro-symmetry is respected.

In such mesoscopic sample, we have calculated the molecular polarizability in the laboratory frame  $\alpha(\omega, \omega)$ . This 3x3 matrix is almost diagonal (because in the molecular frame  $\alpha(\omega, \omega)$  is also almost diagonal), and varies very little from one molecule to another. Hence, in a good approximation, we can say that  $\alpha(\omega, \omega)$  is constant and diagonal.

The relation between the exciting field and the electromagnetic field is the result of:

- The cause: the electromagnetic field **E**.
- The molecular reaction: the dipole moment induced at the same frequency which is governed by  $\alpha(\omega, \omega)$ .
- The molecule radiation: the effects of the induced dipole to the electromagnetic field felt by a molecule.

At this point we have to make an approximation, which turns out to be consistent later on: we will assume that  $\mathcal{T}^M$  is the same for every molecule in the bulk phase:  $\mathcal{T}$ . This can be understood as a consequence that the source term (**E**) and the molecular reaction ( $\alpha(\omega, \omega)$ ) is the same for all the

# Page 277

molecules of this mesoscopic system. Using this hypothesis, we are able to compute easily the effect of the molecules radiation in this homogeneous sample.

To find the shape of this constant 3x3 tensor,  $\mathcal{T}^M$ , we will imagine that we apply an electromagnetic field along the laboratory z axis. Our mesoscopic sample respects the centro-symmetry (regarding its structure) and **E** is constant throughout this size small compared to the wave length. Hence, the laboratory z axis is different from the x and y because we have applied an electric field along the +z or -z. However, at this scale there is no effect of the wave length: the x and y axis are indistinguishable. In other words, this subsystem (composed of the matter and the light) is  $C_{\infty}$ along the z axis. Therefore, there cannot have a net average vector along the x or y. Thus, the exciting field along the x and y axis is null: otherwise there would be a mesoscopic induced dipole moment along the x or y axis. Hence,  $\mathcal{T} \cdot \hat{z}$  is also along z. By repeating this demonstration when applying an x and y polarization for the electromagnetic field, we found that the tensor  $\mathcal{T}$  shall be diagonal, and with the same value for all the components since the liquid is isotropic.

# Radiation of dipoles: inter-molecular interactions

Assuming that each molecule has the same induced dipole moment due to an external electromagnetic field  $\mathbf{E}$ , we can easily compute the field created by the liquid phase on a given molecule i. Every molecule close to the molecule i produce an electric field due to their induced dipole moment, see Equation III.69. In the laboratory frame centered at the molecule i position, the total electric field produced by all the neighbors at a distance R from the molecule i (contains in the shell denoted  $S_R$ ) is  $\mathbf{F}_i(R)$ :

$$\mathbf{F}_{i}(R) = \sum_{j \in \mathcal{S}_{R}} \frac{1}{4\pi\epsilon_{0}} \frac{3\left(\mathbf{u}_{R}' \cdot \boldsymbol{\mu}\right)\mathbf{u}_{R}' - \boldsymbol{\mu}}{R^{3}} = \frac{1}{4\pi\epsilon_{0}} \int_{0}^{\pi} d\theta \int_{0}^{2\pi} d\phi \frac{3\left(\mathbf{u}_{R}' \cdot \boldsymbol{\mu}\right)\mathbf{u}_{R}' - \boldsymbol{\mu}}{R^{3}} \sin(\theta)\sigma \qquad (\text{IX.47})$$

Where we have assumed a uniform and continuous distribution of dipoles along the sphere of radius R around the molecule *i*.  $\sigma$  is the surface density and  $\mathbf{u}'_R$  the vector from the molecule *j* to the molecule *i*. In the usual spherical coordinate we have  $\mathbf{u}'_R = -\mathbf{u}_R$  so that  $\mathbf{u}_R \cdot \hat{z} = \cos(\theta)$ ,  $\mathbf{u}_R \cdot \hat{x} = \sin(\theta) \cos(\phi)$  and  $\mathbf{u}_R \cdot \hat{y} = \sin(\theta) \sin(\phi)$ .

To compute this field, let us assume that we apply an electromagnetic field **E** along the z direction. Therefore, the exciting field  $\mathbf{e}^M$  will also be in the same direction and since the polarizability in the laboratory frame  $\alpha(\omega, \omega)$  is diagonal, the induced dipole moment  $\boldsymbol{\mu}$  will also be in the z direction. Therefore:  $\mathbf{u}_R \cdot \boldsymbol{\mu} = \cos(\theta)\mu_z$ . Now we can compute the total electric field for the x, y, zdirections at the molecule *i* position.

$$\mathbf{F}_{i} \cdot \hat{x} = \frac{\sigma}{4\pi\epsilon_{0}R^{3}} \int_{0}^{\pi} d\theta \int_{0}^{2\pi} d\phi \left[3\left(\cos(\theta)\mu_{z}\right)\sin(\theta)\cos(\phi) - 0\right]\sin(\theta) \tag{IX.48}$$

$$= \frac{3\sigma}{4\pi\epsilon_0 R^3} \int_0^\pi d\theta \cos(\theta) \sin^2(\theta) \int_0^{2\pi} d\phi \cos(\phi) = 0$$
(IX.49)

The same goes for  $\mathbf{F}_i \cdot \hat{y}$ . For the direction z:

$$\mathbf{F}_{i} \cdot \hat{z} = \frac{\sigma}{4\pi\epsilon_{0}R^{3}} \int_{0}^{\pi} d\theta \int_{0}^{2\pi} d\phi \left[ 3\left(\cos(\theta)\mu_{z}\right)\cos(\theta) - \mu_{z} \right] \sin(\theta)$$
(IX.50)

$$=\frac{3\mu_z\sigma}{2\epsilon_0R^3}\int_0^{\pi}d\theta \left[3\cos^2(\theta)-1\right]\sin(\theta) = \frac{3\mu_z\sigma}{2\epsilon_0R^3}\left(\left[-\cos^3(\theta)\right]_0^{\pi}-\left[-\cos(\theta)\right]_0^{\pi}\right) = 0 \quad (\text{IX.51})$$

Therefore, the field produced by the induced dipole of all the neighbors at a distance R cancels out perfectly. We have proven that if all the molecules are orientated along the z axis, the resultant field is null. Note that for the other electromagnetic field direction (x and y) one should rotate the definition of the  $\theta$  and  $\phi$  axis to keep the same calculus. Thus, in the bulk phase, the neighbor radiation does not impact the exciting field.

## Clausius-Mossotti relation:

We have shown that:

$$\mathcal{T} = \left[1 - \frac{\langle \alpha \rangle}{3\epsilon_0}\right]^{-1} \mathbb{1}$$
(IX.52)

As a reminder, see Appendix IX. A.5, we have:

$$\mathbf{P} = \epsilon_0(\epsilon_r - 1)\mathbf{E} \quad \mathbf{P} = <\boldsymbol{\mu} > = <\alpha(\omega, \omega) \cdot \mathbf{e}^M > = <\alpha(\omega, \omega) > \cdot \mathbf{e}^M = <\alpha(\omega, \omega) > \cdot \mathcal{T} \cdot \mathbf{E} \quad (\text{IX.53})$$

so that:

$$\epsilon_0(\epsilon_r - 1) = <\alpha(\omega, \omega) > \cdot \mathcal{T} \tag{IX.54}$$

$$\rightarrow (\epsilon_r - 1) \left( 1 - \frac{\langle \alpha(\omega, \omega) \rangle}{3\epsilon_0} \right) = \frac{\langle \alpha(\omega, \omega) \rangle}{\epsilon_0}$$
(IX.55)

$$\rightarrow \epsilon_r \left( 1 - \frac{\langle \alpha(\omega, \omega) \rangle}{3\epsilon_0} \right) = 1 + \frac{2 \langle \alpha(\omega, \omega) \rangle}{3\epsilon_0}$$
(IX.56)

$$\rightarrow \epsilon_r = \frac{3\epsilon_0 + 2 < \alpha(\omega, \omega) >}{3\epsilon_0 - < \alpha(\omega, \omega) >}$$
(IX.57)

With is the Clausius-Mossotti relation presented in the main text.

# IX. A.5 Maxwell equations and plane-wave formalism

In this appendix we remind some details regarding the Maxwell equations in vacuum and in the matter.

## Maxwell equations: vacuum phase:

The Maxwell equations in the vacuum are:

$$\boldsymbol{\nabla} \cdot \mathbf{E} = 0 \qquad \qquad \boldsymbol{\nabla} \cdot \mathbf{B} = 0 \qquad (IX.58)$$

$$\nabla \wedge \mathbf{E} = -\frac{\delta}{\delta t} \mathbf{B}$$
  $\nabla \wedge \mathbf{B} = \frac{1}{c^2} \frac{\delta}{\delta t} \mathbf{E}$  (IX.59)

Calculating  $\nabla \wedge \nabla \wedge \mathbf{E}$  leads to the wave equation:

$$\Delta \mathbf{E} - \frac{1}{c^2} \frac{\delta^2}{\delta t^2} \mathbf{E} = 0 \tag{IX.60}$$

It is extremely important to note that the field used in these equations are "mesoscopic": they are somehow averaged not enough to fluctuate at the scale smaller than the wave length. Indeed, what is measured at the experimental scale is not the field at the scale of a molecule but a space average.

A solution to this equation is the use of plane waves with the vacuum dispersion relation:

$$\mathbf{E}^{\omega}(\mathbf{r},t) = \mathbf{E}_{0}^{\omega} \exp[i\mathbf{k}\cdot r - i\omega t] \qquad \qquad \omega^{2} = c^{2}|\mathbf{k}|^{2} = c^{2}k^{2} \qquad (\text{IX.61})$$

Since this equation is linear, we can decompose any electromagnetic field using plane waves that are solutions to the Maxwell equations:

$$\mathbf{E}(\mathbf{r},t) = \int_{-\infty}^{+\infty} \mathbf{E}^{\omega}(\mathbf{r},t) d\omega \qquad (IX.62)$$

Which will also be a solution of the Maxwell equation. Hence, in the following we will assume that the light can be decomposed into this plane wave basis: we will go even further and assume that the fundamental and second harmonic signal are pure plane waves. We know that it is not the case at least for the fundamental because it comes from a femtosecond laser: in the time domain it is a succession of pulses, and in the spatial domain it has a Gaussian shape. However, this approximation will make the analysis of the experimental setup simpler: once the light out of the sample (in the gas phase of the experimental room), its evolution should be well known. The detection is also simple to understand: the different frequencies are separated using the grating of the spectrometer: the intensity observed on a camera or on a photo-multiplicator is proportional to  $|\mathbf{E}^{\omega}|^2$  where the frequency  $\omega$  can be experimentally selected with high accuracy.

# Maxwell equations: dielectric and isotropic material

Hence, we are left with the most difficult part: the light propagation in the liquid phase and at interfaces.

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \qquad \qquad \boldsymbol{\nabla} \cdot \mathbf{B} = 0 \qquad (IX.63)$$

$$\nabla \wedge \mathbf{E} = -\frac{\delta}{\delta t} \mathbf{B} \qquad \nabla \wedge \mathbf{B} = \frac{\epsilon_0}{c^2} \mathbf{J} + \frac{1}{c^2} \frac{\delta}{\delta t} \mathbf{E} \qquad (IX.64)$$

Note that in our system we assume that there is no free charges or current and that the system is non-magnetic. The Gauss-Maxwell equation bears a non-zero term which will act as electric source: the charge density  $\rho$ . In our liquid phase, there is no net charge density: the fluid is composed of neutral molecules. Hence, the charge density as the current are called *bonded* because it arises due to deformation of the electronic molecular cloud.

# Constitutive relations:

With this extra-term, the wave-propagation equation is a bit more difficult to obtain than in the vacuum phase. Usually, the bonded charge is replaced by the *mesoscopic polarization*  $\mathbf{P}$  so that:

$$-\boldsymbol{\nabla} \cdot \mathbf{P}(\mathbf{r}, t) = <\rho > (\mathbf{r}, t) \tag{IX.65}$$

In the main text is presented how to perform this spatial averaging. The mesoscopic polarization is induced by the electromagnetic field itself: it is the reaction of the matter to the incoming field. Hence, we can write in a very general way:  $\mathbf{P}(\mathbf{r},t) = \mathbf{P}[\mathbf{E}](\mathbf{r},t)$ . Using a local approximation, this becomes  $\mathbf{P}(\mathbf{r},t) = \mathbf{P}[\mathbf{E}(\mathbf{r})](\mathbf{r},t)$ . If we assume a linear dependency, we can define a time "response tensor"  $R[\mathbf{r}](t-t_0)$ <sup>6</sup> for every part of space so that:

$$\mathbf{P}(\mathbf{r},t) = \int R[\mathbf{r}](t-t') \cdot \mathbf{E}(\mathbf{r},t')dt' \qquad (IX.66)$$

However, we are looking for plane waves. Hence, this relation becomes simplier <sup>7</sup>:

$$\mathbf{P}^{\omega}(\mathbf{r}) = \epsilon_0 \chi(\omega, \omega) [\mathbf{r}] \mathbf{E}^{\omega}(\mathbf{r}) \tag{IX.67}$$

In plane wave notation and at the linear order, the mesoscopic polarization is given by the product of a tensor, the linear susceptibility tensor, with the electromagnetic field. We find with this simple argument that the induced dipole moment at the linear order oscillated as the same frequency of the electromagnetic field, as in the microscopic quantum calculations. Another definition of  $\chi(\omega, \omega)$ would be to be the proportional factor, if any, which links the induced mesoscopic polarization oscillating at the same frequency as the electromagnetic field.

For the liquid bulk phase of interest, this 3x3 matrix is space-independent and diagonal. Such systems are called *simple linear dielectric*: the induced mesoscopic polarization is proportional to the electric field. In this case, an *effective permittivity*  $\epsilon_r(\omega)$  can be defined so that:

<sup>&</sup>lt;sup>6</sup>Assuming also a causal and time-invariant system

<sup>&</sup>lt;sup>7</sup>A convolution in real-time space is a product in the Fourier space

$$\mathbf{P}^{\omega} = \epsilon_0 (\epsilon_r(\omega) - 1) \mathbf{E}^{\omega} \tag{IX.68}$$

#### **Electric displacement:**

In any case, we can define the *electric displacement*  $\mathbf{D}$ :

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \qquad \qquad \mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E}, \qquad (IX.69)$$

Which can also be defined using the effective permittivity. Using the electric displacement, the Gauss-Maxwell equation becomes:

$$\boldsymbol{\nabla} \cdot \mathbf{D} = 0 \tag{IX.70}$$

Thanks to the charge conservation equation,  $\frac{\delta\rho}{\delta t} + \nabla \cdot \mathbf{J} = 0$ . Combining with the definition of the mesoscopic polarization with respect to the charge density, we have:  $\mathbf{J} = \frac{\delta}{\delta t} \mathbf{P}$ . In the liquid, the light induces oscillating current at each molecular size, often called bonded current. Using this definition, the Ampere-Maxwell equation becomes:

$$\boldsymbol{\nabla} \wedge \mathbf{B} = \frac{\epsilon_0}{c^2} \frac{\delta}{\delta t} \mathbf{P} + \frac{1}{c^2} \frac{\delta}{\delta t} \mathbf{E} = \frac{1}{c^2} \frac{\delta}{\delta t} \mathbf{D}$$
(IX.71)

Therefore, the new wave equation fulfilled by the electric displacement is:

$$\nabla^2 \mathbf{D} - \frac{\epsilon_r}{c^2} \frac{\delta^2}{\delta t} \mathbf{D} = \mathbf{0}$$
(IX.72)

Which leads to the dispersion relation:

$$k^2 = \frac{\epsilon_r \omega^2}{c^2} \tag{IX.73}$$

Hence, we are proven that the building block of the light in the condensed phases are again plane waves but with different dispersion relation. Therefore, in the following we will use the notation **E** not only for the explicit time dependent field, but also the complex amplitude in the time reciprocal space.

The optical index  $n(\omega)$  is defined to be  $n^2 = \epsilon_r$  so that  $k = n\omega/c$ . Thus, in the simple homogeneous linear dielectric media, the electromagnetic waves propagate at the spatial frequency kwhich is different from the vacuum one due to the media polarization. This difference will play an important role at interfaces because it will also impact the shape of the field.

# **Fresnel coefficients:**

When an electromagnetic field crosses an interface that separates two different optical index, the polarization and the direction of propagation are modified, see Figure IX.1. The interface we are


Figure IX.1: Scheme of a gas-liquid interface and the electromagnetic field properties. An incident field arrives with wave vector  $\mathbf{k}^{v}$ . Part is reflected with the same angle  $\theta$  and another part is transmitted with a new wave-vector direction  $\mathbf{k}^{l}$ . The P and S polarization are also defined for the gas and liquid phase.

dealing with presents no net current nor free charges. In this case, the continuity equations at a liquid-air interface are:

$$\mathbf{B}^{v} = \mathbf{B}^{l} \qquad \mathbf{z} \wedge \mathbf{E}^{v} = \mathbf{z} \wedge \mathbf{E}^{l} \qquad \mathbf{z} \cdot \mathbf{D}^{v} = \mathbf{z} \cdot \mathbf{D}^{l} \qquad (IX.74)$$

Where  $\mathbf{z}$  is the unit vector normal to the interface (here z-axis) and the upper index v and l design the value of the fields inside the vacuum or liquid phase respectively. Hence, it means that the electromagnetic field in the interface plane is conserved, and the dielectric displacement is conserved for the out-of-plane direction.

In Surface SHG, the electromagnetic field is applied with a certain angle  $\theta$  with respect to the interface normal, with the corresponding wave vector  $\mathbf{k}^{v}$ . Using the previous set of equations, along with  $\omega \mathbf{B} = \mathbf{k} \wedge \mathbf{E}$  leads to the Snell-Descartes Equations,  $\sin \theta = n \sin \theta'$ .  $\theta'$  is the angle of propagation of the transmitted light in the liquid phase, with the corresponding wave vector  $\mathbf{k}^{l}$ .

To have the electromagnetic field inside the liquid phase, we first have to define the polarization direction of the field. To do so, we define the P polarization which is in the plane formed by the wave vector and  $\mathbf{z}$ , and the S polarization out-of-this plane. Since  $\mathbf{k}^{v}$  and  $\mathbf{k}^{l}$  are not collinear, the P polarization in the gas and in the liquid are not the same while the S remains in the same direction, see Figure IX.1.

Using the previous equations  $^8$  we can obtain the Fresnel coefficient for the P and S polarization:

$$L_S = \frac{\mathbf{E}_S^l}{\mathbf{E}_S^v} = \frac{2\cos\theta}{n\cos\theta' + \cos\theta} \qquad \qquad L_P = \frac{\mathbf{E}_P^l}{\mathbf{E}_P^v} = \frac{2\cos\theta}{n\cos\theta + \cos\theta'} \qquad (IX.75)$$

Hence, the fundamental field value evolves from the gas to the liquid phase, and thus throughout the interface. Importantly, this evolution depends strongly on the fundamental polarization (the ratio between the P and the S direction) and thus should be carefully included to analyze the Surface SHG results.

 $<sup>^8\</sup>mathrm{See}$  for instance p.38 and following of Ref 82

# IX. B | Chapter IV

# IX. B.1 Appendix: More on FROG work flow

In this annex, we will describe in more details how FROG works.

# General description:

First, FROG will open the MD trajectory that should be investigated and built its core objects. Then, during the first part, FROG perform all the "structural" analysis (density, molecular orientation, H-bonds, Radial Distribution Function and electrostatic field environment). If any QM/MM calculation should be performed, it will also write the input files for DALTON. During the second part, FROG will actually not perform much calculation. It will generate submission files intended for a cluster so that the user can launch (and with many options) the many QM/MM calculations that should be performed automatically. The user can loop FROG call during this second part to perform few QM/MM calculation at the time, or all at once. Finally, once all the QM/MM calculations performed, FROG will first read the results obtained from the DALTON output files. Then, it will perform statistical averaging: in time and in space if required. The final result is a small file (from few MB to hundreds of MB) that can be stored in a personal computer.

In order to install FROG on a computer or in a cluster, one should first download the file accessible on Zenodo<sup>9</sup>. The installation should be done as a classical Python package using the "pip" command <sup>10</sup>.

In the following, we will take for example an MD trajectory made of 1200 water molecules and 800 methanol molecules at a liquid-gas interface, see for instance the MD picture at the upper right of Figure IX.2. This is the most complex system we have tried with FROG: at an interface and with several molecule types. Let us say that we want to compute the first hyperpolarizability of the water molecule using the electrostatic embedding scheme. We also want to compute the density, the orientation of the water molecule, and the number of Hydrogen bonds between the methanol and water molecule. All of these quantities shall be returned in function of the molecule altitude so that we can better understand the evolution of the molecular properties with respect to the interface.

# Reading the trajectory and initialization:

In order to work with different MD trajectory type, FROG uses the Python package MDAnalysis [108]. From this, we have the list of the "molecules" <sup>11</sup>, with the atoms coordinates at every time steps. FROG does not try to read the molecule or atomic type from the MD files: we use the input provided

 $<sup>^{9} \</sup>tt https://doi.org/10.5281/zenodo.5998193.$  If this link does not work, please contact me I will be happy to share.

 $<sup>^{10}\</sup>mathrm{Pip}$  may need to install other packages: make sure to have a net connection

<sup>&</sup>lt;sup>11</sup>The molecules are groups of atoms as defined in the MD trajectory, but this can also be a single atom (ions for instance).



Figure IX.2: Scheme of FROG workflow. FROG can use as raw material Molecular Dynamics trajectories from several software and for different geometries: pure liquid or at interfaces. During step I, FROG creates electrostatic environment for every molecule and writes the relative input files for the QM/MM calculation performed by DALTON. Then, the user has to perform the QM calculation on a cluster: FROG helps the user by generating files and by giving feedback in step II. Once all the QM calculations done, FROG reads the obtained values in the DALTON output files during the step III. The returned data contain the distributions of the computed quantities or the spatial evolution of averages.

by the user. For instance, we have to tell to FROG which molecules of the MD are water and which are methanol. This assignment built the most important object FROG called "Molecule Type" (MT). Each MT defined will have its own property and functions provided by a file in the FROG package library. If a user wants to define a new type of molecule, he has to write a new file to define this new MT. Once this is done <sup>12</sup>, FROG will try to initialize the whole procedure with respect to the analysis required.

Indeed, if we want to perform QM/MM calculation for water molecules, we have to be able to describe every molecule within the box: even the methanol ones. Hence, every MT defined should provide a way to describe at the molecule at an electrostatic point of view in order to perform an electrostatic embedding if this type of molecule is in the range. Moreover, if the MT should undergo a QM/MM calculation itself (we would say that this is a "target" MT), it should bear the information required to create the Quantum Hamiltonian (number of electrons and nuclei charges mainly). During this initialization part, FROG also check the coherence of the parameters to avoid future crashes.

This initialization takes very few times usually: about few seconds. However, during this part a lot of information will be printed in the shell in order to keep track of the information really understood by the code. We tried to make FROG talkative so that we can track errors or save the parameters used for a run within the output file itself. For new users, this may also be a good way to understand how the code works and the impact of the different parameters.

# Part I: Building the environment

Now that FROG knows what it should do, it will work in parallel with respect to the number of MD frames. For each frame, it will first compute the properties which do not need any QM calculation, as the number of molecules at a given altitude using 2D slices. Some of these structural properties require specific knowledge about the type of the molecule. For example, we have chosen that the "orientation" of a water molecule is given by the projection of its permanent dipole moment in the laboratory frame. For a methanol molecule, we use 6 components: the projection of the C-O vector and the O-H one. These parameters and definitions are provided by the library file attached to the Molecule Type (MT). We can also define inter molecular properties such as the H-bonds: we can define for each MT couple how to define a "Hydrogen bond" in the library files. Depending on the type of analysis required, this may take some time. Typically, for 2 000 molecules, each frame takes a few minutes to be analyzed regarding the structural properties.

If QM/MM calculation should be performed for an MT type, the input files needed for DALTON will be generated there. First, FROG goes over each molecule of this MT and check if they all need

 $<sup>^{12}</sup>$ FROG also checks that it recovers the right number of atoms for every molecule declared in function of its "type" and also checks if the geometries make sense. For instance, for a water molecule, FROG expects three atoms within 1.5 Åfrom one another maximum.

to undergo a Quantum Calculation. For instance, at interfaces we want to have the SHG where the structural  $C_{\infty}$  is broken so only at the interface. However, a good MD simulation of an interface is basically a large bulk phase that does not cover all the space available, see for instance the MD picture at the upper right of Figure IX.2. Hence, almost all the molecules are in the "bulk" phase of the MD trajectory. Therefore, the user can require the QM calculation to be performed for a given type of molecule, and for a given part of space. This way, the number of QM calculations is reduced, and so the computational time, so that we can focus on the most relevant part of space. For every time step, FROG assesses if a molecule fulfills the conditions required: a given molecule can be treated at the QM level for some time steps, and not for others.

For the chosen molecules, FROG builds its environment up to a given distance, see part I of Figure IX.2. Then, according to the input parameters and the information provided by the MT library files, the electrostatic environment is built. This environment is written in the ".pot" file in a directory specific to the time step and molecule number. In this directory, FROG will also write the other files required to perform the QM calculation: the ".dal" file containing the information regarding the general framework (for instance DFT and the response tensor asked) and the ".mol" file containing the information regarding the molecule inside the quantum box (the number of electrons, position of the nuclei and basis set).

At the end of this first part, all the molecules that should be treated at the QM level are referenced. The list of the time steps and molecule number is stored. If there is no need of Quantum calculation, the second part of the run is skipped.

# Part II: Performing the QM calculations

In this part FROG will help you to perform the QM calculation on a calculation cluster. Indeed, using FROG you are able to build very easily a very important set of input files, using large MD or important amount of time steps or both. However, each quantum problem is unique since we have built its own electrostatic environment according to the MD structure. To give an idea, extracting the first hyperpolarizability of one water molecule for a given configuration using a good basis set takes about 15 min on a single core.

Hence, to perform this important set of **independent** calculations, a cluster is necessary. Nowadays, we tend to buy multi-core CPU as they are more versatile: the cluster I have worked on, the "Pôle Scientifique de Modélisation Numérique" (PSMN), has mostly servers with 8, 16 or 32 cores per nodes. Hence, provided the RAM is large enough, we can perform 8, 16 or 32 QM calculations in parallel on each server.

Thus, the second part of FROG is designed to help you deal with these independent calculations: on which queuing list do you want to submit the jobs? How many QM/MM calculations per server? Using the parameters given, FROG generates the submission script during this part: the user has to

send them manually. We have added several parameters and provide advice in the wiki in order to help the user to choose how to perform these calculations. Importantly, the user can try with few (or even one) calculation at the time first.

Then, the user shall wait until the cluster performed the submitted calculations. Once this is done, the user has to launch FROG again. In this case, FROG will check among the QM/MM calculations that should be performed many "succeeded". Indeed, it happens that the QM/MM calculation crashes. One of the reasons can be a RAM overflow (for instance if we send too many calculations on a single server) or a problem with respect to the temporary file accessibility FROG considers that the calculation is "done" if DALTON has reached the end of its procedure.

At the end of this FROG call, the user will have a new set of submission files if there are any calculations not completed yet. Note that between each FROG-run, the user can modify the parameter regarding the submission procedure on the cluster. If FROG considers that all of the QM/MM calculations have been done, it continues to the third and last part.

### Part III: Read the DALTON outputs and averaging

In this last part, FROG will read the results obtained by DALTON and perform statistical averaging. This part run in parallel with respect to the MD frame, as in the first part.

Following the example, it means that for every water molecules that undergo a QM/MM calculation, FROG opens the result file (the .out in part III of Figure IX.2) and reads the first hyperpolarizability tensor at the required frequencies: we can for instance compute at the same time the response at 800 nm and in the static limit. Then, FROG performs a spatial averaging for each frame according to one of these options:

- No discretization: for bulk phases
  - All the molecule properties are treated without any consideration of their position. The distribution of all values, averages and standard deviations are computed using each molecule in the frame for each molecule types.
- Axis-discretization: for interfaces

The user selects one axis of discretization and a number of bins. In this case, the distribution of all values, averages and standard deviations are computed in function of the molecular position with respect to this axis <sup>13</sup>. For instance the z-axis and 100 bin: the MD box for each frame is decomposed into 100 XY-slice <sup>14</sup>. 100 distribution, averaged and standard deviation are computed for each frame corresponding to these 100 slices.

• Layer decomposition: for interfaces

Thank to the Python package Pytim [109], FROG can perform layer-like attribution close to an

 $<sup>^{13}</sup>$ In FROG, the "position" of a molecule is defined within respect to each Molecular Types. This can be defined by the user in the library file if needed. For instance, the "position" of a methanol molecule has been defined to be the one of its Carbon atom.

<sup>&</sup>lt;sup>14</sup>Hence for NPT MD simulations, the assignment may evolve as the MD box size evolves.

interface. Figure IX.3 shows an example for water molecules at the liquid-gas interface: the density of each "layer" are presented in function of the altitude. This geometrical assignation has two advantages. First it allows us to have a robust way to choose which molecules should be computed at the QM/MM level: it would make more sense to study "the response of the first 4 layers of molecules" than "the last 1 nanometer". The second advantage is that it can help us to gain physical insight on the evolution of the properties in function of the position of the molecules. Indeed, as shown in Figure IX.3, the layers of molecules are mixed with respect to the absolute altitude. Hence, in order to have a deep understanding of the molecular property evolution and try to show correlations between different properties this layer assignation is of prime interest.

Once the spatial averaging is performed for all the frame, FROG will merge the results as average over the frames. This last part is s extremely fast and do not need to be parallelized over the frames.

#### Input and Output files:

There are two input files that should be tuned by the user before each run. The main parameter file contains the type of MD trajectory to work with, the composition of the MD and the assignment to the Molecule Types, the list of the structural and optical analysis that should be performed, the requirement in terms of spatial and temporal discretization and the QM/MM parame-If QM/MM calculations are needed, ters. the user shall also provide a "submission template file" to generate the submission script for the QM/MM run on the cluster.

FROG returns several output files, which are all none-human-readable. These pickle files can be



Figure IX.3: Layer assignment of a liquid-gas interface of neat water. The black dashed line is the usual density with respect to the altitude. The layer 1, 2, 3 and 4 are the density of the first, second, third and last layer of molecules when computing from the bulk phase. Each layer spread over few angstrom: this is typically the size occupied by a single water molecule as expected.

read by Python, and we provide several Jupyter Notebook files in the tutorials to help the user to work with them. In short, there is one main output file which holds the main results: the averaged performed over all the time steps. The user can also load the molecular results for every time step quite easily to perform more complex analysis with the raw data. It should be noted that all the parameters used for any FROG run are accessible in this result file: the user can always know how these results were obtained.

### Strengths:

- All the parameters can be modified in main input files. For advanced users, working with the library file by defining the Molecular Type should provide all the liberty needed to define new analysis easily, and to try out many definitions without modifying the main code.
- We have left some rooms for an increase of the electrostatic embedding scheme. Today we have extensively tested PE0 types with point charges. However, we have already coded dipoles and quadrupole electrostatic description for neighbors, and also the polarizable part (PE1). The calculations run, but no real physical test has been performed. Yet, an important part of the code is ready to perform QM/MM calculation with several molecules inside the QM box. However, we have not implemented how to attribute to each molecule its own property once the QM/MM calculation performed on the group of molecules.
- FROG runs in parallel with respect to the number of frames for the first and third parts. Then, the more time-consuming part of FROG is the creation of the electrostatic environment.
- FROG helps you to manage the numerous independent QM/MM calculations you have to perform on the cluster. This part may, of course, be more complete and efficient. However, I have chosen to keep it simple so that a user with few experiences with clusters can really benefit from it <sup>15</sup>.
- FROG proposes native spatial discretization with can be very useful to get first insight on the results. If the user wants more details or develop personal analysis of the molecular data, every molecule's properties at every time step are accessible in the output files.

# **Perspectives:**

FROG is a very young software and some improvement may be done in the future on efficiency, automatization and on functionalities. Here are some ideas:

- FROG does not run in parallel with respect to one frame, *i.e.* not parallelized in space. This can be easily done for many analyses of part I which do not need the knowledge of all the molecule positions. This can also be done with great potential improvement for the part III. We can also parallelize part II, but this part is really light so no great numerical-time gain is expected.
- FROG is slow for the electrostatic environment part, especially for trajectories with an important number of molecules.
- FROG has been written by a none-professional coder. Many sub-functions are home-made, and many procedures are not perfectly efficient numerically speaking. I have focus on writing a first version of the code that is robust, readable by others and can be modulated in the future. Hence, this code would benefit from a rewriting of some part to increase the numerical efficiency using state-of-the-art modules or algorithm.

<sup>&</sup>lt;sup>15</sup>Cluster-veterans will know how to help themselves!

• FROG is a very young software. Even if we have tried to test a lot the software, we have not implemented robust testing routine. Today, to check if a new feature or modification in the code does not impact other part of FROG, I run the tutorials and check the results. In the long-term run, we should have automatic tests: especially if new coders joint the project.

# IX. C | Chapter VI

# IX. C.1 How to get the electric field shined by a dipole at far distance: comparison between 2 formulas

In this part I would like to clarify the use of two formulas that I found either in the "numerical" or "experimental" community. Here, we would like to know what is the electromagnetic field shine in a particular direction, for instance toward a detector placed at a position  $\mathbf{D}$ , by an oscillating dipole. For the two cases, we will compute the results in the static limit and by assuming that the detector  $\mathbf{D}$  is placed very far away from the emitter.

# First formula: electromagnetic fields generated by a dipole

Be a dipole n,  $\mathbf{P}^{n}$ , at the position  $\mathbf{r}^{n}$ . The electromagnetic field shine at a position  $\mathbf{r}^{m}$ ,  $\mathbf{E}^{n,m}$ , is in the static limit:

$$\mathbf{E}^{n,m} = T^{nm} \cdot \mathbf{P}^n \qquad \qquad T^{nm}|_{kl} = \left[ 3\mathbf{r}_k^{nm} \ \mathbf{r}_l^{nm} - |\mathbf{r}^{nm}|^2 \delta_{kl} \right] \ |\mathbf{r}^{nm}|^{-5}$$

Where  $\mathbf{r}^{nm} = \mathbf{r}^m - \mathbf{r}^n$ .

Let us assume that the position  $\mathbf{r}^m$  is the detector one: **D**. **D** is far away for the position  $\mathbf{r}^n$ . So:  $\mathbf{r}^{nD} = \mathbf{D} - \mathbf{r}^n \simeq \mathbf{D}$ . Then,

$$T_{kl}^{nD} = T_{kl}^{D} = \begin{bmatrix} 3\mathbf{D}_k \ \mathbf{D}_l - |\mathbf{D}|^2 \delta_{kl} \end{bmatrix} \ |\mathbf{D}|^{-5}$$
$$\mathbf{E}^{n,m} = \mathbf{E}^{n,D} = T^D \cdot \mathbf{P}^n$$
$$\mathbf{E}^{n,D} = \begin{bmatrix} 3 \left( \mathbf{D} \cdot \mathbf{P}^n \right) \mathbf{D} - |\mathbf{D}|^2 \mathbf{P}^n \end{bmatrix} |\mathbf{D}|^{-5}$$

### Second formula:

In the experimental team, we usually rely on this formula:

$$\mathbf{E}^{n,D} \propto [(\mathbf{D} \wedge \mathbf{P}^n) \wedge \mathbf{D}]$$

Let us develop:

$$\begin{split} \mathbf{E}_{k}^{n,D} \propto \epsilon_{kij} \left( \mathbf{D} \wedge \mathbf{P}^{n} \right)_{i} \ \mathbf{D}_{j} \propto \epsilon_{kij} \left( \epsilon_{ilm} \mathbf{D}_{l} \ \mathbf{P}_{m}^{n} \right) \mathbf{D}_{j} \\ \propto -\epsilon_{ikj} \epsilon_{ilm} \mathbf{D}_{l} \ \mathbf{P}_{m}^{n} \ \mathbf{D}_{j} \ \propto - \left( \delta_{kl} \delta_{jm} - \delta_{km} \delta_{jl} \right) \mathbf{D}_{l} \ \mathbf{P}_{m}^{n} \ \mathbf{D}_{j} \\ \propto - \left( \mathbf{D}_{k} \ \mathbf{P}_{m}^{n} \ \mathbf{D}_{m} - \mathbf{D}_{j} \ \mathbf{P}_{k}^{n} \ \mathbf{D}_{j} \right) \\ \propto \mathbf{D}_{j} \ \mathbf{D}_{j} \ \mathbf{P}_{k}^{n} - \mathbf{P}_{m}^{n} \ \mathbf{D}_{m} \ \mathbf{D}_{k} \\ \mathbf{E}^{n,D} \propto |\mathbf{D}|^{2} \mathbf{P}^{n} - (\mathbf{D} \cdot \mathbf{P}^{n}) \mathbf{D} \end{split}$$

These formulas are different in general, but we just want the polarization normal to the electric

field propagation:  $\mathbf{E}^{n,D} \cdot \mathbf{D}^{\perp 16}$ . Therefore, terms in the direction of  $\mathbf{D}$  can be neglected, and thus the 2 formulas become:

$$\mathbf{E}^{n,D} \cdot \mathbf{D}^{\perp} = -|\mathbf{D}|^{-3} \mathbf{P}^n \cdot \mathbf{D}^{\perp} \qquad \qquad \mathbf{E}^{n,D} \cdot \mathbf{D}^{\perp} \propto |\mathbf{D}|^2 \mathbf{P}^n \cdot \mathbf{D}^{\perp} \qquad (IX.76)$$

Therefore, the two formulas are equivalent up to a phase factor.

<sup>&</sup>lt;sup>16</sup>As a reminder:  $\mathbf{D} \cdot \mathbf{D}^{\perp} = 0$ .

# IX. C.2 How to compute $\langle \mathscr{B}^2 \rangle$ for a given $\beta$ ?

In this appendix, I present in detail the link between the first hyperpolarizability in the molecular,  $\beta$ , and in the laboratory,  $\mathscr{B}$ , frames. We will see how to build the average value of  $\mathscr{B}$  using distribution for  $\beta$  with only one component, than for several components. The aim of this appendix is to give to the reader more insight about the  $\mathscr{B}$  behavior in the liquid phase: how does the centro-symmetry impact this laboratory hyperpolarizability for a given molecular  $\beta$  response?

#### Some important elements to have in mind:

First, let us remind some basics about the different objects.

The  $\beta$  is a 3x3x3 matrix expressed in the molecular frame. In this frame,  $\beta$  shall respect the molecule symmetry. As water has a  $C_{2v}$  symmetry: only the *ccc*, *caa* and *cbb* components can be non-zero (plus their permutations) <sup>17</sup>.

To link the molecular and laboratory frame, we can use the rotational matrix R of the molecule. Hence, the hyperpolarizability in the laboratory frame can be expressed as:

$$\mathscr{B}_{ijk} = \sum_{abc} R_{ia} R_{jb} R_{kc} \beta_{abc} \tag{IX.77}$$

The incoherent HRS signal is given by the averaged of  $\mathscr{B}^2$  over all the molecule configurations: both its orientation in the laboratory frame **and** its  $\beta$  value. In this appendix we focus on the orientation averaging, and in the main text, Section V. B on the effect of the  $\beta$  fluctuations.

We know that the molecule is in the liquid phase: it has not net orientation in the laboratory frame. There are several consequences on the  $\mathscr{B}$  tensor. First, it has no net average for any component:  $\langle \mathscr{B} \rangle = 0$ . Second, this tensor is invariant over laboratory axis renaming. For instance:  $\langle \mathscr{B}^2 \rangle_{xxx} = \langle \mathscr{B}^2 \rangle_{yyy} = \langle \mathscr{B}^2 \rangle_{zzz}, \langle \mathscr{B}^2 \rangle_{zxx} = \langle \mathscr{B}^2 \rangle_{zyy} = \langle \mathscr{B}^2 \rangle_{xyy} = \cdots$  The physical argument is that the laboratory frame is somehow ill-defined: we could rename the x axis the y, rotate the frame or move it in any direction. The system is still an infinite homogeneous bulk liquid: the way we define our laboratory frame should not matter. Finally, we have a way to compute the  $\mathscr{B}$  distribution assuming a constant  $\beta$  tensor.

### **Rotational averaging: Euler matrix**

To do so, we can use an Euler formalism, see for instance [78] Let us define the molecular frame in the laboratory one using 3 angles  $\theta$ ,  $\psi$  and  $\phi$ .  $\theta$  is the angle projecting the *c* molecular axis to the *z* one and is defined within  $[0, \pi]$ .  $\psi$  and  $\phi$  are the 2 other angles defined within  $[0, 2\pi]$ . Using these definitions, the rotational matrix is given by:

 $<sup>^{17}</sup>$ In the liquid state, this symmetry is broken due to the environment: this property is true only for the averaged values.

$$R(\theta, \psi, \phi) = \begin{pmatrix} \cos(\psi)\cos(\phi) - \cos(\theta)\sin(\psi)\sin(\phi) & \sin(\psi)\cos(\phi) + \cos(\theta)\cos(\psi)\sin(\phi) & \sin(\theta)\sin(\phi) \\ -\cos(\psi)\sin(\phi) - \cos(\theta)\sin(\psi)\cos(\phi) & -\sin(\psi)\sin(\phi) + \cos(\theta)\cos(\psi)\cos(\phi) & \sin(\theta)\sin(\phi) \\ & \sin(\theta)\sin(\psi) & -\sin(\theta)\cos(\psi) & \cos(\theta) \\ & & (IX.78) \end{cases}$$

Where  $R_{ia}$  defined in Equation IX.77 is the *i*-column and *a*-line.

For a random orientation of the molecule in the laboratory frame, the molecular frame explore every  $\theta$ ,  $\psi$  and  $\phi$  angles with a associated probability  $\sin(\theta)$ . For instance, to compute the average of  $\mathscr{B}$  assuming a **constant**  $\beta$  tensor:

$$<\mathscr{B}>_{ijk} = \frac{1}{8\pi^2} \int_0^{\pi} d\theta \sin(\theta) \int_0^{2\pi} d\psi \int_0^{2\pi} d\phi \left( \sum_{ijk} R(\theta, \psi, \phi)_{ia} R(\theta, \psi, \phi)_{jb} R(\theta, \psi, \phi)_{kc} \beta_{abc} \right)$$
(IX.79)

$$=\frac{1}{8\pi^2}\sum_{ijk}\left(\int_0^{\pi}d\theta\sin(\theta)\int_0^{2\pi}d\psi\int_0^{2\pi}d\phi R(\theta,\psi,\phi)_{ia}R(\theta,\psi,\phi)_{jb}R(\theta,\psi,\phi)_{kc}\right)\beta_{abc} \quad (\text{IX.80})$$

$$\langle \mathscr{B} \rangle_{ijk} = \frac{1}{8\pi^2} \sum_{ijk} 0 \times \beta_{abc}$$
(IX.81)

Which is zero when performing the calculation, as the centro-symmetry imposes, for any tensor  $\beta_{abc}$ .

#### Rotational averaging: $\mathcal{G}$

Yet, it may be interesting to study not only the average value in the laboratory frame, but the tensorial distribution of  $\mathscr{B}$ , noted  $\mathcal{D}[\mathscr{B}]$ . This tensorial distribution possesses 3x3x3 distributions defined in  $\mathbb{R}$ .  $\mathcal{D}[\mathscr{B}]_{ijk}(y)$  is the probability of finding the ijk component of the first hyperpolarizability in the laboratory frame equal to y. Note that due to the centro-symmetry, this probability must be the same as finding the value at -y:  $\mathcal{D}[\mathscr{B}]_{ijk}(y) = \mathcal{D}[\mathscr{B}]_{ijk}(-y)$ .

We would like to construct these distributions not for a particular set of  $\beta$ , but for any. Hence, one way to do it is to define a Green-like tensorial distribution  $\mathcal{G}$ .  $\mathcal{G}_{ijk,abc}$  is the  $\mathcal{D}[\mathscr{B}]_{ijk}$  distribution obtained using  $\beta_{abc} = 1$  and all other  $\beta$  component equal to zero. While this definition is quite simple, the analytical calculation is more difficult than the averaged proposed previously.

Let us take an example:  $\mathcal{G}_{zzz,aaa}$ . This is the distribution of  $\mathscr{B}_{zzz}$  over all the possible molecular orientation for a  $\beta_{aaa} = 1$ . If a molecule orientation is given by the  $(\theta, \psi, \phi)$  angles, we have its  $\mathscr{B}_{zzz}$  by combining Equation IX.77 with Equation IX.78:  $\mathscr{B}_{zzz} = (\sin(\theta)\sin(\psi))^3$ . Hence, one can give a formal equation for  $\mathcal{G}_{zzz,aaa}(y)$ :

$$\mathcal{G}_{zzz,aaa}(y) = \frac{1}{8\pi^2} \int_0^\pi d\theta \sin(\theta) \int_0^{2\pi} d\psi \int_0^{2\pi} d\phi \,\,\delta\left[y - (\sin(\theta)\sin(\psi))^3\right] \tag{IX.82}$$

Where  $\delta$  is the Dirac function. The integrals aim to average over all the possible orientation of the molecule in the laboratory frame.

To solve this problem, we have decided to use a numerical approach. We discretized the possible Euler sphere using many points and compute this distribution iteratively. In other words, we compute the value of  $R(\theta, \psi, \phi)_{ia}R(\theta, \psi, \phi)_{jb}R(\theta, \psi, \phi)_{kc}$  for all (ijk, abc) couples and over many  $(\theta, \psi, \phi)$  angles. When we found a value between y and y + dy for a particular (ijk, abc) couple, we increase the value of the distribution  $\mathcal{G}_{ijk,abc}(y)$  by  $\sin(\theta)$ .

### $\mathcal{G}$ results

Figure IX.4, IX.5 and IX.6 present some of the  $\mathcal{G}$  distributions using this Euler-based method. As expected, these curves are even due to centro-symmetry. In Figure IX.4 several  $\beta$  components effects on the  $\mathscr{B}_{zzz}$  distribution are tested:  $\beta_{aaa}$ ,  $\beta_{bbb}$  and  $\beta_{ccc}$ . All these distributions are in top of each other as expected due to centro-symmetry.

Figure IX.5 and IX.6 present different typical distribution depending on the  $\beta$  component for the  $\mathscr{B}_{zzz}$  and  $\mathscr{B}_{zxx}$  component respectively. All of these curves are extremely picked near zero and are continuous. Moreover, there are monotonous over [-1, 0] and [0, 1]: the probability of finding a large  $|\mathscr{B}|$  is smaller than finding small  $|\mathscr{B}|$  for any ijk, abc couple.



Figure IX.4:  $\mathcal{G}_{zzz,abc}$  distribution for abc=aaa, bbb and ccc using the Euler sampling method. All the curves are in top of each other.

To have a better idea of the impact of these curves on the experimental measurement, which is  $\langle \mathscr{B}^2 \rangle$ , let us compute  $\langle \mathcal{G}(y) \times y^2 \rangle$ . The results are presented in Table IX.1 for some ijk, abc couple. For instance,  $\langle \mathcal{G}(y) \times y^2 \rangle_{zzzaaa} = 0.14$  means that the value obtained for  $\langle \mathscr{B}^2 \rangle_{zzz}$  using  $\beta_{aaa} = 1$ , and the rest of the  $\beta$  component are zero, is 0.14.

We recover that the major contribution in this formalism to the  $\langle \mathscr{B}^2 \rangle_{zzz}$  is indeed due to the  $\beta_{ccc}$  components <sup>18</sup> and that the  $\langle \mathscr{B}^2 \rangle_{zxx}$  to the  $\beta_{cbb}$ . If we compute the depolarization ratio D

<sup>&</sup>lt;sup>18</sup>Or  $\beta_{aaa}$  or  $\beta_{bbb}$  due to centro-symmetry. For simplicity we will use  $\beta_{ccc}$  to keep close to the water case.



Figure IX.5:  $\mathcal{G}_{zzz,abc}$  distribution for abc=ccc, cbb and cba using the Euler sampling method.



Figure IX.6:  $\mathcal{G}_{zxx,abc}$  distribution for abc=ccc, cbb and cba using the Euler sampling method.

for molecular hyperpolarizability composed of one component  $\beta_{ccc} = 1$ , we found  $D = \langle \mathscr{B}^2 \rangle_{zxx}$ /  $\langle \mathscr{B}^2 \rangle_{zzz} = 0.2$  as expected.

# Comparison explicit Vs analytical sampling:

We have also computed these  $\mathcal{G}$  functions using FROG. Figure IX.7 and IX.8 presents the  $\mathcal{G}_{zyxaaa}$  and  $\mathcal{G}_{zzzaaa}$  respectively obtained by the Euler-based method and the one by FROG. Both should represent the same quantity: the distribution of  $\mathcal{B}$  using  $\beta_{aaa} = 1$  and the other component are zero. This has been done in order to gain confidence on our methods and on the explicit sampling using by FROG.

Using the Euler method, the sampling is made using the 3 angles  $\theta$ ,  $\psi$  and  $\phi$  and their associ-

$\mathcal{G}_{ijk,abc}$	$< \mathcal{G}(y)  imes y^2 >$
zzz,aaa	0.14
zzz,bbb	0.15
zzz,ccc	0.14
zzz,ccc	0.14
$_{\rm zzz,cbb}$	0.03
zzz,cba	0.01
zxx,ccc	0.03
zxx,cbb	0.09
zxx,cba	0.03

Table IX.1: Averaged value of  $\mathcal{G}(y) \times y^2$  for some  $\mathcal{G}$  components. The higher the results, the more the relative  $\beta_{abc}$  component would contribute to the  $\langle \mathscr{B}^2 \rangle_{ijk}$  value.

ated rotational matrix, see Equation IX.78, and by exploring all the possible angle value. About 500 million angle couples are sampled over the Euler sphere.

Using FROG, an MD simulation of bulk water provides the orientation of the molecule in the laboratory frame. The projection of the  $\beta$  into the laboratory frame is not done using the Euler-based rotational matrix: the molecular frame is directly written in terms of the laboratory one to obtain the rotational matrix. This procedure is more numerically demanding: it requires an MD simulations and more post-treatment than the Euler-based method. About 1.5 million configurations are used for the FROG sampling.

Both methods return the same  $\mathcal{G}$  for all the ijk, abc tried. Note that the distributions are not normed in this representation because the Euler and FROG method do not use exactly the same discretization.



Figure IX.7:  $\mathcal{G}_{zyxaaa}$  distribution obtained for the FROG and Euler method. The distributions are not normed in this representation because the Euler and FROG methods do not use exactly the same discretization.



Figure IX.8:  $\mathcal{G}_{zzzaaa}$  distribution obtained for the FROG and Euler method. The distributions are not normed in this representation because the Euler and FROG methods do not use exactly the same discretization.

#### How to use $\mathcal{G}$ for complex $\beta$ ?

Yet, we are interested in molecules with several non-zero  $\beta$  components: all the components may be non-zero and do not respect any geometrical symmetry. For instance, using  $\beta$  obtained by our QM/MM results, all the  $\beta$  components are non-zero.

Then, we can wonder how we can use the formalism defined previously to tackle this problem: what is the distribution of  $\mathscr{B}$ , for a particular  $\beta$  with several non-zero components? Looking back at the previous equation, the formal answer is obvious for this distribution noted  $\mathcal{D}[\mathscr{B}|\beta]$ :

$$Y_{ijk}[\beta](\theta,\psi,\phi) = \sum_{ijk} R_{ia}(\theta,\psi,\phi) R_{jb}(\theta,\psi,\phi) R_{kc}(\theta,\psi,\phi) \beta_{abc}$$
(IX.83)

$$\mathcal{D}[\mathscr{B}|\beta]_{ijk}(y) = \frac{1}{8\pi^2} \int_0^\pi d\theta \sin(\theta) \int_0^{2\pi} d\psi \int_0^{2\pi} d\phi \,\,\delta\left[y - Y_{ijk}[\beta](\theta,\psi,\phi)\right] \tag{IX.84}$$

However, using the previous  $\mathcal{G}$  to compute this quantity is numerically difficult. For the sake of example, let us use 3 components for  $\beta$ : *ccc*, *caa* and *cba*. Using our previous formalism:

$$\mathcal{D}[\mathscr{B}|\beta]_{ijk}(y) = \int_{-1}^{1} \mathcal{G}_{ijk,ccc}(y_1) dy_1 \int_{-1}^{1} \mathcal{G}_{ijk,caa}(y_2) dy_2 \int_{-1}^{1} \mathcal{G}_{ijk,cba}(y_3) dy_3 \ \delta\left(y - [y_1 \times \beta_{ccc} + y_2 \times \beta_{caa} + y_3 \times \beta_{cba}]\right)$$
(IX.85)

Which is doable for few components, but the complexity is extremely high for many  $\beta$  components <sup>19</sup>. The reasons is that the first hyperpolarizability value in the laboratory frame is obtained by the sum of all the  $\beta$  components once reoriented in the laboratory frame. Thus, the different component in this frame can be positive or negative: the total value is not obvious to predict. It seems that the

<sup>&</sup>lt;sup>19</sup>If N points are used to discretize a  $\mathcal{G}_{ijk,abc}$  distribution, the complexity is  $N^{18}$ , or  $N^{10}$  with the Kleinman symmetry.

way to arrange the  $\mathcal{G}$  to get the  $\mathcal{D}[\mathscr{B}|\beta]$  depends on the  $\beta$  tensor in a very complex way. Therefore, to overcome this challenge, we decided to still use the Euler-based method to obtain the  $\mathcal{D}[\mathscr{B}|\beta]$ , but without using the previously calculated  $\mathcal{G}$ . To do so, we use the Equation IX.84 for a given  $\beta$  and sample the Euler sphere directly.

The  $\beta$  used is:  $\beta_{caa} = -1.9$ ,  $\beta_{aca} = -2.0$ ,  $\beta_{cbb} = 2.7$ ,  $\beta_{bcb} = 2.3$ ,  $\beta_{ccc} = 4.0^{-20}$ . Figure IX.9 and IX.10 present the obtained  $\mathcal{D}[\mathscr{B}|\beta]_{zyx}$  and  $\mathcal{D}[\mathscr{B}|\beta]_{zzz}$  respectively. We have performed these calculations using the analytical (Euler) or explicit (FROG) sampling, both returns the same results. These distributions are quite counterintuitive: they are non-monotonous and seem almost non-continuous.



Figure IX.9:  $\mathcal{D}[\mathscr{B}|\beta]_{zyx}$  distribution obtained for the FROG and Euler method for a particular  $\beta$ , see the main text. The distributions are not normed in this representation because the Euler and FROG methods do not use exactly the same discretization.

This confirms that the impact of a complex  $\beta$ , with several non-zero components with different signs, on the average  $\mathscr{B}$  is not trivial.

# **Conclusions:**

To conclude this appendix, we have seen that:

- The  $\mathscr{B}$  distribution obtained for a single  $\beta$  component,  $\mathcal{G}$ , is non-trivial. Some components of  $\beta$  seem to have more impact than others on the final  $\langle \mathscr{B}^2 \rangle$  value, see Table IX.1.
- As for the  $\mathcal{G}$ , no  $\mathcal{D}[\mathscr{B}|\beta]$  analytical solution has been proposed: only numerical resolution for a particular  $\beta$  has been done so far. The results obtained using an Euler-based formalism and explicit calculation using FROG matches.
- When using complex  $\beta$  component, the  $\mathcal{D}[\mathscr{B}|\beta]$  is extremely peaky and even non-continuous. The impact on the different  $\beta$  component in  $\mathcal{D}[\mathscr{B}|\beta]$  is non-trivial.

 $<sup>^{20}\</sup>beta_{aac} = \beta_{aca}$  and  $\beta_{bbc} = \beta_{bcb}$ .



Figure IX.10:  $\mathcal{D}[\mathscr{B}|\beta]_{zzz}$  distribution obtained for the FROG and Euler method for a particular  $\beta$ , see the main text. The distributions are not normed in this representation because the Euler and the FROG methods do not use exactly the same discretization.

For all of these reasons, we have proposed in the main text the calculation of the  $\langle \mathscr{B}^2 \rangle$  directly instead of the distributions  $\mathcal{D}[\mathscr{B}|\beta]$ . Yet, numerically speaking, we have the  $\mathcal{D}[\mathscr{B}]$  using the QM/MM results and FROG, *i.e.* using an explicit sampling over the molecular orientation and  $\beta$  values. The results is presented in Figure V.11 and remind in Figure IX.11.



Figure IX.11: First hyperpolarizability distribution in the laboratory frame for the water in the liquid phase at the QM/MM level in the zero-frequency limit. The distribution is not normalized: a total of 156 250 molecules have been used. This distribution takes into account correlation between  $\beta$  component if any.

If we compare the distribution for the  $\mathscr{B}_{zzz}$  component obtained for a single  $\beta$  (the average one) in Figure IX.10 and the one for all the  $\beta$  fluctuations in Figure IX.11, we observe that: (i)

including the  $\beta$  fluctuations lead to continuous and monotonous distribution and (ii) the width of the distribution with the  $\beta$  fluctuation is larger. This second point is confirmed by the calculation of  $\langle \mathscr{B}^2 \rangle$  in Section V. C, see Table V.2.

# IX. C.3 Are the $\beta$ component independent?

In the main text, we assumed that the different component of the first hyperpolarizability in the molecular frame can be assumed independent if they are not related by Kleinman symmetry. We found that such approximation is not true because the individual SHG intensity obtained is not the same with and without this approximation.

To go further, we have plotted the correlation between some  $\beta$  component, shown in Figure IX.12, IX.13 and IX.14. The  $\beta$  of 156 250 water molecules at the QM/MM level in the bulk phase have been computed for an exciting frequency of 800 nm. Figure IX.12 shows the correlation between the component  $\beta_{bbb}$  and  $\beta_{bba}$ : in this case there is no correlation since the correlated probability is spherical. This is not the case between the  $\beta_{ccc}$  and  $\beta_{caa}$  components: for instance it is more likely to have large  $\beta_{caa}$  if  $\beta_{ccc}$  is also large. Hence, some hyperpolarizability components are indeed correlated.



Figure IX.12: Correlation distribution of  $\beta_{bbb}$  and  $\beta_{bba}$ . On the right the distribution of  $\beta_{bbb}$ . On the top the distribution of  $\beta_{bba}$ . The total number of molecules used is 156 250. The correlation plot shows no correlation between  $\beta_{bbb}$  and  $\beta_{bba}$ .

To conclude this appendix, we have also plotted the correlation between the  $\beta_{cbb}$  and  $\beta_{bcb}$  components in Figure IX.14. Since the response calculation have been made at 800 nm, these quantities are not strictly equal, but extremely correlated. That is why in the main text we have assumed that the  $\beta$  components equal within the Kleinman approximation (*i.e.* the couple *ijk* and *abc* are linked by permutation) should be considerate fully dependent:  $P(\beta_{ijk} = Y \text{ and } \beta_{abc} \neq Y) = 0$ .



Figure IX.13: Correlation distribution of  $\beta_{ccc}$  and  $\beta_{caa}$ . On the right the distribution of  $\beta_{caa}$ . On the top the distribution of  $\beta_{ccc}$ . The total number of molecules used is 156 250. The correlation plot shows a correlation between  $\beta_{ccc}$  and  $\beta_{caa}$ .



Figure IX.14: Correlation distribution of  $\beta_{cbb}$  and  $\beta_{bcb}$ . On the right the distribution of  $\beta_{bcb}$ . On the top the distribution of  $\beta_{cbb}$ . The total number of molecules used is 156 250. The correlation plot shows a very strong correlation between  $\beta_{cbb}$  and  $\beta_{bcb}$ .

# IX. C.4 Y study

In this part we will see some property for the Y tensor

# Real and Imaginary part:

Let us see some interesting property regarding the real and imaginary part of  $Y_{ijk,abc}$ :

$$Y_{ijk,abc} = \sum_{n,m} \mathscr{B}^n_{ijk} \mathscr{B}^m_{abc} \exp\left[i\Delta \mathbf{q} \cdot (\mathbf{r}_n - \mathbf{r}_m)\right]$$
(IX.86)

$$Y_{abc,ijk} = \sum_{n,m} \mathscr{B}^m_{ijk} \mathscr{B}^n_{abc} \exp\left[i\Delta \mathbf{q} \cdot (\mathbf{r}_m - \mathbf{r}_n)\right]$$
(IX.87)

$$Y_{abc,ijk} = \sum_{n,m} \mathscr{B}^n_{ijk} \mathscr{B}^m_{abc} \exp\left[-i\Delta \mathbf{q} \cdot (\mathbf{r}_n - \mathbf{r}_m)\right]$$
(IX.88)

$$Y_{abc,ijk} = \sum_{n,m} \mathscr{B}^n_{ijk} \mathscr{B}^m_{abc} \exp\left[i\Delta \mathbf{q} \cdot (\mathbf{r}_n - \mathbf{r}_m)\right]^*$$
(IX.89)

$$Y_{abc,ijk} = \left(\sum_{n,m} \mathscr{B}^n_{ijk} \mathscr{B}^m_{abc} \exp\left[i\Delta \mathbf{q} \cdot (\mathbf{r}_n - \mathbf{r}_m)\right]\right)^*$$
(IX.90)

$$Y_{abc,ijk} = Y^*_{ijk,abc} \tag{IX.91}$$

Note that this property has been used to obtain the equation on the SHS intensity.

# $Y_{ijk,abc}^{\mathbf{incoh}}$ :

First the 'incoherent' term  $Y_{ijk,abc}^{\text{incoh}}$ . Two cases should be considerate: ijk = abc, and  $ijk \neq abc$ ,

$$Y_{ijk,ijk}^{\text{incoh}} = \sum_{n} (\mathscr{B}_{ijk}^{n})^2 \neq 0$$
(IX.92)

In this case, the  $Y_{ijk,ijk}^{\text{incoh}}$  can be non-zero. For the second terms, this is less obvious. We observe numerically that:

$$ijk \neq abc$$
 with permutation,  $Y_{ijk,abc}^{\text{incoh}} = \sum_{n} \mathscr{B}_{ijk}^{n} \mathscr{B}_{abc}^{n} = 0$  (IX.93)

(IX.94)

If we consider that  $\mathscr{B}_{ijk}^n$  is independent, this makes sense because  $\langle \mathscr{B} \rangle$  is always null. However, as shown in the previous Appendix IX. C.3, some components are not completely independent. Hence, I have not been able to prove it.

 $Y^{{\operatorname{coh}}}_{ijk,abc}$ :

If every  $\mathscr{B}^n$  are independent from the  $\mathscr{B}^m$  of another molecule, then  $P(\mathscr{B}^m_{abc}|\mathscr{B}^n_{ijk}) = P(\mathscr{B}^m_{abc})$ :

$$Y_{ijk,abc}^{\text{coh}} = \sum_{n} \mathscr{B}_{ijk}^{n} \left\{ \sum_{m \neq n} \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot (\mathbf{r}_{n} - \mathbf{r}_{m})\right] \right\}$$
(IX.95)

$$\propto \int P(\mathscr{B}_{ijk}^{n}) \mathscr{B}_{ijk}^{n} \int P(\mathscr{B}_{abc}^{m} | \mathscr{B}_{ijk}^{n}) \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot (\mathbf{r}_{n} - \mathbf{r}_{m})\right] d\mathscr{B}_{ijk}^{n} d\mathscr{B}_{abc}^{m}$$
(IX.96)

$$\propto \int P(\mathscr{B}_{ijk}^n) \mathscr{B}_{ijk}^n \int P(\mathscr{B}_{abc}^m) \mathscr{B}_{abc}^m \exp\left[i\Delta \mathbf{q} \cdot (\mathbf{r}_n - \mathbf{r}_m)\right] d\mathscr{B}_{ijk}^n d\mathscr{B}_{abc}^m \tag{IX.97}$$

And:

$$-Y_{ijk,abc}^{\text{coh}} = \sum_{n} (-\mathscr{B}_{ijk}^{n}) \left\{ \sum_{m \neq n} \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot (\mathbf{r}_{n} - \mathbf{r}_{m})\right] \right\}$$
(IX.98)

$$\propto \int P(-\mathscr{B}_{ijk}^{n})(-\mathscr{B}_{ijk}^{n}) \int P(\mathscr{B}_{abc}^{m}|\mathscr{B}_{ijk}^{n}) \mathscr{B}_{abc}^{m} \exp\left[i\Delta \mathbf{q} \cdot (\mathbf{r}_{n} - \mathbf{r}_{m})\right] d\mathscr{B}_{ijk}^{n} d\mathscr{B}_{abc}^{m} \quad (\text{IX.99})$$

Due to centro-symmetry:  $P(-\mathscr{B}_{ijk}^n) = P(\mathscr{B}_{ijk}^n)$ . Hence:

$$-Y_{ijk,abc}^{\rm coh} \propto \int P(\mathscr{B}_{ijk}^n) (-\mathscr{B}_{ijk}^n) \int P(\mathscr{B}_{abc}^m | \mathscr{B}_{ijk}^n) \mathscr{B}_{abc}^m \exp\left[i\Delta \mathbf{q} \cdot (\mathbf{r}_n - \mathbf{r}_m)\right] d\mathscr{B}_{ijk}^n d\mathscr{B} \mathscr{B}_{abc}^m \quad (\text{IX.100})$$
$$-Y_{ijk,abc}^{\rm coh} = Y_{ijk,abc}^{\rm coh} = 0 \qquad (\text{IX.101})$$

Thus,  $Y_{ijk,abc}^{coh}$  measured indeed if the hyperpolarizability in the laboratory frame of different molecules are correlated.

### Form of the intensity equation:

In the general case, the intensity is given by:

$$I_V \propto Y_{yyy,yyy} \cos^4 \gamma + (2\text{Re}Y_{yyy,yxx} + 4Y_{yxy,yxy}) \cos^2 \gamma \sin^2 \gamma + Y_{yxx,yxx} \sin^4 \gamma$$
(IX.102)

$$+4\text{Re}Y_{yyy,yxy}\cos^3\gamma\sin\gamma+4\text{Re}Y_{yxx,yxy}\cos\gamma\sin^3\gamma\tag{IX.103}$$

$$I_H \propto Y_{zyy,zyy} \cos^4 \gamma + (2\text{Re}Y_{zyy,zxx} + 4Y_{zxy,zxy}) \cos^2 \gamma \sin^2 \gamma + Y_{zxx,zxx} \sin^4 \gamma$$
(IX.104)

$$+4\text{Re}Y_{zyy,zxy}\cos^3\gamma\sin\gamma+4\text{Re}Y_{zxx,zxy}\cos\gamma\sin^3\gamma\tag{IX.105}$$

Yet, the odd terms in cos and sin power are not observed experimentally and for a good reason: they are zero due to the centro-symmetry of the liquid phase if we assume that  $\Delta \mathbf{q} = 0$ .

# IX. C.5 Deviation of $\mathscr{B}$ for a single frame

We expect to have  $\langle \mathscr{B} \rangle = 0$  since it is a centro-symmetric medium. However, we sample over a discrete number of configurations for each frame (about 15 000): the average value is not perfectly zero. As we increase the size of the MD box or the number of frames, the average get closer and closer to zero. We can also observe small fluctuations on the standard deviation of  $\mathscr{B}$  over the MD frame.



Figure IX.15: Average *B* for several components in function of the MD frame.



Figure IX.16: Standard deviation for  $\mathscr{B}$  for several components in function of the MD frame.

# IX. C.6 Coherent calculation with the same $\beta$ for all molecules

As for the incoherent section, we compare the results obtained by the QM/MM results, where each molecule can have a different  $\beta$ , with the one obtained by setting the molecular hyperpolarizability of all the molecules to the average one:  $\beta^n = \langle \beta \rangle$ .

The numerical procedure is extremely similar as the one developed before: we are still using the first hyperpolarizability in the laboratory frame of each molecule  $\mathscr{B}^n$  as the key quantity. However, in this case we use the same first hyperpolarizability for each molecule so that:

$$\mathscr{B}_{ijk}^{n} = \sum_{abc} R_{ia}^{n} R_{jb}^{n} R_{kc}^{n} < \beta >_{abc} \qquad (\text{IX.106})$$

Thus, when computing the coherent part,  $< \mathscr{B}^n \mathscr{B}^m >$ , only the orientational correlation can affect the results: not the potential correlation between  $\beta^n$  and  $\beta^m$ . As we are using the averaged  $\beta$ , it respects the molecular  $C_{2v}$  symmetry:  $<\beta>_{caa}=-1.9, <\beta>_{aca}=-2.0, <\beta>_{cbb}=$ 2.7,  $<\beta>_{bcb}=2.3, <\beta>_{ccc}=4.0$  and the other



Figure IX.17: Evolution of  $\langle C^n \rangle$  for the  $\beta^n = \langle \beta \rangle$  in log scale. Several components are plotted corresponding to V-related outcoming polarization contributions.

component are null. Since the type of calculation only requires the structure, and no QM/MM calculations, we have been able to increase the statistics: the calculation is made over 100 frames instead of 10. We use the same box containing about 15 000 molecules.

The obtained averaged  $\langle C^n \rangle$  is plotted for several components related to the V outcoming polarization in Figure IX.17. Compared to the results using QM/MM  $\beta$ , we have far fewer fluctuations of C for large r. This may be due to the fact that we have less phase space to explore if we fix the molecular hyperpolarizability. Indeed, the coherent part for  $\beta^n = \langle \beta \rangle$  is the joint probability of finding molecule n and m within a certain relative orientation. The coherent part of QM/MM  $\beta^n$  is the joint probability of finding molecule n and m within a certain relative orientation and with a certain  $\beta^n$  and  $\beta^m$ tensor value (where every  $\beta$  tensor has about 10 degrees of liberty).

Therefore, the obtained averaged  $< \mathscr{B}^n > \mathrm{in}$ 



Figure IX.18: Evolution of  $\langle C^n \rangle$  for several components corresponding to V-related outcoming polarization contributions. The dashed lines are obtained using  $\beta^n = \langle \beta \rangle$  and in full line the results for QM/MM  $\beta^n$ . Every curve is normalized with respect to its maximum in [0, 7 Å] to easier the comparison.

the  $\beta^n = \langle \beta \rangle$  case is about 0.01 a.u. (instead

of 0.3 a.u., see Figure IX.15) and its standard deviation about 1 or 2 a.u. (instead of 3 to 6 a.u., see Figure IX.16).

Therefore, we have gained an important amount of precision: we can try to check if there is orientational correlation for large r. However, no signal above the noise appears in  $\langle C^n \rangle$  for distance larger than 5-6 Å as for the QM/MM results. Therefore, we will again focus on the short-range correlation range: up to the third solvation layer.

In Figure IX.18 is plotted the averaged Cnormalized by its maximal value within  $r \in$ [0, 7Å] for components related to the V outcoming polarization. In dashed line is presented the value obtained for  $\beta^n = <\beta >$  and in full line the results for QM/MM  $\beta^n$ . We can observe that there is a poor agreement between the coherent contribution for all the components. For instance, the yyy, yyy first layer contribution for the  $\beta^n = \langle \beta \rangle$  is too large compared to the QM/MM results. This indicates that on top of the orientational correlation between the central molecule and its first shell, the first hyperpolarizability in the molecule frame are correlated. This makes complete sense if we have in mind the electrostatic effect on the molecular first hyperpolarizability: the closest molecule produces the largest electrostatic field, and thus impact the most  $\beta$ .



Figure IX.19: SHS intensity depending on the incoming polarization angle  $\gamma$  for the V and H outcoming polarization, in blue and red respectively. The blue dot is the V experimental results and the red stars the H. The full line is the predicted signal of the incoherent plus coherent part using QM/MM  $\beta$ . The dashed line the predicted signal of the incoherent plus coherent part using  $\beta^n = \langle \beta \rangle$ .

To conclude this part we have plotted the predicted SHS intensity in Figure IX.19 for the  $\beta^n = \langle \beta \rangle$  and QM/MM case along with the experimental signal. The error for the  $\beta^n = \langle \beta \rangle$  is first due to the incoherent term, see Table V.2, and second due to the coherent term as emphasis by Figure IX.18. Therefore, the  $\beta^n = \langle \beta \rangle$  is not a good approximation for water to compute its SHS response.

# IX. D | Chapter VI

# IX. D.1 Distribution of $\beta^{dq}$ in bulk

Using FROG, we also have access to the independent distribution of all the components for all hyperpolarizabilities. In Figure IX.20 is plotted the distributions of four components: one (zyxx) forbidden by the molecular  $C_{2v}$  symmetry, and three which can be non-zero in averaged. For all the components, we can compute the average value and the standard deviation, using the classic formula:  $\sigma[X] = \sqrt{\langle X^2 \rangle - \langle X \rangle^2}$ . These results are shown in the main part. However, we could also fit every distribution using a Gaussian shape:

$$P(\beta_{abcd}^{dq} = Y) = N \exp\left[-\frac{1}{2} \left(\frac{Y - \langle \beta_{abcd}^{dq} \rangle}{\sigma[\beta_{abcd}^{dq}]}\right)^2\right], \qquad (IX.107)$$

and find  $N, < \beta_{abcd}^{dq} >$  and  $\sigma[\beta_{abcd}^{dq}]$ . For these distributions in the bulk phase, the two definitions of the average and standard deviation matches.



Figure IX.20: Distribution of four  $\beta^{dq}$  components in the liquid phase at the DFT/CAM-B3LYP (d-aug-ccpVTZ) level for 800 nm. Every distribution is fitted using a Gaussian shape in dashed lines. The values are presented in atomic units.

# IX. E | Chapter VII

# IX. E.1 Fundamental power control

One has to keep in mind several aspects regarding the fundamental intensity:

- The larger the intensity, the larger the SHG signal. It means the acquisition can be quicker and the ratio signal over noise is better.
- The intensity inside the cell is significantly lower than the one after the power control. Several components absorbed the fundamental frequency, for instance the x60 objective absorbs approximately half of the incoming power, see Figure IX.27.
- Since the objective absorbs the fundamental, we cannot shine too strong an intensity: it would damage it.



Figure IX.21: Output power in function of the polarizer angle just after the power control, red dots, or just before the objective, bleu square. The fit is made using a  $\cos^2$ function according to the Malus law.

For these reasons, the intensity is also measured just before the microscope, see Figure IX.21. The difference between the output power and the one measured before the objective is due to the telescope. At the focalization point of the first lens, we have added a pinhole of 50  $\mu$ m to filter potential non-Gaussian mode of the fundamental beam: an important part of the fundamental power is lost here.

Typical intensities used are from 100 mW to 200 mW. Higher non-linear terms can occur if the power is too high. Therefore, we should check with several fundamental power that the second order behavior is retrieved.

# IX. E.2 Derivation of the interferometry:

#### Analytical derivation:

In this section, we will derive the equation needed to measure the  $\Delta z$  distance using spectral interferometry.

First, let us assume a plane wave propagating along the z-direction:

$$\mathbf{E}(z,t) = \mathbf{E}_0 \exp\left[i(\omega t - kz)\right] \qquad \qquad k = n(\lambda) \frac{2\pi}{\lambda}$$

Such approximation may be quite false for the x60 objective at the focalization point. However, we are performing these acquisitions usually far away from the focalization point: this plane wave approximation is thus not that false.

Now, let assume 2 interfaces distanced by  $\Delta z$ . The reflected signal at a negative z is given by  $\mathbf{E}_r(z,t)$ . Assuming different amplitude of the reflected signal for the 2 interfaces r1 and r2:

$$\mathbf{E}_{r}(z,t) = \mathbf{E}_{0} \{ r_{1} \exp \left[ i(\omega t - kz) \right] + r_{2} \exp \left[ i(\omega t - k(z + 2\Delta z)) \right] \}$$
(IX.108)

$$\mathbf{E}_{r}(z,t) = r_{1}\mathbf{E}_{0}\exp\left[i(\omega t - kz)\right]\left\{1 + \frac{r_{2}}{r_{1}}\exp\left[-ik2\Delta z\right]\right\}$$
(IX.109)

(IX.110)

Note that  $r_1$  is the reflection coefficient of the first interface, and  $r_2$  is the reflection coefficient of the second interface times the transmission coefficient of the first interface. However, since we are working with complicated interfaces (sphere), we will keep the  $r_1$  and  $r_2$  notation.

The intensity of the reflected signal is:

$$I_r(z,t) = I_0 r_1^2 \left\{ 1 + \frac{r_2}{r_1} \exp\left[-ik2\Delta z\right] \right\} \left\{ 1 + \frac{r_2}{r_1} \exp\left[-ik2\Delta z\right] \right\}^*$$
(IX.111)

$$I_r(z,t) = r_1^2 I_0 \left\{ 1 + \frac{r_2^2}{r_1^2} + \frac{r_2}{r_1} \exp\left[-ik2\Delta z\right] + \frac{r_2}{r_1} \exp\left[ik2\Delta z\right] \right\}$$
(IX.112)

$$I_r(z,t) = r_1^2 I_0 \left\{ 1 + \frac{r_2^2}{r_1^2} + 2\frac{r_2}{r_1} \cos(k2\Delta z) \right\}$$
(IX.113)

(IX.114)

The intensity collected depend on the wavelength. There are extrema which can be more or less easy to detect depending on the shape of the initial spectra (a pure flat spectrum is easiest then an initial very noisy and oscillating one) and the ratio of  $r_1/r_2$ . For instance, if the light is fossilized on a given interface, the other interface reflects less signal. Hence, it is not a good idea to focalize the light close to an interference is we want to see the extrema. Therefore, we tend to set the focalization point below the 2 interfaces.

Now, let us say that the extrema are located at the wave vector  $\{k_1, k_2...k_n\}$ . The extrema



Figure IX.22: Optical index of SiO2 deposits with respect to the wavelength.

respect:

$$\cos(k_i 2\Delta z) = \pm 1 \tag{IX.115}$$

$$k_i 2\Delta z = q\pi, q \in \mathbb{Z} \tag{IX.116}$$

$$\frac{4n(\lambda_i)}{\lambda_i}\Delta z = q \tag{IX.117}$$

Thus, when the wavelength of these extrema is found,  $\{\lambda_1, \lambda_2...\lambda_n\}$ , one can plot  $\frac{n(\lambda)}{\lambda}$  to extract the distance of interest  $\Delta z$ , see Figure ExpFig: min and max wafer lambda for fit delta z for instance.

### **Optical index:**

Here we should emphasize that the optical index dependency with respect to the wavelength,  $n(\lambda)$ , should be known to extract the  $\Delta z$ . Among the several ways to characterize this optical index, one can use the Cauchy formula at the order 2:

$$n(\lambda) = n_0 + \frac{B}{\lambda^2} \tag{IX.118}$$

Such fit is made for SiO2 deposit, which is used to calibrate our method, see Figure IX.22. The data used for this fit are from https://refractiveindex.info/?shelf=main&book=SiO2& page=Rodriguez-de\_Marcos. We use the same Cauchy formula for water, but with  $n_0=1.32$  and B=3165 nm<sup>-2</sup>.



### IX. E.3 Guanaco: numerical treatment of the spectral interferences:

Figure IX.23: In full red line, the raw spectrum of the SiO<sub>2</sub> deposit reflection on the x10 objective. In dashed lines the smoothed spectrum. The dots are the positions of the minima and maxima as detected by Guanaco. In insert the linear regression of the positions of the local minima and maxima found in the main figure from the averaged SiO<sub>2</sub> spectrum. The slope leads to the deposit thickness according to Equation IX.119:  $\Delta z = 997.7$  nm.

To calibrate our measurement, we have used a  $SiO_2$  deposit on a Si surface. This deposit is 1  $\mu$ m thick, with a relative uncertainty of 10 nm. This sample is expected to be extremely flat, with a homogeneous deposit of SiO2: the error made on the measurement of the SiO2 deposit should only depend on our optical source, detection, and the numerical treatment. The obtained spectrum is plotted in Figure IX.23.

The derivations of the spectral interferometry are reminded in the Annex IX. E.2. We will assume that this formalism works both in the case of the SiO2 deposit and the lamella-sphere system. Annex IX. E.2 leads to a final equation which links the wavelength of the minima and maxima  $\{\lambda_1, \lambda_2...\lambda_n\}$  of the interference spectrum to the inter-plane distance  $\Delta z$ :

$$\frac{4n(\lambda_i)}{\lambda_i}\Delta z = q \; ; \; q \in \mathbb{Z} \tag{IX.119}$$

Where  $n(\lambda_i)$  is the optical index of the material between the two interfaces (for the wafer the SiO<sub>2</sub> optical index, for the lamella sphere it is either the air or liquid optical index) and q an integer. Hence, we should first detect the spectrum extrema position, see for instance Figure IX.23, and then, plot  $\frac{n(\lambda_i)}{\lambda_i}$  in function of the relative extrema number q, see insert of Figure IX.23, the slope

#### is $4\Delta z$ .

In more details, we have automatized the extrema detection using the scipy.signal.find\_peaks function from the scipy module of Python. The main issue is that working with the raw spectra is that it is more difficult to make the analysis procedure automatic and robust compared to smooth one. Once the spectrum is smoothed (using a sliding averaging over several wavelength measurement), the other parameters used to find the extrema are extremely robust as the fit of  $\frac{n(\lambda_i)}{\lambda_i}$ . Hence, the major numerical error is in the smoothing choice. In Figure IX.24 we plotted the effect of the smoothing of the measured  $\Delta z$ . The typical error made due to this wavelength averaging is 20 nm. We have confirmed this by looking at other spectra from



Figure IX.24: Measured  $SiO_2$  thickness depending on the smoothing parameter used. The fluctuation are due to the numerical analysis default and are about 20 nm.

other systems – data not shown. Therefore, we measured a thickness of  $998 \pm 20$  nm for the SiO<sub>2</sub> wafer, for an expected value of  $1000 \pm 5$  nm.

We should also emphasize that this wafer is quite difficult to measure with our method: there are only few oscillation within the spectral range of our acquisition (from 350 nm to 650 nm because of the reflection range of the dichroic mirror). In the lamella-sphere setup with a typical  $\Delta z$ about tens of  $\mu$ m, we observed way more oscillations, see Figure VII.12 or VII.19. Hence, for the system of interest, the lamella sphere, the smoothing of the spectra has even a smaller impact due to the large number of extrema observed. However, this also points out that this method cannot be used to measure the lamella-sphere distance below the optical wavelengths. Therefore, this spatial interferometry measurement is used to characterize our confinement setup and to provide an initial absolute  $\Delta z$  value: we will rely on the piezzo displacement to achieve the confinement below the micron.

Other errors can arise due to the white source. Not because the total intensity of the spectra varies in time, but that the shape of the intensity varies. For instance, that  $I(\lambda_1)/I(\lambda_2)$  is not constant for close wavelengths. This is another advantageous of the spectral interferometry measurement: we are not impacted by the fluctuation of the total intensity. Hopefully, we do not observe large fluctuations within the spectra. Finally, we can also increase the time used to measure acquire the spectra to reduce the spectrum fluctuations.

# IX. E.4 Optical components characteristics



Figure IX.25: Transmittance and reflectivity according to the constructor Thorlabs of the dichroic mirror: DMLP650R.



Figure IX.26: Transmittance according to the constructor Nikon of the x10 objective: CFI Plan Achro 10X MRL00102.



Figure IX.27: Transmittance according to the constructor Nikon of the x60 objective: CFI Plan Apochromat VC 60XC WI, MRD07602.


Figure IX.28: PMT efficiency depending on the wavelength. We are using the 175 model.

## Bibliography

- [1] Patrick Norman, Kenneth Ruud, and Trond Saue. <u>Principles and practices of molecular</u> properties: Theory, modeling, and simulations. John Wiley & Sons, 2018.
- [2] Damien Laage and James T Hynes. A molecular jump mechanism of water reorientation. Science, 311(5762):832–835, 2006.
- [3] David P Shelton. Long-range orientation correlation in water. <u>The Journal of Chemical</u> Physics, 141(22):224506, 2014.
- [4] Julien Duboisset and Pierre-François Brevet. Salt-induced long-to-short range orientational transition in water. Physical Review Letters, 120(26):263001, 2018.
- [5] Simone Pezzotti, Daria Ruth Galimberti, and Marie-Pierre Gaigeot. 2d h-bond network as the topmost skin to the air-water interface. <u>The journal of physical chemistry letters</u>, 8(13):3133– 3141, 2017.
- [6] Guillaume Le Breton and Laurent Joly. Molecular modeling of aqueous electrolytes at interfaces: Effects of long-range dispersion forces and of ionic charge rescaling. <u>The Journal of</u> Chemical Physics, 152(24):241102, 2020.
- [7] Laurent Joly, Robert H Meißner, Marcella Iannuzzi, and Gabriele Tocci. Osmotic transport at the aqueous graphene and hbn interfaces: Scaling laws from a unified, first-principles description. ACS nano, 15(9):15249–15258, 2021.
- [8] Sarah Hocine, Remco Hartkamp, Bertrand Siboulet, Magali Duvail, Benoit Coasne, Pierre Turq, and Jean-François Dufrêche. How ion condensation occurs at a charged surface: A molecular dynamics investigation of the stern layer for water–silica interfaces. <u>The Journal of</u> Physical Chemistry C, 120(2):963–973, 2016.
- [9] Colin D Bain. Studies of adsorption at interfaces by optical techniques: ellipsometry, second harmonic generation and sum-frequency generation. <u>Current opinion in colloid & interface</u> science, 3(3):287–292, 1998.
- [10] Turgut Battal, Gemma C Shearman, Dimitrina Valkovska, Colin D Bain, Richard C Darton, and Julian Eastoe. Determination of the dynamic surface excess of a homologous series of cationic surfactants by ellipsometry. Langmuir, 19(4):1244–1248, 2003.

- [11] Elizabeth Donohue, Sina Khorsand, Gabriel Mercado, Kristen M. Varney, Paul T. Wilder, Wenbo Yu, Alexander D. MacKerell, Patrick Alexander, Que N. Van, Ben Moree, Andrew G. Stephen, David J. Weber, Joshua Salafsky, and Frank McCormick. Second harmonic generation detection of Ras conformational changes and discovery of a small molecule binder. <u>Proc.</u> <u>Natl. Acad. Sci. U.S.A.</u>, 116(35):17290–17297, 2019.
- [12] Alicia C. McGeachy, Emily R. Caudill, Dongyue Liang, Qiang Cui, Joel A. Pedersen, and Franz M. Geiger. Counting charges on membrane-bound peptides. <u>Chem. Sci.</u>, 9:4285–4298, 2018.
- [13] Giuseppe Licari, Joseph S. Beckwith, Saeideh Soleimanpour, Stefan Matile, and Eric Vauthey. Detecting order and lateral pressure at biomimetic interfaces using a mechanosensitive secondharmonic-generation probe. Phys. Chem. Chem. Phys., 20:9328–9336, 2018.
- [14] Hong-tao Bian, Ran-ran Feng, Yuan Guo, and Hong-fei Wang. Specific na+ and k+ cation effects on the interfacial water molecules at the air/aqueous salt solution interfaces probed with nonresonant second harmonic generation. J. Chem. Phys., 130(13):134709, 2009.
- [15] Heather C Allen, Nadia N Casillas-Ituarte, M Roxana Sierra-Hernandez, Xiangke Chen, and Cheng Y Tang. Shedding light on water structure at air-aqueous interfaces: ions, lipids, and hydration. Phys. Chem. Chem. Phys., 11(27):5538–5549, 2009.
- [16] Marialore Sulpizi, Mathieu Salanne, Michiel Sprik, and Marie-Pierre Gaigeot. Vibrational sum frequency generation spectroscopy of the water liquid–vapor interface from density functional theory-based molecular dynamics simulations. J. Phys. Chem. Lett., 4(1):83–87, 2012.
- [17] Yixing Chen, Halil I. Okur, Nikolaos Gomopoulos, Carlos Macias-Romero, Paul S. Cremer, Poul B. Petersen, Gabriele Tocci, David M. Wilkins, Chungwen Liang, Michele Ceriotti, and Sylvie Roke. Electrolytes induce long-range orientational order and free energy changes in the H-bond network of bulk water. Sci. Adv., 2(4):1–9, 2016.
- [18] Paul E Ohno, HanByul Chang, Austin P Spencer, Yangdongling Liu, Mavis D Boamah, Hongfei Wang, and Franz M Geiger. Beyond the gouy-chapman model with heterodyne-detected second harmonic generation. J. Phys. Chem. Lett., 10(10):2328–2334, 2019.
- [19] Laetitia Dalstein, Kuo-Yang Chiang, and Yu-Chieh Wen. Direct quantification of water surface charge by phase-sensitive second harmonic spectroscopy. <u>The Journal of Physical Chemistry</u> Letters, 10(17):5200–5205, 2019.
- [20] Thanh Tung Pham, Alban Jonchère, Jean François Dufrêche, Pierre François Brevet, and Olivier Diat. Analysis of the second harmonic generation signal from a liquid/air and liquid/liquid interface. J. Chem. Phys., 146(14):144701, 2017.
- [21] Lisa B Dreier, Christoph Bernhard, Grazia Gonella, Ellen HG Backus, and Mischa Bonn. Surface potential of a planar charged lipid-water interface. what do vibrating plate methods,

second harmonic and sum frequency measure? <u>The Journal of Physical Chemistry Letters</u>, 9(19):5685–5691, 2018.

- [22] Charlotte Bouquiaux, Frédéric Castet, and Benoît Champagne. Unravelling the effects of cholesterol on the second-order nonlinear optical responses of di-8-anepps dye embedded in phosphatidylcholine lipid bilayers. <u>The Journal of Physical Chemistry B</u>, 125(36):10195– 10212, 2021.
- [23] R Kramer Campen, Allison K Pymer, Satoshi Nihonyanagi, and Eric Borguet. Linking surface potential and deprotonation in nanoporous silica: Second harmonic generation and acid/base titration. J. Phys. Chem. C, 114(43):18465–18473, 2010.
- [24] Siyun Xu, Sirui Xing, Shin-Shem Pei, Vladislav Ivaništšev, Ruth Lynden-Bell, and Steven Baldelli. Molecular response of 1-butyl-3-methylimidazolium dicyanamide ionic liquid at the graphene electrode interface investigated by sum frequency generation spectroscopy and molecular dynamics simulations. J. Phys. Chem. C, 119(46):26009–26019, 2015.
- [25] Arianna Marchioro, Marie Bischoff, Cornelis Lütgebaucks, Denys Biriukov, Milan Předota, and Sylvie Roke. Surface characterization of colloidal silica nanoparticles by second harmonic scattering: Quantifying the surface potential and interfacial water order. <u>The Journal of</u> Physical Chemistry C, 123(33):20393–20404, 2019.
- [26] Jonathan H Raberg, Jenel Vatamanu, Stephen J Harris, Christina HM van Oversteeg, Axel Ramos, Oleg Borodin, and Tanja Cuk. Probing electric double-layer composition via in situ vibrational spectroscopy and molecular simulations. <u>J. Phys. Chem. Lett.</u>, 10(12):3381–3389, 2019.
- [27] Simone Pezzotti, Daria Ruth Galimberti, and Marie-Pierre Gaigeot. Deconvolution of bil-sfg and dl-sfg spectroscopic signals reveals order/disorder of water at the elusive aqueous silica interface. Phys. Chem. Chem. Phys., 21(40):22188–22202, 2019.
- [28] Aurélie Bruyère. <u>Génération de second harmonique sur des films moléculaires chiraux</u>. PhD thesis, Université de Lyon, 2016.
- [29] Lucile Sanchez. <u>Génération de second harmonique de milieux diélectriques: du tensioactif en solution à la microparticule de silice optiquement piégée</u>. PhD thesis, Université de Lyon, 2018.
- [30] Anders Osted, Jacob Kongsted, Kurt V. Mikkelsen, Per Olof Åstrand, and Ove Christiansen. Statistical mechanically averaged molecular properties of liquid water calculated using the combined coupled cluster/molecular dynamics method. J. Chem. Phys., 124(12):124503, 2006.
- [31] Philip Kaatz, Elizabeth A Donley, and David P Shelton. A comparison of molecular hyperpolarizabilities from gas and liquid phase measurements. <u>J. Chem. Phys.</u>, 108(3):849–856, 1998.

- [32] BF Levine and CG Bethea. Effects on hyperpolarizabilities of molecular interactions in associating liquid mixtures. JCP, 65(6):2429–2438, 1976.
- [33] Tárcius N. Ramos, Sylvio Canuto, and Benoît Champagne. Unraveling the electric fieldinduced second harmonic generation responses of stilbazolium ion pairs complexes in solution using a multiscale simulation method. J. Chem. Information Model., 60(10):4817–4826, 2020.
- [34] B. F. Levine and C. G. Bethea. Second and third order hyperpolarizabilities of organic molecules. JCP, 63(6):2666-2682, 1975.
- [35] Elsa C.Y. Yan, Yan Liu, and Kenneth B. Eisenthal. New method for determination of surface potential of microscopic particles by second harmonic generation. <u>JPCB</u>, 102(33):6331–6336, 1998.
- [36] Tatsuya Joutsuka, Tomonori Hirano, Michiel Sprik, and Akihiro Morita. Effects of thirdorder susceptibility in sum frequency generation spectra: a molecular dynamics study in liquid water. Physical Chemistry Chemical Physics, 20:3040–3053, 2018.
- [37] Simone Pezzotti, Daria Ruth Galimberti, Y Ron Shen, and Marie-Pierre Gaigeot. Structural definition of the bil and dl: A new universal methodology to rationalize non-linear  $\chi^{(2)}(\omega)$  sfg signals at charged interfaces, including  $\chi^{(3)}(\omega)$  contributions. <u>Physical Chemistry Chemical</u> Physics, 20(7):5190–5199, 2018.
- [38] Remco Hartkamp, Anne Laure Biance, Li Fu, Jean François Dufrêche, Oriane Bonhomme, and Laurent Joly. Measuring surface charge: Why experimental characterization and molecular modeling should be coupled. COCIS, 37:101–114, 2018.
- [39] Emily Ma, Paul E. Ohno, Jeongmin Kim, Yangdongling Liu, Emilie H. Lozier, Thomas F. Miller, Hong-Fei Wang, and Franz M. Geiger. A new imaginary term in the second-order non-linear susceptibility from charged interfaces. JPCL, 12(24):5649–5659, 2021. PMID: 34110833.
- [40] HanByul Chang, Paul E. Ohno, Yangdongling Liu, Emilie H. Lozier, Naomi Dalchand, and Franz M. Geiger. Direct measurement of charge reversal on lipid bilayers using heterodynedetected second harmonic generation spectroscopy. <u>JPCB</u>, 124(4):641–649, 2020. PMID: 31903764.
- [41] Yann Foucaud, Bertrand Siboulet, Magali Duvail, Alban Jonchere, Olivier Diat, Rodolphe Vuilleumier, and Jean-François Dufrêche. Deciphering second harmonic generation signals. Chemical Science, 12(45):15134–15142, 2021.
- [42] Sophie Brasselet and Joseph Zyss. Multipolar molecules and multipolar fields: probing and controlling the tensorial nature of nonlinear molecular media. JOSA B, 15(1):257–288, 1998.
- [43] Daniel Borgis, Luc Belloni, and Maximilien Levesque. What does second-harmonic scattering measure in diluted electrolytes? <u>The Journal of Physical Chemistry Letters</u>, 9(13):3698–3702, 2018.

- [44] Julien Duboisset, Fabien Rondepierre, and Pierre-François Brevet. Long-range orientational organization of dipolar and steric liquids. <u>The Journal of Physical Chemistry Letters</u>, 11(22):9869–9875, 2020.
- [45] Antonin Pardon, Oriane Bonhomme, Clotilde Gaillard, Pierre-François Brevet, and Emmanuel Benichou. Nonlinear optical signature of nanostructural transition in ionic liquids. <u>Journal of</u> Molecular Liquids, 322:114976, 2021.
- [46] Roger G Horn and JN Israelachvili. Direct measurement of forces due to solvent structure. Chemical Physics Letters, 71(2):192–194, 1980.
- [47] Roger G Horn and Jacob N Israelachvili. Direct measurement of structural forces between two surfaces in a nonpolar liquid. The Journal of Chemical Physics, 75(3):1400–1411, 1981.
- [48] B Cross and J Crassous. Rheological properties of a highly confined film of a lyotropic lamellar phase. The European Physical Journal E, 14(3):249–257, 2004.
- [49] Thomas Becker and F Mugele. Nanofluidics: Molecularly thin lubricant layers under confinement. Molecular simulation, 31(6-7):489–494, 2005.
- [50] Chloé Barraud. <u>Nanorhéologie de fluides complexes aux interfaces</u>. PhD thesis, Université Grenoble Alpes (ComUE), 2016.
- [51] Marc de Wergifosse, Frédéric Castet, and Benoît Champagne. Frequency dispersion of the first hyperpolarizabilities of reference molecules for nonlinear optics. <u>J. Chem. Phys.</u>, 142(19):194102, 2015.
- [52] Pierre Beaujean and Benoît Champagne. Coupled cluster evaluation of the second and third harmonic scattering responses of small molecules. <u>Theoretical Chemistry Accounts</u>, 137(4):1– 11, 2018.
- [53] Tina D Poulsen, Peter R Ogilby, and Kurt V Mikkelsen. A quantum mechanical method for calculating nonlinear optical properties of condensed phase molecules coupled to a molecular mechanics field: A quadratic multiconfigurational self-consistent-field/molecular mechanics response method. J. Chem. Phys., 115(17):7843–7851, 2001.
- [54] Kristian O Sylvester-Hvid, Kurt V Mikkelsen, Patrick Norman, Dan Jonsson, and Hans Ågren. Sign change of hyperpolarizabilities of solvated water, revised: Effects of equilibrium and nonequilibrium solvation. The Journal of Physical Chemistry A, 108(41):8961–8965, 2004.
- [55] Chungwen Liang, Gabriele Tocci, David M. Wilkins, Andrea Grisafi, Sylvie Roke, and Michele Ceriotti. Solvent fluctuations and nuclear quantum effects modulate the molecular hyperpolarizability of water. Physical Review B, 96(4):1–6, 2017.

- [56] Gabriele Tocci, Chungwen Liang, David M Wilkins, Sylvie Roke, and Michele Ceriotti. Second-harmonic scattering as a probe of structural correlations in liquids. <u>The journal of</u> physical chemistry letters, 7(21):4311–4316, 2016.
- [57] Kazuya Shiratori, Shoichi Yamaguchi, Tahei Tahara, and Akihiro Morita. Computational analysis of the quadrupole contribution in the second-harmonic generation spectroscopy for the water/vapor interface. The Journal of chemical physics, 138(6):064704, 2013.
- [58] Grazia Gonella, Cornelis Lütgebaucks, Alex GF De Beer, and Sylvie Roke. Second harmonic and sum-frequency generation from aqueous interfaces is modulated by interference. <u>The</u> Journal of Physical Chemistry C, 120(17):9165–9173, 2016.
- [59] Kurt V. Mikkelsen, Yi Luo, Hans Ågren, and Poul Jørgensen. Sign change of hyperpolarizabilities of solvated water. J. Chem. Phys., 102(23):9362–9367, 1995.
- [60] Kerry Garrett, XochitlA Sosa Vazquez, Shawn B Egri, Jacob Wilmer, Lewis E Johnson, Bruce H Robinson, and Christine M Isborn. Optimum exchange for calculation of excitation energies and hyperpolarizabilities of organic electro-optic chromophores. J. Chem. Theory Comput., 10(9):3821–3831, 2014.
- [61] Silvio Osella, N Arul Murugan, Naresh K Jena, and Stefan Knippenberg. Investigation into biological environments through (non) linear optics: A multiscale study of laurdan derivatives. J. Chem. Theory Comput., 12(12):6169–6181, 2016.
- [62] Nanna Holmgaard List, Hans Jørgen Aagaard Jensen, and Jacob Kongsted. Local electric fields and molecular properties in heterogeneous environments through polarizable embedding. Physical Chemistry Chemical Physics, 18(15):10070–10080, 2016.
- [63] Giuseppe Licari, Lukasz Cwiklik, Pavel Jungwirth, and Eric Vauthey. Exploring fluorescent dyes at biomimetic interfaces with second harmonic generation and molecular dynamics. Langmuir, 33(14):3373–3383, 2017.
- [64] Marc De Wergifosse, Edith Botek, Evelien De Meulenaere, Koen Clays, and Benoît Champagne. Oniom investigation of the second-order nonlinear optical responses of fluorescent proteins. J. Phys. Chem. B, 122(19):4993–5005, 2018.
- [65] Claire Tonnelé, Benoît Champagne, Luca Muccioli, and Frédeéric Castet. Nonlinear optical contrast in azobenzene-based self-assembled monolayers. <u>Chemistry of Materials</u>, 31(17):6759– 6769, 2019.
- [66] Charlotte Bouquiaux, Claire Tonnelé, Frédéric Castet, and Benoît 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(11):2101–2109, 2020.

- [67] Christian B Nielsen, Ove Christiansen, Kurt V Mikkelsen, and Jacob Kongsted. Density functional self-consistent quantum mechanics/molecular mechanics theory for linear and nonlinear molecular properties: Applications to solvated water and formaldehyde. J. Chem. Phys., 126(15):154112, 2007.
- [68] Marcelo Hidalgo Cardenuto and Benoît Champagne. The first hyperpolarizability of nitrobenzene in benzene solutions: investigation of the effects of electron correlation within the sequential qm/mm approach. Phys. Chem. Chem. Phys., 17(36):23634–23642, 2015.
- [69] S. Plimpton. Fast Parallel Algorithms for Short Range Molecular Dynamics. <u>Journal of</u> Computational Physics, 117(June 1994):1–19, 1995.
- [70] Guillaume Le Breton, Oriane Bonhomme, Emmanuel Benichou, and Claire Loison. First hyperpolarizability of water in the bulk phase: long-range electrostatic effects included via the second hyperpolarizability. ChemXiv, 2022.
- [71] Guillaume Le Breton, Oriane Bonhomme, Pierre-François Brevet, Emmanuel Benichou, and Claire Loison. First hyperpolarizability of water at the air-vapor interface: a qm/mm study questions standard experimental approximations. PCCP, 23:24932–24941, 2021.
- [72] Casper Steinmann, Peter Reinholdt, Morten Steen Nørby, Jacob Kongsted, and Jógvan Magnus Haugaard Olsen. Response properties of embedded molecules through the polarizable embedding model. International Journal of Quantum Chemistry, 119(1):e25717, 2019.
- [73] Kestutis Aidas, Celestino Angeli, Keld L Bak, Vebjørn Bakken, Radovan Bast, Linus Boman, Ove Christiansen, Renzo Cimiraglia, Sonia Coriani, Pål Dahle, et al. The dalton quantum chemistry program system. <u>Wiley Interdisciplinary Reviews: Computational Molecular</u> Science, 4(3):269–284, 2014.
- [74] O. Bonhomme, L. Sanchez, E. Benichou, and P.F. Brevet. Multistep micellization of standard surfactants evidenced by second harmonic scattering. <u>JPCB</u>, 125(38):10876–10881, 2021. PMID: 34530611.
- [75] Antonin Pardon, Oriane Bonhomme, Clotilde Gaillard, Pierre-François Brevet, and Emmanuel Benichou. Nonlinear optical signature of nanostructural transition in ionic liquids. <u>J. Mol.</u> Liq., 322:114976, 2021.
- [76] Pierre-François Brevet. <u>Surface second harmonic generation</u>. PPUR presses polytechniques, 1997.
- [77] Robert W Boyd. Nonlinear optics. Academic press, 2007.
- [78] Julien Duboisset. <u>Génération de second harmonique de biomolécules: des acides aminés aux</u> protéines. PhD thesis, Université Claude Bernard-Lyon I, 2009.

- [79] Akihiro Morita. <u>Theory of Sum Frequency Generation Spectroscopy</u>, volume 97. Springer, 2018.
- [80] Richard M Martin. <u>Electronic structure: basic theory and practical methods</u>. Cambridge university press, 2006.
- [81] José-Philippe Pérez, Robert Carles, and Robert Fleckinger. <u>Electromagnétisme: fondements</u> et applications avec 300 exercices et problèmes résolus. Dunod, 2002.
- [82] Richard Taillet. Optique physique: Propagation de la lumière. De Boeck Supérieur, 2015.
- [83] Julien Guthmuller. Étude théorique de molécules conjuguées d'intérêt biologique pour l'optique non linéaire. PhD thesis, Lyon 1, 2006.
- [84] Shambhu Ghimire, Anthony DiChiara, Emily Sistrunk, Pietro Agostini, Louis Dimauro, and David A. Reis. Observation of high-order harmonic generation in a bulk crystal. <u>Nature</u> Physics, 7:138–141, 02 2011.
- [85] Nicolas Tancogne-Dejean, Oliver D. Mücke, Franz Kärtner, and Angel Rubio. Ellipticity dependence of high-harmonic generation in solids originating from coupled intraband and interband dynamics. Nature Communications, 8, 12 2017.
- [86] Guillaume Le Breton, Angel Rubio, and Nicolas Tancogne-Dejean. High-harmonic generation from few-layer hexagonal boron nitride: Evolution from monolayer to bulk response. <u>Physical</u> Review B, 98(16):165308, 2018.
- [87] Tomasz Seidler, Katarzyna Stadnicka, and Benoît Champagne. Linear and second-order nonlinear optical properties of ionic organic crystals. <u>The Journal of chemical physics</u>, 141(10):104109, 2014.
- [88] Claire Loison, Mehmet Nail Nasir, Emmanuel Benichou, Françoise Besson, and Pierre-François Brevet. Multi-scale modeling of mycosubtilin lipopeptides at the air/water interface: structure and optical second harmonic generation. <u>Physical Chemistry Chemical Physics</u>, 16(5):2136– 2148, 2014.
- [89] Houxian Chen, Menglin Liu, and Tianying Yan. Molecular multipoles and (hyper) polarizabilities of water by ab initio calculations. Chemical Physics Letters, 752:137555, 2020.
- [90] Pierre Beaujean and Benoît Champagne. Coupled cluster evaluation of the frequency dispersion of the first and second hyperpolarizabilities of water, methanol, and dimethyl ether. <u>The</u> Journal of Chemical Physics, 145(4):044311, 2016.
- [91] Pierre Beaujean and Benoît Champagne. Coupled cluster investigation of the vibrational and electronic second and third harmonic scattering hyperpolarizabilities of the water molecule. JCP, 151(6), 2019.

- [92] Rick A Kendall, Thom H Dunning Jr, and Robert J Harrison. Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions. <u>The Journal of chemical</u> physics, 96(9):6796–6806, 1992.
- [93] Andrzej J Sadlej and Miroslav Urban. Medium-size polarized basis sets for high-levelcorrelated calculations of molecular electric properties: Iii. alkali (li, na, k, rb) and alkalineearth (be, mg, ca, sr) atoms. <u>Journal of Molecular Structure: THEOCHEM</u>, 234:147–171, 1991.
- [94] Peter Reinholdt, Jacob Kongsted, and Jógvan Magnus Haugaard Olsen. Polarizable density embedding: A solution to the electron spill-out problem in multiscale modeling. <u>J. Phys.</u> Chem. Lett., 8(23):5949–5958, 2017. PMID: 29178794.
- [95] Casper Steinmann, Peter Reinholdt, Morten Steen Nørby, Jacob Kongsted, and Jógvan Magnus Haugaard Olsen. Response properties of embedded molecules through the polarizable embedding model. Int. J. Quantum Chem., 119(1):e25717, 2019.
- [96] Marc de Wergifosse, Vincent Liégeois, and Benoît Champagne. Evaluation of the molecular static and dynamic first hyperpolarizabilities. <u>International Journal of Quantum Chemistry</u>, 114(14):900–910, 2014.
- [97] David P Shelton and Julia E Rice. Measurements and calculations of the hyperpolarizabilities of atoms and small molecules in the gas phase. Chemical Reviews, 94(1):3–29, 1994.
- [98] Pau Besalú-Sala, Sebastian P Sitkiewicz, Pedro Salvador, Eduard Matito, and Josep M Luis. A new tuned range-separated density functional for the accurate calculation of second hyperpolarizabilities. Physical Chemistry Chemical Physics, 22(21):11871–11880, 2020.
- [99] Jeff R Hammond, Niranjan Govind, Karol Kowalski, Jochen Autschbach, and Sotiris S Xantheas. Accurate dipole polarizabilities for water clusters n= 2–12 at the coupled-cluster level of theory and benchmarking of various density functionals. <u>The Journal of chemical physics</u>, 131(21):214103, 2009.
- [100] Martin J Paterson, Ove Christiansen, Filip Pawłowski, Poul Jørgensen, Christof Hättig, Trygve Helgaker, and Paweł Sałek. Benchmarking two-photon absorption with cc3 quadratic response theory, and comparison with density-functional response theory. <u>The Journal of</u> chemical physics, 124(5):054322, 2006.
- [101] Kazuya Shiratori and Akihiro Morita. Molecular theory on dielectric constant at interfaces: A molecular dynamics study of the water/vapor interface. <u>The Journal of chemical physics</u>, 134(23):234705, 2011.
- [102] M Alejandra Sánchez, Tanja Kling, Tatsuya Ishiyama, Marc-Jan van Zadel, Patrick J Bisson, Markus Mezger, Mara N Jochum, Jenée D Cyran, Wilbert J Smit, Huib J Bakker, et al.

Experimental and theoretical evidence for bilayer-by-bilayer surface melting of crystalline ice. Proceedings of the National Academy of Sciences, 114(2):227–232, 2017.

- [103] Antonin Pardon. <u>Sonder la structuration de liquides ioniques par optique non-linéaire: du</u> volume à l'interface liquide-liquide. PhD thesis, Université de Lyon, 2021.
- [104] John E Sipe. New green-function formalism for surface optics. JOSA B, 4(4):481–489, 1987.
- [105] Maarten T.P. Beerepoot, Arnfinn Hykkerud Steindal, Nanna Holmgaard List, Jacob Kongsted, and Jógvan Magnus Haugaard Olsen. Averaged Solvent Embedding Potential Parameters for Multiscale Modeling of Molecular Properties. <u>Journal of Chemical Theory and Computation</u>, 12(4):1684–1695, 2016.
- [106] Erik Rosendahl Kjellgren, Jógvan Magnus Haugaard Olsen, and Jacob Kongsted. Importance of accurate structures for quantum chemistry embedding methods: Which strategy is better? Journal of Chemical Theory and Computation, 14(8):4309–4319, 2018.
- [107] Alireza Marefat Khah, Peter Reinholdt, Jógvan Magnus Haugaard Olsen, Jacob Kongsted, and Christof Hattig. Avoiding electron spill-out in qm/mm calculations on excited states with simple pseudopotentials. <u>Journal of Chemical Theory and Computation</u>, 16(3):1373–1381, 2020.
- [108] Naveen Michaud-Agrawal, Elizabeth J Denning, Thomas B Woolf, and Oliver Beckstein. Mdanalysis: a toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem., 32(10):2319–2327, 2011.
- [109] Marcello Sega, György Hantal, Balázs Fábián, and Pál Jedlovszky. Pytim: A python package for the interfacial analysis of molecular simulations, 2018.
- [110] Remco Hartkamp, Anne-Laure Biance, Li Fu, Jean-François Dufrêche, Oriane Bonhomme, and Laurent Joly. Measuring surface charge: Why experimental characterization and molecular modeling should be coupled. <u>Current Opinion in Colloid & Interface Science</u>, 37:101–114, 2018.
- [111] Shaowei Ong, Xiaolin Zhao, and Kenneth B Eisenthal. Polarization of water molecules at a charged interface: second harmonic studies of the silica/water interface. <u>Chemical Physics</u> Letters, 191(3-4):327–335, 1992.
- [112] Emily Ma, Paul E Ohno, Jeongmin Kim, Yangdongling Liu, Emilie H Lozier, Thomas F Miller III, Hong-Fei Wang, and Franz M Geiger. A new imaginary term in the second-order nonlinear susceptibility from charged interfaces. <u>The Journal of Physical Chemistry Letters</u>, 12(24):5649–5659, 2021.
- [113] Emily Ma and Franz M Geiger. Divalent ion specific outcomes on stern layer structure and total surface potential at the silica: water interface. <u>The Journal of Physical Chemistry A</u>, 125(46):10079–10088, 2021.

- [114] Luc Belloni, Daniel Borgis, and Maximilien Levesque. Screened coulombic orientational correlations in dilute aqueous electrolytes. <u>The Journal of Physical Chemistry Letters</u>, 9(8):1985– 1989, 2018.
- [115] Douwe Jan Bonthuis, Stephan Gekle, and Roland R Netz. Dielectric profile of interfacial water and its effect on double-layer capacitance. Physical review letters, 107(16):166102, 2011.
- [116] Yanbo Xie, Li Fu, Thomas Niehaus, and Laurent Joly. Liquid-solid slip on charged walls: the dramatic impact of charge distribution. Physical Review Letters, 125(1):014501, 2020.
- [117] Jérôme Crassous, Elisabeth Charlaix, and Jean-Luc Loubet. Nanoscale investigation of wetting dynamics with a surface force apparatus. Physical review letters, 78(12):2425, 1997.
- [118] Emily E Meyer, Kenneth J Rosenberg, and Jacob Israelachvili. Recent progress in understanding hydrophobic interactions. <u>Proceedings of the National Academy of Sciences</u>, 103(43):15739–15746, 2006.
- [119] Timothy S Groves, Carla S Perez-Martinez, Romain Lhermerout, and Susan Perkin. Surface forces and structure in a water-in-salt electrolyte. <u>The Journal of Physical Chemistry Letters</u>, 12(6):1702–1707, 2021.
- [120] Saravana Kumar, Peter Cats, Mohammed B Alotaibi, Subhash C Ayirala, Ali A Yousef, René van Roij, Igor Siretanu, and Frieder Mugele. Absence of anomalous underscreening in highly concentrated aqueous electrolytes confined between smooth silica surfaces. <u>Journal of Colloid</u> and Interface Science, 622:819–827, 2022.
- [121] Frédéric Restagno, Jérôme Crassous, Elisabeth Charlaix, Cécile Cottin-Bizonne, and Michel Monchanin. A new surface forces apparatus for nanorheology. <u>Review of scientific instruments</u>, 73(6):2292–2297, 2002.
- [122] Chloé Barraud, Benjamin Cross, Cyril Picard, Fréderic Restagno, Lilianne Léger, and Elisabeth Charlaix. Large slippage and depletion layer at the polyelectrolyte/solid interface. <u>Soft</u> <u>matter</u>, 15(31):6308–6317, 2019.
- [123] Richard Villey, Emmanuelle Martinot, Cécile Cottin-Bizonne, Magali Phaner-Goutorbe, Liliane Léger, Frédéric Restagno, and Elisabeth Charlaix. Effect of surface elasticity on the rheology of nanometric liquids. Physical review letters, 111(21):215701, 2013.
- [124] Stefan HJ Idziak, Cyrus R Safinya, Robert S Hill, Keith E Kraiser, Marina Ruths, Heidi E Warriner, Suzi Steinberg, Keng S Liang, and Jacob N Israelachvili. The x-ray surface forces apparatus: structure of a thin smectic liquid crystal film under confinement. <u>Science</u>, 264(5167):1915–1918, 1994.
- [125] P Kékicheff, J Iss, P Fontaine, and A Johner. Direct measurement of lateral correlations under controlled nanoconfinement. Physical Review Letters, 120(11):118001, 2018.

- [126] Ashis Mukhopadhyay, Jiang Zhao, Sung Chul Bae, and Steve Granick. An integrated platform for surface forces measurements and fluorescence correlation spectroscopy. <u>Review of scientific</u> instruments, 74(6):3067–3072, 2003.
- [127] Hilton B de Aguiar, Joshua D McGraw, and Stephen H Donaldson Jr. Interface-sensitive raman microspectroscopy of water via confinement with a multimodal miniature surface forces apparatus. Langmuir, 35(48):15543–15551, 2019.
- [128] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. <u>Phys. Rev.</u>, 136:B864–B871, Nov 1964.
- [129] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. Phys. Rev., 140:A1133–A1138, Nov 1965.
- [130] Mel Levy. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. <u>Proceedings of the</u> National Academy of Sciences, 76(12):6062–6065, 1979.
- [131] Lieb Elliott H. Density functionals for coulomb systems. <u>International Journal of Quantum</u> Chemistry, 24(3):243–277.
- [132] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. <u>Physical review</u>, 136(3B):B864, 1964.
- [133] Seymour H Vosko, Leslie Wilk, and Marwan Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. <u>Canadian Journal</u> of physics, 58(8):1200–1211, 1980.
- [134] Axel D Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. Physical review A, 38(6):3098, 1988.
- [135] Chengteh Lee, Weitao Yang, and Robert G Parr. Development of the colle-salvetti correlationenergy formula into a functional of the electron density. Physical review B, 37(2):785, 1988.
- [136] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp). <u>Chemical physics letters</u>, 393(1-3):51–57, 2004.
- [137] Axel D Becke. A new mixing of hartree–fock and local density-functional theories. <u>The Journal</u> of chemical physics, 98(2):1372–1377, 1993.