Moleculaire dynamica simulaties om zeoliet nanogroei en gedrag van vloeistoffen te bestuderen

Molecular dynamics simulations as a tool for understanding zeolite nanogrowth and liquid structure

Marc Van Houteghem

Promotoren:
Prof. dr. ir. Veronique Van Speybroeck
Prof. dr. Michel Waroquier

Proefschrift ingediend tot het behalen van de graad van
Doctor in de ingenieurswetenschappen: toegepaste natuurkunde

Vakgroep Toegepaste Fysica
Voorzitter: Prof. dr. ir. Christophe Leys

Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar: 2013-2014

Dit onderzoekswerk werd uitgevoerd in het Centrum voor Moleculaire Modellering, onder leiding van prof. dr. ir. Veronique Van Speybroeck en prof. dr. Michel Waroquier.

Leden van de examencommissie

Voorzitter

prof. dr. ir. Luc Taerwe (Universiteit Gent)

Leescommissie

prof. dr. Michel Waroquier (Universiteit Gent, co-promotor)

dr. ir. An Ghysels (Universiteit Gent)

prof. dr. Denis Constales (Universiteit Gent)

prof. dr. Evert Jan Meijer (Universiteit Amsterdam)

Overige leden

prof. dr. ir. Veronique Van Speybroeck (Universiteit Gent, promotor)

prof. dr. Dimitri Van Neck (Universiteit Gent)

prof. dr. ir. Toon Verstraelen (Universiteit Gent)

prof. dr. ir. Karen Hemelsoet (Universiteit Gent)

prof. dr. Jan Ryckebusch (Universiteit Gent)

Contents

Nederlandstalige samenvatting

Moleculaire dynamica (MD) simulatie is een modelleringstechniek waarbij opeenvolgende configuraties van het moleculair systeem worden gegenereerd door de beweginsvergelijkingen van Newton te integreren. Deze bewerkingen worden uitgevoerd met behulp van computers. Een noodzakelijk ingrediënt in het recept van MD voor het integreren van de beweginsvergelijkingen van Newton zijn de krachten tussen de deeltjes. Deze krachten worden berekend vanuit de potentiële energie van het systeem. Het resultaat van een MD simulatie is een traject dat op elk tijdsstip de atomaire posities en snelheden van het moleculair systeem bevat. Daarenboven kunnen andere grootheden berekend worden, zoals de druk of temperatuur.

Er zijn veel beschikbare methodes om de potentiële energie van het systeem te bekomen. Het grootste onderscheid wordt gemaakt op basis van het al dan niet expliciet beschrijven van de elektronische structuur van het moleculair systeem. Indien dit niet wordt gedaan spreekt men van **moleculaire mechanica (MM)** methodes waarbij gebruik gemaakt wordt van krachtvelden die de potentiële energie van het systeem voorstellen door geparametriseerde analytische functies in functie van interne vrijheidsgraden. Als de elektronische structuur van de atomen in rekening wordt gebracht, wordt de potentiële energie berekend door het expliciet oplossen van het kwantummechanisch (**QM**) veeldeeltjessysteem.

In een MD simulatie worden een aantal microscopische toestanden beschouwd, bepaald door de bewegingsvergelijkingen en de beschrijving van de interactie tussen de atomaire kernen op basis van klassieke fysica of kwantummechanica. De relatie tussen *alle* mogelijke atomaire configuraties voor een gegeven verzameling van atomen en de overeenkomstige potentiële energie voor elke microtoestand wordt het **potentiële energie oppervlak (PES)** genoemd. De **rode draad doorheen deze thesis** is dat MD een krachtige techniek is om macroscopische grootheden te berekenen door middel van een complexe bemonstering van het PES, i.e. een voldoende aantal microscopische toestanden van het PES worden verkend door het traject van het systeem zodat statistisch gerechtvaardigde resultaten worden bekomen. In dit doctoraatswerk worden de vibrationele spectra van zeolietbouwstenen en eigenschappen die vloeistofstructuur karakteriseren (radiale distributiefuncties en structuurfactoren) berekend met resultaten afkomstig van MD simulaties. De afleiding van macroscopische grootheden vanuit microscopische informatie is gebaseerd op statistische fysica. Het Centrum voor Moleculaire Modellering (CMM) is actief in het gebied van het ontwikkelen van statistische programma's voor het analyseren van de resultaten van MD simulaties (*Paper IV*).

In het eerste deel van deze thesis wordt MM MD aangewend voor het analyseren van zeoliet nanogroei met behulp van **vibrationele spectra** (hoofdstuk 2, *Paper I*). Zeolieten zijn microporeuze anorganische materialen, meestal met aluminosilicaat componenten, die kristalstructuren vertonen met poriën en kanalen die groot genoeg zijn om de diffusie van kleine moleculen toe te laten. Als gevolg van deze structurele karakteristieken, hebben zeolieten wijdverspreide toepassingen als ionenwisselaars, adsorptiematerialen en katalysatoren. Wegens de directe link tussen de kristalstructuur en de fysische eigenschappen is er een grote interesse in de structurele chemie van zeolieten vanuit zowel de academische wereld als de industrie. Vibrationele spectroscopie is een generieke term voor verschillende types van spectra (infraroodspectra, inelastische neutronverstrooiingsspectra); de techniek is gebaseerd op **Fourier transformaties** die data van MD simulaties transformeert van het tijdsdomein naar het frequentiedomein. Spectroscopie kan gebruikt worden om zeoliet bouwstenen te karakteriseren. Voor het begrijpen van de transformatie van silica eenheden naar microporeus kristallijn materiaal, is vooral de correlatie tussen de grootte of de structuur van de silica eenheden belangrijk, alsook de positie van de absorptiepieken in het spectrum. We identificeren pieken in het vibrationele spectrum van zeoliet oligomeren en ringen met specifieke **interne coördinaten**, en focusseren ons op de invloed van de parameters **ringgrootte** en **topologie** op het vibrationele spectrum. Hiervoor werd een nieuw protocol ontwikkeld – de **atomaire snelheidsprojectie methode** – die ons in staat stelt om signalen met de interne vrijheidsgraden van de zeoliet nanodeeltjes te vereenzelvigen. Met deze methode is het mogelijk om te voorspellen welke specifieke modes worden beïnvloed door een veranderende grootte of topologie van silica nanodeeltjes die belangrijk zijn voor zeolietsynthese. Op die manier kan moleculaire modellering assisteren in het begrijpen van de mechanismen die plaatsvinden gedurende de initiële fases van de zeoliet nanogroei. De atomaire snelheidsprojectie methode wordt toegepast op silicaliet-1, en de interne coördinaten worden geconstrueerd op basis van de **rekbeweging** tussen twee atomen: de rekbewegingen zelf en lineaire combinaties ervan kunnen dienen als interne coördinaten. Het besluit is dat de analyse van vibrationele spectra zeer nuttig is om meer inzicht te verwerven in het beginstadium van zeolietsynthese op moleculair niveau omdat het spectrum van eigenmodes van interne coördinaten kan variëren bij verschillende moleculaire systemen. Op het einde van dit eerste deel wordt een afweging gemaakt van de MD versus de statische methodologie.

In het tweede deel wordt de **structuur van vloeistoffen** bestudeerd met QM MD (hoofdstuk 3, *Paper II* en *Paper III*). Hierbij wordt de potentiaal beschreven op basis van kwantummechanische methoden om ook niet-covalente interacties correct in kaart te brengen. Meer specifiek, maken we gebruik van **dichtheidsfunctionaaltheorie (DFT)**, die het veeldeeltjesprobleem formuleert in termen van de elektronendichtheid om de grondtoestandsenergie te berekenen, aangevuld met empirische correctietermen voor de dispersiekrachten. De invloed van deze correctietermen zal specifiek bestudeerd worden.

Vloeistoffen hebben geen vaste structuur zoals wordt waargenomen bij vaste stoffen. De intermoleculaire krachten, waarvan de bekendste de **waterstofbruggen** zijn, leiden tot karakteristieke ruimtelijke correlaties. Het is belangrijk de microscopische structuur van vloeistoffen in zo gedetailleerd mogelijk te kennen omdat deze ook de macroscopische eigenschappen bepaalt, bv. thermodynamische functies. Een uitstekende grootheid om de vloeistofstructuur te bestuderen is de **radiale distributiefunctie (RDF)**, wat een relatieve waarschijnlijkheidsdichtheid is om een atoom op een afstand van een ander atoom te vinden in de oorsprong. Initieel worden de RDFs van vloeibaar methanol met waterstofbruggen bestudeerd in simulaties waarvan het aantal deeltjes, het volume en de temperatuur constant zijn (**NVT**) (*Paper II*). Verder wordt er een correctieschema opgezet voor de **basis set superpositie fout (BSSE)**, en wordt de invloed van deze correctie op de RDF geëvalueerd. De BSSE is een gevolg van het gebruik van niet complete basissets waarin de atomaire en moleculaire orbitalen worden geëxpandeerd, en veroorzaakt een artificiële versterking van de intermoleculaire interacties en een artificiële verkorting van de intermoleculaire afstanden. Het nieuw voorgestelde **correctiemodel**, dat moet **geparametrizeerd** worden, bestaat uit een klassieke krachtveld energieterm die dient toegevoegd te worden in latere MD simulaties. Initieel wordt enkel de intermoleculaire waterstofbruginteractie in methanol gekalibreerd omdat geanticipeerd kan worden dat de BSSE het grootst zal zijn voor dit type van interactie. Aangezien het in rekening brengen van dispersie onontbeerlijk is om de RDF correct te reproduceren, wordt een gecombineerde studie van dispersie en BSSE uitgevoerd. Door de introductie van deze correcties, met het model voor BSSE en dispersie, worden de afwijkingen op de amplitude van de eerste piek van de relevante partiële RDFs van de atomen die deelnemen aan de waterstofbrug gereduceerd van 12%-16% tot 0.4%-1%, in vergelijking met het experiment. Dit toont aan dat DFT MD, gecorrigeerd voor BSSE, in staat is om de vloeistofstructuur van methanol in een NVT ensemble te reproduceren.

Gebaseerd op het werk op methanol, wordt de studie van vloeistofstructuur uitgebreid op vier niveaus (hoofdstuk 3, *Paper III*). De eerste uitbreiding bestaat erin dat we niet één, maar vijf vloeistoffen beschouwen: **methanol**, **chloroform**, **acetonitrile**, **tetrahydrofuraan** en **ethanol**. De keuze van deze solventen is gemotiveerd door de relevantie voor de industrie en hun classificatie (niet-polair, polair aprotisch, polair protisch). Van deze vijf komen waterstofbruggen enkel voor in methanol en ethanol. Ten tweede worden, naast NVT simulaties, ook simulaties uitgevoerd in een ensemble waarbij het aantal deeltjes, de druk en de temperatuur constant zijn (**NpT**). Veel experimenten worden uitgevoerd onder constante druk en temperatuur, zodoende dat simulaties in het isotherme-isobare ensemble dichter aanleunen bij experimentele data. De derde uitbreiding bestaat erin dat de parametrizatie van het correctiemodel voor BSSE wordt veralgemeend en toegepast op alle solventen. Op deze manier worden er drie MD simulaties per solvent uitgevoerd (NVT, NpT en NpT gecorrigeerd voor BSSE), waarvan telkens de RDFs bepaald worden en, indien mogelijk, worden vergeleken met RDFs die afgeleid zijn op basis

van experimentele data. De enige grootheid die experimenteel kan worden opgemeten is de totale **structuurfactor (SF)**, die gerelateerd is aan de RDF door middel van een **Fourier transformatie**. Alle andere grootheden, zoals de partiële RDFs, kunnen niet rechtstreeks opgemeten worden. Partiële RDFs worden bekomen door een inverse Fourier transformatie van de partiële SFs. De partiële SFs worden op hun beurt berekend door een iteratief protocol dat neutronverstrooiingsexperimenten combineert met MD simulaties gebaseerd op empirische interatomaire potentialen. Op die manier cumuleren RDFs, die worden bepaald op basis van experimentele data, onnauwkeurigheden die gegenereerd worden in de verschillende stappen van de procedure. De vierde en **belangrijkste uitbreiding** is de beoordeling van de vloeistofstructuur in de reciproke ruimte. Een Fourier transformatie die RDFs naar SFn converteert, en omgekeerd, vereist wiskundige bewerkingen maar de nauwkeurigheid ervan kan beter gecontroleerd worden. De studie van SFn spitst zich toe op tetrahydrofuraan omdat hiervoor de meeste experimentele data beschikbaar is. Speciale aandacht gaat uit naar het onderscheid van de **inter- en intramoleculaire contributies** in de totale SF, wat met experimentele methodes zeer complex is. In dit werk vinden we dat de intramoleculaire bijdragen de intermoleculaire interacties domineren in de totale SF. Gezien vanuit experimenteel oogpunt het niet eenvoudig is om de dominerende intramoleculaire fluctuaties te scheiden van het totale spectrum, is de experimentele data ontoereikend om op accurate wijze intermoleculaire partiële RDFs te berekenen. Deze thesis toont ook aan dat de dominantie compleet is in die zin dat de intramoleculaire interacties zeer groot zijn voor zowel grote als kleine golfgetallen. Verder wordt er aangetoond dat het signaal van de intermoleculaire verstrooiing snel klein wordt bij grotere golfgetallen. Dit alles suggereert dat de hoofdoorzaak van fouten de extractie is van de intermoleculaire totale SF van de totale gemeten SF. De afwijking tussen de theoretische en de experimentele SFn is vooral groot in het golfgetalgebied waar de intermoleculaire contributies zich manifesteren. Gezien de totale bijdrage van de intermoleculaire contributies klein is, zal een kleine relatieve fout in de (berekende) intramoleculaire SF niet onderscheidbaar zijn van een relatief grote fout in de intermoleculaire SF. De gevoeligheden van kleine perturbaties in de SFn op de RDFs worden onderzocht door middel van de drie types van MD simulaties. Het besluit is dat de SFn hoogst ongevoelig zijn aan veranderingen in de partiële intermoleculaire RDFs. Hierdoor wordt er expliciet aangetoond dat de afleiding van partiële intermoleculaire RDFs vanuit een klein aantal SFn, zoals het geval is bij modellen die RDFs afleiden uit experimentele data, slecht geconditioneerd is.

De **hoofdconclusie** van deze thesis is dat MD een **veelzijdige** en **krachtige** moleculaire modelleringstechniek is omdat het kan toegepast worden op zeer **diverse problemen** en in staat is om informatie te verschaffen die met andere technieken niet altijd kan bekomen worden. In deze thesis wordt MD toegepast op vaste stoffen (zeolieten) en vloeistoffen (solventen), twee zeer verschillende fases van de materie. Het zou zeer moeilijk zijn om het inzicht in de moleculaire mechanismen van de zeolietvorming, zoals besproken wordt in

hoofdstuk 2, te bekomen met experimentele technieken. Dezelfde conclusie kan getrokken worden in verband met het werk op de structuur van vloeistoffen, waarvan het onderscheid tussen de inter- en intramoleculaire contributies het meest prominente voorbeeld is.

English summary

Molecular dynamics (MD) simulation is a modeling technique in which successive configurations of the molecular system are generated by integrating Newton's equations of motion. These calculations are executed with computers. A necessary ingredient in the recipe of MD when Newton's equations are integrated, are the forces between the particles. These forces are computed from the potential energy of the system. The result of an MD run is a trajectory that contains at each time step the atomic positions and velocities of the molecular system. In adddition, other quantities can be calculated, such as pressure or temperature.

There is a plethora of methods available to obtain the potential energy of the system, and a main classification is made whether or not the methodology explicitely calculates the electronic structure problem for a given set of atomic coordinates. If this is not the case we refer to the methods as **molecular mechanics (MM)** methods which use force-fields to represent the potential energy of the system by parametrized analytical functions in function of internal degrees of freedom. If the electronic structure of the atoms is taken into account, the potential energy is computed by explicitely solving the many-body electronic structure problem using **quantum mechanics (QM)**.

In an MD run a given number of microscopic states are considered, determined by the equations of motion and by using a potential energy which is either described classically or based on quantum mechanics. The relation between *all* possible atomic arrangements for a given collection of atoms and the corresponding potential energy for each microstate is called the **potential energy surface (PES)**. The **central point of this thesis** is that MD is a powerful tool to compute macroscopic properties by means of a complex sampling of the PES, i.e. sufficient microscopic states of the PES are explored by the system trajectory to determine statistically justified macroscopic results. In this work, vibrational spectra of zeolite buildig blocks and properties characterizing liquid structure (radial distribution functions and structure factors) are calculated from the output obtained from MD simulations. The derivation of macroscopic properties from microscopic information is based on statistical physics. The Center for Molecular Modeling (CMM) is active in developing software to analyse output from MD simulations (*Paper IV*).

In the first part of this thesis MD, based on classical force fields is used to analyse **zeolite** nanogrowth with **vibrational spectra** (chapter 2, *Paper I*). Zeolites are microporous inorganic materials, mostly with aluminosilicate components, which exhibit crystal structures containing pores and cages lare

enough to permit the diffusion of small molecules. As a result of these structural characteristics, zeolites find widespread application as ion exchangers, adsorbents and catalysts. Because of the direct link between their crystal structure and physical properties, there is intense interest in their structural chemistry from both academia and industry. Vibrational spectroscopy is a generic term for several types of spectra (IR, INS); it is a **Fourier-transform** based technique which receives input from MD simulations and transforms information from the time domain to the frequency domain. Spectroscopy can be used to characterize zeolite building blocks. In the understanding of the transformation of silica species into microporous crystallites, the correlation between the size or the structure of the silica species and the position of the absorption peaks in the spectrum is especially important. We identify peaks in vibrational spectra of zeolite oligomers and rings with specific **internal coordinates**, and focus on the influence of the parameters **ring-size** and **topology** on the vibrational spectrum. For this purpose, a new protocol – the **atomic velocity projection method** – has been developed which enables to link signals with the internal degrees of freedom of zeolite nanoparticles. With this method it is possible to predict which particular modes are influenced by changing size and topology of the silica nanoporticles important for zeolite synthesis. As such, molecular modeling assists in the understanding of the mechanisms during the early stages of the zeolite growth. The atomic velocity projection method is applied to silicalite-1, and the internal coordinates are constructed from **stretches** between two atoms: the stretches themselves can serve as internal coordinates or linear combinations of them. Concluding, the analysis of vibrational spectra is a very useful tool in understanding the initial stages of zeolite formation, as the spectral behavior of eigenmodes of internal coordinates can vary with different molecular systems. MD can then be used to simulate various structures which are key components during zeolite growth. At the end of this first part, an assessment is made of the MD versus static methodology.

In the second part the **structure of liquids** is studied with QM MD (chapter 3, *Paper II* and *Paper III*). The potential is determined from quantum mechanical methods in order to also describe the non-covalent interactions correctly. More specifically, we will rely on **density functional theory (DFT)**, which formulates the many-electron problem in terms of the electron density to compute the ground-state energy, but will also add empirical corrections for the dispersion interactions. The influence of these correction terms will be studied specifically. Liquids do not have a structure in the sense that solids have a structure. The intermolecular forces, of which the most famous is the **hydrogen bonding**, lead to characterisitic spatial correlations. It is important to know the microscopic structure of liquids in as much detail as possible because it determines their macroscopic properties, e.g. thermodynamic functions. An excellent tool to study liquid structure is the **radial distribution function (RDF)**, which is a relative probability density of finding an atom at a distance from another atom at the origin. Initially, we study the RDFs of the hydrogen-bonded liquid methanol in simulations where the number of par-

ticles, the volume and the temperature are kept constant (**NVT**) (*Paper II*). Moreover, a correction scheme is set up for the **basis set superposition error (BSSE)**, and evaluate the influence of this correction on the RDF. The BSSE results from using an incomplete basis set for expanding the atomic and molecular orbitals, causing an artificial strengthening of the intermolecular interactions and artificial shortening of intermolecular distances. The new proposed **correction model**, which needs to be **parametrized**, is a classical force-field energy term and can be added in subsequent MD simulations. At first, only the intermolecular hydrogen bond interaction in methanol is calibrated, as it is expected that the BSSE will be the largest for this type of interaction. Since taking into account **dispersion** is indispensable in any attempt to correctly reproduce the RDF, a combined study of BSSE and dispersion is performed. Upon introduction of these corrections (model for BSSE and dispersion), the discrepancies between theory and experiment on the first peak height of the relevant partial RDFs of the atoms that participate in hydrogen-bonding are reduced from 12%-16% to 0.4%-1%. This shows that accurate DFT MD corrected for BSSE is able to reproduce the liquid structure of methanol in an NVT ensemble.

Based on the work performed on methanol, the study of liquid structure is extended on four levels (chapter 3, *Paper III*). First, we consider five liquids: **methanol**, **chloroform**, **acetonitrile**, **tetrahydrofuran** and **ethanol**. The choice of these solvents is motivated by their importance for the industry and their classification (non-polar, polar aprotic, polar protic). Of these five cases, hydrogen bonding occurs only in methanol and ethanol. Second, besides the NVT MD runs, the simulations are performed in an ensemble where the number of particles, the pressure and the temperature are constants (**NpT**). Many experimental measurements are made under conditions of constant temperature and pressure, and so simulations in the isothermal-isobaric ensemble are most directly relevant to experimental data. The third extension is the generalisation of the parametrization procedure of the correction model for BSSE and the application to all solvents. This way, we arrive at three MD simulations per solvent (NVT, NpT and NpT corrected for BSSE), of which all the RDFs are determined and, where possible, compared with RDFs derived from experimental data. The only property that is a direct experimental observable is the total **structure factor (SF)**, which is related to the RDF through a **Fourier transform**. All other quantities, including the partial RDFs, are not directly available from experiment. Partial RDFs are obtained after an inverse Fourier transform of the partial SFs. These latter properties are in turn determined form an iterative protocol combining neutron scattering experiments with isotopic substitution and MD runs based on empirical interatomic potentials. As such, RDFs derived from experiment are not free from cumulating inaccuracies generated in the various steps in the procedure. The fourth and **most important extension** is the assessment of liquid structure in reciprocal space. A Fourier transform to convert RDFs to SFs, and vice versa, requires some mathematical manipulations but the accuracy of these computations is

better controlled. The work on SFs focuses on tetrahydrofuran as a case study, as most benchmark data are available in this case. Special attention is given to disentangling the total SF into **inter- and intermolecular contributions**, which experimentally poses a lot of complications. We find that the intramolecular contributions largely dominate the intermolecular interactions to the total structure factor. As from experimental point of view the subtraction of the dominant intramolecular fluctuations from the total spectrum is not a trivial task to perform with a high accuracy, the experimental data is incomplete to extract a correct solution. Furthermore, this thesis reveals that the dominance is complete in the sense that the intramolecular terms cover the whole momentum range, including the low wave vector part where we expected that it is mainly controlled by intermolecular contributions. Another aspect revealing from our computation, is that the signal from intermolecular scattering fades out quite fast with increasing momentum transfer. These facts suggest that the first and probably the most crucial source for generating errors is the extraction of the intermolecular total structure factor from the total measured scattered amplitude. The discrepancy between theoretical and experimental SFs is large, especially in the wave vector part in which the intermolecular contribution manifests itself. Since the total contribution of the intermolecular RDF to the SF is small, a small relative error in the (computed) intramolecular SF will not be distinguishable from a relative large error in the intermolecular SF. The sensitivities of small perturbations in the SFs on the RDFs are examined by means of the three types of MD simulation. The conclusion is that SFs are almost insensitive to changes in partial intermolecular RDFs. This way, we explicitely show that the derivation of partial intermolecular RDFs from a small number of SFs, as is the case with models that derive RDFs from experimental data, is ill-conditioned.

The **main conclusion** of this thesis is that MD is a **versatile** and **powerful** molecular modeling technique as it can be applied to a very **diverse set of problems** and is able to provide **information otherwise unattainable**. In this thesis, MD has been applied to solid state materials (zeolites) and liquids (solvents), two very distinct forms of matter. The insights into the molecular mechanisms of structuring of silica nanoparticles, as discussed in chapter 2, would be very hard to obtain with experimental techniques. The same conclusion can be drawn concerning the work on the structure of fluids, the disentanglement of the inter- and intramolecular contributions being the most prominent example.

List of abbreviations

BO	Born-Oppenheimer
BSSE	basis set superposition error
CGF	contracted gaussian function
CMM	center for molecular modeling
COK	centrum voor oppervlaktechemie en katalyse
CP	counterpoise
DFT	density functional theory
EtOH	ethanol
FF	force field
GPW	gaussian- and plane waves
GTO	gaussian type orbital
HF	Hartree-Fock
IC	internal coordinate
INS	inelastic neutron scattering
IR	infrared
IZA	international zeolite association
LCAO	linear combination of atomic orbitals
MD	molecular dynamics
MeCN	acetonitrile
MeOH	methanol
MM	molecular mechanics
NDIS	neutron diffraction with isotopic substitution
NMA	normal mode analysis
NMR	nuclear magnetic resonance
PES	potential energy surface
PW	plane wave
QM	quantum mechanical
RDF	radial distribution function
SF	structure factor
TCM	chloroform
THF	tetrahydrofuran
ZSM-5	zeolite socony mobil-five

I
Molecular dynamics simulations as a tool for understanding zeolite nanogrowth and liquid structure

1

General Introduction and goal of this thesis

1.1 Computer simulations as a tool for studying molecular systems

Each paper that passed through my hands during the last four years, had a common message expressed in different versions: 'Computer simulation has become a powerful tool for obtaining detailed insights into the structure, dynamics, and function of molecular systems.'

Computer simulations act as a bridge between microscopic length and time scales and the macroscopic world of the laboratory: we set up a model for the interactions between atoms and molecules, and obtain predictions of bulk properties. As such, a direct route is provided from the microscopic details of a system (charges and masses of the atoms, the interactions between them, molecular geometry, etc.) to macroscopic properties of experimental interest (thermodynamic properties, information about the structure, diffusion coefficients, and so on). There are only a handful of non-trivial, exactly soluble problems in statistical mechanics. Exactly soluble means that a complete specification of the microscopic properties of a system leads directly to a set of interesting results or macroscopic properties. Essentially exact results can be obtained for a given model system without having to rely on approximate theories. If the calculated properties of a model system are not satisfactory, the model is inadequate and the estimate of the intermolecular interactions needs to be improved.

Simulations act as a bridge in another sense: between theory and experiment. We may test a hypothesis by conducting a simulation on a molecular system using a model based on that theory, comparing the outcome with experimental results.

We may also carry out simulations on the computer that are difficult or impossible in the laboratory, e.g. experiments under extremes of temperature and pressure, while a computer simulation of, say, materials with conditions as they appear at the core of the earth, would be perfectly feasible. Quite subtle details of molecular motion and structure might be difficult to probe experimentally, but can be extracted readily from a computer simulation. In addition, we can use computers to predict the properties of materials that have not yet been made.

Ultimately we may want to make direct comparisons with experimental measurements made on specific solid state materials, liquids or gases, in which case a good model of the molecular interactions is essential. In any case we want to reduce the amount of fitting and guesswork in this process to a minimum. On the other hand, we may be interested in phenomena of a rather generic nature, or we may simply want to discriminate between good and bad theories. However, in many cases, it is not necessary and even not realistic to strive for a perfect molecular model; we should temper our ambitions in developing a model that suits our goal, containing the essential physics to solve our specific problem.

Computer experiments have become standard practice, and not only in the field of studying molecular interactions but to the extent that they now provide the first and the last test of a new theoretical result. However, as experienced over the last four years, results from the computer as such offers no understanding, only numbers. The useful information has to be extracted from it, which is the easy part. Actually knowing which information you want to obtain and interpret it correctly is the hard part, and a form of art on its own.

1.2 Molecular modeling

Molecular modeling is the science of studying molecular structure by means of an extensive number of theoretical and numerical models. The computations encompass ab initio and semi-empirical quantum mechanical methods, molecular mechanics, path-integral techniques, Monte Carlo, free energy and solvation methods, (bio)chemical information and databases, plus many other established procedures. The broad research field of modeling molecular systems by computers is truly multidisciplinary. It is used by biologists who describe the cellular picture, chemists who fill in the atomic and molecular details, physicists extend these views to the electronic level and the underlying forces, mathematicians analyze and formulate appropriate numerical models and algorithms, and computer scientists provide the support for running large computer programs [1]. Combining these disciplines is something the researcher in molecular modeling can only hope to achieve after years of experience in the field.

Although molecular models represent a highly simplified version of the *real*

molecular environment, they are able to discern patterns and add insights that we otherwise are unable to notice. However, the art lies in the fact that the developed and applied models have to be appropriate for the problems that we are faced with.

1.3 Molecular dynamics: the method explored in this thesis

In this thesis the focus lies on studying molecular systems with the help of molecular dynamics simulations (MD). Ever since the first MD simulation by Adler and Wainwright [2], the number of different techniques, methods and models for this type of computer simulation and the number of applications for it, grows extensively. And, although this branch of science is still relatively new, MD has evolved into an important and widely used theoretical tool that allows researchers in chemistry, physics and biology to model the behavior of many different types of systems, including gases, liquids, solids, surfaces and clusters.

What is MD? In an MD simulation, the classical equations of motion governing the microscopic time evolution of a many-body system are solved numerically, i.e. the force between the particles is computed allowing the integration with Newton's equations of motion, $m_i \ddot{\boldsymbol{r}}_i(t) = \boldsymbol{F}_i = -\nabla_i V(\boldsymbol{r}^N)$ [3], to move particles to new positions and identify the velocities at these locations. The MD methodology can provide a window into the microscopic dynamical behavior of the individual atoms that make up a molecular system. From this information, the microscopic mechanisms can be 'observed' and dynamical properties can be calculated, e.g. absorption spectra, reaction rate constants, and transport properties. In addition to providing a dynamical picture, MD can also be employed as a means of sampling from a statistical mechanical ensemble and determining equilibrium properties, e.g. average thermodynamic quantities (pressure, temperature, etc.) or structural functions like the radial distribution function.

Central theme of this thesis. In the MD-technique, a sufficient region of phase space, i.e. the $6N$-dimensional space spanned by all particle coordinates and momenta for a system of N particles, is explored by the system trajectory to compute a macroscopic property. One can look at MD as a dynamic computation, where a given number of successive microstates are considered, determined by the equations of motion and the description of the particle interaction – with quantum mechanics and/or classical physics – under consideration. The quality of the results of the MD simulation then depends on the accuracy of this description. The conversion of the very detailed information of many atomic positions and momenta to derive macroscopic properties is the merit of statistical physics. The basic concepts of this theory are outlined in Chapter 2 of the standard work of Allen and Tildesley [4]. On the other hand,

thermodynamic quantities can also be obtained based on static computations, considering only one point on the potential energy surface (PES) – the PES is a hypersurface defined by the potential energy of a collection of atoms over all possible atomic arrangements. Typically, this approach is employed when the computational cost is rather high. This may be the case for very large systems or when an electronic structure method is used at a very high level of theory. However, the method considers one point only in the potential energy landscape or the potential energy surface. This is, from the point of view of statistical physics, rather limited. For example, when a system has multiple important conformers, then it is a good practice to sample more than just one of these configurations to arrive at a more statistically justified result. Macroscopic properties are always ensemble averages over a representative statistical ensemble, so the knowledge of a single structure, even if it is the structure of the global energy minimum, is then often not sufficient. Instead, the final result is more reliable when a representative ensemble at a given temperature is generated, taking into account multiple local minima on the PES, to compute macroscopic properties. In this perspective, the **central theme of this PhD thesis** is

> to go beyond the concept of considering one point of the PES, as in static point methods, by means of a more complex sampling of phase space using the MD technique.

MD as a tool for studying zeolite nanogrowth. Firstly, MD is used to compute vibrational spectra of silica oligomers and rings, which play an important role in zeolite synthesis. This research was inspired by the long term collaboration between the CMM and the COK. The objective of this project is to enhance the interpretation of vibrational spectra, based on a molecular dynamical approach, to obtain a better understanding of the mechanisms present during the early stages of the zeolite growth. For this an efficient protocol, based on linking spectra patterns with atomic motions, is developed, as explained in Chapter 2 (*Paper I*, *Paper IV*). Two methods are conventionally used for the calculation of vibrational spectra linking spectral patterns with atomic motions: normal-mode analysis (NMA) and MD simulations. The first technique is based on a static approach and the latter is a Fourier-based technique which receives input from MD simulations. NMA is computationally less demanding but is restricted to the harmonic approximation which allows only small deviations from a local or global minimum on the PES. The restriction stems from the fact that NMA is in essence a second-order approximation of the minima in the PES. As a consequence, only one minimum can be calculated and only a small part of the energy surface is considered, which limits the technique when multiple minima are present. By applying MD, this problem is then circumvented. The results of this part are described in chapter 2 of this thesis.

MD as a tool for studying liquid structure. Secondly, MD is used to compute radial distribution functions (RDFs) in methanol (*Paper II*) and

a set of classes of other important solvents (*Paper III*), which is illustrated in Chapter 3. The RDF is a histogram of discretized pair separation distances which shows how many pairs there are in each interval. Here, a pair can be e.g. a pair of atoms or molecules. In the second part of Chapter 3 (*Paper III*), the structure factor (SF) is introduced as a complementary tool to interpret RDFs better. A SF is related to the RDF through a Fourier transform, and in this work SFs are employed to analyse the accuracy of models predicting or extracting information on liquid structure. The RDF and SF are structural properties of the fluid that do not depend on the time evolution of the system, i.e. they are static equilibrium averages. Such an average can be accomplished by MD simulation, and in this perspective it is obvious that this is not achievable by static point methods. The accurate derivation of RDFs is not straightforward and within the framework of this thesis some models are introduced to describe accurately RDFs using molecular dynamics simulations.

Versatility of MD. In the work related to zeolites, data from MD simulations are used to gain insight in the generated spectra. Simulations are done with molecular mechanics (MM) and the MM-model applied as input for the MD simulation of the zeolite building blocks is not investigated. On the contrary, in the work on RDFs the computer simulations were completed using a quantum mechanical approach, and the input for the MD simulation is altered in order to improve the agreement with experimental RDFs. However, regardless whether the focus is on analyzing data from an MD simulation (*Paper I*, *Paper IV*), or rather improving the simulation itself (*Paper II*, *Paper III*), MD remains the central theme throughout this whole doctoral dissertation. In Chapter 2, the power of this approach is contrasted with the limited scope of static methods, as used in *Paper V* where NMR- and Raman spectra are computed.

1.4 MM and QM methods

The computational methods that are used to do MD on a molecular system are divided in two main categories: (i) MM methods that do not explicitly consider electrons in the system, also commonly referred to as force-fields (FF), and (ii) quantum mechanical (QM) methods (or ab initio methods) that explicitly consider electrons in the systems, which have their roots in the theory of quantum mechanics, and involve the solution of the Schrödinger equation at some level of approximation. In both approaches it is assumed that the Born-Oppenheimer (BO) approximation is valid and that the atomic nuclei can be treated as classical point particles. There are also computational methods that are a combination of MM- and QM methods, they are mostly referred to as QM/MM-methods.

In both the MM as well as in the QM classes, there is a plethora of possibilities available. All of them have their advantages and disadvantages, but the

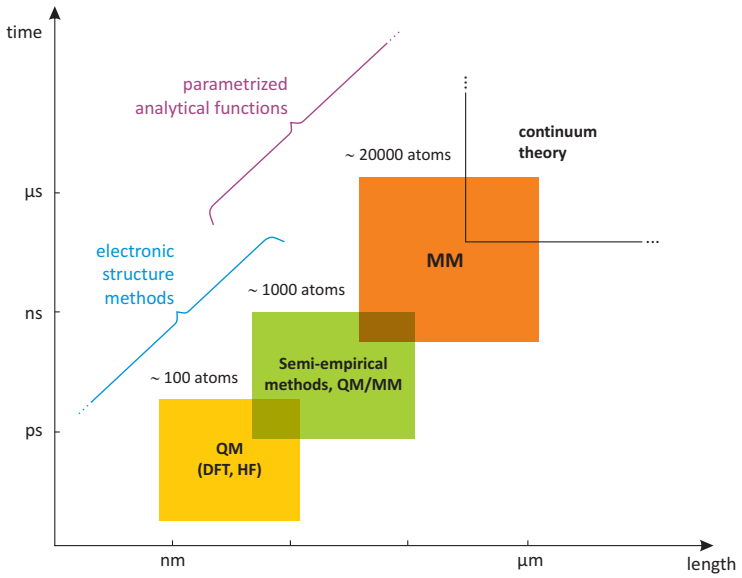


Figure 1.1: Length-time diagram for QM and MM methods.

accessible time- and length-scale covered by MD, as shown in Fig. 1.1, is always important. It is clear that the more detailed a simulation technique operates, the smaller the accessibility of long times and large length scales becomes. With quantum simulations the electronic structure problem is solved at every time step of the simulation and is therefore very expensive. It may only be applied for smaller systems or for phenomena that occur on smaller time scales. FFs approximate electronic distributions in a rather coarse-grained fashion by assigning an analytical form to the potential energy expression in which atomic charges represent the electrostatic interactions. The required/targeted accuracy is a factor which must be kept in mind at all times. Also polarization effects may be added. More information on how FFs are constructed is given in the next section. No single method of calculation is likely to be ideal for all applications, and it depends on the pursued objectives and the molecular system itself which method is best suited.

In the preceding discussion, it has been assumed that the particles can be treated as classical point particles, according to Newton's equations of motion. This treatment is usually justifiable, but for systems where hydrogen/proton motion is important, e.g. at low temperature, such an approximation is not valid. Whenever these so called nuclear quantum effects may be important, a quantum formalism, such as the Feynman imaginary time path integrals [5], is required. This type of MD is in general not performed at the CMM, but it is important to mention this issue because the same topic of *Paper II* and *Paper*

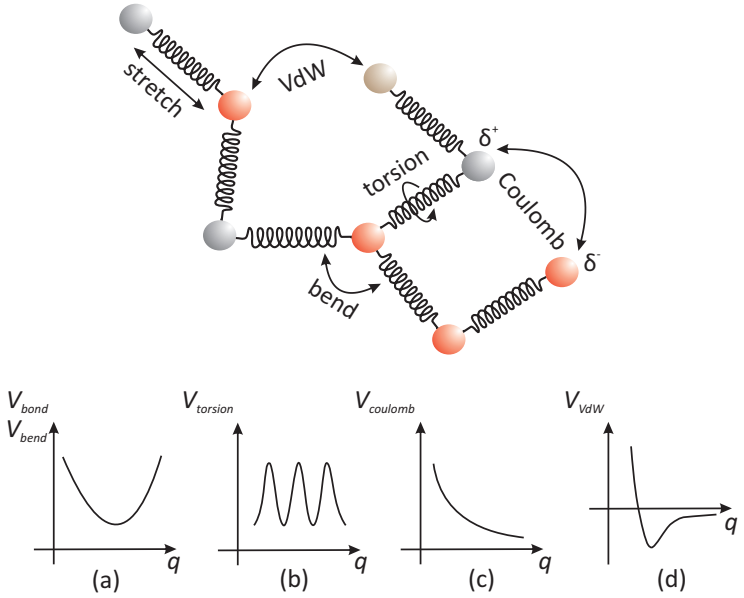


Figure 1.2: Pictorial representation of the contributing potential energy terms in a common force-field.

III has also been investigated by path integral MD. [6–9]

1.4.1 MM-methodology

In *Paper I*, FFs have been used extensively, therefore a general outline on how FFs are constructed is presented here. MM is based on interatomic potentials, which are parametrized analytical functions that represent the potential energy of a system as a function of internal coordinates q. A FF is the analytical representation of the potential energy of a system in a given conformation as a sum of terms which account for different molecular interactions (Fig. 1.2): mechanical-like strains arising from equilibrium bond length and angle deviations, torsion flexibility, Van der Waals and electrostatics. A typical FF has the following form:

$$V_{\mathrm{FF}} = V_{\mathrm{covalent}} + V_{\mathrm{non\text{-}covalent}}, \tag{1.1}$$

with

$$V_{\mathrm{covalent}} = V_{\mathrm{bond}} + V_{\mathrm{bend}} + V_{\mathrm{torsion}} \tag{1.2}$$

$$V_{\mathrm{non\text{-}covalent}} = V_{\mathrm{VdW}} + V_{\mathrm{Coulomb}} \tag{1.3}$$

The contributions to the potential energy have the following mathematical form:

$$V_{\text{bond}} = \sum_{i=1}^{N_r} \frac{1}{2} K_{r,i} \left(r_i - r_{0,i}\right)^2 \tag{1.4}$$

$$V_{\text{bend}} = \sum_{j=1}^{N_\theta} \frac{1}{2} K_{\theta,j} \left(\theta_j - \theta_{0,j}\right)^2 \tag{1.5}$$

$$V_{\text{torsion}} = \sum_{k=1}^{N_\phi} \frac{1}{2} K_{\phi,k} \left[1 - \cos\left(m_k \left(\phi_k - \phi_{0,k}\right)\right)\right] \tag{1.6}$$

$$V_{\text{VdW}} = \frac{1}{2} \sum_{i \neq j}^{N} \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] \tag{1.7}$$

$$V_{\text{Coulomb}} = \frac{1}{2} \sum_{i \neq j}^{N} \frac{Q_i Q_j}{4\pi\epsilon_0 r_{ij}} \tag{1.8}$$

Here, N is the total number of atoms in the system, Q_i is the charge of atom i and r_{ij} is the distance between atom i and atom j, N_r is the total number of stretch bonds, N_θ the number of bending angles, and N_ϕ the number of dihedral angles. There are several parameters in these equations which need to be determined: force constants ($K_{r,i}$, $K_{\theta,j}$, $K_{\phi,k}$), rest values ($r_{0,i}$, $\theta_{0,j}$, $\phi_{0,k}$) and multiplicity factors (m_k).

The potential terms V_{bond} and V_{bend} have a generic form as in Fig. 1.2(a) (Eq. (1.4) and Eq. (1.5)), V_{torsion} like Fig. 1.2(b) (Eq. (1.6)), V_{Coulomb} commonly looks like Fig. 1.2(c) (Eq. (1.8))because of the $1/r$-factor and V_{VdW} typically shows behaviour as in Fig. 1.2(d) (Eq. (1.7)) due to the combination of the attractive r^{-6}- and the repulsive r^{-12}-contributions. An important property of such a model is that the topology (how the atoms are connected) is fixed and that chemical reactions can not take place. Currently also reactive FFs are developed but this is out of the scope of this thesis. [10]

1.4.2 QM-methodology

In *Paper III* and *Paper IV*, MD is studied based on quantum mechanical methods. This technique known as ab initio MD uses forces calculated from the potential energy, which is computed with first principles methods. This is sometimes referred as 'on the fly' electronic structure calculations. Up to some years ago this methodology was not applicable as the true bottleneck relies in the calculation of the PES. However, recent advances in electronic structure theory as well as readily available high-speed computers have begun to upgrade the ab initio MD approach as a viable tool. The most important element in an

ab initio MD calculation is the representation of the electronic structure. Calculation of the exact ground-state electronic wavefunction is intractable, and approximations must be used in some way. The electronic structure theory employed should be reasonably accurate, yet not too computationally demanding for the current application. Ab initio MD is not tied to any particular approach and the strength or weakness of a particular ab initio MD scheme is intimately connected to the strength or weakness of the chosen electronic structure method. Over the years a variety of different approaches such as density functional theory [11, 12] (DFT), Hartree-Fock (HF), post-HF (e.g. Møller-Plesset 2 (MP2)), semi-empirical, etc., were combined with MD, and this list is certainly incomplete. Density functional theory has become a very attractive method in computational chemistry as it allows to go beyond Hartree-Fock accuracy, provided a good functional is chosen. On the other hand this method is still computationally feasible. For a detailed explanation of this approach we refer to standard textbooks. [13–15]

1.5 Software

Computational chemistry is performed using both available program packages such as Gaussian, CP2K, ORCA, ... and in-house developed software. The CMM has developed a set of software tools which may be found on the website and are freely available within an academic environment. In this thesis, CP2K is extensively used, and to a lesser extent Gaussian is employed.

Computationally expensive routines like molecular simulations need to be run with software that is written in compiled languages (Fortran, C, C++, Java, etc.) because the execution speed of the software is of essential importance. The program CP2K [16] is an example of such software. For all the rest, i.e. preparing input, post-processing data, scripting, etc., a high-level language will do just fine. The latter is slower in performance because they are not compiled all the way down to binary machine code. However, a high-level language, e.g. Python, is easier to use and aims at a more hands-on approach.

1.5.1 CP2K

CP2K is a program to perform atomistic and molecular simulations of solid state, liquid, molecular, and biological systems. It is an open-source project hosted at `http://www.cp2k.org` and is written in Fortran 90. It is a very modular program, and the goal of CP2K is to provide an infrastructure allowing the user to perform various kinds of atomic simulations. The vast majority of simulations done for this dissertation were completed using CP2K, although other standard computational chemistry software have also been used (*Paper IV*), e.g. Gaussian [17, 18].

Some key features of CP2K are: a variety of model Hamiltonians (classical, semi-empirical, DFT and combinations (QM/MM)), MD and Monte Carlo, various statistical ensembles (NVE, NVT, NpT), free energy and PES tools, computation of various properties (vibrational, NMR, EPR (Electron Paramagnetic Resonance), XAS (X-ray absorption Spectroscopy)).

1.5.2 Python

Python is an interpreted, interactive, object-oriented programming language. It is commonly defined as an object-oriented scripting language – a definition that blends support for object-oriented programming with an overall orientation toward scripting roles. A whole chapter can be devoted on the reasons for using Python (Chapter of 1 of ref. [19]):

- It is easy to learn and the syntax is clear.
- The language is hands-on (practical matters like extracting data from an output file becomes easier) and more advanced at the same time (object-oriented programming), allowing to stick with just one language for a very broad range of applications.
- There are a number of specialized libraries available for Python. The most important ones for this work are NumPy, SciPy and MatPlotLib. These extensions can be invoked by Python at anytime when there is need for a number-crunching component that runs at C speed. This way, Python can act as a control language when these compiled extensions are linked into the Python program.
- The software developed at the CMM is usally written in Python. The compatibility between the new Python program and the available CMM-software [20] is then enhanced by a vast amount. Especially the MolMod library, the Yaff library, the MD-TRACKS package [21] and the ZEOBUILDER GUI [22] toolkit have been used extensively in this work.

1.5.3 MD-Tracks

MD-TRACKS (*Paper IV*) was written by Toon Verstraelen [23]. The project was initiated little before my arrival at the CMM and was finalised during my master thesis. My contribution to MD-TRACKS was mainly testing the code, and reporting bugs and unexpected behaviour.

MD simulations can produce a massive amount of output – the trajectory data. This data includes the time-dependent atomic coordinates, forces, cell parameters, atomic velocities, derivatives of the dipole moment, different types of energies, etc. MD-TRACKS is a collection of analysis scripts to process this vast quantity of data and to perform advanced statistical operations.

The analysis is performed with commands that operate on a binary trajectory database. The software is compatible with multiple simulation codes and consecutive commands can be cominbed.

MD-TRACKS was employed extensively during the course of my PhD, for inheriting low-level source code and as a user. MD-TRACKS has been proven very useful numerous times, although recently focus has shifted towards already existing formats for processing large amounts of numerical data, HDF5 being the most prominent one.

MD-TRACKS led to a peer-reviewed publication [21] (*Paper IV*). My contribution was mainly concentrated on testing and validating the source code and tracing bugs.

2

MD as a tool for studying zeolite nanogrowth

2.1 Introduction on zeolites

The history of zeolites began some 250 years ago when the Swedish scientist Axel Fredrik Crönsted discovered a mineral that swelled when exposed to heat. This new family of minerals was called zeolite, derived from the Greek words "zeo" (to boil) and "lithos"(stone). Zeolites are aluminosilicates where connected SiO_4 and AlO_4 tetrahedra give rise to microporous crystalline structures containing cavities and channels of molecular dimension. The Si and Al atoms are often indicated with the symbol T. An empirical formula for an aluminosilicate zeolite is $M^{n+}_{x/n}Al_xSi_{1-x}O_2 \cdot yX$, where x can vary from 0 – 0.5 and M^{n+} are inorganic or organic cations and X represents neutral guest molecules or included species. [24–27] Upon introduction of Al into the silica framework, the $Al^{3+} \leftrightarrow Si^{4+}$ substitution gives an overall negative charge to the framework, and requires the presence of extraframework cations within the structure to keep the overall framework neutral.

Zeolite structures are designated by a 3 capital letter code according to rules set by the Commision of the International Zeolite Association (IZA). The 5$^{\text{th}}$ edition of the *Atlas of Zeolite Framework Types* [28], published by the IZA describes 133 zeolite structures, and regular updates are found on the website of the IZA [29]. The codes are generally derived from the names of the type materials, e.g. the code for the **MFI** structure (topology) stems from the abbreviation ZS**M**-5 (**fi**ve), where this latter name originates from the full name **Z**eolite **S**ocony **M**obil-five. The ZSM-5 zeolite is made of several pentasil units linked together by oxygen bridges, where a pentasil unit consists of five-membered rings (Fig. 2.2). While a first distinction between different zeolites can be made based on the topology, another criterion is the chemical

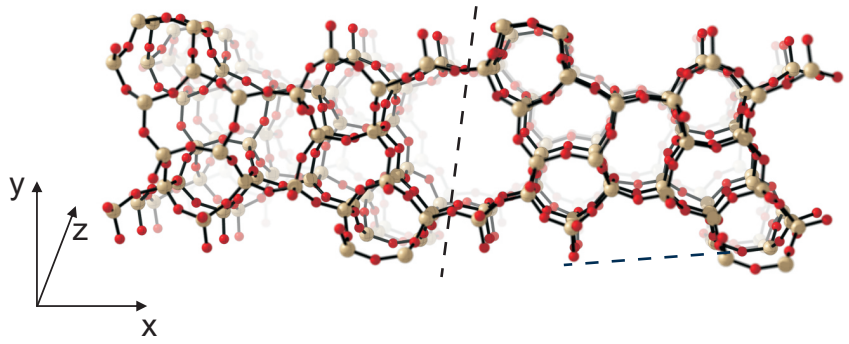


Figure 2.1: Two silicalite-1 unit cells, where the unit cell is repeated in the x-direction. The distinction between the two unit cells is indicated by the dashed black vertical line. If several of these unit cells are linked together, a three-dimensional nanoporous network of channels arises. The blue horizontal dashed line indicates the possible formation of another channel if a another unit cell is connected. It is the formation of such channels, amongst others properties, that makes the zeolite material very interesting for applications, e.g. for diffusion of guest molecules.

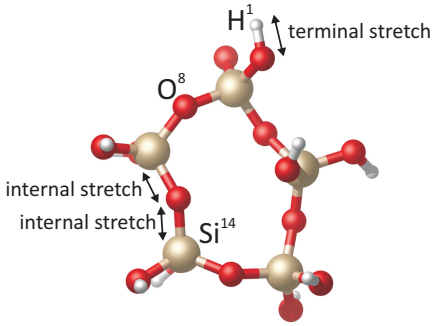


Figure 2.2: Optimised five-membered silica cluster – the fundamental building block for the MFI-type zeolite.

composition of the framework. For example, both zeolites ZSM-5 and silicalite-1 belong to the so-called pentasil group. The difference between ZSM-5 and silicalite-1 is the aluminum content, which is nihil for the latter, i.e. silicalite-1 is the full silica version of the aluminosilicalite ZSM-5.

The field of applications of zeolites is very broad and the depth of its present knowledge varies substantially. They are extensively used in a variety of applications which may be classified primarily as follows: adsorption, catalysis and ion exchange. The number of application fields is immense: oil refining, petrochemistry, biogas industry, construction, heating and refrigeration, waste-,litter- and sewage-treatment, aquaculture, nuclear-, mining- and steel-industry, agricultural use, animal food industry, (bio)medical applications, radioprotection and optical electronics, the list can go on. [30–32] Moreover, in every domain there are again numerous applications. An attempt to sum them all would result in an incomplete list, but a good start is the *Handbook of Zeolite Science and Technology* [24], and references therein.

2.2 Synthesis of zeolites

Zeolite scientists have been, and still are, particularly creative: so far some 40 zeolites are discovered in nature, and approximately 130 zeolites are synthesized [30]. The opportunity to create zeolitic structures with different framework topology and different chemical composition allows to broaden their application in more and more diverse areas. Zeolites are synthesized in aqueous media with temperatures between 100 °C and 200 °C in the presence of organic and/or inorganic cations and a mobilizing agent. The crystallization process is determined by a large number of variables, and influences the kinetics and the final phases formed. Despite the efforts to fully comprehend the synthesis of zeolites, the connection between the initial conditions and the final structure formed is not yet fully understood due to the complexity of the synthesis mechanism and the metastable nature of zeolites [33]. Reviews on the synthesis of zeolites are given in refs. [25, 26, 34–38].

The formation of silicalite-1 is a frequently studied process, and Fourier transform infrared (IR) spectroscopy has been a key method to study colloidal solutions and solid samples. The most important characteristic in these investigations is the adsorption band at $550\,\mathrm{cm}^{-1}$ in the (final) silicalite-1 crystal. This signal is associated with the presence of five-membered rings in pentasil zeolites, as the signal is assigned to five-ring vibrations [39], and is regarded as the spectroscopic signature of the MFI-type zeolite (MFI “fingerprint”).

Important contributions in the field of zeolite formation have been made by the Center for Surface Chemistry and Catalysis (COK). The COK is a research group that focuses on experimental work and was an important partner for the work done in *Paper I* and *Paper V*. Based on the clear solution synthesis of silicalite-1 [40], the COK performed IR spectroscopy measurements on clear

solution mixtures and extracted nanoparticles [41–44]. In the mixture (in situ) the adsorption band initially occurs at $650\,\mathrm{cm}^{-1}$, but shifts and broadens to $600\,\mathrm{cm}^{-1}$ with time, and for the final silicalite-1 crystal the MFI fingerprint at $550\,\mathrm{cm}^{-1}$ is observed. Investigation on the extracted nanoparticles (ex situ) learns that a shift toward higher wavenumbers in parallel with a decrease in particle size takes place. Infrared measurements on two different populations of unique nanoparticles were done: (i) on so-called precursors (a discrete multiple of an elementary unit of specific size, between 33 and 35 Si atoms) and (ii) on nanoslabs, which are composed of ca. 400 Si atoms. For the former an absorption band at $590\,\mathrm{cm}^{-1}$ is found, while for the latter a high IR activity at $570\,\mathrm{cm}^{-1}$ is seen. After condensation of the nanoslabs into a crystalline structure, the band at $550\,\mathrm{cm}^{-1}$ is again recovered.

2.3 A better understanding of the synthesis of zeolites with theoretical spectra

Given the benefits zeolites offer to modern day society, understanding how zeolites nucleate and grow is of fundamental scientific and technological importance. Moreover, insight into the molecular mechanisms of structuring of silica can lead to the development of new hierarchical materials.

Based on the experimental work by the COK, a paper was published in 2008 by the CMM, where molecular dynamics was used to study infrared spectra from the initial stages of self-organization of elementary building blocks through to nanosized silicalite-1 crystals [45]. A combined experimental-theoretical effort makes it possible to search for an unambiguous fingerprint in the IR spectra once crystallinity appears. New is the introduction of the peak shift as a measure for nanoparticle assembly: at what stage(s) does a shift occur and why? Focus lies on the formation of silicate-1 and five-membered rings. In this paper it is found that, as the particles become bigger in size, the fingerprint band lowers in frequency from $650\,\mathrm{cm}^{-1}$ to $550\,\mathrm{cm}^{-1}$, as shown in Fig. 2.3. An isolated five-membered ring (5T) reveals an IR-active vibration around $650\,\mathrm{cm}^{-1}$. By connecting pentasil rings, slightly larger building blocks are constructed, but no substantial changes for the peak position have been observed. The situation changes if a more condensed structure is formed, such as the silica octamer, referred to as $8\mathrm{T}(4 \times 5)$, in which four five-membered rings form a tiny cage-like structure: a sudden shift by approximately $50\,\mathrm{cm}^{-1}$ is observed. The red shift becomes even more substantial when the particles grow larger. Therefore the IR peak shift can be tracked down from loosely connected rings, over condensed units, MFI precursors and nanoslabs towards the formation of one-,two- and three-dimensional growth to a full crystal, where the MFI fingerprint of $550\,\mathrm{cm}^{-1}$ is recovered.

Aim. The shift of the IR fingerprint band forms the subject of *Paper I*. While ref. [45] is a step forward in the understanding of the transformation of

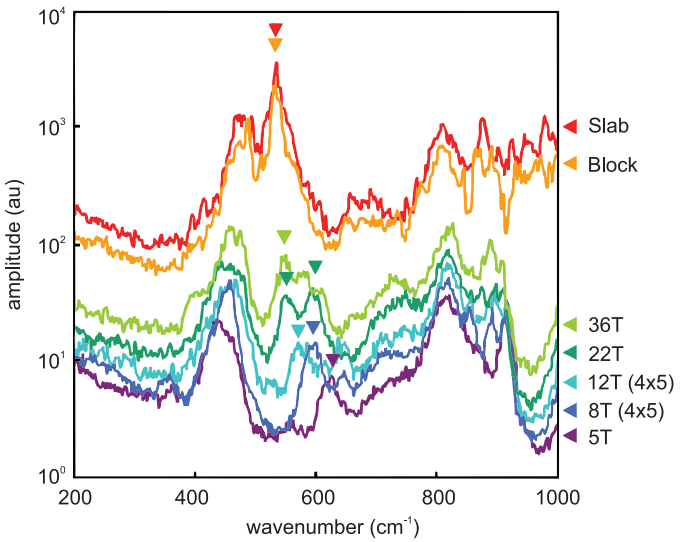


Figure 2.3: IR spectra derived from molecular dynamics for silica building blocks of different sizes important for the formation of the zeolite silicalite-1. The curves which are shifted upwards belong to larger building blocks. Reprinted from ref. [45]. Copyright 2008. American Chemical Society.

soluble silica species into microporous crystallites by identifying sudden change in the IR spectra, it is clear that the responsible mechanisms and processes can further be elucidated. Unraveling the spectra of zeolite nanoparticles, derived from molecular dynamics, can aid in this procedure. The **goal of *Paper I*** can be formulated as:

> identifying peaks in spectra of zeolite building blocks with specific well-chosen internal degrees of freedom, i.e.: link the spectra with particular motions of internal coordinates (ICs). In particular, focus lies on the identification of particular modes which are influenced by changing size and topology of the silica building blocks.

Whereas the focus in ref. [45] lies on five-membered rings and the MFI framework, *Paper I* extends the picture of zeolite formation by also considering essential nanoparticles for other types of zeolite: four-, five-, and six-rings and all types of connections between them.

2.4 Spectra derived from MD trajectories

Vibrational spectroscopy (IR, Raman, inelastic neutron scattering (INS), etc.) can provide valuable information about the chemical state of a molecule and

the structure and dynamics of its environment. Assignment of vibrational modes is an important step in the interpretation of these data as fingerprints of conformations. Modeling and numerical simulations can be very helpful in this respect.

While standard textbooks on molecular vibrations and spectroscopy give a quantum mechanical treatment on the subject, the computation of power spectra (or spectral density) from classical MD requires a different approach [46]: (i) an MD run generates the atomic trajectories for the system of interest from a set of initial atomic coordinates and momenta and a given expression for the potential surface, (ii) a central dynamical variable $x(t)$ is calculated from the trajectories, its choice depends on the type of the desired power spectrum. In *Paper I*, the INS- and IR-spectra are considered, for which the central dynamical variable $x(t)$ is the atomic velocity and the time derivative of the dipole moment, respectively.

General power spectrum. The band shape function $I(\omega)$, or in the case of any dynamical variable $x(t)$ the power spectrum or spectral density, is, by virtue of the Wiener-Khintchine theorem, related to the Fourier transform of its autocorrelation function $C(t)$ (Chapter 22 of ref. [47], [48]):

$$I(\omega) = \int_{-\infty}^{+\infty} C(t) \exp(-i\omega t) \mathrm{d}t, \tag{2.1}$$

with

$$C(t) = < x(t)x(t+\tau) >, \tag{2.2}$$

where ω is the angular frequency and the average $<>$ indicates an average over an ensemble appropriate to the problem. In Chapter 10 of ref. [49] it is shown from linear respons theory, that this power spectrum $I(\omega)$ can be expressed as a time average:

$$I(\omega) = \lim_{t\to\infty} \frac{1}{t} \left| \int_0^t x(t) \exp(-i\omega t) \mathrm{d}t \right|^2 . \tag{2.3}$$

The derivation is based on the ergodic principle, i.e. the ensemble average of $C(t)$ equals the time average.

Inelastic Neutron Scattering spectrum. The definition of the INS spectrum is obtained by plugging the atomic velocities into Eq. (2.3) for the variable $x(t)$. If the atomic velocity history is obtained in an MD calculation, the INS spectrum $I_{INS}(\omega)$ can then be computed as:

$$I_{INS}(\omega) = \lim_{t\to\infty} \frac{1}{t} \sum_{\substack{i=1 \\ \alpha=x,y,z}}^{N} \left| \int_0^t v_{i,\alpha}(t) \exp(-i\omega t) \mathrm{d}t \right|^2 . \tag{2.4}$$

IR spectrum. The IR spectrum is acquired in a similar way as the INS signal using an autocorrelation function of the time derivative of the dipole moment (Eq. 2.1).

The INS- and IR-spectra are based on a Fourier transform which converts information from the time domain to the frequency domain. Both spectra are directly dependent on the velocities of the atoms $\dot{R}_{i,\alpha} = v_{i,\alpha}$ so that the positions of peaks in INS- and IR-spectra coincide, but the intensities will vary. In *Paper I* focus is set on the INS-spectra since charges in force fields are not (yet) well defined in order to obtain reliable qualitative spectral amplitudes. Moreover, cancellation effects frequently appear in the time-varying dipole moment vector. It implies that the amplitudes in an IR spectrum are smaller and faster affected by relative noise. On the contrary, the INS spectrum contains only a summation of positive contributions (Eq. (2.4)) and the relative noise on the spectrum decreases. It also provides information about all vibrations of the system, i.e. all IR and Raman active modes. So in this thesis exclusively INS-spectra are considered, but extension to the IR power spectrum is possible.

2.5 Atomic velocity projection method

2.5.1 Concept and formulation

In *Paper I*, a protocol is presented to identify signals in vibrational spectra of silica oligomers. The method is based on the projection of the atomic velocity vectors on the tangential directions of the trajectories belonging to a predefined set of ICs. In this way only contributions of atomic motions along these ICs are taken into consideration. These ICs are linear combinations of the Cartesian coordinates and are unaffected by translations or rotations of the molecule as a whole. This new methodology – the velocity projection method – makes a detailed analysis of vibrational spectra (Eq. (2.4)) possible by establishing a one-to-one correspondence between a spectral signal and a proper IC. The method is applied to the spectra of zeolite oligomers and rings, and should serve as a basic understanding of which modes and which peaks in the vibrational spectra are influenced by size of the ring and topology. The essence of the method is shown in Fig. 2.4 and Fig. 2.5.

The velocity projection method consists of projecting the Cartesian velocity vector $v_{i,\alpha}$ of each atom i on the atomic tangent vector of the internal coordinate $q(R_{i\alpha}(t))$ (Fig. 2.4). If one considers only one IC, the definition of the tangent vector of atom i at each time step is given by :

$$J_{i\alpha} = \frac{\partial q}{\partial R_{i\alpha}} \tag{2.5}$$

where α runs over the Cartesian x, y, z-components of atom i. Obviously, i must be one of the atoms that specifies the IC $q(R_{i\alpha})$. Only the tangential

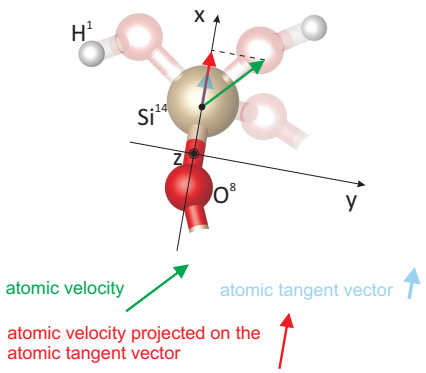
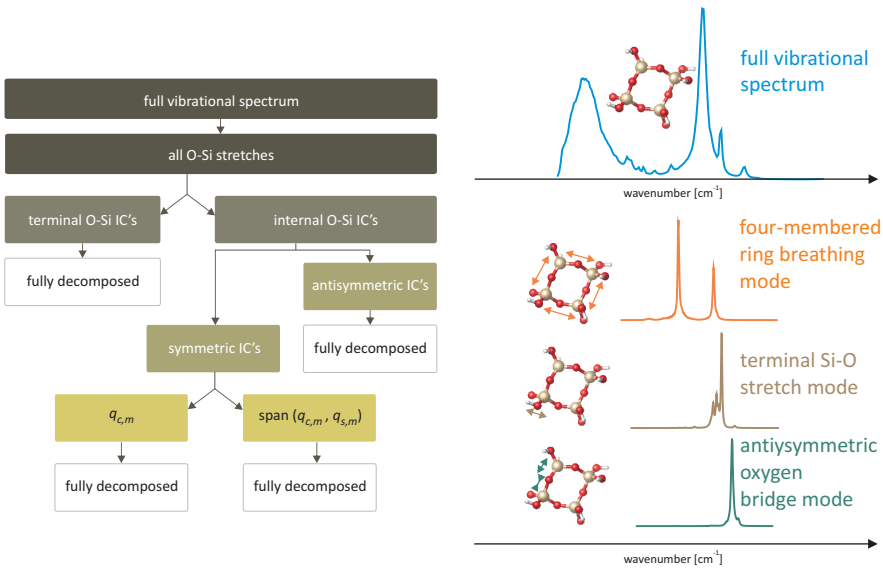


Figure 2.4: Illustration of the atomic velocity projection method applied to the Si^{14}-atom. The IC used for this projection is the O-Si stretch.



(a) Flowscheme to classify different ICs used to decompose vibrational zeolite spectra.

(b) Visualization of the decomposition of the full vibrational spectrum of the single four-membered ring into partial INS spectra restricted to changes of a particular IC.

Figure 2.5: Classification of ICs and comparison between the full and partial INS spectra.

velocity component $(v_{//})_{i\alpha}$ gives a non-vanishing contribution to the change of the IC and this parallel velocity is the component that should be identified with this particular IC. The velocity component $(v_{//})_{i\alpha}$ of each atom alongside the IC is then used to compute the Fourier transform (Eq. (2.4)) at each time time step to construct a partial INS spectrum restricted to those vibrations inducing changes of the IC of interest. A comparison between the full and partial INS spectra makes a full analysis of the vibrational spectrum possible (Fig. 2.5(b)): spectral peaks can be linked to IC's and corresponding (internal) degrees of freedom. When a peak in the projected spectrum coincides with its counterpart in the total spectrum, one can assume that there is no motion along other orthogonal coordinates that contributes to this peak. Such a peak is then completely resolved.

2.5.2 Analysis using ICs.

The peaks in the full spectra can be (partially) decomposed by a proper selection of ICs, so a suitable selection of the relevant ICs is important. This can be done for the various zeolite building blocks under study. Linear combinations of various ICs can be constructed to form new ICs, and all of them can then be classified (Fig. 2.5(a)): each class has its specific internal motion and generates a partial INS spectrum for each building block. Focus is set on ICs constructed from stretches between two atoms: the stretches themselves can serve as ICs or linear combinations of them. This way we arrive at the scheme of Fig. 2.5(a), where the symmetric and antisymmetric IC's are formed by respectively adding and subtracting two stretching Si-O bonds of one particular Si-O-Si bridge (internal stretches, see Fig. 2.2).

In *Paper I*, it is found that the peaks due to the symmetric stretch mode are not completely resolved. In order to gain insight into the normal modes of this particular IC of a molecular ring system, a model is considered of n identical and equally spaced masses on a circle. The normal modes are obtained as the eigenvectors of the circulant matrix C:

$$C_{jk} = c_{(j-k) \bmod n}, \tag{2.6}$$

for j,k=1, ..., n where the notation $(i) \bmod n$ implies integer arithmetic modulo n. The eigenvectors of the circulant matrix can be used to create new IC's on which we can project the atomic velocities, based on the symmetric stretch IC's. There are n symmetric stretches in the circular molecular system and $q_{ss,k}$ is short for the k^{th} symmetric stretch IC. With the circulant matrix being the model for the Hessian, we can construct these new internal coordinates $q_{c,m}$ and span$(q_{c,m}, q_{s,m})$ (Fig. 2.5(a)) for each eigenvector m by making linear combinations of the symmetric stretches $q_{ss,k}$, where the coefficients are the elements of the eigenvector. The IC span$(q_{c,m}, q_{s,m})$ is introduced as the subspace spanned by the vectors $q_{c,m}$ and $q_{s,m}$ as they only differ in phase and one can not discriminate between them.

2.6 Effects of ring size and topology

We find that the spectra of the terminal O-Si and the antisymmetric internal O-Si IC's are not influenced by the topology and not by the size of the considered rings: the spectra of the elementary 4,5- and 6-ring are the same (see Fig. 2.6). In addition, when other rings are attached to these basic n-rings, the same conclusions can be drawn. When the atomic velocities are projected on the symmetric O-Si IC's, the resulting spectra differ for the 4-,5- and 6-ring. The spectral shifts of the spectra of IC $q_{c,0}$ are tiny (Fig. 2.6), while those corresponding with span$(q_{c,m}, q_{s,m})$ are rather substantial (see *Paper I*). The same trend can be observed in connected ring systems, e.g. when the spectra of the four- and five-ring of the 7T(4,5) system are compared.

It is recommended to refer to the spectra of the elementary rings when investigating the effect of connecting multiple n-rings. The most important conclusions are that the $q_{c,0}$ IC is the most sensitive mode to changes in molecular topology (see Fig. 2.7). It is the most global mode: we associate it to the breathing mode as it is determined by adjacent stretches which are not in antiphase, giving this mode a more collective character. The eigenmode of the span$(q_c, 2, q_s, 2)$ IC on the other hand is the most local mode, as the successive stretches move in antiphase with each other. The topology dependence here is minimal. About the connectivity we found that the size of the added n-rings does not play a crucial role; it is rather the way how these rings are connected to each other that causes the observed spectral changes.

When one projects the atomic velocities on the orthogonal complement of all stretch IC's present in a molecular structure, we found that the contribution of the stretch modes to the spectra is spectrally well separated from other contributions like bending and dihedral angles. This confirms that the spectral region in which we are interested for zeolite synthesis can be almost completely resolved by considering changes in stretch modes.

2.7 Applicability

The atomic velocity projection method is implemented, tested and applied with Python software. The core of the code is intimately linked with other CMM-software (MD-TRACKS, MolMod package). The methodology is generally applicable, and in principle the software should be extendable to other molecular systems. In the current version, however, convenient use has been made of the fact that the software is only applied to zeolites, which makes the practical implementation less complex. E.g. since only silica, oxygen and hydrogen atoms appear in the chemical formula of silicalite-1, the atomic numbers can be hard coded instead of having to consult a database containing information for all the elements of the periodic table.

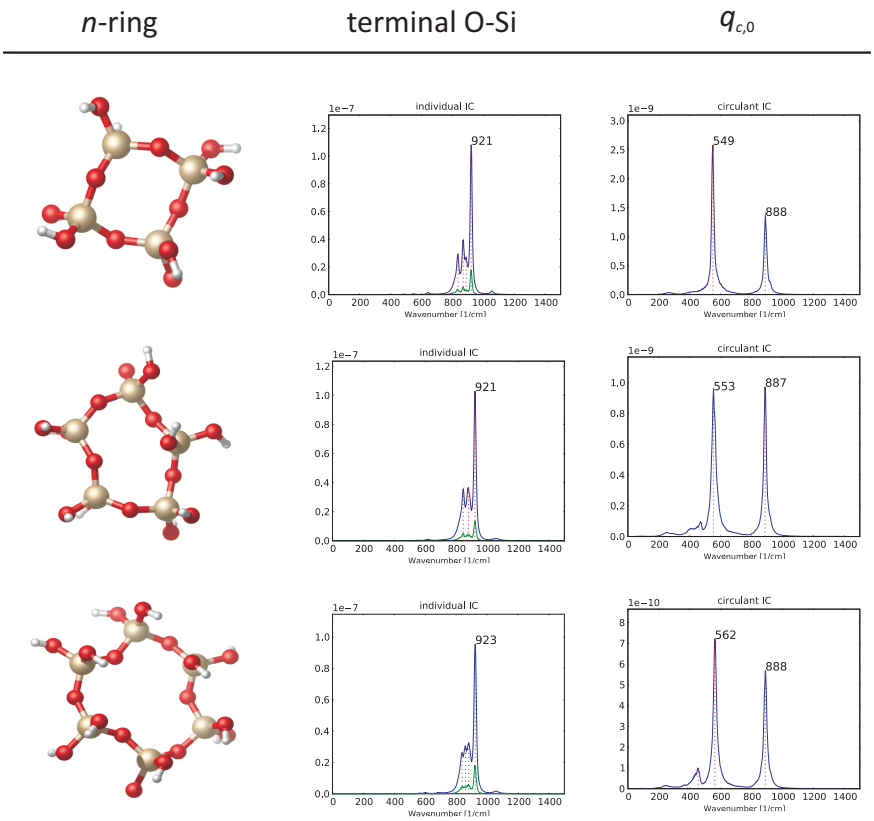


Figure 2.6: Projection of atomic velocites on the terminal O-Si and on the circulant symmetric $q_{c,0}$ IC for the four-, five, and six-ring. The green curve represents the standard deviation of the average of the spectra of all the terminal IC's in each n-ring.

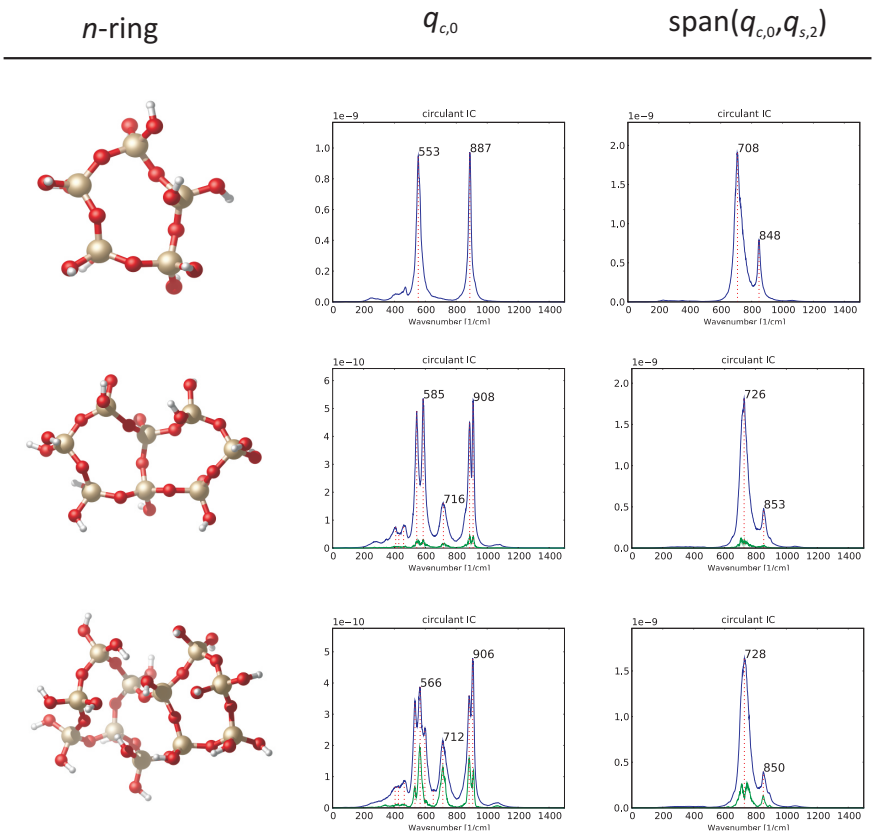


Figure 2.7: Projection of atomic velocites on the circulant symmetric $q_{c,0}$ and span$(q_c, 2, q_s, 2)$ IC's for connected five-rings. The green curve represents the standard deviation of the average of the spectra of the seperate five-rings.

The methodology can be applied to other molecular systems, potentially with an extension in the software, provided some basic requirements are met, e.g. the power spectrum $I(\omega)$ can be expressed in terms of a variable $x(t)$ (Eq. 2.3), where $x(t)$ is derivable from MD simulations.

Two limitations must be mentioned: (i) the particular IC on which the projection has been performed, is in general a function of time. The fluctuations due to the time dependency of the ICs are also computed with the Fourier transform but should be small enough to neglect. (ii) The circulant matrix as a model for the Hessian, and its eigenvectors as a solution of the NMA eigenvalue equation as projection coefficients, can only be used for single closed ring systems. Therefore, it is not expected to completely resolve spectra in systems where fused rings are present. If multiple rings are attached to each other, then one has to distinguish between them. This is still useful because the resulting spectra can be directly compared with spectra from other n-ring structures.

2.8 Assessment of the MD approach versus static methodology

Paper V [50] is a combination of experimental work and theoretical modeling. In this work, silica solution (sol) - gel processing is studied with Raman and nuclear magnetic resonance (NMR) spectroscopy. Careful control of the sol and its transformation to gel allows synthesis of a wide variety of silica materials with specific properties depending on synthesis conditions. A better understanding of the synthesis process is important for the synthesis of new zeolites (see Section 2.2). The paper studies the reacion pathway of tetramethylorthosilicate $Si(OCH_3)_4$ (TMOS) in the presence of solvent (methanol) and an acidic catalyst (H_2SO_4) in an aqueous environment: which silica nanoparticles are formed and how? It is a continuation of earlier work on tetraethylorthosilicate $Si(OC_2H_5)_4$ (TEOS) [51], also a collaboration between the CMM and the COK.

Molecular modeling provides a valuable aid in unraveling the molecular aspects and underlying mechanism of the silica condensation process by predicting theoretical Raman vibrational frequencies and ^{29}Si NMR chemical shifts on silicalite oligomers.

Raman spectra can be determined from MD as the Fourier transform of the autocorrelation function of the time variation of the electric polarizability matrix.

When placed in a magnetic field, NMR active nuclei absorb electromagnetic radiation at a frequency characteristic of the isotope. The resonant frequency, energy of the absorption, and the intensity of the signal are proportional to the strength of the magnetic field. In NMR spectroscopy, use is made of the fact

that the nuclear magnetic moment has a resonance frequency shift depending on the molecular environment. [52, 53] The chemical shift provides information about the structure of the molecule and the signals are usually reported relative to a reference signal. Tetramethylsilane $Si(CH_3)_4$ (TMS) is the accepted internal standard for calibrating chemical shift for ^{29}Si NMR spectroscopy in organic solvents (where TMS is soluble).

In *Paper V*, Raman vibrational frequencies and NMR shifts are determined with static quantum chemical computations. The calculations are performed on silicalite chain structures from monomer up to octamer and some cyclosilicates (three- and four-rings) in vacuo, i.e. the gas phase. The general workflow for such type of computational modeling consists of two steps: (i) find the optimized structure that matches a local minimum in the PES, (ii) determine the desired molecular property on this optimised structure. This workflow has inherent limitations as it relies on static calculations. To perform a comparison with the experiment, some approximations are necessarily made:

(i) On the PES, there are multiple local minima, corresponding to different conformers for a given oligomer. Finding all these minima is difficult, as there is no prior knowledge of the shape of the PES. We then necessarily need to be satisfied with a limited number of such local minima. From a practical point of view, it is sometimes difficult to converge to a local minimum with static methods, even for a single (large) molecule. Examples of a tetramer and octamer molecule are given in Fig. 2.8.

(ii) When the molecule is larger, it is possible to predetermine a given conformation (e.g. "linear" (Fig. 2.8(a)) or "curled up" (Fig. 2.8(b))) and force the convergence process to a certain local minimum. This, however, relies heavily on intuitive insight.

(iii) As the computations are performed in vacuo, there is little room to take into account external parameters, such temperature, pressure, ambient conditions. Usually, as in *Paper V*, the resulting discrepancy between experiment and theoretical results is lifted by applying a scaling factor. This scaling factor is determined by fitting the theoretical to the experimental spectra. Besides the fact that this way an agreement between theory and experiment is forced upon, an assessment of the influence of external parameters seperately is not evident.

The molecular dynamics technique resolves these issues by a more a complex sampling of phase space and with the possibility of performing MD in different ensembles.

The combined experimental and theoretical study led to a peer-reviewed publication (*Paper V*).

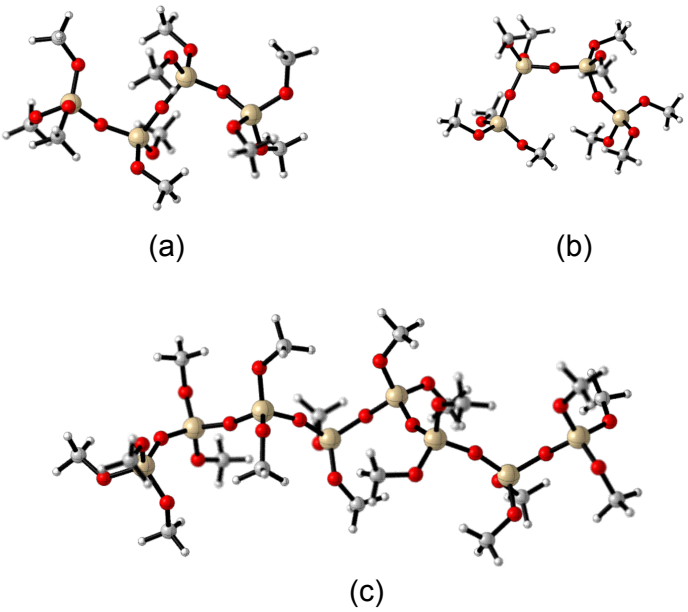


Figure 2.8: Some of the lowest energy conformers of tetramer (linear (a) and curled up (b) configuration), and octamer(c).

3

MD as a tool for studying liquid structure

In this chapter the MD-methodology is based on quantum-mechanical principles for the generation of MD trajectories. A profound study of the liquid structure makes a combined QM MD approach indispensable. The feasibility of Density Functional Theory (DFT) methods in MD simulations has nowadays been proved in many applications and we will rely on this methodology in the further course of this work. Although special care needs to be taken into account for the dispersion interactions.

3.1 Applied QM MD methodology

There is a plethora of electronic structure methods available to calculate the ab initio forces in an MD simulation. The MD procedure is not tied to any particular electronic structure approach, and the strength or weakness of a particular ab initio MD study is intimately connected to the strength or weakness of the underlying approximate electronic structure method chosen.

General eigenvalue equation. Whatever the choice for a specific electronic structure method may be, one always ends up with a complicated system of coupled integro-differential equations comprised in the single-electron eigenvalue equation:

$$\hat{H}_{sp}\psi_i(\boldsymbol{r}) = \epsilon_i\psi_i(\boldsymbol{r}), \tag{3.1}$$

where $\hat{H}_{sp}$ is the single-particle hamiltonian, $\{\epsilon_i\}$ are the eigenvalues and the ψ_i are the single-electron wavefunctions of orbital i. The exact form of the operator $\hat{H}_{sp}$ and the interpretation of the orbitals depends on the specific electronic structure method. E.g. in HF theory the ψ_i are one-electron (or single-particle) orbitals, while in DFT they represent singe-particle Kohn-Sham

orbitals, i.e. they play an indirect role and are introduced only to construct the ground state density. Whatever the structure of the single particle Hamiltonian, the same mathematical techniques can be used to solve them. The conceptual machinery that is practiced for solving Eq. (3.1) is important for this work.

The LCAO ansatz as a solution. To this end, we first introduce the linear-combination-of-atomic-orbitals (LCAO) ansatz, which is by far the most practical way to make the iterative self-consistent field procedure for solving Eq. (3.1) computationally accessible. This expansion of molecular orbitals in atomic orbitals is originally introduced in the framework of HF theory. The same LCAO concept is used to solve the DFT Kohn-Sham equations [54] To clarify the LCAO ansatz, we will make use of the generic term *molecular orbitals*, keeping in mind that the interpretation and their terminology is different for various electronic structure methods.

In the LCAO expansion of the molecular orbitals, a scheme introduced by Roothaan in 1951 [55], a set of n predefined basis functions $\{\phi_\mu\}$ are introduced and linearly expanded as

$$\psi_i(\boldsymbol{r}) = \sum_{\mu=1}^{n} c_{\mu i}\phi_\mu(\boldsymbol{r}), \tag{3.2}$$

where ψ_i is the i^{th} molecular orbital, $c_{\mu i}$ are the coefficients of the linear combination, ϕ_μ is the μ^{th} atomic orbital, and n is the number of atomic orbitals. In practice, the molecular orbitals are expressed as linear combination of a finite set (a basis set) of prescribed basis functions, which are in fact simple analytic functions. Popular is the primitive Gaussian χ^{GTO}, or Gaussian-type orbital (GTO), which is local in space:

$$\chi^{\text{GTO}}(\boldsymbol{r}) = (x-x_0)^{n_x}(y-y_0)^{n_y}(z-z_0)^{n_z}e^{-\alpha(\boldsymbol{r}-\boldsymbol{r_0})^2}, \tag{3.3}$$

where n_x, n_y, n_z are integer numbers and their sum classifies the GTOs as s-, p-functions, etc., $\boldsymbol{r}_0 = (x_0, y_0, z_0)$ represents the nucleus of the atom so that they are centered at the nuclear positions and α represents the orbital exponent which determines how compact or diffuse the resulting function is. A major advantage is that they are easier to handle numerically because the product of two GTOs located at different atoms is another GTO. A major disadvantage is that they are associated with a particular atom, which can cause a basis set superposition error (BSSE) in MD simulations, see Section 3.2. It is exactly this effect which is studied in *Paper III* and *Paper IV*, i.e.: a model is presented and applied to a series of relevant solvents to correct for BSSE. Quantum-chemical software uses contracted Gaussian functions (CGF), i.e. linear combinations of primitive Gaussians are used as basis functions:

$$\phi_\mu^{\text{CGF}}(\boldsymbol{r}) = \sum_{k=1}^{l} d_{k\mu}\chi_k^{GTO}(\boldsymbol{r}), \tag{3.4}$$

where l is the length of the contraction, i.e. the number of primitive GTOs. Such basis functions will have its coefficients and exponents fixed.

Through the LCAO expansion we have translated the non-linear optimization problem (Eq. (3.1)), into a linear one with the coefficients $\{c_{\mu i}\}$ being the only variables, which can be expressed in the language of standard linear algebra and can be coded into efficient computer programs. "Solving" Eq. (3.1) then means that the coefficients $c_{\mu i}$ need to be found.

DFT and the GPW-method. The MD-runs in *Paper III* and *Paper IV* were done using DFT [56, 57]. DFT formulates the many-electron problem in terms of the electron density $n(\boldsymbol{r})$

$$n(\boldsymbol{r}) = \sum_{\substack{i \\ \text{occ}}} |\psi_i(\boldsymbol{r})|^2, \tag{3.5}$$

to compute the ground-state energy E_0. More specific, the Gaussian- and Plane Waves method [58] (GPW) was used in this work. The GPW-method is a DFT-technique which uses a hybrid type basis set in the sense that gaussian basis functions are used as a primary basis set, but an auxiliary plane wave (PW) basis set is employed additionally. Gaussian basis functions and PWs each have their own advantages and disadvantages, and the GPW method tries to combine the best of both worlds. In a PW basis set, the basis functions are of the form:

$$\phi_\mu^{\text{CGF}}(\boldsymbol{r}) \sim e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{3.6}$$

Given their periodicity and infinite extent, these functions are naturally suitable for periodic simulations. One advantage of PWs is that several types of integrals are much easier to evaluate in reciprocal space, and using fast Fourier-transformation techniques, it is possible to switch easily between a real and reciprocal representation of certain quantities. Two disadvantages are (i) they are poor at representing the wavefunction near the atomic nuclei where the electron density is at its highest, as they are delocalized in space, (ii) the number of plane waves needed to arrive at an acceptable accuray is usually very high. Within the GPW-method, a Gaussian type orbitals basis set is used for the real space representation and the auxiliary PW basis to compute the long-range periodic electrostatic interactions in the reciprocal space. It is important to observe that plane waves are not centered at the position of a nucleus, so that BSSE with PWs does not occur.

3.2 The BSSE and a correction scheme

The Basis Set Superposition Error (BSSE) is a direct consequence of using incomplete basis sets. The errors are more pronounced for smaller basis sets. Consider two monomers A and B, whose geometry optimization and binding energy have been computed in the basis sets a and b respectively, which are both

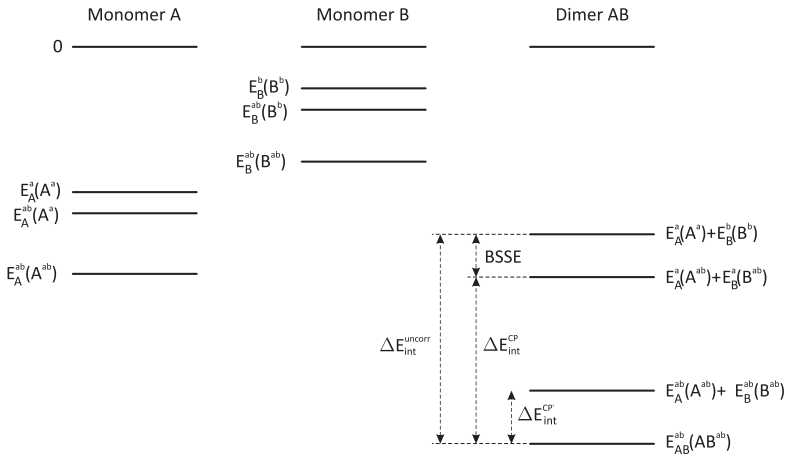


Figure 3.1: Schematic overview of the energy levels of monomers A and B, the uncorrected interaction energy $\Delta E_{\mathrm{int}}^{\mathrm{uncorr}}$ of the bimolecular complex AB, the CP-corrected energy $E_{\mathrm{int}}^{\mathrm{CP}}$ of AB and the BSSE.

incomplete. An artificial stabilization takes place as monomer A approaches monomer B since monomer A utilizes the extra basis functions from monomer B to describe its electron distribution. Applying the following convention for the binding energy $E_{\mathrm{fragment}}^{\mathrm{basis\ set}}(\mathrm{geometry}^{\mathrm{basis\ set}})$, we easily see that $E_{\mathrm{A}}^{\mathrm{ab}}(\mathrm{A}^{\mathrm{a}}) < E_{\mathrm{A}}^{\mathrm{a}}(\mathrm{A}^{\mathrm{a}})$. The energy is lowered by using each other's basis functions. This causes an artificial strengthening of the intermolecular interactions and artificial shortening of intermolecular distances due to the overlap between the basis functions a and b.

When forming a dimer AB, the uncorrected interaction energy is given by:

$$\Delta E_{\mathrm{int}}^{\mathrm{uncorr}} = E_{\mathrm{AB}}^{\mathrm{ab}}(\mathrm{AB}^{\mathrm{ab}}) - E_{\mathrm{A}}^{\mathrm{a}}(\mathrm{A}^{\mathrm{a}}) - E_{\mathrm{B}}^{\mathrm{b}}(\mathrm{B}^{\mathrm{b}}). \tag{3.7}$$

A better approach is the consideration of the energy of the monomer A in the dimer basis ab (and likewise for monomer B), yielding following expression:

$$\Delta E_{\mathrm{int}}^{\mathrm{CP}} = E_{\mathrm{AB}}^{\mathrm{ab}}(\mathrm{AB}^{\mathrm{ab}}) - E_{\mathrm{A}}^{\mathrm{ab}}(\mathrm{A}^{\mathrm{a}}) - E_{\mathrm{B}}^{\mathrm{ab}}(\mathrm{B}^{\mathrm{b}}). \tag{3.8}$$

The difference between Eq. (3.7) and (3.8) is called the counterpoise (CP) correction as introduced by Boys and Bernardi [59]. The interaction energy is negative while $\Delta E_{\mathrm{int}}^{\mathrm{CP}} < \Delta E_{\mathrm{int}}^{\mathrm{uncorr}}$. The BSSE correction thus reduces the overstabilization of the uncorrected interaction energy: $\mathrm{BSSE} = \Delta E_{\mathrm{int}}^{\mathrm{CP}} - \Delta E_{\mathrm{int}}^{\mathrm{uncorr}} > 0$ (see Fig. 3.1). The counterpoise correction as expressed in Eq. (3.8) can even be refined if we take into account geometry relaxation effects. The monomer energies in Eq. (3.8) are computed in the extended basis set ab but based on geometries optimized in the individual monomer basis functions. However, the geometries of monomers A and B slightly change as they approach each other

and form a bimolecular complex. Instead of $E_{\mathrm{A}}^{\mathrm{ab}}(\mathrm{A}^{\mathrm{a}})$ we can rather introduce $E_{\mathrm{A}}^{\mathrm{ab}}(\mathrm{A}^{\mathrm{ab}})$ or even $E_{\mathrm{A}}^{\mathrm{ab}}(\mathrm{AB}^{\mathrm{ab}})$, where in the latter the monomer geometry is cut from the optimized AB dimer structure:

$$\Delta E_{\mathrm{int}}^{\mathrm{CP'}} = E_{\mathrm{AB}}^{\mathrm{ab}}(\mathrm{AB}^{\mathrm{ab}}) - E_{\mathrm{A}}^{\mathrm{ab}}(\mathrm{A}^{\mathrm{ab}}) - E_{\mathrm{B}}^{\mathrm{ab}}(\mathrm{B}^{\mathrm{ab}}). \tag{3.9}$$

In Eq. (3.9) the BSSE correction is even larger than in Eq. (3.8), since $E_{\mathrm{A}}^{\mathrm{ab}}(\mathrm{A}^{\mathrm{ab}}) < E_{\mathrm{A}}^{\mathrm{ab}}(\mathrm{A}^{\mathrm{a}})$. It has been argued in the literature that in some cases such a procedure could lead to an overestimation of the BSSE error. [60] In particular, in smaller basis sets the risk for overcorrecting is much larger. For that reason we prefer to consider Eq. (3.8) as the CP-corrected interaction energy in the further applications in this work.

This CP-method will form an important cornerstone in the model that is proposed in *Paper II* and *Paper III*.

3.3 The radial distribution function and the structure factor

In this section we focus on the structure of liquids and more specifically the spatial distribution of the liquid molecules in equilibrium. This property is established by intermolecular radial distribution functions. Experimental information on liquid structure can be extracted from neutron diffraction measurements. A direct observable is the structure factor $S(k)$.

Definition of the radial distribution function (RDF). The RDF $g(r)$ of a fluid describes how the liquid density varies as a function of the distance from a reference molecule. It reflects the correlations in the distribution of the molecules of the fluid arising from the forces they exert on each other. The radial distribution function gives the probability of finding a pair of molecules at a distance r apart, relative to the expected probability for a completely random distribution at the same density, and is defined as:

$$g(r) = \frac{\mathrm{d}n(r)}{\mathrm{d}n_0(r)}. \tag{3.10}$$

This way, $g(r)$ represents the number of particle pairs $\mathrm{d}n(r)$ in the spherical shell with radius r and $r + \mathrm{d}r$, with r the interparticle distance (relative coordinate), relative to the number of particle pairs $\mathrm{d}n_0(r)$, obtained in an uncorrelated ideal gas. With ρ_0 as the density of particles in the ideal gas case, $\mathrm{d}n_0(r)$ may be expressed as $\rho_0 4\pi r^2 \mathrm{d}r$. Fig. 3.2 is an academic example of an RDF, showing the relation to a model two-dimensional liquid (figure adopted from ref. [61] and ref. [62]).

If there is more than one atom type present in multicomponent materials, it is useful to split the total pair correlation function $g(r)$ into several terms

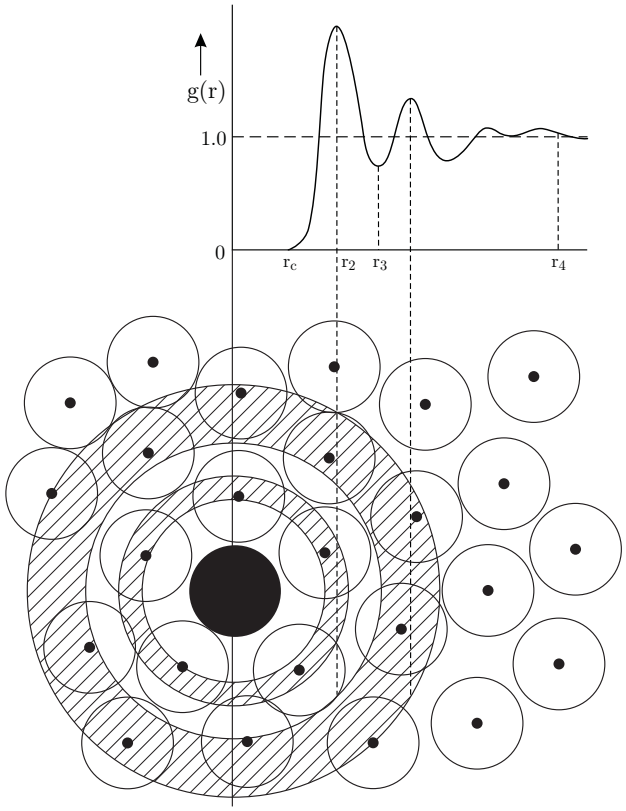


Figure 3.2: A two-dimensional liquid and its RDF, showing the relation between the liquid structure and the RDF description. r_c is the closest distance of approach of two particles, r_2 is the most probable nearest neighbour separation, r_3 is the extent of the first coordination shell, r_4 the end of short range order.

$g_{\alpha\beta}(r)$, one for each pair of atom types. This was first suggested by Faber and Ziman [63]. The partial atom pair correlation function $g_{\alpha\beta}(r)$ then describes the correlations of a distinct $\alpha\beta$ atom pair of two different molecules (inter) or within the same molecule (intra).

Usefulness of an RDF. The RDF plays a central role in theories of liquids for several reasons: (i) The form of $g(r)$ provides considerable insight into what is meant by the structure of a liquid. (ii) Numerical results for $g(r)$ can be compared with theoretical predictions and thus serve as a criterion to test a particular theory. [3] (iii) If it assumed that the total potential energy of the N-body system is pair-wise additive, then all the thermodynamic functions of the system can be written in terms of integrals $g(r)$. Examples of such derivation can be found in refs. [4,47,64,65]. (iv) They can be compared with experimental data, albeit in an indirect manner, as explained in the next paragraph.

The structure factor (SF) and connection with RDF. The RDF and partial RDFs (between atom types) can not be considered as primary experimental observables [66], as they can not be measured in a *direct* way, but are *derived* from experimental data. X-ray and neutron scattering methods provide the experimental means to probe the atomic structure of liquids, i.e. experiments do not measure pair interactions in coordinate space directly but rather measure the sum of all contributing pair interactions in reciprocal space. The resulting property is the SF $S(k)$ and is expressed in terms of the wavevector k, which represents the momentum transfer between the incident and scattered radiation beams. This SF contains then information about the structure of atoms and molecules relative to each other in k-space. A three-dimensional Fourier transform then links the scattering pattern $S(k)$ with the distribution functions $g_{\alpha\beta}(r)$ of pairs of atoms α and β of the system. [61], [47], [67]

If $g_{\alpha\beta}(r)$ is the partial pair correlation function of atom types α and β, then the partial SF $S_{\alpha\beta}$ may be expressed as:

$$S_{\alpha\beta}(k) - 1 = 4\pi\rho \int_0^\infty \mathrm{d}r r^2 \frac{\sin(kr)}{kr} (g_{\alpha\beta}(r) - 1), \tag{3.11}$$

i.e. the correlation between the α and the β chemical species in reciprocal space is related to the correlations between the different chemical species in the coordinate space. Here, ρ is the average particle number density. The factor 4π in Eq. (3.11) arises because the average is taken over all possible angles by integrating out the angle-dependency. Note that Eq. (3.11) is only valid for an isotropic and homogeneous liquid.

Similarly to the RDF [63], the total structure factor is a weighted sum of partial SFs. The two most popular experimental techniques for measuring SFs are the X-ray and neutron scattering methods. [61] If the SF is measured with the latter technique, the sum of partial SFs is weighted by the varying interaction potentials of the neutrons with the atoms:

$$S_{\mathrm{tot}}(k) = \sum_{\alpha \leq \beta} (2 - \delta_{\alpha\beta}) c_\alpha b_\alpha c_\beta b_\beta \left[S_{\alpha\beta}(k) - 1 \right], \tag{3.12}$$

where $c_\alpha = N_\alpha/N$ is the atomic fraction of type α atoms, N_α the number of type α atoms, N the total number of atoms in the system, and b_α is the constant coherent scattering length which characterizes the interaction strength between the atoms and the incident radiation. Note that the scattering length varies from isotope to isotope. The Kronecker $\delta_{\alpha\beta}$ avoids double counting of pairs of atoms of the same type. The total experimental $S(k)$ may be compared with the SF from the MD simulations by applying Eq. (3.12) on the simulated partial SFs. The partial SFs are not directly available from experiment. They are nevertheless derived from the total experimental SF via fitting of Eq. (3.12). For experimental total SFs obtained by X-ray scattering, the coherent scattering length in Eq. (3.12) is replaced by the k-dependent form factor $f_\alpha(k)$.

MD as a tool to compute RDFs and SFs Structural properties, like an RDF or SF, do not depend on the time evolution of the system: they are static equilibrium averages. Such averages can then be obtaind by MD simulations, or equally well by Monte Carlo simulations.

3.4 The influence of BSSE on the RDF

Having introduced the concepts of BSSE and the RDF, one can ask: what is the influence of BSSE on the RDF in MD simulations? In order to predict the RDF as accurate as possible, the theoretical model should be free from approximations which could influence the theoretical RDF. Within this context, it is not yet established how important BSSE corrections could be in the reproduction of RDFs.

The effect of BSSE has already been studied extensively on hydrogen-bonded systems, focusing on binding energies and properties that depend on energy derivatives: (i) It has been shown that the calculation of the BSSE corrections can change their potential energy curves and surfaces, [68–73] and can alter the bond distances. [68, 70, 71, 74, 75] (ii) It is well known that BSSE can be a significant fraction of the binding energy of the water dimer, which could affect liquid simulations. [58] (iii) Simon et al. analyzed the effect of BSSE on the geometries and vibrational frequencies of 15 hydrogen-bonded systems. It is shown that the intermolecular H-bond distance increases when correcting for BSSE, whereas the intermolecular H-bond stretching frequency decreases, and that the observed changes depend on the level of theory. [76]

The effect of BSSE on structural and dynamical properties of the liquid has not been investigated previously. In fact, there is little information available concerning the performance of a Gaussian basis set in the context of ab initio MD simulations of liquids. So the influence of BSSE on RDFs - and SFs- remains an open question.

The first intermolecular peak is a signature for the liquid structure. Often, this peak in the RDF of hydrogen-bonded liquids, e.g. water and methanol

[77–87], is overestimated in DFT-MD compared to experiment. The search for plausible explanations of this quasi systematic overestimation forms the onset of a long debate, but until now it is not yet unraveled and what suits for one particular case does not suit for another case. It has been found in the literature that taking into account the dispersion interaction aids in removing the discrepancy between simulation and experiment.

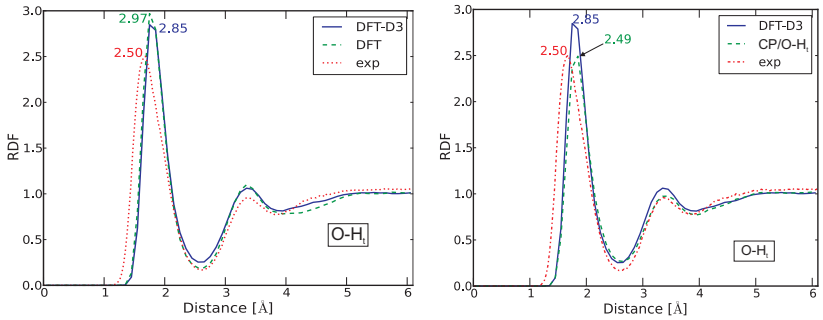
Dispersion interactions are indispensable to predict correct densities and RDFs. Dispersion is a long-range van der Waals interaction between electron clouds arising from instantaneous transition dipoles. These dispersion interactions are a quantum-mechanical electron correlation effect that is completely absent when considering classical charge distributions. Dispersion is not accounted for in DFT, but many models exist to describe them. Most of them work with empirical pairwise interatomic potentials of the C^6R^{-6} form with damping terms, which can differ from model to model. The most popular disperions terms are the D-corrections of Grimme. [88, 89] Most of the published work on the subject of artefacts that influence the RDF focus on the impact of dispersion on water RDFs. One should be very careful to extrapolate the conclusions from these investigations to other cases.

Aim. Having introduced localized basis sets, DFT, BSSE, RDFs, we are now ready to formulate the aim of *Paper II*.

In the work of *Paper II* the RDFs of liquids are studied with the help of MD simulations, based on DFT. The MD simulations are run in the canonical ensemble (constant number of paricles N, constant volume V, constant temperature T). Since it is not yet clear what the net impact really is from each ingredient on the RDF, it is important to get detailed insight into the various factors contributing to the overall result of the RDF. The research focuses on the role of BSSE on the RDFs. The **goal of *Paper II*** can be formulated as

> establishing the effect of BSSE on RDFs derived from DFT-MD using localized Gaussian basis sets. An efficient protocol is presented to compensate for this BSSE in the form of a classical correction term that can be added in subsequent MD simulations. Since taking into account dispersion is indispensable in any attempt to reproduce correctly the RDF of methanol, a combinded study of BSSE and dispersion is performed.

The correction model is pairwise between certain atoms in different molecules and is a systematic correction of the BSSE. This parametrized classical force-field energy term is calibrated by fitting reference BSSE data points computed with the counterpoise method. Hence the correction term is denoted as the CP-model. Initially we study a more straightforward case, i.e. methanol, which is a hydrogen-bonded liquid, in the NVT ensemble. It is expected that the BSSE contribution will be the most significant in the intermolecular interaction between the oxygen and the hydroxyl hydrogen. Therefore, the CP-model is



(a) DFT,DFT-D3,exp. (b) DFT-D3,CP/O-H_t,exp.

Figure 3.3: Comparison RDFs methanol before and after the application of the CP model. We distinguish between the hydroxyl hydrogen (H_t) and the methyl hydrogen (H_m), as they are situated in different chemical environments.

initially applied on the intermolecular hydrogen bond. Afterwards (*Paper III*), more complex case studies are examined. In principle the BSSE correction method is generally applicable on any type of interaction where BSSE is significant. It is tested to what extension the extra correction term will improve the RDFs.

3.5 Effect of application of the CP model

The relative importance of BSSE versus the dispersion interaction (DFT-D3) is thoroughly assessed in *Paper II*. The dispersion energy is of the same order of magnitude as the BSSE, but has an opposite sign. Combinations of the two types of corrections – a correction term for BSSE and dispersion – leads to three types of DFT MD simulations that are carried out. For the simulations with and without dispersion, without correction for BSSE, errors on the first peak height of the RDFs between 12% and 16% are found (Fig. 3.3(a)). A classical correction model for BSSE is proposed and its parameters were derived from counterpoise (CP) computations on methanol dimers. This CP model can then be added to the DFT energy in subsequent MD simulations. We confirmed that an exponential decaying functional form for the CP model works well. Upon introduction of the correction, the errors on the heights of the first peak of the relevant intermolecular RDFs participating in the hydrogen bond are reduced to 0.4% and 1%, see Fig. 3.3(b). A small shift of the peak position compared to experiment is still observed, which could be due to the added repulsive interaction of the CP model and the limitations of the BLYP functional.

We focused on hydrogen bridges in methanol by applying the CP model to the intermolecular interaction between the oxygen and the hydroxyl hydrogen,

but the results should be extendable to other types of interaction where BSSE is important. However, including more interaction terms in the CP model complicates the optimization of the parameters because of the inherent parameter correlation and ill-conditioning of the set of equations that needs to be solved. It needs furthermore to be tested in how far the CP model is transferable to other solvents. In the discussion on SFs, focus is set on THF as total experimental SFs of this solvent are available, and the group of Soper *et al.* has performed a lot of work in extracting partial intermolecular RDFs from this experimental data.

3.6 Extension to NpT, other solvents and analysis of structure factors

Extension to NpT. Just as one may wish to specifiy the temperature in a MD simulation, so it may be desired to maintain the system at a constant pressure. Many experimental measurements are made under conditions of constant temperature and pressure, and so simulations in the isothermal-isobaric ensemble are most directly relevant to experimental data. Certain structural rearrangements may be achieved more easily in an isobaric simulation than in a simulation at constant volume. A macroscopic system maintains constant pressure by changing its volume. A simulation in the isothermal-isobaric ensemble also maintains constant pressure by changing the volume of the simulation cell. [65]

Extension to other solvents. The structure of five liquids will be investigated with DFT-based molecular dynamics simulations: methanol (MeOH), chloroform (TCM), acetonitrile (MeCN), tetrahydrofuran (THF) and ethanol (EtOH). The availability of first principles MD simulations on liquids to derive their liquid structure is less abundant in literature. The choice of the five solvents were based on (i) their importance for the industry, and (ii) their classification: non-polar (TCM), polar aprotic (MeCN, THF), and polar protic (MeOH and EtOH). Among these five cases, hydrogen bonding occurs only in methanol and ethanol. Therefore, we can no longer anticipate that the BSSE will be dominant for one particular interaction, and the calibration procedure for the parameters in the CP model necessarily must be generalised.

Extension to analysis of structure factors. An RDF can not be determined directly from experiment, so this property can not be considered as primary experimental information about the liquid structure. The only property that is a direct experimental observable is the total structure factor, which is related to the RDF through a Fourier transform. All other quantities, including the partial RDFs, are not directly available from experiment. Partial RDFs are obtained after an inverse Fourier transform of the partial SFs. These latter properties are in turn determined form an iterative protocol combining neutron scattering experiments with isotopic substitution and MD runs with empirical

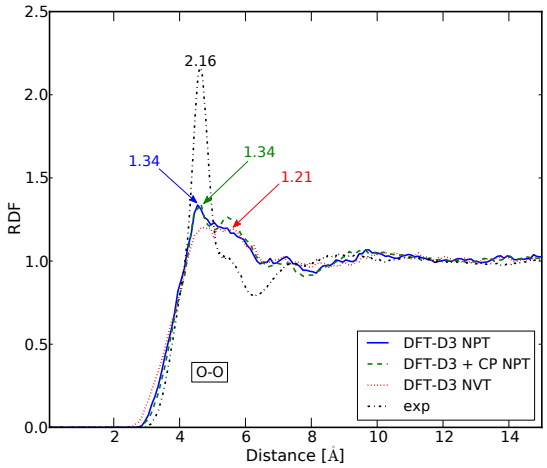


Figure 3.4: Intermolecular partial O-O RDF of THF. The RDF derived from experimental data is taken from ref. [94]. The cause of the discrepancy between results from theoretical simulations (blue, green and red curve) and the experimental model (black curve) is unraveled in *Paper III*.

interatomic potentials. As such, RDFs derived from experiment are not free from cumulating inaccuracies generated in the various steps in the procedure, and the final results are absolutely not free from any bias. [90–93] The most important extension is the assessment of liquid structure in reciprocal space. The work on SFs focuses on THF as a case study, as this solvent is particularly interesting because the RDFs derived from experiment [94] reveal some prominent peaks completely missing in any MD simulation, of which an example is shown in Fig. 3.4. The interpretation of the experimentally obtained total structure factor is a difficult task as the measured scattering intensity is the result of a weighted summation of different pairwise spatial correlation functions formed by all atoms in the system (Eq. (3.12)). Disentangling the various correlations is from the experimental point of view very tedious. In particular the separation of the experimental total structure factor into intramolecular and intermolecular contributions is extremely complex.

Aim. MD modeling has proven to be very effective in helping to guide assignments of peaks in the experimentally derived RDFs, though these last quantities are not direct observables and obtained after some error-prone manipulations. Caution is advised in using them as reference quantities to judge the quality of the theoretical predictions. The **goal of *Paper III*** is twofold, and can be formulated as

(i) assess the evolution of the density that results from NpT MD simulations of MeOH, TCM, MeCN, THF and EtOH and generalize

the calibration procedure as is is suggested in *Paper II*, (ii) compute (partial) RDFs and their counterparts, the (partial) SFs, and disentangle the impact of inter- and intramolecular contributions on the SFs of THF. In addition, we examine the sensitivities of small perturbations in the SFs on the RDFs.

Disentangling intra- from intermolecular contributions poses experimentally a lot of problems. Theoretically, a Fourier transform to convert RDFs to SFs, and vice-versa, requires some mathematical manipulations, but the accuracy of these computations is better controlled. Moreover, theory can asssist in giving the correct interpretation of the features that are experimentally observed in the SFs.

3.7 Study of RDFs and SFs with NpT dynamics

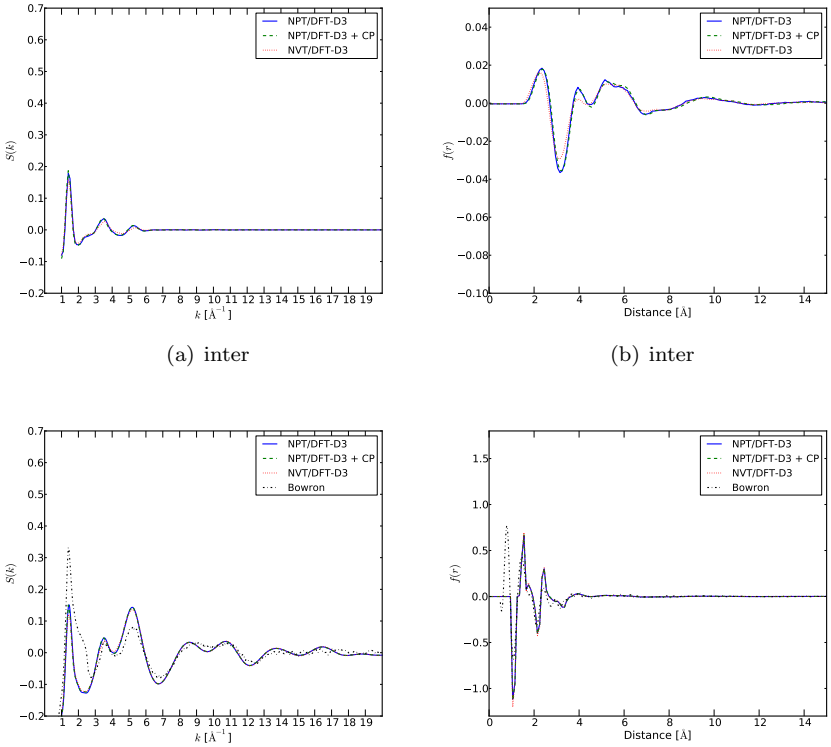
This work comprises the study of the structure of five liquids (MeOH, TCM, MeCN, THF and EtOH) by means of DFT-based MD simulations in both a NVT and NpT ensemble.

First the influence of BSSE on the various liquid properties has been investigated. For this purpose the BSSE correction scheme, as proposed in *Paper II*, has been extended to all intermolecular interactions. MD simulations in a NpT ensemble show that they have a beneficial effect on the reproduction of the density in all five liquids under study. Only in the case of TCM a serious underestimation of the density is observed. We ascribe this deficiency to the inappropriate description of the employed DFT functional BLYP-D3 in describing the intermolecular interactions in the absence of hydrogen bonds, as confirmed in the literature in NpT dynamics on water. [77]

Second, properties have been computed from the simulations for a structural determination of the five liquids. In first instance all partial RDFs for the three methods under consideration (NpT/DFT-D3, NpT/DFT-D3 + CP and NVT/DFT-D3) have been calculated, and compared with experiment if available (see Fig. 3.4). BSSE corrections turn out to overstructure the liquid in NpT simulations, but for the majority of the liquids under study they remain minor and of no real importance in the further discussion on the overall performance of the RDFs. In the liquids TCM and THF – where hydrogen bonds are not occurring – the experimental agreement is dramatic. Peaks appear at short distances, which are completely missing in any MD simulation (Fig. 3.4). As experimental partial RDFs are not directly measurable, these properties are not free from cumulating inaccuracies generated in the various steps in the procedure. The conclusion from this thesis is that a fair assessment can only be performed in reciprocal space, where the total structure factor, as a direct experimental observable, can be compared with theory.

Since we dispose of all theoretical partial RDFs belonging to THF, we are able to compute all partial structure factors by performing a Fourier transform following Eq. (3.11). A weighted summation of these structure factors, as given in Eq. (3.12), yields the theoretical prediction of the total structure factor, which can directly be compared with the experimentally measured $S_{\mathrm{exp}}(k)$ (labeled as "Bowron"). The curves for the NpT/DFT-D3, NpT/DFT-D3 + CP and NVT/DFT-D3 simulations are compared with the experiment in Fig. 3.5 for THF. Comparing Fig. 3.5(a) with Fig. 3.5(c), it is observed that the contributions of the intramolecular interactions to the total SF is immense and that the signal from intermolecular scattering fades out quite fast with increasing momentum transfer. Our MD simulations succeed in a very satisfactory reproduction of S_{tot} in the deuterated counterpart of THF (see *Paper III*). On the contrary the proton sample displays a shoulder at about 2 $\mathrm{Å}^{-1}$ (Fig. 3.5(c)), which is not present in the MD simulations but which causes a prominent peak at 0.8 Å in the total pair correlation function $f(r)$ (Fig. 3.5(d)), where this latter property is computed with the inverse Fourier transform. This is an indication that the shoulder structure in the SF originates from intramolecular interactions, and the conclusion is that even the first large peak in the spectrum is composed of both inter- and intramolecular contributions. This implies that the dominance appears to be complete in the sense that the intramolecular terms cover the whole momentum range, including the low k region where we expected that it is mainly controlled by intermolecular contributions. If the shoulder structure is not filtered out in the derivation of the intermolecular total structure factor, it affects the partial SFs with undesirable consequences in the RDFs. These facts suggest that the first and probably the most crucial source for generating errors is the extraction of the intermolecular total structure factor from the total measured scattered amplitude. Comparing the absolute values of the total pair correlation function $f(r)$ of the inter and the inter+intra molecular contributions(Fig. 3.5(b) and Fig. 3.5(d) respectively), one can observe in a more intuitive way that the intramolecular contributions are dominant.
Another striking feature of Fig. 3.5 is that the discrepancy between theoretical and experimental SFs is large, especially in the wave vector part in which the intermolecular contribution manifests itself. Since the total contribution of the intermolecular RDF to the SF is small, a small relative error in the (computed) intramolecular SF will not be distinguishable from a relative large error in the intermolecular SF. Such small errors on the intramolecular SF are surely present, which can be noticed in Fig. 3.5(c) from the difference between theoretical and experimental SF at large k.

Furthermore, we have found that SFs are very insensitive to changes to partial intermolecular RDFs. This is noted in partial SFs where the type of MD hardly has any affect, while the corresponding RDFs exhibit clear structural differences. This explicitly shows that the derivation of partial intermolecular RDFs from a small number of structure factors is ill-conditioned. The experimentally derived partial structure factor in THF also displays large damping



(a) inter (b) inter

(c) inter + intra (d) inter + intra

Figure 3.5: Total SF for THF C_4H_8O obtained as a weighted sum of the partial SFs limited to intermolecular contributions (upper left) and to inter- and intramolecular contributions (bottom left). Also given are the corresponding total pair correlation functions $f(r)$, computed as an inverse Fourier transform of the total SF (right column). Units are barn/sr/atom.

oscillations in the low k region, which are signatures of an insufficient filtering of the intramolecular contributions, resulting to spurious and prominent peaks in the RDFs.

Concluding, the theoretical simulations give evidence that the extraction of intermolecular partial RDFs and even partial SFs from experimental data should be regarded as error-prone.

4

Conclusions and perspectives

In molecular dynamics simulations, the potential energy surface is sufficiently sampled to extract statistically averaged macroscopic properties from microscopic information. This thesis explores MD to advance the understanding of the molecular mechanisms that lead to the formation of zeolites and the structure of liquids.

In modern day society, there are numerous applications of zeolites. These nanoporous materials occur in nature, but can also be synthesized in the laboratory. Molecular modeling can assist in the comprehension on how zeolites nucleate and grow from nanoparticles through zeolite precursors to nanocrystals before the macroscopic crystals are formed.

In this work, focus is set on vibrational spectra which are computed based on the atomic velocities obtained from MD simulations. In the first part of the thesis, the velocity projection method has been developed to assign (parts of) spectral signals in vibrational spectra to particular changes in internal coordinates. The peaks in the full spectrum can then be (partially) decomposed by a proper selection of internal coordinates (ICs). In principle, one is free to construct any type of ICs, and initially focus is set on ICs constructed from stretches between atoms. This approach has been applied for the various zeolite building blocks under study, and this way this thesis contributes in the comprehension of the molecular mechanisms that occur during the first stages of zeolite synthesis. The velocity projection method is applied on nanoparticles that are suggested to appear in the formation of the silicalite-1 zeolite. It is tested to what extent the restriction to the single class of stretches is accurate enough, and it is found that there is an obvious spectral distinction between the contribution of the stretch modes to the spectrum and that from other contributions, like bending or dihedral angles. This confirms that the spectral region in which we are interested for zeolite synthesis can be almost completely resolved by considering changes in stretch modes. Extra attention has been given to the size and the connectivity of the elementary building blocks, and

how these parameters influence the spectrum. We find that especially the topology is an important parameter that determines the spectral properties of zeolites. Concluding, vibrational spectra are very useful in understanding the process of nanogrowth as the spectral behavior of eigenmodes of ICs can vary with different molecular systems.

Although this thesis contributes to the understanding in the elementary steps in silica organization during zeolite nanogrowth, there is still room for improvement. A possible extension could be the application of the velocity-projection method to the ZSM-5 material, and later to other zeolite topologies. It would be interesting to shift the focus of this study from inelastic neutron scattering (INS) spectra to infrared (IR) spectra, which would require a more qualitative force field since the charges in force fields have to be very well-defined in order to obtain reliable spectral amplitudes. Inversely, theoretical spectra can assist in the development of reliable force-fields. In an attempt to decompose the spectra of the symmetric ICs further, the circulant matrix is introduced in this thesis as a model for the Hessian of the considered ring systems. If we want to extend the applications to systems which are not restricted to a ring structure, a more general approach is desired. A way to visualize vibrational modes that correspond to peaks in vibrational spectra would also be very instructive.

The second part of this thesis deals with liquids; more specific, the structure of liquids was investigated. The understanding of liquids and their properties is very important as they are present everywhere in daily life. The organisation of molecules in a liquid is such that they are not independent from each other (like ideal gases), nor is there a long-range order (like in solids). This makes liquids particularly interesting for studying their structure, and MD can attribute a great amount to the understanding of the molecular organisation in liquids. A thorough comprehension of structural properties, as a characteristic of liquids, is important as they form a bridge between the microscopic structure of the liquid and the macroscopic properties, such as thermodynamic functions.

First, focus is set on methanol. Methanol is the smallest alcohol consisting of an hydroxyl and methyl group, and an intermolecular hydrogen-bond is formed between the hydroxyl and methyl group of different molecules. At first instance, the intermolecular partial radial distribution functions (RDFs) are computed in the NVT-ensemble with MD based on first principles. In subsequent simulations, the effect of a newly proposed parametrized correction model for the basis set superpostion error (BSSE) is examined. We find that the discrepancies in RDFs between simulation and the ones computed with empirical input, are substantially smaller when using the correction model for BSSE. Second, as a continuation of this work, the study on the structure of liquids is extended on four levels: (i) not only methanol is considered, but also chloroform, acetonitrile, tetrahydrofuran and ethanol, (ii) the MD runs are performed in the NpT ensemble, (iii) the correction model for BSSE is generalised, and (iv) the structure factor (SF) is considered as an instrument

to assess the accuracy of models predicting or extracting information on liquid structure. This latter extension is the most important one. The study of the structure factor focused on tetrahydrofuran, for which interesting discrepancies are observed between theoretical RDFs and RDFs derived from experiment. As appears from this work, theoretical simulations help in estimating the impact of inter- and intramolecular contributions on the RDF and SF separately. Intramolecular contributions dominate the entire behavior of the total SF, so the intramolecular contributions should be projected out, as only the intermolecular terms are of physical interest. Through a combined study of RDFs and SFs, it is explicitely shown that the extraction of partial intermolecular RDFs from a small number of SFs, as is the case for models that extract partial RDFs from experimental (total) data, is ill-conditioned. A correct interpretation of reproduction of structural quantities (RDFs and SFs) is vital, as it has an important impact on derived properties, e.g. the energy of the molecular system, the pressure, the chemical potential, the isothermal compressibility.

The work on liquids has provided more insight into the comprehension of structural behaviour of liquids under different conditions (NVT, NpT). Still, there are more variables that require attention, e.g. longer NpT simulation times and performing MD simulations on other liquids. A similar analysis of the structure factor can be performed on chloroform where also spurious spikes in the RDFs, derived from experimental data, are noticed. The appearance of these peaks probably have the same source of error as as in the case of tetrahydrofuran, so a profound analysis of the structure factors of chloroform would be highly valuable.

Molecular modeling is a relatively new branch of science. There is much more to explore in the world of molecular dynamics, and the topics covered in this thesis are no exception. As the exploration of the field of molecular modeling and simulation continues to grow, together with the ever rising expansion of computer power, society will benefit more and more from this floroushing branch of science.

Bibliography

[1] T. Schlick. *Molecular Modeling and Simulation*. Springer, 2006.

[2] B. J. Alder and T. E. Wainwright. *J. Chem. Phys.*, 27(5):1208, 1957.

[3] D. Frenkel and B. Smit. *Understanding Molecular Simulation*. Academic Press, Elsevier, 2006.

[4] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 1987.

[5] R. P. Feynman and A. R. Hibbs. *Quantum Mechanics and Path Integrals*. McGraw-Hill, 1965.

[6] G. S. Fanourgakis, G. K. Schenter, and S. S. Xantheas. *J. Chem. Phys.*, 125(14):141102, 2006.

[7] B. Guillot and Y. Guissani. *J. Chem. Phys.*, 108(24):10162, 1998.

[8] M. E. Tuckerman, D. M. Marx, M. L. Klein, and M. Parrinello. *Science*, 275:817, 1997.

[9] E. Schwegler, J. C. Grossman, F. Gygi, and G. Galli. *J. Chem. Phys.*, 121(11):5400, 2004.

[10] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard. *J. Phys. Chem. A*, 105(41):9396, 2001.

[11] P. Hohenberg and W. Kohn. *Phys. Rev.*, 136(3B):B864, 1964.

[12] P. Hohenberg and L. J. Sham. *Phys. Rev.*, 140(4A):A1133, 1965.

[13] C. J. Cramer. *Essentials of Computational Chemistry: Theories and Models*. Wiley: Cicester, 2005.

[14] D. Marx and J. Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009.

[15] W. Koch and M. C. Holthausen. *A Chemist's guide to Density Functional Theory*. Wiley-VCH, 2002.

[16] CP2K developers group. `http://www.cp2k.org`.

[17] M. Frisch *et al.* Gaussian 03. Gaussian, Inc. Wallingford CT, 2003.

[18] M. Frisch *et al.* Gaussian 09. Gaussian, Inc. Wallingford CT, 2009.

[19] M. Lutz. *Learning Python*. O'Reilly, 2009.

[20] Center for Molecular Modeling. `http://www.molmod.ugent.be/software`.

[21] T. Verstraelen, M. Van Houteghem, V. Van Speybroeck, and M. Waroquier. *J. Chem. Inf. Model.*, 48(12):2414, 2008.

[22] T. Verstraelen, V. Van Speybroeck, and M. Waroquier. *J. Chem. Inf. Model.*, 48(7):1530, 2008.

[23] T. Verstraelen. *Molecular Modeling of the Synthesis of Zeolites and Related Nanoporous Materials*. PhD thesis, Ghent University, 2009.

[24] S. M. Auerbach, K. A. Carrado, and P. K. Dutta. *Handbook of Zeolite Science and Technology*. Marcel Dekker, Inc., 2003.

[25] D. W. Breck. *Zeolite Molecular Sieves. Structure Chemistry and use*. Wiley, New York, 1974.

[26] R. M. Barrer. *Hydrothermal Chemistry of Zeolites*. Academic Press, London, 1982.

[27] S. Bhatia. *Zeolite Catalysis: Principles and Applications*. CRC Press, Boca Raton, 1990.

[28] C. Baerlocher, W. M. Meier, and D. H. Olson. *Atlas of Zeolite Framework Types*. Elsevier, fifth revised edition, 2001.

[29] International Zeolite Association. `http://www.iza-structure.org`.

[30] M. Guisnet and J.-P. Gilson. *Catalytic Science Series: Volume 3 – Zeolites for Cleaner Technologies*. Imperial College Press, 2002.

[31] J. Weitkamp and L. Puppe. *Catalysis and Zeolites: Fundamentals and Applications*. Springer, 1999.

[32] A. Galarneau, F. Di Renzo, F. Fajula, and J. Vedrine. *Zeolites and Mesporous Materials at the Dawn of the 21st Century*. Elsevier, 2001.

[33] C. Martínez and J. Pérez-Pariente, editors. *Zeolites and ordered porous solids: fundamentals and applications*. 3rd FEZA school on zeolites, Polytechnic University of Valencia, 2011.

[34] C. S. Cundy and P. A. Cox. *Chem. Rev.*, 103(3):663, 2003.

[35] C. S. Cundy and P. A. Cox. *Microp. Mesop. Mater.*, 82(1–2):1, 2005.

[36] P. A. Jacobs and J. A. Martens. *Synthesis of High-Silica Aluminosilicate Zeolites*, volume 33 of *Studies in Surface Science and Catalysis*. Elsevier, 1987.

[37] R. Szostak. *Molecular Sieves, principles of synthesis and indentification.* Van Nostrand Reinhold, New York, 1989.

[38] R. Szostak. *Handbook of Molecular Sieves.* Van Nostrand Reinhold, New York, 1992.

[39] J. N. Watson, L. E. Iton, R. I. Keir, J. C. Thomas, T. L. Dowling, and J. W. White. *J. Phys. Chem. B*, 101(48):10094, 1997.

[40] B. J. Schoeman, J. Sterte, and J. E. Otterstedt. *Zeolites*, 14(2):110, 1994.

[41] R. Ravishankar, C. E. A. Kirschhock, P.-P. Knops-Gerrits, E. J. P. Feijen, P. J. Grobet, P. Vanoppen, F. C. De Schryver, G. Miehe, H. Fuess, B. J. Schoeman, P. A. Jacobs, and J. A. Martens. *J. Phys. Chem. B*, 103(24):4960, 1999.

[42] C. E. A. Kirschhock, R. Ravishankar, F. Verspeurt, P. J. Grobets, P. A. Jacobs, and J. A. Martens. *J. Phys. Chem. B*, 103(24):4965, 1999.

[43] C. E. A. Kirschhock, R. Ravishankar, L. Van Looveren, P. A. Jacobs, and J. A. Martens. *J. Phys. Chem. B*, 103(24):4972, 1999.

[44] C. E. A. Kirschhock, S. P. B. Kremer, P. J. Grobet, P. A. Jacobs, and J. A. Martens. *J. Phys. Chem. B*, 106(19):4897, 1999.

[45] D. Lesthaeghe, P. Vansteenkiste, T. Verstraelen, A. Ghysels, C. E. A. Kirschhock, J. A. Martens, V. Van Speybroeck, and M. Waroquier. *J. Phys. Chem. C*, 112(25):9186, 2008.

[46] P. H. Berens and K. R. Wilson. *J. Chem. Phys.*, 74(9):4872, 1981.

[47] D. McQuarrie. *Statistical mechanics.* Harper & Row, 1976.

[48] D. W. Noid, M. L. Koszykowski, and R. A. Marcus. *J. Chem. Phys.*, 67(2):404, 1977.

[49] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, 1965.

[50] A. Depla, E. Verheyen, A. Veyfeyken, M. Van Houteghem, K. Houthoofd, V. Van Speybroeck, M. Waroquier, C. E. A. Kirschhock, and J. A. Martens. *J. Phys. Chem. C*, 115(22):11077, 2011.

[51] A. Depla, D. Lesthaeghe, T. S. van Erp, A. Aerts, K. Houthoofd, F. Fan, C. Li, V. Van Speybroeck, M. Waroquier, C. E. A. Kirschhock, and J. A. Martens. *J. Phys. Chem. C*, 115(9):3562, 2011.

[52] M. Kaupp, M. Bühl, and V. G. Malkin. *Calculation of NMR and EPR Parameters.* Wiley-VCH, 2004.

[53] C. N. Banwell. *Fundamentals of Molecular Spectroscopy.* McGraw-Hill, 1983.

[54] K. Capelle. *Braz. J. Phys.*, 36(4A):1318, 2006.

[55] C. C. J. Roothaan. *Rev. Mod. Phys.*, 23(2):69, 1951.

[56] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press, 1989.

[57] R. G. Parr and W. Yang. *Ann. Rev. Phys. Chem.*, 46:701, 1995.

[58] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. *Comput. Phys. Commun.*, 167(2):103, 2005.

[59] S. F. Boys and F. Bernardi. *Mol. Phys.*, 19(4):553, 1970.

[60] F. van Duijneveldt, J. van Duijneveldt - van de Rijdt, and J. van Lenthe. *Chem. Rev.*, 94(7):1873, 1994.

[61] G. W. Neilson, A. K. Adya, and S. Ansell. *Annu. Rep. Prog. Chem., Sect. C*, 98(2):273, 2002.

[62] J. L. Finney and A. K. Soper. *Chem. Soc. Rev.*, 23(1):1, 1994.

[63] T. E. Faber and J. M. Ziman. *Phil. Mag.*, 11(109):153, 1965.

[64] J.-P. Hansen and I.R. McDonald. *Theory of Simple Liquids*. Academic Press, Elsevier, 2005.

[65] A. R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education Limited, 2001.

[66] Z. Steinczinger and L. Pusztai. *Condens. Matter Phys.*, 15(2):23606, 2012.

[67] A. K. Soper. *ISRN Physical Chemistry*, 2013:279463, 2013.

[68] T. van Mourik. *J. Phys. Chem. A*, 112(44):11017, 2008.

[69] D. Tzeli, I. D. Petsalakis, and G. Theodorakopoulos. *J. Phys. Chem. A*, 111(36):8892, 2007.

[70] A. E. Shields and T. van Mourik. *J. Phys. Chem. A*, 111(50):13272, 2007.

[71] L. F. Holroyd and T. van Mourik. *Chem. Phys. Lett.*, 442(1–3):42, 2007.

[72] P. Hobza and Z. Havlas. *Theor. Chem. Acc.*, 99(6):372, 1998.

[73] P. Salvador, B. Paizs, M. Duran, and S. Suhai. *J. Comput. Chem.*, 22(7):765, 2001.

[74] R. Crespo-Otero, L. A. Montero, W.-D. Stohrer, and J. M. De la Vega. *J. Chem. Phys.*, 123(13):134107, 2005.

[75] D. Tzeli and A. A. Tsekouras. *Chem. Phys. Lett.*, 496(1-3):42, 2010.

[76] S. Simon, J. Bertran, and M. Sodupe. *J. Phys. Chem. A*, 105(17):4359, 2001.

[77] J. Schmidt, J. VandeVondele, I.-F. William Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy. *J. Phys. Chem. B*, 113(35):11959, 2009.

[78] J.-W. Handgraaf, T. S. van Erp, and E. J. Meijer. *Chem. Phys. Lett.*, 367(56):617, 2003.

[79] J.-W. Handgraaf, E. J. Meijer, and M.-P. Gaigeot. *J. Chem. Phys.*, 121(20):10111, 2004.

[80] M. E. Martín, M. L. Sánchez, F. J. del Valle, and M. A. Aguilar. *J. Chem. Phys.*, 116(4):1613, 2002.

[81] T. Ishiyama and A. Morita. *J. Chem. Phys.*, 131(24):244714, 2009.

[82] T. Ishiyama, V. V. Sokolov, and A. Morita. *J. Chem. Phys.*, 134(2):024509, 2011.

[83] S. Patel and C. L. Brooks. *J. Chem. Phys.*, 122(2):024508, 2005.

[84] M. L. Sánchez, M. E. Martín, M. A. Aguilar, and F. J. del Valle. *J. Comput. Chem.*, 21(9):705, 2000.

[85] M. J. McGrath, I.-F. Kuo, and J. I. Siepmann. *Phys. Chem. Chem. Phys.*, 13:19943, 2011.

[86] I.-C. Lin, A. P. Seitsonen, M. D. Coutinho-Neto, I. Tavernelli, and U. Rothlisberger. *J. Phys. Chem. B*, 113(4):1127, 2009.

[87] J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli. *J. Chem. Phys.*, 120(1):300, 2004.

[88] S. Grimme. *J. Comput. Chem.*, 25(12):1463, 2004.

[89] S. Grimme. *J. Comput. Chem.*, 27(15):1787, 2006.

[90] G. Lippert, J. Hutter, P. Ballone, and M. Parrinello. *J. Phys. Chem.*, 100(15):6231, 1996.

[91] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, C. J. Mundy, J. VandeVondele, J. Hutter, F. Mohamed, and M. Krack. *J. Phys. Chem. A*, 110(2):640, 2006.

[92] R. L. McGreevy. *J. Phys.: Condens. Matter*, 13(46):R877, 2001.

[93] A. K. Soper. *Phys. Rev. B*, 72(10):104204, 2005.

[94] D. T. Bowron, J. L. Finney, and A. K. Soper. *J. Am. Chem. Soc.*, 128(15):5119, 1996.

II
Papers

Paper I

Atomic Velocity Projection Method: A New Analysis Method for Vibrational Spectra in Terms of Internal Coordinates for a Better Understanding of Zeolite Nanogrowth

M. Van Houteghem, T. Verstraelen, D. Van Neck, C. Kirschhock, J.A. Martens, M. Waroquier, V. Van Speybroeck

Journal of Chemical Theory and Computation, 7 (4), 1045-1061, 2011

JCTC Journal of Chemical Theory and Computation ARTICLE
pubs.acs.org/JCTC

Atomic Velocity Projection Method: A New Analysis Method for Vibrational Spectra in Terms of Internal Coordinates for a Better Understanding of Zeolite Nanogrowth

Marc Van Houteghem,[†] Toon Verstraelen,[†] Dimitri Van Neck,[†] Christine Kirschhock,[‡] Johan A. Martens,[‡] Michel Waroquier,[†] and Veronique Van Speybroeck*[,†]

[†]Center for Molecular Modeling, QCMM Alliance Ghent-Brussels, Ghent University, Technologiepark 903, B-9052 Zwijnaarde, Belgium

[‡]Center for Surface Chemistry and Catalysis, Leuven University, Kasteelpark Arenberg 23, B-3001 Heverlee, Belgium

Supporting Information

ABSTRACT: An efficient protocol is presented to identify signals in vibrational spectra of silica oligomers based on theoretical molecular dynamics (MD) simulations. The method is based on the projection of the atomic velocity vectors on the tangential directions of the trajectories belonging to a predefined set of internal coordinates. In this way only contributions of atomic motions along these internal coordinates are taken into consideration. The new methodology is applied to the spectra of oligomers and rings, which play an important role in zeolite synthesis. A suitable selection of the relevant internal coordinates makes the protocol very efficient but relies on intuition and theoretical insight. The simulation data necessary to compute vibrational spectra of relevant silica species are obtained through MD using proper force fields. The new methodology—the so-called velocity projection method—makes a detailed analysis of vibrational spectra possible by establishing a one-to-one correspondence between a spectral signal and a proper internal coordinate. It offers valuable perspectives in understanding the elementary steps in silica organization during zeolite nanogrowth. The so-called velocity projection method is generally applicable on data obtained from all types of MD and is a highly valuable alternative to normal-mode analysis which has its limitations due to the presence of many local minima on the potential energy surface. In this work the method is exclusively applied to inelastic neutron scattering, but extension to the infrared power spectrum is apparent.

INTRODUCTION

Zeolites are microporous inorganic materials, mostly with aluminosilicate components, which exhibit crystal structures containing pores and cages large enough to permit the diffusion of small molecules. Zeolites are indispensable in many industrial applications, e.g., in heterogeneous catalysis, absorption and molecular separation, and ion exchange. These applications and the extension of their application field in many domains are a motivation for further investigation aiming at an even deeper insight into the behavior of a zeolite.[1−3]

Understanding how zeolites nucleate and grow is of fundamental scientific and technological importance. Insight into the molecular mechanisms of structuring of silica can lead to the development of new hierarchical materials promising high potential for optimization of processes in catalysis and molecular separation. In addition, controlled zeolite synthesis could open new fields of application, such as optical electronics,[4] bioimplants,[5] etc. A variety of efforts have already been made to elucidate the early stages of zeolite growth.[6−9] The formation of the siliceous zeolite silicalite with MFI topology is one of the best studied cases.[10−20] Colloidal silicalite-1 is synthesized from hydrolysis of tetraethylorthosilicate (TEOS) as a monomeric silica source in aqueous tetrapropylammonium hydroxide (TPAOH) at room temperature.[21−24] In this study we will focus on these early stages of zeolite formation from a modeling perspective using vibrational spectroscopy: infrared (IR) and inelastic neutron scattering (INS) spectroscopy. These are important identification tools in zeolite synthesis, but the spectra are sometimes difficult to interpret. In ref 25 it was found that at the beginning of TEOS hydrolysis, small oligomers are formed which grow in number and size as the reaction progresses to form nanoparticles. Comparison between simulated and experimental IR patterns has illustrated how the silica contained in the colloidal nanoparticles evolve with time, leading from small five-ring oligomers toward successively more condensed five-ring species.[22,24,26]

Two methods are conventionally used for the calculation of vibrational spectra linking spectral patterns with atomic motions: normal-mode analysis (NMA) based on static approaches and a Fourier-based technique which receives input from molecular dynamics (MD) simulations. NMA is computationally less demanding but is restricted to the harmonic approximation which allows only small deviations from a local or global minimum on the potential energy surface (PES). The restriction stems from the fact that NMA is in essence a second-order approximation of the minima in the PES. As a consequence, only one minimum can be calculated and just a small part of the

Received: September 20, 2010
Published: March 03, 2011

surface is taken into account, which limits the technique when multiple minima are present.[27] Static approaches can go beyond the harmonic approximation. We refer in particular to the work of Scribano and Benoit[28] where the vibrational self-consistent field method (VSCF) is combined with the single-to-all (STA) approach succeeding in the construction of a sparse PES at high ab initio levels of theory. With the second technique, data from MD simulations are used to calculate spectra. In the past many papers have already been published in literature aiming at identifying signals in the Raman or IR spectrum as signatures for specific vibrations characterizing the structure of the molecular system under study. It is not the intention to give a complete survey of all works published in this area, we limit ourselves to those related with vibrations in silica particles. In particular we refer to ref 29 as a pioneering work in which vibrational spectra have been interpreted by means of a normal-mode analysis in terms of symmetry coordinates predicted by point group theory. Other papers analyzed the vibrational eigenmodes with the help of MD to identify signals as signatures of ring structures in silica.[30] Their focus is mainly devoted to the description of symmetric stretching modes, such as breathing modes in rings, as these are recognized to be Raman active. The vibrational eigenmodes are projected onto the coherent breathing modes of the bridging oxygen atoms. They could in many cases provide an unambiguous interpretation for the origin of the Raman lines. Also the group of Smirnov has been very active in this domain.[31−37] Power spectra belonging to a suitably chosen symmetric internal coordinate describing the breathing vibration of a specific ring are constructed and were found to be able to discriminate the various ring structures. To get a better insight into the participation of atoms of different types in the vibrational spectra, a power spectrum for each kind of atom was computed by Fourier transformation of the atomic velocity autocorrelation functions. Some MD techniques (e.g., ab initio MD) are computationally expensive while others are not (e.g., based on force fields). So it depends on the chosen MD method whether or not a large part of the PES can be scanned. The reliability of such numerical simulation methods to reproduce experimental results strongly depends on the method to calculate the PES. Accurate potentials are necessary to describe the atomic and molecular forces of the molecule under study and to reproduce the structural properties and dynamical data. In order to obtain a reliable spectrum it is essential to scan a large part of the PES and thus to simulate during a representative time frame (∼ns range). In view of this, only force field-based methods are viable to calculate the PES as quantum mechanically based methods appear to be too expensive for larger systems and for longer time scales, although serious progress is made on this issue by the group of Parrinello[38,39] by combining Car−Parrinello and Born−Oppenheimer MD. Most of the existing protocols to derive vibrational spectra from the data collected during MD simulations are applicable regardless of whether force fields or ab initio MD are used. Despite many attempts to interpret the vibrational spectra and to define the internal mode contributing to a certain band in the vibrational spectrum further model development remains desirable. Very recently Jacob and Reiher[40] developed a model in which the frequencies at which the bands in the vibrational spectra appear, and the total intensities of these bands can be interpreted in terms of localized modes. This method seems to be very appropriate in analyzing spectra of large molecules like polypeptides and proteins, but the protocol is constrained to static quantum chemical calculations.

In this paper we present a completely new methodology for the interpretation of vibrational spectra but based on molecular dynamical approaches. In the new model the atomic velocities are projected on properly selected directions fully determined by the internal mode from which we want to know its impact on the spectrum. Numerical applications of the new method are based on classical MD simulations. For the silica building blocks under investigation in this work, we use a silica-derived force field based on the gradient curves method[41] which was previously successfully applied to investigate the MFI fingerprint in zeolites.[26]

The infrared power spectrum is computed by taking the Fourier transform of the autocorrelation function of the time variation of the electric dipole moment. Similarly, the velocity power spectrum or INS spectrum is determined by the Fourier transform of the atomic velocities, which is easier to compute after a MD run as it only needs the instantaneous atomic velocities at each time step. Compared to NMA, this dynamical approach is more flexible in the treatment of nonharmonic problems, although also in static approaches progress is made in going beyond the harmonic approximation as already stated.

The first goal of this paper is to identify peaks in vibrational spectra of zeolite building blocks with specific internal degrees of freedom. In particular we were interested in how the size of rings and/or connectivity affects signals in the vibrational spectrum. For this purpose a new method has been developed enabling us to link the spectrum with the internal degrees of freedom of zeolite nanoparticles. This method provides a suitable tool that predicts which particular modes are influenced by changing size and topology of the silica building blocks. As such, it constitutes a significant progress in understanding the mechanisms during the early stages of the zeolite growth, and it gives complementary and valuable information in addition to the usual protocol of spectra analysis where experimentally measured spectra are compared with fully theoretical spectra extracted from simulations of the zeolite frameworks.[34−37,42] The basic understandings of zeolite spectra and the lattice dynamics of zeolites are well-known.[43−45] But a computational method for an unambiguous validation and confirmation of peak assignments in smaller silica particles can still be improved. In addition, it can in turn be helpful in the development and/or refinements of models to compute atomic forces, e.g., force fields. If a MD run does not yield the correct vibrational spectrum with regard to experiment, in the sense that some peaks are missing or not at the correct position, the new methodology makes it possible to specify the internal mode that is poorly described. Relevant terms in the force field can be adjusted accordingly to remove the discrepancies.

In ref 26 the MFI topology was studied in terms of shifts of peaks as MFI-structured nanoparticles grow. As the particles become bigger in size, the fingerprint band lowers in frequency from 650 to 550 cm^{-1}. An isolated five-membered ring (5T) reveals an IR-active vibration around 650 cm^{-1}. By connecting pentasil rings, slightly larger building blocks are constructed, but no substantial changes for the peak position have been observed. The situation changes if a more condensed structure is formed, such as the silica octamer, referred to as 8T (4 × 5) (see Figure 1l); a sudden shift by approximately 50 cm^{-1} is observed. The red shift becomes even more substantial when the particles grow larger. This 550 cm^{-1} band is regarded as the spectroscopic

Journal of Chemical Theory and Computation ARTICLE

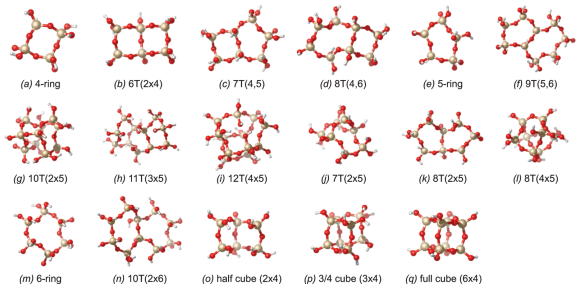


Figure 1. Simulated molecular structures of silica particles consisting of combinations between various four-, five-, and six-rings.

signature of MFI-type zeolites and is also a typical feature of similar structures built from five-membered rings, the so-called pentasil zeolites.

In this paper we apply a new methodology—the so-called velocity projection method—to a large variety of building blocks encountered in zeolite synthesis, including four-, five-, and six-rings and all types of connections between them (Figure 1). The analysis presented here should serve as a basic understanding of which modes and which peaks in the vibrational spectra are influenced by the particular interconnectivity and topology. The lack of experimental vibrational spectra of such small silica particles in the literature strengthens the importance of such a theoretical investigation. For larger species experimental data are available.

■ COMPUTATIONAL SECTION

The input of the computed spectra is generated from MD simulations. Interatomic interactions between framework atoms were represented by an in-house developed force field. It was especially designed for zeolites and was calibrated at the post-Hartree−Fock MP2/6-311+g(d,p) level of theory, with the gradient curves method (GCM).[41] This is a novel technique that facilitates the development of transferable force-field models. It makes extensive use of regularization techniques[46] and of generic energy terms based on series expansions to obtain an optimal bias−variance trade-off during the fitting procedure. The force field was previously thoroughly benchmarked to reproduce the MFI fingerprint and IR-band shifts in other related small silica nanoparticles. All optimizations and simulations have been carried out with the CP2K[47] program package.[26]

All initial geometries were built with the in-house developed software program package ZEOBUILDER.[48] Geometry optimizations were performed on all structures using the conjugate gradient method. With these optimized coordinates as input, an initial equilibration MD run of 5 ps was carried out. The actual production run which provided the data for analysis was an MD run of 1 ns at a temperature of 300 K with a Nose thermostat in the NVT statistical ensemble. To process the data, the MD-TRACKS program[49] was used.

From linear response theory, if the dipole moment history is obtained in an MD calculation, then the infrared spectrum $I_{IR}(\omega)$ can be computed as[50−53]

$$I_{IR}(\omega) = \lim_{t \to \infty} \frac{1}{t} \sum_{\alpha=x,y,z} \left| \int_0^t \frac{d\mu_\alpha}{dt} \exp(-i\omega t) dt \right|^2 \quad (1)$$

where ω is the angular frequency, and μ_α is the Cartesian components of the dipole moment. This is equivalent to the Fourier transform of the autocorrelation function of the time derivative of the dipole moment.[53,54] The dipole moment is defined as

$$\boldsymbol{\mu} = -\int (\boldsymbol{r} - \boldsymbol{R}_c)\rho(\boldsymbol{r}) d\boldsymbol{r} + \sum_{i=1}^{N} Z_i(\boldsymbol{R}_i - \boldsymbol{R}_c) \quad (2)$$

where $\rho(\boldsymbol{r})$ is the electron density, Z_i represents the nuclear charge of atom i, $\boldsymbol{R}_i$ is the position of atom i, and N the total number of atoms. The reference point $\boldsymbol{R}_c$ is the center of charge for charged systems and is arbitrary for neutral systems. In molecular mechanics, one preferentially uses an effective charge Q_i for each atom, and the dipole moment is written as

$$\boldsymbol{\mu} = \sum_{i=1}^{N} Q_i(\boldsymbol{R}_i - \boldsymbol{R}_c) \quad (3a)$$

$$\frac{d\boldsymbol{\mu}}{dt} = \sum_{i=1}^{N} Q_i \boldsymbol{v}_i \quad (3b)$$

In this work fixed charges were used. The velocity power spectrum or inelastic neutron scattering spectrum (INS) can be obtained in a similar way as the infrared signal using an autocorrelation function:

$$I_{INS}(\omega) = \lim_{t \to \infty} \frac{1}{t} \sum_{\substack{i=1 \\ \alpha=x,y,z}}^{N} \left| \int_0^t \boldsymbol{v}_{i,\alpha}(t) \exp(-i\omega t) dt \right|^2 \quad (4)$$

dx.doi.org/10.1021/ct100538c | J. Chem. Theory Comput. 2011, 7, 1045–1061

Journal of Chemical Theory and Computation ARTICLE

The IR and INS spectra are based on a Fourier transform which transforms information from the time domain to the frequency domain. One can easily see that both spectra are directly dependent on the velocities of the atoms $\boldsymbol{R}_{i\alpha} = \boldsymbol{v}_{i\alpha}$ so that the positions of peaks in IR and INS spectra coincide, but the intensities will vary.

The Fourier transform of the velocity autocorrelation provides information about the density of vibrational modes as a function of energy and hence reveals the underlying frequencies of the molecular system. The IR spectrum shows which internal modes are infrared active. If some of the IR active modes are localized on specific functional groups, then the corresponding bands can then serve as a spectral fingerprint of the groups. However, we focus in this paper on the INS spectra since charges in force fields are not (yet) well-defined in order to obtain reliable qualitative spectral amplitudes. Moreover, cancellation effects frequently appear in the time-varying dipole moment vector (eq 3a). It implies that the amplitudes in an IR spectrum are smaller and faster affected by relative noise. On the contrary, the INS spectrum contains only a summation of positive contributions (eq 4), and the relative noise on the spectrum becomes smaller. It also provides information about all vibrations of the system, i.e., all IR and Raman active modes.

METHOD

Velocity Projection Method. In order to associate peaks in the full INS spectrum to particular motions of internal coordinates we developed a method where the atomic velocities are projected to selective directions defined by the specific nature of the internal coordinate. An outline of this projection method is sketched below.

We make use of internal coordinates in order to assign a peak to one particular degree of freedom . An internal coordinate (IC) defines the location of the atoms in a molecule relative to the other atoms in the molecule. Some examples of IC's are bond distances, bending angles, dihedrals, Urey−Bradley stretches,[55] a linear combination of stretches, etc. Each internal coordinate q can be expressed as a function of the atomic Cartesian coordinates $\boldsymbol{R}_{i\alpha}$ and depends on time: $q = q(\boldsymbol{R}_{i\alpha}(t))$. The velocity projection method then consists of projecting the Cartesian velocity vector $\boldsymbol{v}_{i\alpha} = \dot{\boldsymbol{R}}_{i\alpha}$ of each atom i on the atomic tangent vector of the internal coordinate q at each time step. The IC's themselves are overcomplete and are, in contrast with NMA, not orthogonal to each other. In this section, the different classes of IC's that were used in this work will be illustrated. If one considers only one IC, then the definition of the tangent vector of atom i at each time step is given by

$$\boldsymbol{J}_{i\alpha} = \frac{\partial q}{\partial \boldsymbol{R}_{i\alpha}} \tag{5}$$

where α runs over the Cartesian x, y, and z components of atom i. Obviously, i must be one of the atoms that specifies the IC $q(\boldsymbol{R}_{i\alpha})$. The velocity vector belonging to atom i can be decomposed into its tangential and normal component with respect to the atomic tangent vector $\boldsymbol{J}_{i\alpha}$:

$$\boldsymbol{v}_{i\alpha} = (\boldsymbol{v}_{\perp})_{i\alpha} + (\boldsymbol{v}_{//})_{i\alpha} \tag{6}$$

Only the tangential component gives a nonvanishing contribution to the change of the internal coordinate:

$$\delta q = \sum_i \boldsymbol{J}_i \cdot \delta \boldsymbol{R}_i \tag{7}$$

and with a suitable normalization can be set as

$$(\boldsymbol{v}_{//})_{i\alpha} = \frac{(\sum_{j\beta} \boldsymbol{v}_{j\beta}\boldsymbol{J}_{j\beta})\boldsymbol{J}_{i\alpha}}{\sum_{j\beta} \boldsymbol{J}_{j\beta}^2} \tag{8}$$

while the normal component lets the IC unchanged in first order (for small time steps):

$$(\boldsymbol{v}_{\perp})_{i\alpha} = \boldsymbol{v}_{i\alpha} - (\boldsymbol{v}_{//})_{i\alpha} \tag{9}$$

The parallel velocity is the component that should be identified with this particular internal coordinate. Note that the IC is a function of time, and we project on a time dependent vector. To compute the actual spectrum, the Cartesian components of the velocities of every atom at each time step are plugged into eqs 1 and 4. The fluctuations due to the time dependency of the IC's are also computed with the Fourier transformation but should be small enough to neglect. Accordingly, an autocorrelation function (eq 4) is computed with the instantaneous $(\boldsymbol{v}_{//})_{i\alpha}$ velocity projections. In this way a partial INS spectrum has been constructed restricted to those vibrations inducing changes of the IC of interest. A comparison between the full and partial INS spectra makes a full analysis of the vibrational spectrum possible: spectral peaks can be linked to IC's and corresponding (internal) degrees of freedom. When a peak in the projected spectrum coincides with its counterpart in the total spectrum, one can assume that there is no motion along other orthogonal coordinates that contributes to this peak. Such a peak is then completely resolved.

We note that the normalization of the tangential component (eq 8) is not unique and that a separate projection of the atomic velocity to the tangent vector belonging to that atom leads to a somewhat different normalization factor:

$$(\boldsymbol{v}_{//})_{i\alpha} = \frac{(\sum_{\beta} \boldsymbol{v}_{j\beta}\boldsymbol{J}_{j\beta})\boldsymbol{J}_{i\alpha}}{\sum_{\beta} \boldsymbol{J}_{j\beta}^2} \tag{10}$$

For computational reasons we have given preference to the normalization in (eq 8) to avoid numerical inaccuracies generated by tangent vectors $\boldsymbol{J}_i$ of very small amplitude, as would be the case with the choice of solution 10.

Internal Coordinates. *Analysis Using Internal Coordinates.* At this point it is useful to outline the exact workflow. The standard procedure to obtain a vibrational spectrum from MD data is to compute the Fourier transform of the Cartesian atomic velocities (eqs 1 and 4) at each time step. In this work we project the Cartesian atomic velocities onto IC's to obtain the velocity component $(\boldsymbol{v}_{//})_{i\alpha}$ of each atom alongside the IC (eq 8 and eq 9). Of these projected velocities, which are by definition equal to or smaller than the atomic Cartesian velocities, the Fourier transform is also computed at each time step to obtain the spectrum of the velocities for each IC separately. It can happen that, due to symmetry, some classes of IC's generate almost identical spectra; then the average was taken over them. In the next section the different classes of internal coordinates that were used in this work are discussed.

Classes for Internal Coordinates. The peaks in the full spectra can be (partially) decomposed by a proper selection of IC's. This can be done for the various zeolite building blocks under study in

Journal of Chemical Theory and Computation ARTICLE

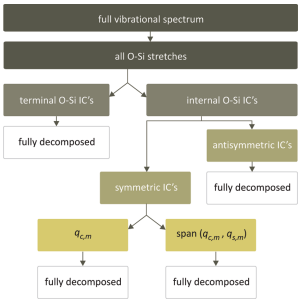


Figure 2. Flowscheme to classify different IC's used to decompose vibrational zeolite spectra.

this work and displayed in Figure 1. The IC's or linear combinations of them are then classified. Each class has its specific internal motion and generates a partial INS spectrum for each building block.

We will follow the following protocol. For all structures we compute the full vibrational spectrum based on the (nonprojected) atomic velocities as resulting from the MD simulation. For each individual IC we then project the atomic velocities according to the outlined procedure trying to link spectral peaks to these IC's. Linear combinations of various IC's can be constructed to form new IC's. With a well-chosen linear combination, one is able to fully assign a peak to this new specific internal coordinate. The general workflow for combining IC's to further unravel the spectra and to be able to link them with a spectral peak is shown in Figure 2. In this paper we focus on IC's constructed from stretches between two atoms: the stretches themselves can serve as IC's or linear combinations of them. The stretching vibration is one of the most straightforward internal motions to describe. In an earlier study by van Santen et al.,[45] it was stated that the lattice dynamics of amorphous silica can be correctly described by ignoring the Si−O−Si bending and solely focusing on the Si−O stretchings (see Figure 3). It implies that the wavenumber regime of 500−1200 cm^{-1} is sufficient for studying the zeolite nanogrowth. As the studied species are however much smaller, we will validate in how far only stretches are applicable for fully resolving the spectrum. The higher the connectivity, the higher the rigidity and thus preventing the bending mode. Mixing of the Si−O stretching and Si−O−Si bending modes in the case of small oligomers may be assumed to be of minor importance because their modes are spectrally well separated.

Looking at Figure 3 we can discriminate between two types of O−Si stretches. The first one is the stretch mode, which occurs along the O−Si bond, where the oxygen atom is connected with a terminal hydrogen atom. This will be referred to as the O−Si terminal stretch. The second type consists of the two O−Si stretches in the Si−O−Si bridge. Such O−Si stretch is called an internal O−Si stretch. We make an approximation in the sense that the interaction between the symmetric and antisymmetric modes can be neglected, which is quite reasonable in view of the

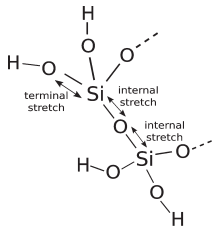


Figure 3. Schematic representation of the Si−O stretches present in the studied zeolite structures.

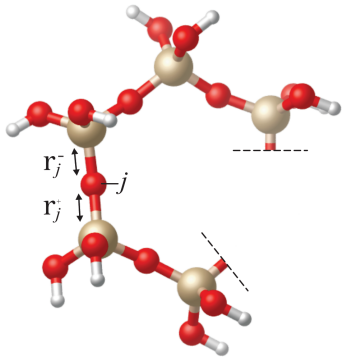


Figure 4. Ring stretching Si−O bonds defining symmetric and antisymmetric IC's.

large difference between the spectral values of these modes and the magnitude of the coupling constant.[45]

One can choose a particular Si−O stretch as IC, but in practice, linear combinations of stretch modes are more appropriate choices for IC's. In this way we separate the 2*N* ring stretching Si−O bonds, present in a *N*-membered ring, into two classes of IC's (symmetric and antisymmetric). For one particular Si−O−Si bridge *j*:

$$q_{ss,j} = \frac{1}{\sqrt{2}}(r_j^- + r_j^+) \tag{11}$$

and

$$q_{as,j} = \frac{1}{\sqrt{2}}(r_j^- - r_j^+) \tag{12}$$

as visualized in Figure 4 .

The coordinate transformation between all stretching internal coordinates and the N symmetric and antisymmetric IC's is orthogonal and of dimension $2N \times 2N$:

$$\begin{bmatrix} \boldsymbol{q}_{ss} \\ \boldsymbol{q}_{as} \end{bmatrix} = \mathbf{T}\mathbf{r} \tag{13}$$

$$\mathbf{r} = (2N \times 1) \text{ matrix (Si} - \text{O stretches)}$$

$$\mathbf{T} = 2N \times 2N \text{ matrix and orthogonal}$$

The force constant matrix in $\{\boldsymbol{q}_{ss},\boldsymbol{q}_{as}\}$ space becomes

$$\mathbf{U} = \frac{1}{2}(\boldsymbol{q}_{ss}^T \boldsymbol{q}_{as}^T)\mathbf{C}_q \begin{bmatrix} \boldsymbol{q}_{ss} \\ \boldsymbol{q}_{as} \end{bmatrix} \tag{14}$$

with $\mathbf{C}_q = \mathbf{T}\mathbf{C}_r\mathbf{T}^T$, which is not block diagonal even with a diagonal structure of $\mathbf{C}_r$:

$$\mathbf{C}_q = \begin{bmatrix} \mathbf{C}_{q_{ss}} & \mathbf{C}_{q_{coupl}} \\ \mathbf{C}_{q_{coupl}} & \mathbf{C}_{q_{as}} \end{bmatrix} \tag{15}$$

In the model proposed it is assumed that the coupling between the symmetric and antisymmetric stretch modes is very weak and may be omitted: $\mathbf{C}_{q_{coupl}} = 0$. This seems to be justified since the frequency bands of both classes are well separated, as will be demonstrated later in the discussion section.

It is interesting to examine the symmetric stretch mode in more detail. The peaks due to the symmetric stretch are not completely resolved, and they lie in the wavenumber region which is interesting for obtaining insight into zeolite synthesis (the antisymmetric stretch peaks are completely resolved and lie in a less interesting area). In order to gain insight into the normal (symmetric stretch) modes of a molecular ring system, we consider a simplified model of n identical and equally spaced masses on a circle. Due to its periodicity, the normal modes of such a system can be obtained as the eigenvectors of a so-called circulant matrix C:

$$C_{jk} = c_{(j-k) \bmod n} \tag{16}$$

for $j, k = 1, ..., n$, where the notation $(i) \bmod n$ implies integer arithmetic modulo n. Since the Hessian of the model system is a symmetric matrix ($C_{jk} = C_{kj}$), the coefficients must also obey

$$c_{(i) \bmod n} = c_{(-i) \bmod n} \tag{17}$$

where we used the notation $-i = n - i$. The circulant matrix C for an even n then has the form:

$$C = \begin{bmatrix} c_0 & c_1 & \cdots & c_{n/2} & c_{n/2-1} & \cdots & c_1 \\ c_1 & c_0 & \cdots & & & & \\ \vdots & & \ddots & & & & \\ c_{n/2} & & & \ddots & & & \\ c_{n/2-1} & & & & \ddots & & \\ \vdots & & & & & & \\ c_1 & & \cdots & & & c_1 & c_0 \end{bmatrix} \tag{18}$$

The form of C for n being odd is analogous. In the remainder we drop the explicit "mod n" notation for the $c_{(i)}$ indices, but it is understood that they should be interpreted in arithmetic modulo n. The eigenvectors of the circulant matrix can be used to create new IC's on which we can project the atomic velocities, based on the symmetric stretch IC's $q_{ss,k}$. With the circulant matrix being the model for the Hessian, we can construct these new internal coordinates, i.e., linear combinations of symmetric stretches, to further resolve the spectra of silica rings.

The eigenvalues and the k^{th} component of the normalized eigenvectors are (see Appendix Section):

$$\lambda^{(m)} = \sum_k c_{(k)} \exp\left(\frac{-2\pi imk}{n}\right) = \sum_k c_{(k)} \cos\left(\frac{2\pi mk}{n}\right) = \lambda^{(-m)} \tag{19}$$

and

$$V_k^{(m)} = \frac{1}{\sqrt{n}} \exp\left(\frac{2\pi imk}{n}\right) \tag{20}$$

Looking at the eigenvalue spectrum one can distinguish between even n (four-, six-ring, etc.) and odd n (five-ring, etc.). For even n the eigenvalues (eigenvectors) become

$$\begin{cases} \lambda^{(0)}(V^{(0)}) : \text{nondegenerate} \\ \lambda^{(m)}(V^{(m)}, V^{(-m)}) \text{ for } m = 1, ..., \frac{n}{2} - 1 : \text{two-fold degenerate} \\ \lambda^{(n/2)}(V^{(n/2)}) : \text{nondegenerate} \end{cases}$$

For odd n they are

$$\begin{cases} \lambda^{(0)}(V^{(0)}) : \text{nondegenerate} \\ \lambda^{(m)}(V^{(m)}, V^{(-m)}) \text{ for } m = 1, ..., \frac{n-1}{2} : \text{two-fold degenerate} \end{cases}$$

It is practical to switch to real eigenvectors for the two-fold degenerate eigenspaces, by taking linear combinations:

$$W_k^{(m)} = \frac{1}{2}(V_k^{(m)} + V_k^{(-m)}) = \cos\left(\frac{2\pi}{n} mk\right) \tag{21a}$$

$$W_k^{(-m)} = \frac{1}{2i}(V_k^{(m)} - V_k^{(-m)}) = \sin\left(\frac{2\pi}{n} mk\right) \tag{21b}$$

It is now possible to construct a new internal coordinate for each eigenvector m by making linear combinations of the symmetric stretches $q_{ss,k}$, where the coefficients are the elements of the eigenvector. For the circular systems under consideration, the new internal coordinates become

$$q_{c,m} = \sum_k W_k^{(m)} q_{ss,k} = \sum_k \frac{V_k^{(m)} + V_k^{(-m)}}{2} q_{ss,k} = \sum_k \cos\left(\frac{2\pi}{n} mk\right) q_{ss,k} \tag{22a}$$

$$q_{s,m} = \sum_k W_k^{(-m)} q_{ss,k} = \sum_k \frac{V_k^{(m)} - V_k^{(-m)}}{2i} q_{ss,k} = \sum_k \sin\left(\frac{2\pi}{n} mk\right) q_{ss,k} \tag{22b}$$

Journal of Chemical Theory and Computation ARTICLE

where n is the number of symmetric stretches in the circular molecular system and $q_{ss,k}$ is short for the k^{th} symmetric stretch IC. Since $q_{c,m}$ and $q_{s,m}$ only differ in phase we could not discriminate between them. Therefore it is useful to consider the subspace spanned by these vectors. This subspace is referred to as span($q_{c,m}$,$q_{s,m}$). At each time step of the simulation we project the atomic velocities on the set of vectors which spans this subspace. For topologically symmetric systems the INS spectra of $q_{c,m}$, $q_{s,m}$ and span($q_{c,m}$,$q_{s,m}$) coincide (e.g., Figure 1b, n, and q and a, e, and m).

Note that this technique can only be used for single closed ring systems, and it is not expected to completely resolve spectra in systems with fused rings. If multiple rings are attached to each other, then one has to distinguish between them to make use of the circulant matrix as a model for the Hessian. This is still useful because the resulting spectra can be directly compared with spectra from other n-ring structures. In this way the immediate effect of a different topology on a specific structure can be studied.

RESULTS AND DISCUSSION

Figure 1 gives an overview of the three-dimensional (3D) optimized geometries of the simulated structures. Some of them are key components of the MFI structure, and others have been suggested to occur in the early stages of other zeolite structures.[26] The initial geometry as input for the MD simulation does not need to coincide with the global minimum as all possible structures on the PES will be visited during the MD run. The most elementary structures are a single four-, five- and six-membered ring (Figure 1a, e, and m), and their spectra may be regarded as reference. Larger building blocks can be constructed by connecting these elementary rings with other n-ring structures. These new structures yield new (partial) INS spectra which can be compared with the reference spectra, and such an investigation provides us information on the influence of the topology (i.e., how atoms are linked to each other) on the vibrational spectra. In the synthesis of zeolites with MFI topology it has been found that the five-membered ring (5T) (Figure 1e), also called the pentasil ring, is of special importance.[56−58] By connecting these pentasil rings larger building blocks can be constructed. If three T-atoms are added, then two 8T(2 × 5) (Figure 1k) and three 11T(3 × 5) (Figure 1h) sideways five-membered rings arise. If two sides of two different five rings are shared (sharing three T-atoms), then the 7T five ring is formed, 7T(2 × 5) (Figure 1j). The five ring 8T(4 × 5) (Figure 1l) is a silica octamer, where the rings form a cagelike structure. The species where two five rings do not share but are connected with an oxygen bridge is referred to as 10T(2 × 5) (Figure 1g). In this topology an extra six ring is present. Finally the silica dodecamer structure forms the 12T(4 × 5) (Figure 1i).[26]

In aqueous silicate solutions the cubic octamer[59,60] often has been observed experimentally and proposed as a building block. Therefore we also consider this full cube (6 × 4) (Figure 1q), a half cube (2 × 4)(Figure 1o), and a three-fourths cube (3 × 4) (Figure 1p).

Physical Meaning of IC's. The most basic molecular structures are the four-, five-, and six-rings (Figure 1a, e, and m, respectively). Their spectra may serve as a reference for other topologies where n_1-rings are attached to n_2-rings. In this way, the vibrational spectra can be applied as a tool for studying topological differences in zeolite structures (e.g., during nanogrowth).

Looking at the diagram of Figure 2, we can distinguish between two types of IC's: those that can be identified as single stretches and those that show a weighted mixture of stretch modes. For the first type of IC's it is straightforward to assign a physical motion to them since they coincide with real stretches. For the second type of IC's it is less straightforward, but they also represent a real physical combined vibration.

Symmetric and antisymmetric ring vibrations have been defined in eqs 11 and 12. When the atomic velocities are projected on these IC's, we gain knowledge on the magnitude of these combined modes relative to the full vibrational spectrum.

The antisymmetric motions represent modes where the O atom oscillates between two Si atoms. Their spectral lines are located at frequencies larger than 800 cm^{-1} and are discussed in subsection Classes for Internal Coordinates.

The characteristic internal modes of a molecular system are determined by the eigenvalue equation of the Hessian:

$$HE = \omega^2 \mathbf{M}\mathbf{E} \tag{23}$$

Here H is the Hessian which contains the second derivatives of the potential energy with respect to the Cartesian coordinates, $\mathbf{M}$ is the mass matrix, ω^2 represents the eigenvalues, and $\mathbf{E}$ represents the $3N \times 3N$ matrix which contains the eigenvectors of the different modes (N = number of atoms). As shown in the Internal Coordinates Section, we introduced the circulant matrix as a model for the Hessian corresponding to the symmetric IC's and determined its eigenvectors and eigenvalues. The physical meaning of the combined modes of IC's $q_{c,m}$ and $q_{s,m}$ is interesting for further survey. For the four-, five-, and six-ring we have the following IC's (see eqs 22a and 22b) ($\alpha = 2\pi/n$):

- Four-ring: $q_{c,0}$, span($q_{c,1}$,$q_{s,1}$), $q_{c,2}$ ($\alpha = 90°$).
- Five-ring: $q_{c,0}$, span($q_{c,1}$,$q_{s,1}$), span($q_{c,2}$,$q_{s,2}$) ($\alpha = 72°$).
- Six-ring: $q_{c,0}$, span($q_{c,1}$,$q_{s,1}$), span($q_{c,2}$,$q_{s,2}$), $q_{c,3}$ ($\alpha = 60°$).

They all can be associated to a real physical mode. Consider, for example, the four-ring. When projecting on the corresponding IC's, the resulting spectral peaks agree with the molecular breathing motions, as depicted in Figure 5, in which an outward arrow means that the particular bond stretches out, whereas an inward arrow points to a shrinking motion of the bond. From this figure it is clear that $q_{c,1}$ and $q_{s,1}$ actually represent the same breathing mode, only shifted in phase by 90°. Thus, whenever a peak arises when projecting on a particular IC, we know with which vibrational eigenmode it corresponds. As a special combined mode we define in this work the breathing mode, in which all the present Si−O stretches are elongated or shrunk simultaneously. Here the mode $q_{c,0}$ represents this breathing mode.

Elementary Building Blocks: Four-, Five-, and Six-Rings. The spectra of all structures considered in this paper for all IC's are taken up in the Supporting Information. In this section some relevant spectra of the four-, five-, and six-rings are shown and will be further explained. We will focus on the spectral shifts that occur when spectra of the four-, five-, and six-rings for the same type of IC are compared. The conclusions that are drawn from these elementary structures are transferable to more complex molecular structures. Also for the sake of completeness INS and IR spectra for the four-, five-, and six-membered ring systems are given in the Supporting Information. They can serve as comparative material when analyzing velocity power spectra associated to specific IC's.

Terminal O−Si and Antisymmetric IC's. The spectra of the terminal O−Si and antisymmetric IC's are shown in Figure 6 for

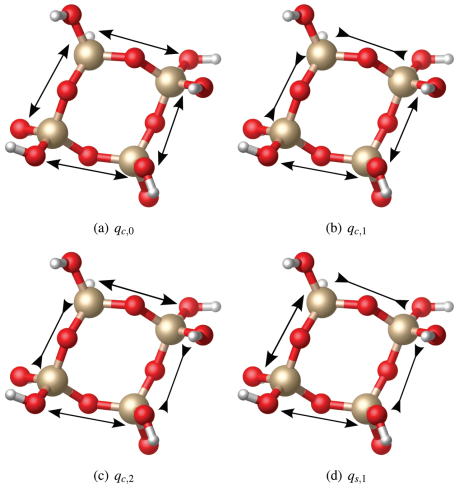


Figure 5. Visualization of the physical breathing modes when projecting atomic velocities on eigenvectors of the circulant matrix as a model for the Hessian.

the three elementary structures. For these classes of stretches multiple IC's are present. Per category the average spectrum is taken, as the spectra belonging to each IC of the same class are in essence not deviating much from each other. Also the standard deviation (stdev) of this average is given (lower curve). For both types of IC's the spectra manifestly show little influence on the size of the ring (same position, relative amplitudes, and shape).

Circulant Symmetric IC's. As outlined in the Method Section, the symmetric IC's are analyzed making use of the properties of the circulant matrix as a model for the Hessian. The spectra of the velocity projections on the $q_{c,m}$ and $q_{s,m}$ IC's of the four-, five-, and six-ring are shown in Figures 7 and 8. From these figures, some interesting features related to ring size can be derived.

(i) Comparing the first eigenmode, $q_{c,0}$, of the four, five-, and six-ring we see that the spectra of this breathing mode only exhibit a very small spectral shift which is almost negligible. The spectrum of the four-ring has peaks at 549 and 888 cm^{-1}, and for the five- and six-rings, we get 553 and 887 cm^{-1} and 562 cm^{-1} and 888 cm^{-1}, respectively. A small third peak emerges at about 453 cm^{-1} when increasing the size of the ring. It is at this stage not clear to which motion this third vibrational peak could be associated. The two peaks in the four-ring spectrum probably belong to the breathing modes of the ring Si and O atoms with a growing impact on the stretches. As the size of the ring increases, a collective nonplanar vibration could emerge. Although intuitive to some extent, these explanations remain speculative as so far no visualization program is available which could remove this ambiguity.

(ii) The peaks of the second mode of the four-ring, span($q_{c,1}$, $q_{s,1}$), make a shift compared to the same mode of the five- and six-rings. The two peaks with the four-ring occur at 642 and 869 cm^{-1}, the five- and six-rings show peaks at 617 and 876 cm^{-1} and 602 and 881 cm^{-1}, respectively. There is some peak drift, and the interpeak distance grows with increasing ring size (red shift for the left peak, blue shift of the right peak). They are still relatively small but big enough to be measurable and to conclude that there is a size dependence, which is more pronounced than in the $q_{c,0}$ mode.

(iii) The above made conclusions are also valid when the eigenmode span($q_{c,2}$,$q_{s,2}$) of the five- and six-rings are compared.

(iv) The interpretation of the spectra for these elementary building blocks, however, is also valid when combinations of these structures are made, as can be seen from the figures taken up in Figure 12 and the Supporting Information, e.g., 7T(4,5), 8T(4,6), and 9T(5,6). There is a very small topology dependence for the $q_{c,0}$ mode, and the influence of the ring size is larger for higher order modes.

We can discuss now how far the above features agree with what has been published in literature. The independence of spectra of the $q_{c,0}$ breathing coordinate on the ring size contradicts results of previous modeling studies. These works have demonstrated a

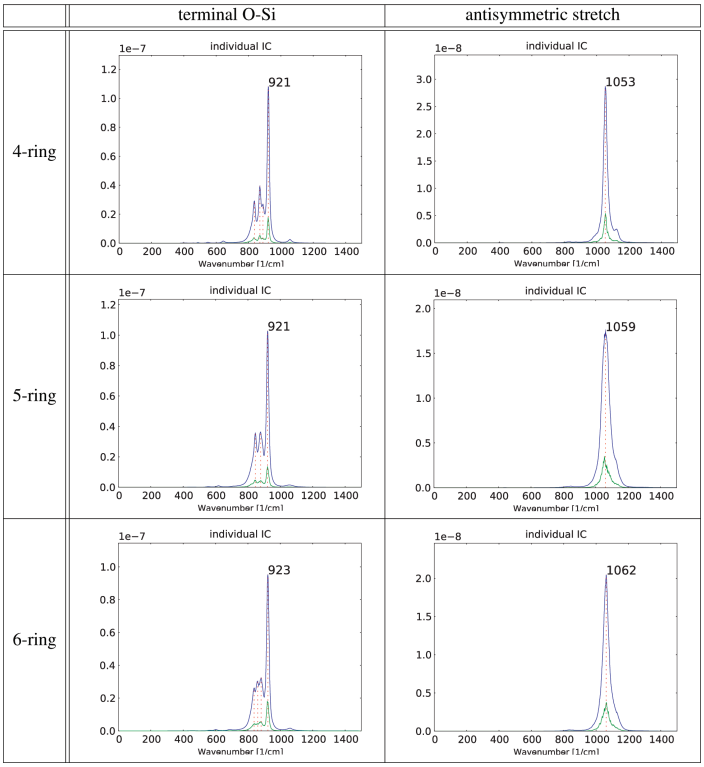


Figure 6. Projection of atomic velocities on the terminal O−Si and antisymmetric stretch IC's.

downward shift of the "ring breathing frequencies" with the increase of ring size. In the paper of Smirnov et al.[31] the internal coordinate measuring the ring pore-opening has been put equal to

$$P(t) = \frac{1}{N/2}\sum_{i=1}^{N/2} R_i(t) \tag{24}$$

where $R_i(t)$ stands for the deviation of the i^{th} ring diameter from the mean value. The ring diameter is defined as the distance between oxygen atoms on the opposite sides of the ring. In order to comment on this issue, we have computed the velocity power spectrum of this vibrational coordinate $P(t)$ making use of the MD results. These calculations could reveal the origin of the apparent discrepancies between the results presented in this work and those of literature. We also did the same exercise for the silica, instead of the oxygen atoms. The results are surprising and are displayed in Figure 9. Some peaks occurring in the $P(t)$ velocity power spectra for the silica atoms perfectly coincide with the $q_{c,O}$ in both four-, five-, and six-ring spectra. More specifically, four-ring: 550 and 892 cm^{-1}; five-ring: 550 and 883 cm^{-1}; and six-ring: 558 and 892 cm^{-1}. They do not show any ring size dependence. On the contrary, the oxygen $P(t)$ velocity power spectra do not show any resemblance with the $q_{c,O}$, except perhaps for the spectral signal at 453 cm^{-1} appearing in six-membered rings. The downward shift of the ring breathing

Journal of Chemical Theory and Computation ARTICLE

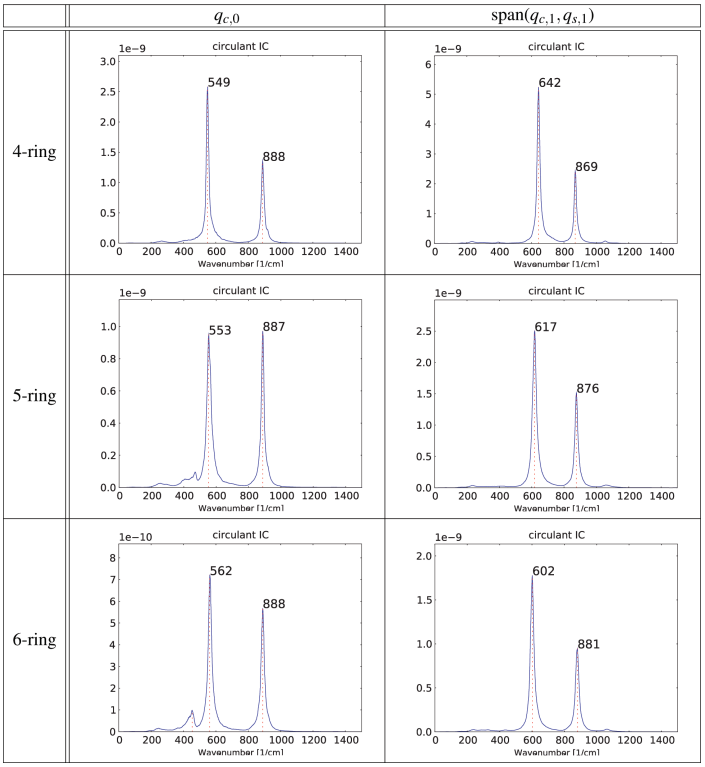


Figure 7. Projection of atomic velocities on the circulant symmetric IC's $q_{c,0}$ and span($q_{c,1}$,$q_{s,1}$) for the four-, five-, and six-ring.

frequencies observed in literature is also reproduced in our $P(t)$ velocity power spectra for the oxygen atoms: 483, 475, and 458 cm^{-1} in going from four- to six-rings. They represent the lowest frequency peak in the spectrum. What can we conclude from this comparative study? Which peak corresponds really with the breathing mode where all ring diameter vibrations are supposed to be in phase? Since the pore-opening vibrational coordinate $P(t)$ does not take into account the phase of all ring diameter vibrations, the $P(t)$ velocity power spectra show admixtures of various ring modes, as shown in Figure 5 in case of a four-ring.

For completeness, we also introduced another internal coordinate being the sum of all O−O distances of neighboring oxygen bridges as suggested in ref 32. We adapted the velocity projection method on this internal coordinate, and the resulting spectra are given in the Supporting Information. These spectra reveal a striking resemblance with previous $P(t)$ velocity power spectra both for the oxygen as for the silica atoms.

The presence of identical peaks in the three different types of power spectra (three different internal coordinates, namely $q_{c,0}$, $P(t)$ and $\sum R_{O-O}$) points manifestly toward a common vibrational mode, which represents probably the "true" breathing mode. However, to remove any ambiguity, we should be able to visualize the vibration corresponding to each peak in the power spectrum. But this is not an easy task and falls outside the scope of this work. Another aspect concerns the role of the force field used to perform the MD runs. It should be stressed that the fact that all

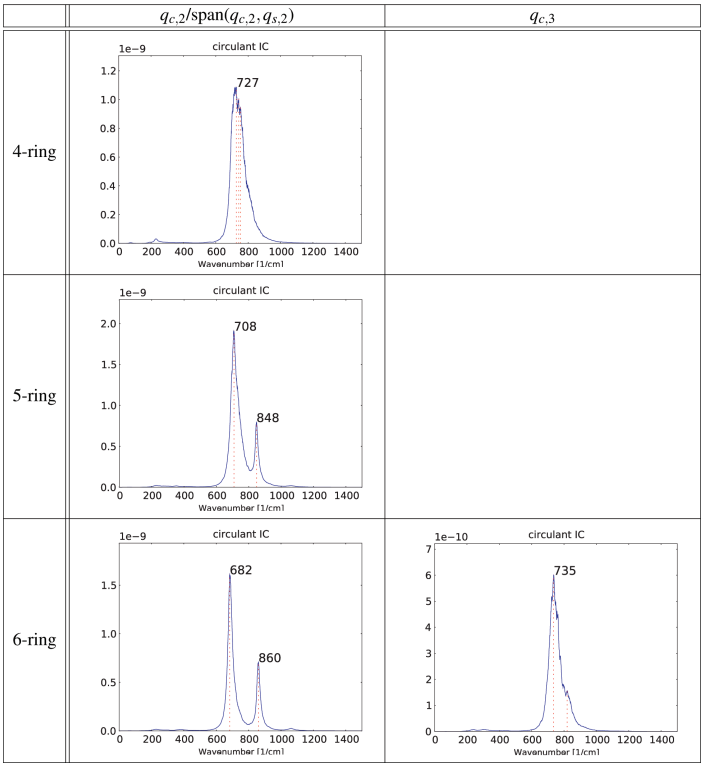


Figure 8. Projection of atomic velocities on the circulant symmetric IC's $q_{c,2}$/span($q_{c,2}$,$q_{s,2}$) and $q_{c,3}$ for the four-, five-, and six-ring.

peaks observed in the $q_{c,0}$ spectra are also retrieved in the standard "non-projected" velocity power spectra corresponding with pore-opening IC's is a property independent of the choice of the force field. Use of another force field can reproduce the spectral lines at somewhat different wavenumbers and affects probably too the shift of a peak when increasing the ring size, but the general features, as sketched above, remain unaffected.

Projection on Orthogonal Basis of IC's and the Orthogonal Complement. In the previous section only one class of internal coordinates (O−Si stretches and linear combinations of them) has been investigated. In this section we want to see to what extent the restriction to this single class of stretches is accurate enough. In other words what is the impact of the bending and dihedral motions on the vibrational spectra, or, more precisely, what is the impact of the remainder on the spectra after projecting out all tangent vectors belonging to stretches? The projection technique consists of determining all tangential atomic velocity vectors inducing a change in a particular IC. Modes induced by the normal components of the velocities are not considered as far as they do not belong to the entire class of IC's (stretches). For that reason we develop a method able to construct the so-called orthogonal complement of the basis of a given set of internal coordinates containing all the O−Si stretches (terminal O−Si + internal O−Si IC's). To obtain this orthogonal complement, the singular value decomposition of the

Journal of Chemical Theory and Computation ARTICLE

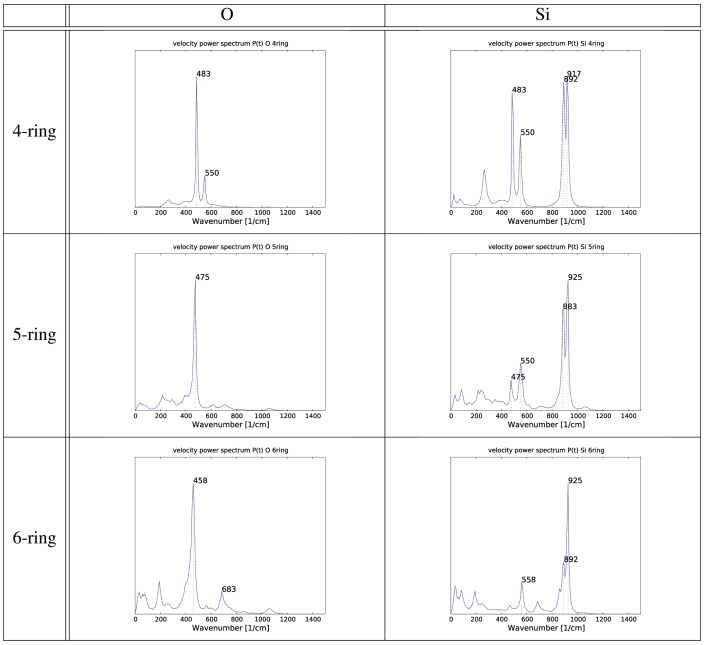


Figure 9. Velocity power spectra $P(t)$ for oxygen atoms as well as for silica and for the four-, five-, and six-rings.

matrix **A** holding the tangent vectors of all the IC's is computed at every time step:

$$\mathbf{A}_{ki} = \frac{\partial q_k}{\partial x_i} \tag{25}$$

where $k = 1, ..., K$ is an index characterizing the IC's, and the subscript $i = 1, ..., 3N$ denotes the Cartesian coordinates. This procedure decomposes **A** into three matrices:[61]

$$\mathbf{A} = \mathbf{UWV}^T \tag{26}$$

The matrices **U** and **V** are orthogonal, and **W** has singular values and is diagonal. The first K columns of **V** form an orthogonal basis for the IC's at each time step, while columns K to $3N$ construct the basis for the orthogonal complement at each time step. In a second step the atomic velocities are projected on both orthogonal bases. The result for the four-, five-, and six-rings is shown in Figure 10. The two full INS spectra are clearly well separated. The green spectrum ranging from 500 to 1200 cm^{-1} covers all O−Si stretches. The blue spectrum involves all IC modes belonging to the orthogonal complement, and for the three-ring structures, all peaks are located below 500 cm^{-1}. This is a nice result as it demonstrates that our methodology works quite well in projecting out all bending and dihedral motions. The conclusion is that the frequency region in question between 500−1200 cm^{-1} is adequately probed by stretch IC's and that a one-to-one comparison with the full spectrum is possible. Analysis of the lower frequency spectrum should be done with care. The blue spectrum in Figure 10 grouping all IC's, excluding stretches, is a global spectrum. We did not project on the tangential vector of an individual IC belonging to the class of bending and dihedral angles. In principle the same procedure as in the previous subsection could be performed, but in view of the ultimate goal of this study, such investigation was expected to give little added value. Nevertheless additional information was extracted, which could be of interest. To illustrate with an example, we refer to the peak at around 25 cm^{-1} which is prominently present in the low-frequency spectra of the three-ring systems. This peak is probably due to a collective puckering mode of the ring, but as already stressed, a full treatment of the velocity projection protocol on

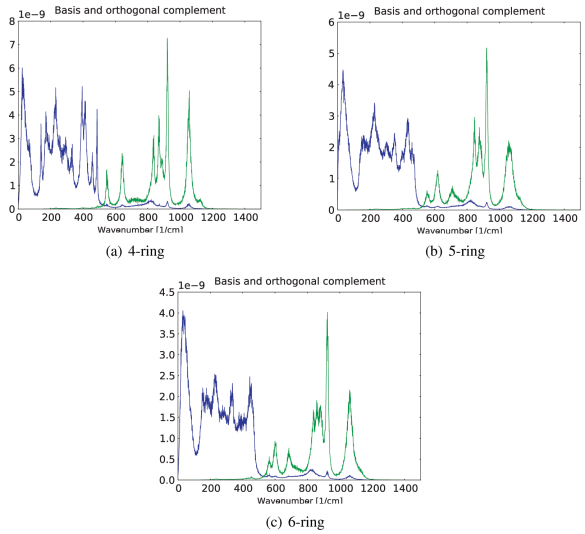


Figure 10. INS spectra of the projection of the atomic velocities on the basis of all O−Si IC's (green curve) and on the orthogonal complement of these IC's (blue curve).

the tangential vector of this puckering angle will reveal the real type of this specific mode.

Influence of the Connectivity. So far we studied the elementary four-, five-, and six-membered rings. We examined the influence of the sizes by comparing the spectra of them. Now we look at the connectivity (or the topology) by adding *n*-rings to these basic structures and investigate the corresponding fundamental changes in the spectra. In this section we will particularly focus on framework structures that contain five-membered rings (as in ref 26). Also the famous MFI-framework topology is built from five-membered rings. More precisely, in this section we will concentrate on the following 5T ring clusters: 7T(2 × 5); 8T (2 × 5); 8T(4 × 5); and 11T(3 × 5)(see Figure 1). They all show two or more connected five-rings except for the reference five-membered ring. Each mode ($q_{c,0}$, span($q_{c,1}$,$q_{s,1}$), and span-($q_{c,2}$,$q_{s,2}$)) is more or less degenerate due to the appearance of multiple five-rings, hence an average is taken of the corresponding spectra. Spectra belonging to other building blocks are given in the Supporting Information.

As already observed for the elementary four-, five-, and six-membered rings, the ring size will not affect the INS spectra (see Supporting Information) for the terminal O−Si and antisymmetric IC's. The changing topology induced by adding *n*-rings to elementary rings has no influence at all on the spectra; no peak shifts occur, and the shapes remain unaltered. This observation was more or less expected for the terminal O−Si stretches but rather unexpected and even surprising for the antisymmetric stretch mode.

On the contrary, significant changes in the vibrational INS spectra occur for the various symmetric stretches. Spectra of the projected velocities on the circulant symmetric IC's of the five-membered ring systems under study are shown in Figure 11.

All spectra are referred to with respect to the reference spectral lines observed for the elementary five-membered ring. Two connected five-rings sharing two atoms (8T(2 × 5), Figure 1k) cause a splitting of the two main peaks with a slight blue and red shift for the $q_{c,0}$ and span($q_{c,1}$,$q_{s,1}$) mode. In principle two degenerate $q_{c,0}$ modes should exist. The slight coupling between these modes (for two rings share two silica atoms) gives rise to the observed small splitting. A third peak is manifestly present in the breathing mode $q_{c,0}$ and could be resolved by a suitable selection of an additional IC. The other connected five-ring structures lead to similar features. The 11T(3 × 5) structure with three connected five-rings shows the same pattern as the 8T(2 × 5), with the difference that now three five-rings are attached to each other which is reflected in the projected spectrum. In the 7T(2 × 5) structure the situation becomes more complex because now three Si atoms are shared. Since in the 8T(4 × 5) structure even more five-rings are shared, the spectra become more distinct from the reference single five-ring spectra. For the span($q_{c,2}$,$q_{s,2}$) mode, the spectra show a remarkable resemblance in all structures. Another issue concerns

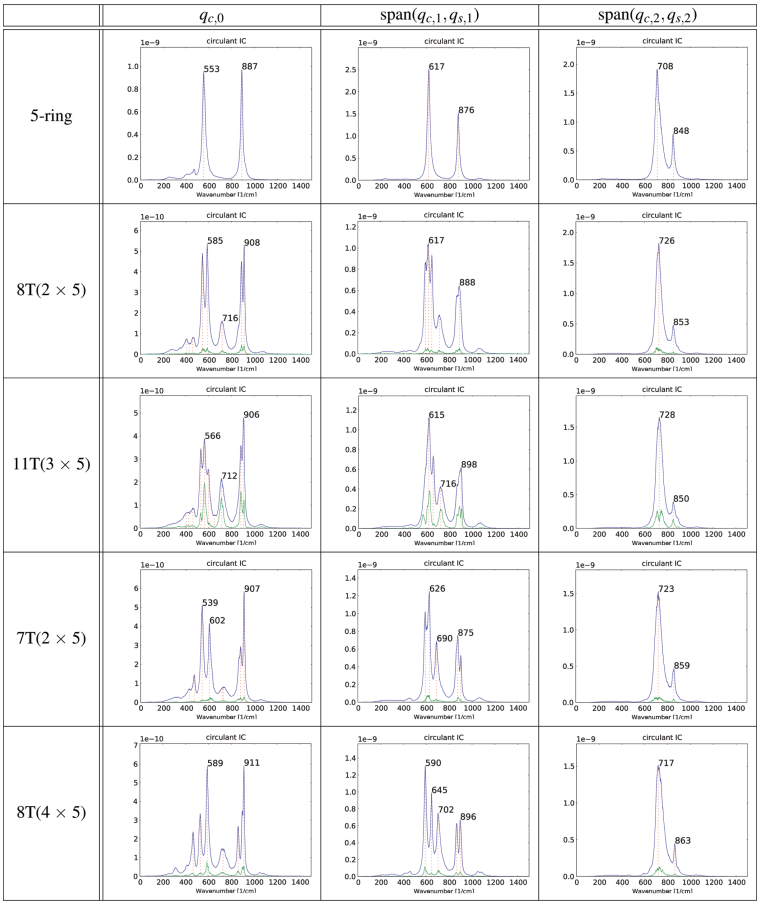


Figure 11. INS spectra belonging to atomic velocities projected on the circulant symmetric q_{ss} IC's for connected five-rings. Green curve: standard deviation.

the way spectra are affected by connecting rings of different sizes. We compare in Figure 12 the INS spectra generated by connecting a five-membered ring with, respectively, a four- and a six-membered ring. Other spectra belonging to connected rings are shown in the Supporting Information. A common feature is the observation that when an n_2-ring is attached to a basis n_1-ring

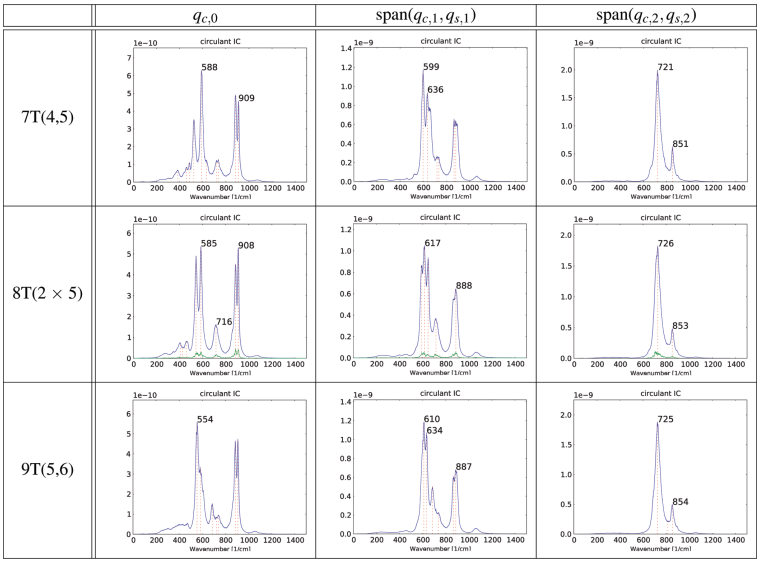


Figure 12. INS spectra belonging to atomic velocities projected on the circulant symmetric $q_{c,0}$, span($q_{c,1}$,$q_{s,1}$) and span($q_{c,2}$,$q_{s,2}$) IC's for the five-ring connected with, respectively, a four-, five-, and six-ring.

($n_1 = 4-6$), the peaks of the $q_{c,0}$ and span($q_{c,1}$,$q_{s,1}$) mode of the basis n_1-ring seem to split up. One can distinguish between symmetrical ($n_1 = n_2$, e.g., five-ring attached to five-ring) and antisymmetrical splitting ($n_1 \neq n_2$, e.g., four-ring attached to five-ring). These effects are a logical result from the fact that the peaks of the modes of the n_1- and n_2-rings occur at the same frequency. When the two rings are attached, a resonance effect arises which causes one peak to turn into two floating modes: one peak with a slightly higher and one peak with a slightly lower wavenumber around the resonance frequency. When $n_1 = n_2$ the peaks show more or less the same amplitude, when $n_1 \neq n_2$ the difference in amplitude is much larger. Therefore it is useful to introduce the nomenclature of symmetrical and antisymmetrical splitting. An example is the 7T(2 × 5): the 553 and 887 cm^{-1} peaks of the five-ring split into 539, 602, 876, and 907 cm^{-1}, respectively. Just as in the case of connecting five-rings, the span($q_{c,2}$,$q_{s,2}$) mode remains the same.

The $q_{c,0}$ breathing mode is the most sensitive to connectivity differences. When more rings are attached to the elementary n_1-rings, it is obvious that more peaks appear, since the ring systems cannot be regarded as independent systems anymore and mixing of modes occurs which results in extra peaks. This coupling gives rise to a splitting of the original peaks or the emerging of additional peaks. The $q_{c,0}$ mode is the most collective mode because the breathing mode is determined by stretches which are not in an immediate antiphase (see Figure 5), giving this mode a more global character. The eigenmode of the span($q_{c,2}$,$q_{s,2}$) IC on the other hand is the most local mode, as the successive stretches move in antiphase with each other (Figure 5). The topology dependence here is minimal.

It is important to note that the $q_{c,0}$ breathing mode is not affected by the size of the ring (see subsection "Circulant Symmetric IC's"), but is affected by the particular way in which it is connected to other rings.

CONCLUSION AND PERSPECTIVES

In this paper we have used the velocity projection method to assign (parts of) spectral peaks in zeolite vibrational spectra to particular changes in IC's. The focus was on essential zeolite building units, which can be assumed to play important roles during initial states of zeolite formation and during zeolite growth in general. The analysis of vibrational spectra can be a very useful tool in understanding the process, as the spectral behavior of eigenmodes of IC's can vary with different molecular systems. MD can then be used to simulate various structures which are key components during zeolite growth. The atomic velocities obtained from these MD runs can then be projected on a well chosen set of IC's.

Journal of Chemical Theory and Computation ARTICLE

Here, we focused on stretches, linear combinations of them, and eigenvectors of the circulant matrix as IC's. The circulant matrix was proposed as a model for the Hessian of ring molecular structures.

We found that the spectra of the terminal O−Si and the antisymmetric internal O−Si IC's were not influenced by the topology and by the size of the considered rings; the spectra of the elementary four,five-, and six-ring are the same. In addition, when other rings are attached to these basic *n*-rings, the same conclusions can be drawn.

When the atomic velocities are projected on the symmetric O−Si IC's, the resulting spectra differ for the four-,five-, and six-ring. The spectral shifts of the spectra of IC $q_{c,0}$ are tiny, while those corresponding with span($q_{c,m}q_{s,m}$) are rather substantial. The same trend can be observed in connected ring systems. We should, however, be very careful in drawing conclusions, as it is unclear what the role is of the force field in this specific field. It is recommended to refer to the spectra of the elementary rings when investigating the effect of connecting multiple *n*-rings. The most important conclusions are that the $q_{c,0}$ IC is the most sensitive mode to changes in molecular topology. It is the most global mode. We associate it to the breathing mode as it is determined by adjacent stretches which are not in antiphase. All the peaks occurring in the projected velocity $q_{c,0}$ power spectra are retrieved in power spectra belonging to other internal coordinates defining the ring-opening vibration.

About the connectivity we found that the size of the added *n*-rings does not play a crucial role; it is rather the way in which these rings are connected to each other that is the origin of the observed spectral changes.

When one projects the atomic velocities on the orthogonal complement of all stretch IC's present in a molecular structure, we found that the contribution of the stretch modes to the spectra is spectrally well separated from other contributions, like bending and dihedral angles. This confirms that the spectral region in which we are interested for zeolite synthesis can be almost completely resolved by considering changes in stretch modes.

In an attempt to decompose the spectra of the symmetric IC's further we introduced the circulant matrix as a model for the Hessian of the considered ring systems. If we want to extend the applications to systems which are not restricted to a ring structure, another more general approach is desired. A first incentive for this has already been found by linking the displacement (small motions) in IC's with the displacement (small motions) in Cartesian coordinates. We represent the displacement of the IC's by the $3N \times K$ matrix $\mathbf{J}_{ik}$ (N is the number of atoms, K is number of IC's) and the Cartesian displacement by the $3N \times 3N$ matrix $\mathbf{E}_{i\lambda}$. The Cartesian eigenmodes can then be expanded in a redundant set of internal coordinates (they are not orthogonal) with $\alpha_{k\lambda}$ being the matrix which holds the expansion coefficients:

$$\sum_{k=1}^{K} \mathbf{J}_{ik}\alpha_{k\lambda} = \mathbf{E}_{i\lambda} \tag{27}$$

Here λ is a counter for the number of eigenmodes. The coefficients of the matrix $\alpha_{k\lambda}$ can then be used as coefficients for projecting the atomic velocities on the considered internal coordinates. The Cartesian eigenmodes $\mathbf{E}_{i\lambda}$ can only be obtained if the Hessian of the potential energy function is diagonalized, thus the mass-weighted normal mode eigenvalue equation has to be solved. To obtain the Hessian, a specified energy function V has to be provided. First results of this method were obtained, but it is still a work in progress.

APPENDIX

Eigenvalues and the k^{th} component of the normalized eigenvectors. We will show that component k of eigenvector m has the following form:

$$V_k^{(m)} = \exp\left(\frac{2\pi imk}{n}\right) \tag{28}$$

It is easy to see that the set of n orthogonal vectors $V^{(m)}$ with components of the form of eq 28, are eigenvectors of C. The product of C with $V^{(m)}$ can be written as

$$\begin{aligned}\sum_k C_{jk}V_k^{(m)} &= \sum_k c_{(j-k)} \exp\left(\frac{2\pi imk}{n}\right)\\ &= \sum_{k'} c_{(k')} \exp\left(\frac{2\pi im(j-k')}{n}\right)\\ &= \left[\sum_{k'} c_{(k')} \exp\left(\frac{-2\pi imk'}{n}\right)\right] \exp\left(\frac{2\pi imj}{n}\right)\\ &= \lambda^{(m)}V_j^{(m)}\end{aligned} \tag{29}$$

where the dummy summation index k was changed to $k' = (j - k) \bmod n$. It is clear that $V^{(m)}$ is indeed eigenvectors of the symmetric circulant matrix C.

The eigenvalues are real and can straightforwardly be rewritten as

$$\begin{aligned}\lambda^{(m)} &= \sum_k c_{(k)} \exp\left(\frac{-2\pi imk}{n}\right)\\ &= \sum_k \frac{1}{2}(c_{(k)} + c_{(-k)}) \exp\left(\frac{-2\pi imk}{n}\right)\\ &= \sum_k c_{(k)}\frac{1}{2}\left[\exp\left(\frac{-2\pi imk}{n}\right) + \left(\frac{2\pi imk}{n}\right)\right]\\ &= \sum_k c_{(k)} \cos\left(\frac{2\pi mk}{n}\right) = \lambda^{(-m)}\end{aligned} \tag{30}$$

ASSOCIATED CONTENT

Supporting Information. Additional projected velocity power spectra, IR and INS spectra of all the oligomers are given. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
E-mail: veronique.vanspeybroeck@ugent.be.

ACKNOWLEDGMENT

This work is supported by the Fund for Scientific Research−Flanders (FWO), the Research Board of Ghent University (BOF) and BELSPO in the frame of IAP/6/27, the Belgian Prodex office, and ESA. Part of the computational resources and services used in this work were provided by Ghent University. V.

V.S. acknowledges the European Research Council under the European Community's Seventh Framework Programme (FP7-(2007-2013) ERC grant agreement no. 240483). J.A.M. acknowledges the Flemish government for long term structural funding (Methusalem).

REFERENCES

(1) Boronat, M.; Viruela, P.; Corma, A. *J. Am. Chem. Soc.* **2003**, *126*, 3300–3309.
(2) Nascimento, M. *J. Mol. Struct.* **1999**, *464*, 239–247.
(3) Demontis, P.; Suffritti, G. *Chem. Rev.* **1997**, *97*, 2845–2878.
(4) Hedlund, J.; Schoeman, B.; Sterte, J. *Progress in Zeolites and Microporous Materials*; Chon, H., Ihm, S.-K., Uh, Y. S., Eds.; Elsevier: Amsterdam, The Netherlands, 1997; pp 2203− 2210.
(5) Ravishankar, R.; Kirschhock, C.; Schoeman, B.; Vanoppen, P.; Grobet, P.; Storck, S.; Maier, W.; Schryver, F. D.; Martens, J.; Jacobs, P. *J. Phys. Chem. B* **1998**, *102*, 2633–2639.
(6) Dokter, W.; van Garderen, H.; Beelen, T.; van Santen, R.; Bras, W. *Angew. Chem., Int. Ed.* **1995**, *34*, 73–75.
(7) Knight, C.; Wang, J.; Kinrade, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3099–3103.
(8) Auerbach, S.; Monson, P.; Ford, M. *Curr. Opin. Colloid Interface Sci.* **2005**, *10*, 220–225.
(9) Lewis, D.; Richard, C.; Catlow, A.; Thomas, J. *Faraday Discuss.* **1997**, *106*, 451–471.
(10) Burkett, S.; Davis, M. *Chem. Mater.* **1995**, *7*, 920–928.
(11) Corkery, R.; Ninham, B. *Zeolites* **1997**, *18*, 379–386.
(12) Schoeman, B. *Microporous and Mesoporous Mater.* **1998**, *22* 9–22.
(13) Tsay, C.; Chiang, A. *Microporous Mesoporous Mater.* **1998**, *26*, 89–99.
(14) Watson, J.; Brown, A.; Iton, L.; White, J. *J. Chem. Soc., Far. Trans.* **1998**, *94*, 2181–2186.
(15) de Moor, P.; Beelen, T.; van Santen, R.; Beck, L.; Davis, M. *J. Phys. Chem. B* **2000**, *104*, 7600–7611.
(16) Li, Q.; Mihailova, B.; Creaser, D.; Sterte, J. *Microporous and Mesoporous Mater.* **2001**, *43*, 51–59.
(17) Mintova, S.; Olson, N.; Senker, J.; Bein, T. *Angew. Chem., Int. Ed.* **2002**, *41*, 2558–2561.
(18) Aerts, A.; Follens, L.; Haouas, M.; Caremans, T.; Delsuc, M.; Loppinet, B.; Vermant, J.; Goderis, B.; Taulelle, F.; Martens, J.; Kirschhock, C. *Chem. Mater.* **2007**, *19*, 3448–3454.
(19) Kumar, S.; Davis, T.; Ramanan, H.; Penn, R.; Tsapatsis, M. *J. Phys. Chem. B* **2007**, *111*, 3398–3403.
(20) Jin, L.; Auerbach, S.; Monson, P. *J. Phys. Chem. C* **2010**, *114*, 14393–14401.
(21) Kirschhock, C.; Ravishankar, R.; Looveren, L. V.; Jacobs, P.; Martens, J. *J. Phys. Chem. B* **1999**, *103*, 4972–4978.
(22) Kirschhock, C.; Ravishankar, R.; Verspeurt, F.; Grobets, P.; Jacobs, P.; Martens, J. *J. Phys. Chem. B* **1999**, *103*, 4965–4978.
(23) Ravishankar, R.; Kirschhock, C.; Knops-Gerrits, P.; Feijen, E.; Grobet, P.; Vanoppen, P.; Schryver, F. D.; Miehe, G.; Fuess, H.; Schoeman, B.; Jacobs, P.; Martens, J. *J. Phys. Chem. B* **1999**, *103*, 4960–4964.
(24) Verstraelen, T.; Szyja, B.; Lesthaeghe, D.; Declerck, R.; Van Speybroeck, V.; Waroquier, M.; Jansen, A.; Aerts, A.; Follens, L.; Martens, J.; Kirschhock, C.; van Santen, R. *Top. Catal.* **2009**, *52*, 1261–1271.
(25) Petry, D.; Haouas, M.; Wong, S.; Aerts, A.; Kirschhock, C.; Martens, J.; Goskell, S.; Anderson, M.; Taulelle, F. *J. Phys. Chem. C* **2009**, *113*, 20827–20836.
(26) Lesthaeghe, D.; Vansteenkiste, P.; Verstraelen, T.; Ghysels, A.; Kirschhock, C.; Martens, J.; Van Speybroeck, V.; Waroquier, M. *J. Phys. Chem. C* **2008**, *112*, 9186.
(27) In *Normal mode analysis: theory and applications to biological and chemical systems*; Cui, Q., Bahar, I., Ed.; Chapman & Hall: Boca Raton, FL, 2006.
(28) Scribano, Y.; Benoit, D. M. *J. Chem. Phys.* **2007**, *127* 164118.
(29) Bornhauser, P.; Calzaferri, G. *J. Phys. Chem.* **1996**, *100*, 2035–2044.
(30) Pasquarello, A.; Car, R. *Phys. Rev. Lett.* **1998**, *80*, 5145–5147.
(31) Smirnov, K.; Bougeard, D. *Catal. Today* **2001**, *70*, 243–253.
(32) To, T.; Bougeard, D.; Smirnov, K. *J. Raman Spectrosc.* **2008**, *39*, 1869–1877.
(33) Arab, M.; Bougeard, D.; Smirnov, K. *Phys. Chem. Chem. Phys.* **2002**, *4*, 1957–1963.
(34) Jobic, H.; Smirnov, K.; Bougeard, D. *Chem. Phys. Lett.* **2001**, *344*, 147–153.
(35) Smirnov, K.; Bougeard, D. *J. Raman Spectrosc.* **1993**, *24*, 255–257.
(36) Smirnov, K.; Bougeard, D. *J. Phys. Chem.* **1993**, *97*, 9434–9440.
(37) Smirnov, K.; Bougeard, D.; Maire, M. L.; Brémard, C. *Chem. Phys.* **1994**, *179*, 445–454.
(38) Kuhne, T.; Krack, M.; Mohamed, F.; Parrinello, M. *Phys. Rev. Lett.* **2007**, *98*, 066401.
(39) Kuhne, T.; Krack, M.; Parrinello, M. *J. Chem. Theory Comput.* **2009**, *5*, 235–241.
(40) Jacob, C.; Reiher, M. *J. Chem. Phys.* **2009**, *130*, 084106.
(41) Verstraelen, T.; Neck, D. V.; Ayers, P.; Speybroeck, V. V.; Waroquier, M. *J. Chem. Theory Comput.* **2007**, *3*, 1420.
(42) Huang, Y.; Jiang, Z. *Microporous Mater.* **1997**, *12*, 341–345.
(43) Flanigen, E.; Khatami, H.; Szymanski, H. In *Infrared Structural Studies of Zeolite Frameworks*, Advances in Chemistry Series; Flanigen, E. M., Sand, L. B., Eds.; American Chemical Society: Washington, D.C., 1974; Vol. 101.
(44) de Man, A.; van Santen, R. *Zeolites* **1992**, *12*, 269–278.
(45) van Santen, R.; Vogel, D. *Adv. Solid-State Chem.* **1989**, *1*, 151–224.
(46) Wood., S. *J. R. Stat. Soc., B* **2000**, *62*, 413.
(47) CP2K, 2008; http://cp2k.berlios.de.
(48) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Inf. Mod.* **2008**, *48*, 1530–1541.
(49) Verstraelen, T.; Van Houteghem, M.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Inf. Mod.* **2008**, *48*, 2414.
(50) Futrelle, R.; McGinty, D. *Chem. Phys. Lett.* **1971**, *12* 285–287.
(51) McQuarrie, D. In *Statistical mechanics*; Cato, R., Ed.; Harper & Row, 1976.
(52) Berens, P.; Wilson, K. *J. Chem. Phys.* **1981**, *74*, 4872–4882.
(53) Noid, D.; Koszykowski, M.; Marcus, R. *J. Chem. Phys.* **1977**, *67*, 404.
(54) Papoulis, A. In *Probability, random variables and stochastic processes*; York, N., Ed.; McGraw-Hill, 1965.
(55) Urey, H.; Bradley, C. *Phys. Rev.* **1931**, *38*, 1969–1978.
(56) Watson, J.; Iton, L.; Keir, R.; Thomas, J.; Dowling, T.; White, J. *J. Phys. Chem. B* **1997**, *101*, 10094–10104.
(57) Kragten, D.; Fedeyko, J.; Sawant, K.; Rimer, J.; Vlachos, D.; Lobo, R.; Tsapatsis, M. *J. Phys. Chem. B* **2003**, *107*, 10006–10016.
(58) Mohamed, R.; Aly, H.; El-Shahat, M.; Ibrahim, I. *Microporous and Mesoporous Mater.* **2005**, *79*, 7–12.
(59) Haouas, M.; Taulelle, F. *J. Phys. Chem. B* **2006**, *110* 3007–3014.
(60) Moravetski, V.; Hill, J.-R.; Eichler, U.; Cheetham, A.; Sauer, J. *J. Am. Chem. Soc.* **1996**, *118*, 13015–13020.
(61) Press, W.; Teukolsky, S.; Vetterling, W.; Flannery, B. In *Numerical recipes in C: The art of scientific computing*; Press, W., Ed.; Cambridge University Press: New York, 1992.

Paper II

Analysis of the basis set superposition error in molecular dynamics of hydrogen-bonded liquids: application to methanol

M. Van Houteghem, T. Verstraelen, A. Ghysels, L. Vanduyfhuys,
M. Waroquier, V. Van Speybroeck

Journal of Chemical Physics, 137 , 104506, 2012

THE JOURNAL OF CHEMICAL PHYSICS **137**, 104506 (2012)

Analysis of the basis set superposition error in molecular dynamics of hydrogen-bonded liquids: Application to methanol

Marc Van Houteghem, Toon Verstraelen, An Ghysels, Louis Vanduyfhuys, Michel Waroquier, and Veronique Van Speybroeck[a)]
Center for Molecular Modeling, QCMM Alliance Ghent-Brussels, Ghent University, Technologiepark 903, B-9052 Zwijnaarde, Belgium

(Received 18 June 2012; accepted 21 August 2012; published online 13 September 2012)

An efficient protocol is presented to compensate for the basis set superposition error (BSSE) in DFT molecular dynamics (MD) simulations using localized Gaussian basis sets. We propose a classical correction term that can be added *a posteriori* to account for BSSE. It is tested to what extension this term will improve radial distribution functions (RDFs). The proposed term is pairwise between certain atoms in different molecules and was calibrated by fitting reference BSSE data points computed with the counterpoise method. It is verified that the proposed exponential decaying functional form of the model is valid. This work focuses on hydrogen-bonded liquids, i.e., methanol, and more specific on the intermolecular hydrogen bond, but in principle the method is generally applicable on any type of interaction where BSSE is significant. We evaluated the relative importance of the Grimme-dispersion versus BSSE and found that they are of the same order of magnitude, but with an opposite sign. Upon introduction of the correction, the relevant RDFs, obtained from MD, have amplitudes equal to experiment. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4749929]

I. INTRODUCTION

A radial distribution function (RDF) is the elementary tool used to extract the structural information from computer simulations of molecular liquids.[1] The RDF $g(r)$ of a fluid describes how the liquid density varies as a function of the distance from a reference molecule. It reflects the correlations in the distribution of the molecules of the fluid arising from the forces they exert on each other. The RDF is a measure for the probability to find an atom pair separated by a distance r, relative to the probability one would expect for a uniform distribution at the same density.

In this work the RDFs of liquids are studied with the help of molecular dynamics (MD) simulations, based on density functional theory (DFT).[2,3] We focus on a specific category of molecules which interact with each other by means of weak covalent interactions. A well-known example of this type of bonding is the hydrogen bond, which occurs in, e.g., water, methanol or ethanol. In this study, we concentrate on methanol in its liquid form. Methanol is widely used in chemical industry, as a solvent or fuel. It is the smallest alcohol consisting of a hydrogen-bonding hydroxyl group and a hydrophobic methyl group. The characteristic alcohol group allows methanol to form hydrogen bonds that dominate the structural and dynamical behavior of the liquid phase. Because of the hydrogen bonding, the O–H bonds are elongated[4,5] in the liquid phase with respect to the gas phase.[4,6,7]

The liquid phase of methanol is extensively studied. The RDFs of methanol have been determined experimentally[8–11] and have been computed by means of full *ab initio* MD,[5,12–15] mixed empirical and *ab initio* interactions,[16,17] and force-fields.[18–26] Often, the first intermolecular peak in the RDF of methanol as well of water is overestimated compared to experiment.[12,13,15–17,23–25,27–30] The search for plausible explanations of this quasi-systematic overestimation forms the onset of a long debate, but until now it is not yet unraveled and what suits for one particular case does not suit for another case. When computing RDFs from molecular simulations based on first principles (e.g., DFT), several parameters may affect the final result: system-size effects,[13,29,31] basis set superposition error (BSSE), dispersion interaction, differences in exchange-correlation functional,[31–33] differences in basis sets,[31] nuclear quantum effects, etc. Unfortunately, it is not yet clear what the net impact really is from each ingredient on the RDF, and in most cases the corrections work in opposite sense. It means that there is some compensation effect: if we correct the RDF for one missing ingredient, it can be canceled by another. This implies that one should be very careful when correcting for one specific item. It can happen that the final result could be worse and thus it is very important to get detailed insight into the various factors contributing to the overall result of the RDF.

(i) Nuclear quantum effects in liquid water are commonly investigated by methods based on Feynman's path-integral molecular dynamics simulations to calculate finite-temperature equilibrium properties. It is well known that quantum effects become more significant in the intramolecular region. Quantum statistical mechanical simulations are able to reproduce a quantitative agreement with experiment for the heights and broadening of the intramolecular O–H and H–H peaks in liquid water.[34] A general feature is that nuclear quantum effects make the structure of the liquid water softer.[35–40] One notices a broadening and lowering

a)Electronic mail: veronique.vanspeybroeck@ugent.be.

of the RDF compared to the distribution generated from classical molecular dynamics simulations.[34,41–43] An alternative is the Car-Parrinello molecular dynamics method,[44] which is based on the description of the intermolecular interactions at the density functional level of theory. The success of CMPD simulations in water depends how accurate the density based functional describes the Born-Oppenheimer potential energy surface for the description of the hydrogen bond.[28]

(ii) The choice of the exchange-correlation functional does affect the density of the liquid and the radial distribution function. There is a tendency that BLYP produces weaker hydrogen bonds (underbinding) than PBE (overbinding).[27] The global performance of a specific functional cannot be assessed without taking into account the dispersion interaction. Therefore, a quantification of each effect separately is recommended to better understand the global atomic and molecular interactions.

(iii) Dispersion interactions are indispensable to predict correct densities and RDFs, but many models exist to describe them and it is not always clear what impact these differences have on the RDFs. Most models work with empirical pairwise interatomic potentials of the C^6R^{-6} form with damping terms, which can differ from model to model.[45–48] In DFT-D3 of Grimme[47] the empiricism of the model has been largely removed by fitting as much as possible on first principles derived properties. In addition eight-order terms have been added, as well as a three-body correction. In general, more and more molecular dispersion interactions have been proposed in the literature determined from *ab initio* electronic structure calculations and in this way largely reducing the empirical input.[49,50] Their overall effect is that they push the density in the correct direction, but—depending on the individual cases—the RDFs remain overstructured.

This work focuses on the role of BSSE on the RDFs. Most of the published work on this item focused on the impact of dispersion on the radial distribution functions, and as already reported they represent a significant step toward experimental agreement,[27,28] although some discrepancies remain. As most of the simulations have been applied to liquid water, one should be very careful to extrapolate their conclusions to methanol. A combined study of the RDFs in methanol taking into account the two main effects influencing the interatomic potential: BSSE + dispersion, is indispensable in the attempt to reproduce correctly the RDF of methanol.

The BSSE results from the approximation of using a limited number of localized basis functions to describe atomic and molecular orbitals. When two molecules are far apart, they can only use their own basis functions. But when they approach each other, they may lower their energy by using each other's basis functions. This causes an artificial strengthening of the intermolecular interactions and artificial shortening of intermolecular distances due to the overlap between basis functions. The effect is more pronounced for smaller basis sets. Note that BSSE does not occur with plane wave basis sets as they are not associated with any particular atom.

The BSSE is especially prominent in systems with intermolecular hydrogen-bonds,[51–57] and is thus expected to influence the RDF of methanol. It has been shown that the calculation of the BSSE corrections can change their potential energy curves and surfaces,[58–62] and can alter the bond distances.[58,60,61,63,64] It is well known that BSSE can be a significant fraction of the binding energy of the water dimer, which could affect liquid simulations.[31] Simon *et al.* analyze the effect of BSSE on the geometries and vibrational frequencies of 15 H-bonded systems at the B3LYP and MP2 levels of theory, using the 6-31++G(*d,p*) basis set.[53] It is shown that the intermolecular H-bond distance increases when correcting for BSSE, whereas the intermolecular H-bond stretching frequency decreases, and that the observed changes depend on the level of theory. The effect of BSSE on structural and dynamical properties of the liquid has not been investigated previously. In fact, there is little information available concerning the performance of a Gaussian basis set in the context of MD simulations of liquids. So the influence of BSSE on RDFs remains an open question.

New in the model—presented in this work—is the proposal of a systematic correction of the BSSE with a parameterized classical force-field energy term. The parameters are fitted by calculating the BSSE error for a series of fluid configurations. These configurations have been created by *ab initio* MD simulations. Finally, MD simulations are performed including the BSSE correction and dispersion. The RDFs computed from MD runs are compared with the experiment before and after the corrections, and it is assessed to what extent the correspondence with experiment due to these corrections has improved. This geometrical approach to account for BSSE is supported by a very recent paper by Kruse[65] where a semi-empirical correction is introduced, constructed from overlap integrals over Slater functions and which employs computed measures for the incompleteness of the basis set.

In Sec. II, the technical details of the quantum MD simulations and the computational methodology to calculate the RDF and BSSE are given. Section III introduces the new BSSE correction model, i.e., its mathematical form and the calibration of the parameters. Section IV presents the results, which are further discussed in Sec. V. In Sec. VI the most relevant conclusions—in view of future work—are drawn.

II. COMPUTATIONAL DETAILS

A. *Ab initio* MD

In this work, all molecular dynamics calculations were performed with the CP2K code.[66] For the *ab initio* MD, the Quickstep module of CP2K[67] was used, employing the hybrid Gaussian and plane-wave (GPW) density functional method with a BLYP gradient-corrected functional.[68,69] The dual GPW basis set consisted of a triple zeta TZVP Gaussian type orbital basis for the real space representation and an auxiliary plane wave basis expanded to a 400 Ry cutoff to compute the long-range periodic electrostatic interactions in the reciprocal space.[70] The core electron states

are represented by the norm conserving Goedecker-Teter-Hutter (GTH) pseudopotential.[31,71,72] The BLYP functional was chosen because of its good description of the structure and dynamics of water[32] where hydrogen bonds are, as in liquid methanol, the dominant interactions. Furthermore, it has been shown that the BLYP functional gives a proper description of solvation of methanol in water.[73]

For the *ab initio* MD simulations that include the correction for the BSSE with the force-field energy terms, we used term FIST, the molecular mechanics (MM) module of CP2K. Periodic boundary conditions were employed using the minimum image convention. At each timestep the total energy is computed as the sum of the *ab initio* energy and the energy calculated by the classical correction term.

Standard DFT approaches do not account for dispersion interactions.[74,75] Dispersion is a long-range van der Waals interaction between electron clouds arising from instantaneous transition dipoles. These dispersion interactions are a quantum-mechanical electron correlation effect that is completely absent when considering classical charge distributions. To test the effects of dispersion on the fluid structure of methanol, we also performed MD simulations with dispersion-corrected DFT using the DFT-D3 scheme of Grimme for the BLYP functional.[45–47,76] The combination of the two corrections for the BSSE and dispersion then provides us four different potential energy surfaces: no corrections (DFT), corrected for BSSE with the counterpoise (CP) correction (DFT + CP) as explained in Sec. II C, corrected for dispersion (DFT-D3), corrected for BSSE and dispersion (DFT-D3 + CP).

Liquid methanol at room temperature was modeled in a periodic cubic simulation cell containing 129 methanol molecules. The simulation cell size, 20.58 Å, was chosen to correspond with the experimental density of 0.787 kg/m^3. After equilibration, the MD runs lasted 8.4 ps at a temperature of 300 K with a Nosé thermostat in the canonical (NVT) ensemble.[77] The Verlet integrator was used with an integration time step of 1 fs.

B. The radial distribution function

The radial distribution function is defined as

$$g(r) = \frac{\mathrm{d}n(r)}{\mathrm{d}n_0(r)} \tag{1}$$

and represents the number of particle pairs $\mathrm{d}n(r)$ in the spherical shell with radius r and $r + \mathrm{d}r$, with r the interparticle distance (relative coordinate), relative to the number of particle pairs $\mathrm{d}n_0(r)$, obtained in an uncorrelated ideal gas. With ρ_0 as the density of particles in the ideal gas case, $\mathrm{d}n_0(r)$ may be expressed as $\rho_0 4\pi r^2 \mathrm{d}r$.

In practice, the space around a given atom is discretized in concentric spherical shells with width δr. The number of atoms n_i in each shell with finite volume $\delta V_{\mathrm{shell},i}$ is counted,

$$\delta V_{\mathrm{shell},i} = \frac{4}{3}\pi\,(r_i + \delta r)^3 - \frac{4}{3}\pi r_i^3 \simeq 4\pi r_i^2 \delta r. \tag{2}$$

A histogram of the n_i values is built during the MD run and $g(r_i)$ is approximated as $n_i/\rho_0 4\pi r_i^2 \delta r$.

C. Counterpoise correction

In this section we describe a method to correct for BSSE in computations with localized basis sets that are centered at the nuclei. Atom-centered basis sets have a number of advantages. The most important one is that with relatively few basis functions one is able to describe the molecular orbitals to high accuracy. Such basis sets cannot be systematically enlarged in a simple way and the severity of the problem is proportional to the incompleteness of the basis sets.

From the suggested methods to correct for BSSE,[78–83] the most widely used approach is the counterpoise method of Boys and Bernardi.[83] The uncorrected interaction energy of a complex AB can be defined as

$$\Delta E_{int}(\mathrm{AB}) = E^{\mathrm{AB}}(\mathrm{AB}) - E^{\mathrm{A}}(\mathrm{A}) - E^{\mathrm{B}}(\mathrm{B}), \tag{3}$$

where the superscripts denote the basis used and the symbol in parentheses denotes the chemical system considered. Boys and Bernardi suggested to correct Eq. (3) by estimating the amount of artificial stabilization for monomer A as the energy shift caused by a change of basis A → AB,

$$E_{\mathrm{BSSE}}(\mathrm{A}) = E^{\mathrm{AB}}(\mathrm{A}) - E^{\mathrm{A}}(\mathrm{A}), \tag{4}$$

and similarly for monomer B. To evaluate the energy of monomer A in the AB basis, one places all the basis functions of monomer B on the atomic centers of monomer B while omitting the electrons and the nuclear charges of monomer B. The B-basis in the E^{AB} computation and the A-basis in the E^{AB} are called *ghost atoms*. The energy of monomer A in the AB basis is lower than the energy of monomer A in the A-basis ($E_{\mathrm{BSSE}}(\mathrm{A}) < 0$). The CP-correction is then defined as

$$\Delta E_{\mathrm{CP}} = E^{\mathrm{A}}(\mathrm{A}) + E^{\mathrm{B}}(\mathrm{B}) - E^{\mathrm{AB}}(\mathrm{A}) - E^{\mathrm{AB}}(\mathrm{B}). \tag{5}$$

When the estimated errors $E_{\mathrm{BSSE}}(\mathrm{A})$ and $E_{\mathrm{BSSE}}(\mathrm{B})$ are subtracted from the uncorrected interaction energy, we obtain the *counterpoise-corrected* interaction energy,

$$\Delta E_{int}^{\mathrm{CP}}(\mathrm{AB}) = E^{\mathrm{AB}}(\mathrm{AB}) - E^{\mathrm{AB}}(\mathrm{A}) - E^{\mathrm{AB}}(\mathrm{B}). \tag{6}$$

In this derivation we have assumed that the geometries of monomers A and B do not change as they approach each other and form the bimolecular complex.

The CP-computations are carried out on pairs of methanol molecules. These pairs are selected from snapshots created during the MD run. For the CP-computations the same specifications (CP2K software package, basis set, electronic structure method, etc.) were used as those of the MD simulations of the methanol fluid. To accelerate convergence in the CP-computations of the pairs, the box size is decreased from 20.58 Å (in the MD run) to 10.0 Å.

III. DERIVATION OF THE CP FORCE-FIELD CORRECTION MODEL

A. Model

For the CP force field correction term, we consider a pairwise model between certain atoms in two different molecules. Because the BSSE converges by definition to zero at large distances, an exponential decaying model is proposed. The elec-

tron density of a molecule in vacuum exhibits an exponential decay. BSSE follows the same trend as the overlap of the electron density of two molecules and also decays exponentially. The goal of the correction term is to account for the artificial BSSE-stabilization, which may be responsible for the overestimation of the first intermolecular peak of the RDF. Therefore, the CP-model has to be repulsive.

We consider an intermolecular pairwise CP-correction model. To reduce the number of parameters, atom types are introduced. With each pair of atom types t, a unique set of parameters (A_t, B_t) is associated,

$$\Delta E_{\mathrm{CP}}^{\mathrm{mod}} = \sum_{\mu=1}^{N-1} \sum_{\nu=\mu+1}^{N} \sum_{\substack{i \in M_\mu \\ j \in M_\nu}} A_{t(ij)} \exp(-B_{t(ij)} r_{ij}), \qquad (7)$$

where r_{ij} is the distance between the atoms i and j, M_ν is a set of atomic indices of molecule ν, μ and ν are molecule indexes and N is the total number of molecules. An example of an intermolecular pair of atom types t is the pair O–H_m, consisting of the methyl hydrogen of monomer 1 (methanol molecule 1) and the oxygen of monomer 2 (methanol molecule 2). For liquid methanol, $M_\nu = 6 \forall \nu$, and $N = 129$ in our MD simulations.

The model is not to be confused with the functional form as it was introduced by Feller[84] to extrapolate energies, as well as properties, to the complete basis set (CBS) limit.[85–88] In this work, the parameters of the proposed model (i.e., $A_{t(ij)}$ and $B_{t(ij)}$) are fitted to reference data, which will be further explained below, and validated by comparing calculated RDFs with experiment.

B. Parameter calibration

To calibrate the parameters of the model (Eq. (7)), reference data are needed. These training data are obtained from molecular dynamics simulations. A snapshot of the fluid is selected every 400 steps (=400 fs) of the MD simulation (with dispersion, without CP-correction), corresponding to a time τ greater than the velocity auto-correlation time of the centers of mass (COMs) of the molecules to ensure statistically independent samples. The correlation time was computed with the MD-TRACKS program[89,90] and amounted to 86 fs. In this way, 21 frames of the MD simulation are selected. At each snapshot, the distances between the COMs of every possible pair of two methanol molecules are computed. The pairs of molecules for which this distance is smaller than 4 Å, are selected for the computation of reference data. Following this procedure, a total number of 3850 pairs are obtained. We then calculate the CP-correction for each of these pairs with Eq. (5).

The parameters $(A_{t(ij)}, B_{t(ij)})$ are estimated from the training data set of CP-computations. Several fitting procedures, linear and nonlinear, have been considered. After careful consideration, however, only the nonlinear optimized parameters are retained. For this nonlinear fit, the in-house developed software package FFit2[91] is used. FFit2 is a Python library for the calibration of force-field parameters based on electronic structure computations. Specifically, we make use of the submodule MFit2, which can be used to combine several types of calibration schemes for valence force fields.

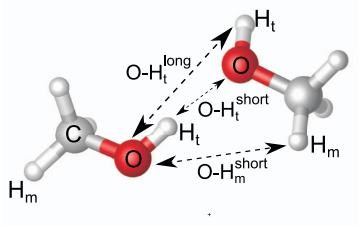


FIG. 1. Labels of the methanol atoms (C,O,H_m,H_t) and the most relevant intermolecular O–H pairs.

In this work, the parameters are optimized with a nonlinear least-squares fitting procedure based on the Levenberg-Marquardt algorithm.[92,93] The Levenberg-Marquardt algorithm is an iterative technique for the minimization of a multivariate function that is expressed as the sum of squares of nonlinear real-valued functions.

Once calibrated parameters $(A_{t(ij)}, B_{t(ij)})$ are obtained, a new MD simulation is performed with the Quickstep and FIST modules of CP2K. The energy in the MD run is computed as the DFT-energy (Quickstep) plus the counterpoise force-field term (FIST).

IV. RESULTS

In this section, the intermolecular RDFs computed from MD runs with and without CP-correction are presented. The relevant RDFs between oxygen and the different types of hydrogen as well the intermolecular hydrogen radial distributions, are discussed in the main paper; all other RDFs are given in the supplementary material.[97]

Figure 1 clarifies the atom types of the methanol molecules as used in this paper. We distinguish between the hydroxyl hydrogen (H_t) and the methyl hydrogen (H_m), as they are situated in different chemical environments. We introduce two distances which depend on the intermolecular geometry: O–H_t^{short} (the shortest intermolecular O–H_t distance between two methanol dimers) and O–H_t^{long} (the longest intermolecular O–H_t distance).

A. RDFs without correction

The RDFs are first computed without correcting for BSSE. Figure 2 shows the results for MD runs with and without dispersion correction. Inspection of Fig. 2 reveals that (i) the peaks of the O–H_t and the H_t–H_t RDF are too high compared to experiment, (ii) the first peaks are positioned at too large distances, and that (iii) the radial distribution is slightly overstructured. The inclusion of dispersion does not improve the correspondence with experiment. The absolute

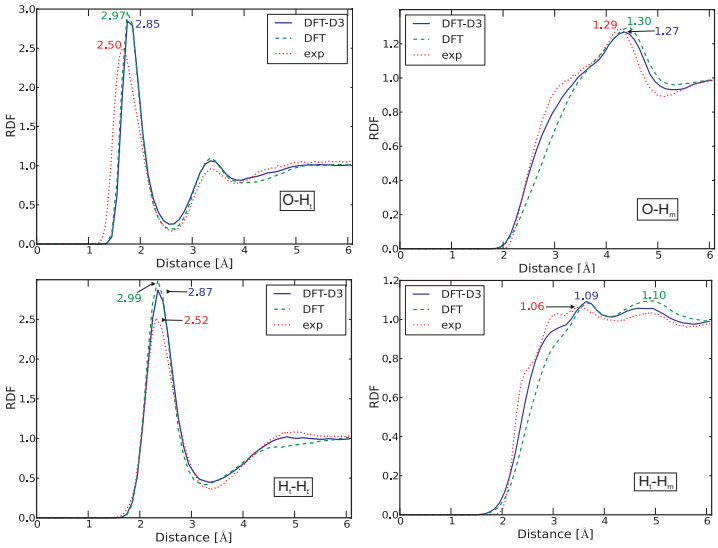


FIG. 2. Radial distribution functions for oxygen-hydrogen and hydrogen-hydrogen with BLYP and BLYP-D3. Experimental curves from Yamaguchi *et al.*[8,9]

and relative characteristic differences in RDFs are numerically reported in Table I.

As can be seen from Fig. 2, the intermolecular peaks of the RDFs involving the hydroxyl hydrogen H_t are overestimated, while those with the methyl hydrogen H_m are not. This immediately suggests that a distinction between H_t and H_m should be made when the parameters (A_t, B_t) of the CP-correction model are fitted.

The deficiency encountered in the position of the first peak, in particular, in the O–H_t RDF, will be discussed later.

B. Energy scan of two methanol monomers

To gain more insight in the pairwise interatomic interaction and to disentangle the contributions of dispersion and BSSE on the total interaction energy, an energy scan is performed of a complex, consisting of two methanol monomers, as a function of the intermolecular separation O–H_t^{short} distance to construct the potential energy surface (PES). First, the geometry of the two methanol molecules is optimized in the gas phase with CP2K, to form a stable dimer at a hydrogen bond distance of about 1.9 Å. Next, the O–H_t^{short} distance is gradually varied with a constrained geometry optimization with a step size of 0.1 Å ranging from 0.6 Å to 4.5 Å to obtain the energy scan in Fig. 3. The residual fluctuations that are visible in the curves are due to the well known egg-box effect.[94–96] The same computational settings are used for the scan as for the MD runs (Sec. II A) for consistency. This energy scan has been performed in four cases: with and without dispersion and with and without CP correction. The goal of this scan is threefold: (i) to study the interaction between two methanol molecules in more detail, (ii) to validate the proposed exponential model for the BSSE, (iii) to assess the relative importance of BSSE versus dispersion interaction. In Fig. 3 the total energy of the dimer E_{AB} is plotted against the O–H_t^{short} distance. All energies are relative with respect to the DFT binding energy of two isolated methanol monomers.

Inclusion of the D3 dispersion correction causes a slight shift to a smaller equilibrium distance. In contrast, BSSE, by means of the CP-correction term, creates the opposite effect

TABLE I. Amplitudes of the first peak of the RDFs for O–H_t and H_t–H_t and the relative error compared to experiment.[8,9]

	Expt.	DFT	Δ_{DFT} (%)	DFT-D3	$\Delta_{DFT\text{-}D3}$ (%)
O–H_t	2.50	2.97	15.7	2.85	12.1
H_t–H_t	2.52	2.99	15.8	2.87	12.2
O–H_m	1.29	1.30	0.7	1.27	1.4
H_t–H_m	1.06	1.09	2.9	1.09	2.7

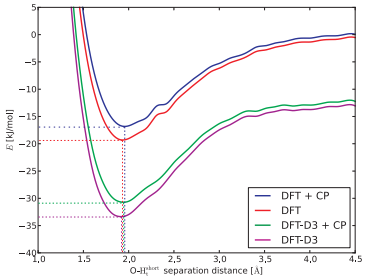


FIG. 3. Scan of a complex formed by two methanol monomers as a function of the separation distance O–H$_t^{short}$. All calculations are performed in gas phase with CP2K and with BLYP as functional.

on the equilibrium distance, but makes the potential some 3 kJ/mol less attractive. To get more insight into the relative importance of the CP correction terms versus the total dispersion energy of the dimer, both energy contributions are plotted in Fig. 4. Since the optimized energy for both monomers is not subtracted, the dispersion curve converges to a value of $2 \times E_{disp}^{mon} = -7.53$ kJ/mol, where E_{disp}^{mon} is the dispersion energy of one optimized monomer. The CP-correction curve converges to zero at a O–H$_t^{short}$ distance of 7.2 Å.[97]

Some interesting conclusions can be drawn: (i) the shapes of the CP-curves support the exponential model for the CP-correction, (ii) the dispersion correction is of the same order of magnitude as the BSSE, but with an opposite sign, (iii) in principle, the exponential trend as observed in Fig. 4 can be used for fitting the ($A_{\text{O–H}_t^{short}}$, $B_{\text{O–H}_t^{short}}$) parameters in Eq. (7). However, many plots as in Fig. 4 can be constructed. One

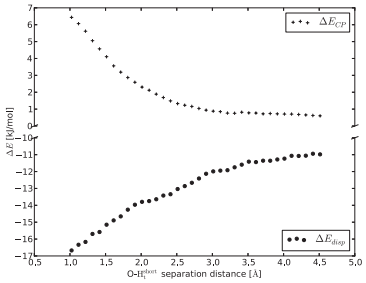


FIG. 4. CP-correction and dispersion energies derived from the $E_{\text{DFT+CP}}$ and $E_{\text{DFT+D3}}$ scans, respectively, in function of the interatomic O–H$_t^{short}$ distance.

could easily introduce many other intermolecular distances (such as O–H$_t^{long}$, O–H$_m^{short}$, etc.) and construct a potential energy scan following the same procedure as done in Fig. 4. Hence, for each pair of atom type t, one could construct a different set of (A_t, B_t) parameters. There is no reason why a particular parameter set is preferable to another. Moreover, there exist multiple dimer configurations with different spatial orientations of the methanol molecules and several paths to separate the two molecules from each other in the constrained geometry optimization. In the scans, performed in this work, only one single path is taken into consideration. It is obvious that this does not lead to a systematic way to derive the parameters. To remedy these shortcomings of the individual potential energy scans, it should be better to set up a fitting procedure that utilizes data from MD, allowing many possible orientations. This latter procedure provides a more statistically averaged result.

C. CP correction as a function of various O–H distances

In order to set up a fitting protocol which is based on MD, we make use of the data set of 3850 pairs of methanol molecules, obtained as described in Sec. III B. Figure 5 displays the logarithmic plot of the CP-energies as a function of the three O–H distances: O–H$_t^{short}$, O–H$_t^{long}$, and O–H$_m^{short}$ as defined in Fig. 1. For each specific pair of atom types t, we can construct an approximate model for the CP-correction,

$$\ln\left(\Delta E_{\text{CP}}(\mu, \nu)\right) \approx -B_{t(ij)} r_{ij} + \ln A_{t(ij)}$$
$$\forall \mu \neq \nu; i \in M_\nu, j \in M_\mu. \tag{8}$$

Equation (8) is a special case of the general model expressed in Eq. (7): only one molecular pair is present and only one kind of interaction is considered. A straight line can be fitted to the logarithmic plots in Fig. 5. With each linear fit, the corresponding Pearson R^2-value can be evaluated as a measure of correlation between the CP-energies and the oxygen-hydrogen distances. It is also clear that it is worth to distinguish between the two kinds of O–H$_t$ distances. The two fits are denoted by FIT/O–H$_t^{short}$ and FIT/O–H$_t^{long}$ and the parameters $A_{\text{O–H}_t}$ and $B_{\text{O–H}_t}$ are listed in Table II. On the contrary no linear correlation between $\ln \Delta E_{\text{CP}}$ and O–H$_m$ distances could be found (Fig. 5(c)). Therefore, this specific interaction is left out of the final fitting procedure. So finally we are going to take two terms into consideration to determine the parameters for the CP-correction model. It is emphasized that in Eq. (8) only one atom pair per dimer is considered to gain insight and cannot be used to correct MD-data as it would be impossible to run a proper simulation with a term that includes only one kind of interaction (i.e.,

TABLE II. A_t,B_t parameters methanol by fitting CP-energies.

Method	Interaction	A_t (kJ/mol)	B_t (1/Å)	R^2
FIT/O–H$_t^{short}$	O–H$_t^{short}$	4.96	0.323	0.857
FIT/O–H$_t^{long}$	O–H$_t^{long}$	11.50	0.450	0.737

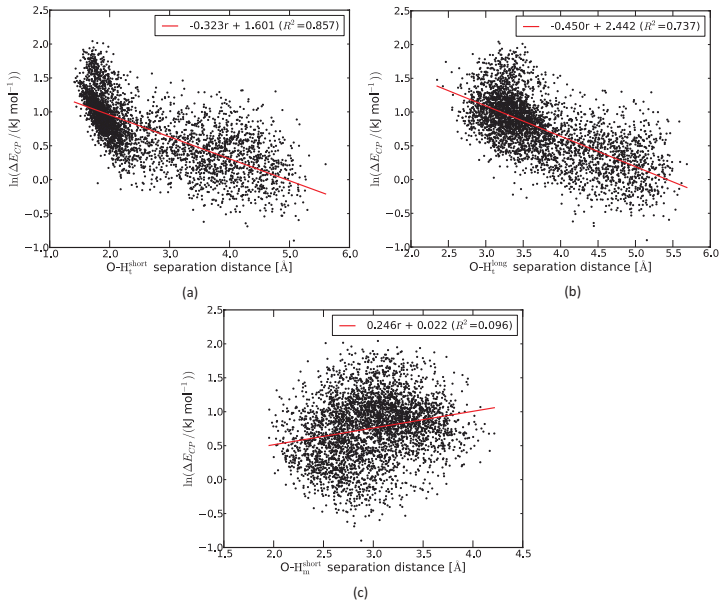


FIG. 5. CP-corrected energies as a function of various O–H distances based on geometries generated from MD: O–H$_t^{short}$, O–H$_t^{long}$, and O–H$_m^{short}$.

O–H$_t^{short}$ or O–H$_t^{long}$). As soon as the two O–H distances for a given pair interchange (short becomes long and vice versa) due to fluctuations on the distances, a discontinuity would appear in the forces. Such events are unphysical and violate energy conservation. For this reason, the linear fits in Fig. 5 are only indicative for the remainder of the analysis.

D. Final protocol for the determination of the CP-correction for the O–H$_t$ interaction

In Sec. IV C, different parameter sets were determined for the pairs O–H$_t^{short}$ and O–H$_t^{long}$, which provided useful insights in the correlation of these pair distances with the CP correction. However, the distinction between O–H$_t^{short}$ and O–H$_t^{long}$ is not compatible with Eq. (7) because H$_t^{short}$ and H$_t^{long}$ are not valid atom types. The distinction between long and short depends on the geometry of the pair, which may interchange during an MD simulation. In order to avoid discontinuities in the inter-molecular forces, the parameters for both O–H$_t^{short}$ and O–H$_t^{long}$ pairs must be identical. In line with Eq. (7), the CP model energy for a single pair of methanol molecules μ and ν (only considering the O–H$_t$ terms) becomes

$$\Delta E_{CP}^{mod}(\mu,\nu) = A_{O-H_t}\exp\left(-B_{O-H_t} r_{O-H_t^{short}}\right) + A_{O-H_t}\exp\left(-B_{O-H_t} r_{O-H_t^{long}}\right). \quad (9)$$

The parameters A_{O-H_t} and B_{O-H_t} are no longer determined by a linear fitting procedure due to the appearance of a non-linear term in the expression of $\ln\{E_{CP}^{mod}(\mu,\nu)\}$. We apply an in-house developed fitting procedure FFit2[91] (Sec. III B). The results are shown in Table III and are deduced from the DFT-D3 MD data set. As initial values for the parameters we have taken $A = 1$ kJ/mol and $B = 1$ Å^{-1} which are reasonable initial values.

TABLE III. A_{O-H_t} and B_{O-H_t} parameters methanol determined with FFit2 based on DFT-D3 MD.

Method	Interaction	Optimized value	
		A (kJ/mol)	B (Å^{-1})
CP/O–H$_t$	O–H$_t$	4.47	0.453

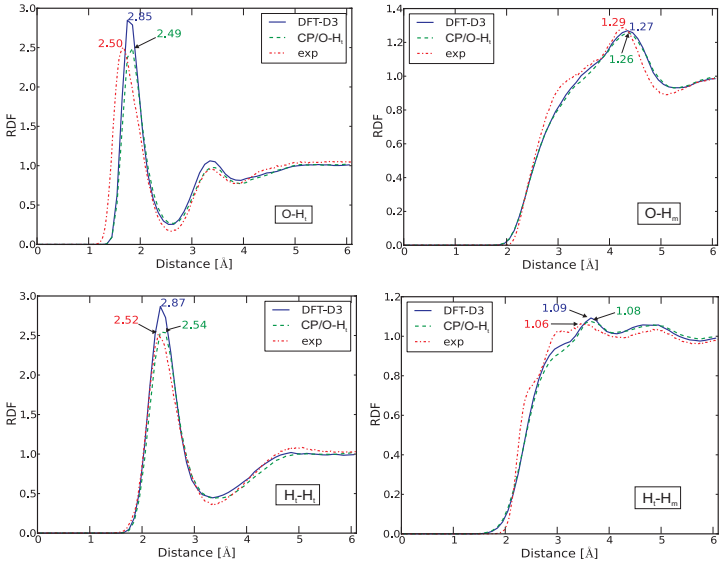


FIG. 6. Comparison RDFs methanol: DFT-D3, CP/O–H_t, and experiment.

E. RDFs with correction

Figure 6 presents the CP-corrected RDFs computed from a MD run following the new CP-protocol and taking into account DFT-D3 dispersion corrections. The DFT-D3 RDF is shown as reference. The inclusion of BSSE according to our CP-model significantly improves the reproduction of the experimental peak for O–H_t and H_t–H_t. The amplitudes are tabulated in Table IV. As expected, the H_t–H_m and the O–H_m RDFs are not affected by adding the extra BSSE model term since the interactions with the H_m atoms were left out of the calibration procedure.

V. DISCUSSION

A first conclusion from the energy scan of the two methanol monomers (Figs. 3 and 4) is that the BSSE and dispersion interaction are almost equal in strength but opposite in sign. Dispersion is attractive and converges very slowly at large distances. At an O–H_t distance of 4.5 Å the dispersion energy still accounts for −11 kJ/mol and the slow increase of the dispersion curve (Fig. 4) demonstrates manifestly its long-range behavior. The CP-energy correction is repulsive and compensates partially the dispersion at short distances. The strongest effects of BSSE are noticed at distances shorter then 2.5 Å, and rapidly decays exponentially to about 1 kJ/mol at intermediate distances (3–3.5 Å), the structure of the RDFs at intermediate and long-range will be mainly determined by the dispersion. This is confirmed by the various RDF curves displayed in Figs. 2 and 6.

TABLE IV. Amplitudes of the first peak of the RDFs for O–H_t and H_t–H_t and the relative error compared to experiment.

	Expt.	DFT-D3	$\Delta_{\text{DFT-D3}}$ (%)	CP/O–H_t	$\Delta_{\text{CP/O–H}_t}$ (%)
O–H_t	2.50	2.85	12.1	2.49	0.4
H_t–H_t	2.52	2.87	12.2	2.54	1.0

The most pronounced effect is the ability of the CP model to reduce the amplitude of the first peak to values which are comparable to experimental predictions (Table IV). In this perspective, it is instructive to compare our results with similar molecular dynamics simulations in the literature, although the influence of BSSE on RDFs has not been investigated in all these works. Special attention goes to the work of Schmidt *et al.* where DFT MD simulations have been carried out for water in near-ambient conditions with almost the same specifications as in our work, but within an isobaric-isothermal ensemble (NPT).[27] The authors used two functionals BLYP and PBE and reported some functional dependence

in the position of the first peak of the oxygen-oxygen RDF in water. BLYP overestimates the position by about 0.1 Å toward larger values due to the tendency of BLYP to produce weaker hydrogen bonds (underbinding) than that predicted by the PBE functional (overbinding). In turn, the height for the first oxygen-oxygen peak is significantly larger for PBE. Inclusion of dispersion following the correction proposed by Grimme,[45] gives a substantial improvement in the RDF but this is not exactly what we observe in methanol. Different Grimme corrections are used in both works, which troubles the comparison. Schmidt *et al.* makes use of DFT-D (D1 and D2) corrections described by damped interatomic potentials of the form C^6R^{-6}, while in this work the more recent DFT-D3 corrections were used. We expect that differences in the D-correction will affect the height and position of the first RDF peak. For the convenience of the reader the most relevant interatomic potentials are included in the supplementary material for the three Grimme parameterizations D1, D2, and D3. We concentrate on the O–H interatomic potential[97] as we could expect that it affects the O–H_t RDF. Minima of this dispersion term are observed around 2.5 Å with an interaction energy of the order of −0.7 kJ/mol for D1 and D3. These features are too moderate to have a significant impact on the RDF. One notices a slight increase of the O–H radial distribution curves—DFT compared with DFT-D3—in the range around 2.5 Å, which is counterintuitive given the attractive nature of the dispersion potential. In the region of the first peak, the dispersion correction to the interatomic O–H potential is small and, as one could expect, its influence on the height of the first peak would be negligible. Nevertheless, one must be very careful in oversimplifying the overall mechanism responsible in the reproduction of the first RDF peak. First, the total dispersion energy in the region of the first peak is of the order of −14 kJ/mol (based on the energy scan in Fig. 4), which contains contributions from multiple pairs. Second, our comments concern the methanol liquid while Schmidt *et al.* deal with water.

The importance of dispersion interactions in predicting structural and dynamical properties of liquid water is also demonstrated by Lin *et al.*[28] In this study both dispersion-corrected atom-centered potentials,[48] as well as empirical DFT-D1 Grimme corrections,[45] are used as two pragmatic ways to describe dispersion interactions in BLYP *ab initio* molecular dynamics simulations. They found an improved agreement of the RDFs with experiment, but significant deviations still persist.

Oxygen-oxygen and oxygen-hydrogen radial distribution functions for liquid water apparently improve when taking into account dispersion interactions. From our MD-BLYP calculations, we conclude that the effect of dispersion in a methanol liquid is less significant. The BSSE corrections turn out to be much more important for a correct description of the height of the first peak in the O–H_t and H_t–H_t RDFs. The deviation with experiment is largely reduced to, respectively, 0.4% and 1% from the experimental value (Table IV). These very small differences can be resolved by noting that the experimental RDF is only extracted indirectly from experiment. It is computed from Monte Carlo simulations according to the empirical potential model refinement (EPSR) procedure. In this iterative technique, a potential energy function is refined by making use of empirical data as input:[98–101] the Fourier transform of the difference between experimental and simulated structure factors is used solely for ensuring the best possible agreement between the simulated and measured partial structure factors.[102] The use of an interaction potential, from which the RDFs are extracted, may generate small discrepancies between this experimentally derived RDF and the "real" RDF. Yamaguchi *et al.* illustrate how a different set of parameters of a single term of the potential can change the resulting site-site RDFs.[9]

TABLE V. Positions of the first intermolecular RDF peak in Å from numerical simulations and experiment.

RDF	Expt.	DFT	DFT-D3	CP/O–H_t
O–H_t	1.68	1.75	1.75	1.85
H_t–H_t		2.35	2.35	2.35

The position of the first peak in the O–H_t and H_t–H_t RDFs is still overestimated with or without BSSE and/or dispersion correction (Table V), compared to experiment. Plausible explanations for this discrepancy are twofold. First, the shift of the peak position towards longer distances is in line with intuitive expectations since the CP model adds a repulsive interaction between the molecules. Second, the differences with experiment in peak positions could be due to the choice of the DFT-functional, BLYP, that was used. Schmidt *et al.* found that, for water, the position of the peaks shifted by about 0.1 Å toward larger values for BLYP as compared to PBE.[27] This shift confirms the tendency of the BLYP functional to produce weaker hydrogen bonds than the PBE functional.[103,104]

One might suspect that the CP correction model may induce artificial distortions of the monomer geometries. We have compared the time averages and spreads of the intramolecular O–H_t distance and C–O–H_t angle (data not shown) between simulations without and with the CP correction model. No significant distortions have been noticed, indicating that the extra intermolecular forces by the model are small compared to the intramolecular forces.

The results so far clearly indicate that the CP model improves the O–H_t RDF, however, it also introduces minor deviations in the O–O RDF.[97] The height of the experimental peak of the O–O RDF has an amplitude of 2.83, while the CP-correction lowers the BLYP-D3 peak from 2.89 to 2.48. This outcome is not surprising since the oxygen atom is covalently bonded to the H_t atom, and hence some correlation exists between the position of the oxygen and the hydroxyl hydrogen. A simultaneous parametrisation of the O–O and the O–H_t interaction could improve the agreement. However, this is a non-trivial task due to the nonlinear character of Eq. (7) in the fitting-procedure. First, the system of equations to be solved in the Levenberg-Marquardt algorithm becomes more ill-conditioned when the parameter space is extended. Second, the parameters A_{O-H_t}, B_{O-H_t} and A_{O-O}, B_{O-O} are heavily correlated: the O–O distance and the O–H_t are not independent variables, causing their parameters to be also

dependent. These problems render it hard to obtain meaningful parameters for the O–O and O–H_t interaction at the same time. Furthermore, parameter correlation impairs the physical meaning of the model parameters in Eq. (7) and reduces the predictive power of the model. When even more interaction terms are considered, the nonlinear optimization problem becomes even more difficult.

VI. CONCLUSION AND PERSPECTIVES

The radial distribution function gives the probability of finding a pair of molecules at a distance r apart, relative to the expected probability for a completely random distribution at the same density. In DFT MD, using localized basis sets, the first intermolecular peak of the RDF of a liquid is often overestimated. In this paper it was investigated for liquid methanol to what extent this error is caused by the BSSE, which follows from the use of finite basis sets in quantum chemical calculations.

The relative importance of BSSE versus the dispersion interaction (DFT-D3) was assessed. The dispersion energy is of the same order of magnitude as the BSSE, but has an opposite sign.

Combinations of the two types of corrections—a correction term for BSSE and dispersion—leads to three types of DFT MD simulations that were carried out. For the simulations with and without dispersion, without correction for BSSE, errors on the first peak height of the RDFs between 12% and 16% were found. A classical correction model for BSSE is proposed and its parameters were derived from counterpoise computations on methanol dimers. This CP model can then be added to the DFT energy in subsequent MD simulations. We confirmed that an exponential decaying functional form for the CP model works well. Because the counterpoise correction correlates stronger with the O–H_t distance, as compared to the O–H_m distance, the final form of the CP model is only based on O–H_t distances. Upon introduction of the correction, the errors on the heights of the first peak of the relevant O–H_t and H_t–H_t RDFs are reduced to 0.4% and 1%, respectively. A small shift of the peak position compared to experiment is still observed, which could be due to the added repulsive interaction of the CP model and the limitations of the BLYP functional.

We focused on hydrogen bridges in methanol by applying the CP model to the intermolecular interaction between the oxygen and the hydroxyl hydrogen, but the results should be extendable to other types of interaction where BSSE is important. However, including more interaction terms in the CP model complicates the optimization of the parameters because of the inherent parameter correlation and ill-conditioning of the set of equations that needs to be solved. It needs furthermore to be tested in how far the CP model is transferable to other protic solvents.

In this paper we concentrated on the effect of BSSE, and correcting for it, on a structural quantity. It would be highly valuable to obtain similar information for other quantities as well, e.g., dynamical properties such as the diffusion coefficient, or vibrational spectra. This way, it would be possible to obtain a complete picture of the influence of BSSE on molecular dynamics simulations that makes use of localized basis sets.

ACKNOWLEDGMENTS

This work is supported by the Fund for Scientific Research - Flanders (FWO), the Research Board of Ghent University (BOF) and BELSPO in the frame of IAP/6/27, the Belgian Prodex office ESA and the European Research Council under the European Community's Seventh Framework Programme (FP7(2007–2013) ERC grant agreement number 240483). The computational resources and services used were provided by Ghent University (Stevin Supercomputer Infrastructure). A.G. and T.V. are post-doctoral researchers funded by the Foundation of Scientific Research - Flanders.

[1] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, 1st ed. (Oxford University Press, 1991), Chap. 2.
[2] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[3] P. Hohenberg and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[4] O. Mó, M. Yáñez, and J. Elguero, J. Chem. Phys. **107**, 3592 (1997).
[5] E. Tsuchida, Y. Kanada, and M. Tsukada, Chem. Phys. Lett. **311**, 236 (1999).
[6] R. Ludwig, ChemPhysChem **6**, 1369 (2005).
[7] J. A. B. da Silva, F. G. B. Moreira, V. M. L. dos Santos, and R. L. Longo, Phys. Chem. Chem. Phys. **13**, 593 (2011).
[8] T. Yamaguchi, K. Hidaka, and A. Soper, Mol. Phys. **96**, 1159 (1999).
[9] T. Yamaguchi, K. Hidaka, and A. Soper, Mol. Phys. **97**, 603 (1999).
[10] T. Yamaguchi, C. J. Benmore, and A. K. Soper, J. Chem. Phys. **112**, 8976 (2000).
[11] A. K. Adya, L. Bianchi, and C. J. Wormald, J. Chem. Phys. **112**, 4231 (2000).
[12] J.-W. Handgraaf, T. S. van Erp, and E. J. Meijer, Chem. Phys. Lett. **367**, 617 (2003).
[13] J.-W. Handgraaf, E. J. Meijer, and M.-P. Gaigeot, J. Chem. Phys. **121**, 10111 (2004).
[14] M. Pagliai, G. Cardini, R. Righini, and V. Schettino, J. Chem. Phys. **119**, 6655 (2003).
[15] M. J. McGrath, I.-F. Kuo, and J. I. Siepmann, Phys. Chem. Chem. Phys. **13**, 19943 (2011).
[16] M. E. Martín, M. L. Sánchez, F. J. del Valle, and M. A. Aguilar, J. Chem. Phys. **116**, 1613 (2002).
[17] M. L. Sánchez, M. E. Martín, M. A. Aguilar, and F. J. del Valle, J. Comput. Chem. **21**, 705 (2000).
[18] M. Haughney, M. Ferrario, and I. R. McDonald, J. Phys. Chem. **91**, 4934 (1987).
[19] L. Bianchi, O. N. Kalugin, A. K. Adya, and C. J. Wormald, Mol. Simulat. **25**, 321 (2000).
[20] J. L. Thomas, D. J. Tobiass, and A. D. MacKerell, J. Phys. Chem. B **111**, 12941 (2007).
[21] L. Zoranić, F. Sokolić, and A. Perera, J. Chem. Phys. **127**, 024502 (2007).
[22] T. Kosztolányi, I. Bakó, and G. Pálinkás, J. Chem. Phys. **118**, 4546 (2003).
[23] T. Ishiyama, V. V. Sokolov, and A. Morita, J. Chem. Phys. **134**, 024509 (2011).
[24] H. Yu, D. P. Geerke, H. Liu, and W. F. V. Gunsteren, J. Comput. Chem. **27**, 1494 (2006).
[25] S. Patel and C. L. Brooks, J. Chem. Phys. **122**, 024508 (2005).
[26] H. Nakano, T. Yamamoto, and S. Kato, J. Chem. Phys. **132**, 044106 (2010).
[27] J. Schmidt, J. VandeVondele, I.-F. W. Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy, J. Phys. Chem. B **113**, 11959 (2009).
[28] I.-C. Lin, A. P. Seitsonen, M. D. Coutinho-Neto, I. Tavernelli, and U. Rothlisberger, J. Phys. Chem. B **113**, 1127 (2009).
[29] J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli, J. Chem. Phys. **120**, 300 (2004).
[30] T. Ishiyama and A. Morita, J. Chem. Phys. **131**, 244714 (2009).
[31] J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello, J. Chem. Phys. **122**, 014515 (2005).
[32] M. Sprik, J. Hutter, and M. Parrinello, J. Chem. Phys. **105**, 1142 (1996).

[33]A. D. Boese, N. L. Doltsinis, N. C. Handy, and M. Sprik, J. Chem. Phys. **112**, 1670 (2000).
[34]G. S. Fanourgakis, G. K. Schenter, and S. S. Xantheas, J. Chem. Phys. **125**, 141102 (2006).
[35]B. Guillot and Y. Guissani, J. Chem. Phys. **108**, 10162 (1998).
[36]M. E. Tuckerman, D. M. Marx, M. L. Klein, and M. Parrinello, Science **275**, 817 (1997).
[37]E. Schwegler, J. C. Grossman, F. Gygi, and G. Galli, J. Chem. Phys. **121**, 5400 (2004).
[38]C. Swalina, Q. Wang, A. Chakraborty, and S. Hammes-Schiffer, J. Phys. Chem. A **111**, 2206 (2007).
[39]E. G. Noya, C. Vega, L. M. Sesé, and R. Ramirez, J. Chem. Phys. **131**, 124518 (2009).
[40]E. Balog, A. L. Hughes, and G. J. Martyna, J. Chem. Phys. **112**, 870 (2000).
[41]J. A. Morrone and R. Car, Phys. Rev. Lett. **101**, 017801 (2008).
[42]G. S. Fanourgakis and S. S. Xantheas, J. Chem. Phys. **128**, 074506 (2008).
[43]H. A. Stern and B. J. Berne, J. Chem. Phys. **115**, 7622 (2001).
[44]R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 24712474 (1985).
[45]S. Grimme, J. Comput. Chem. **25**, 1463 (2004).
[46]S. Grimme, J. Comput. Chem. **27**, 1787 (2006).
[47]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).
[48]I.-C. Lin, M. D. Coutinho-Neto, C. Felsenheimer, O. A. von Lilienfeld, I. Tavernelli, and U. Rothlisberger, Phys. Rev. B **75**, 205131 (2007).
[49]A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).
[50]M. Mantina, A. C. Chamberlin, R. Valero, C. J. Cramer, and D. G. Truhlar, J. Phys. Chem. A **113**, 5806 (2009).
[51]S. Simon, M. Duran, and J. J. Dannenberg, J. Chem. Phys. **105**, 11024 (1996).
[52]K. N. Kirschner, J. B. Sorensen, and J. P. Bowen, J. Chem. Educ. **84**, 1225 (2007).
[53]S. Simon, J. Bertran, and M. Sodupe, J. Phys. Chem. A **105**, 4359 (2001).
[54]S. Scheiner, *Molecular Interactions. From van der Waals to Strongly Bound Complexes*, 1st ed. (Wiley, 1997).
[55]D. Hadži, *Theoretical Treatments of Hydrogen Bonding*, 1st ed. (Wiley, 1997).
[56]F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, Chem. Rev. **94**, 1873 (1994).
[57]A. Bende and S. Suhai, Int. J. Quantum Chem. **103**, 841 (2005).
[58]T. van Mourik, J. Phys. Chem. A **112**, 11017 (2008).
[59]D. Tzeli, I. D. Petsalakis, and G. Theodorakopoulos, J. Phys. Chem. A **111**, 8892 (2007).
[60]A. E. Shields and T. van Mourik, J. Phys. Chem. A **111**, 13272 (2007).
[61]L. F. Holroyd and T. van Mourik, Chem. Phys. Lett. **442**, 42 (2007).
[62]P. Hobza and Z. Havlas, Theor. Chem. Acc. **99**, 372 (1998).
[63]R. Crespo-Otero, L. A. Montero, W.-D. Stohrer, and J. M. D. la Vega, J. Chem. Phys. **123**, 134107 (2005).
[64]D. Tzeli and A. A. Tsekouras, Chem. Phys. Lett. **496**, 42 (2010).
[65]H. Kruse and S. Grimme, J. Chem. Phys. **136**, 154101 (2012).
[66]CP2K, 2011, see http://cp2k.berlios.de.
[67]J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, Comput. Phys. Commun. **167**, 103 (2005).
[68]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[69]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[70]G. Lippert, J. Hutter, P. Ballone, and M. Parrinello, J. Phys. Chem. **100**, 6231 (1996).
[71]S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).
[72]C. Harwigsen, S. Goedecker, and J. Hutter, Phys. Rev. B **58**, 3641 (1998).
[73]E. J. Meijer and T. S. van Erp, Chem. Phys. Lett. **333**, 290 (2001).
[74]P. Hobza, J. Šponer, and T. Reschel, J. Comput. Chem. **16**, 1315 (1995).
[75]S. Kristyán and P. Pulay, Chem. Phys. Lett. **229**, 175 (1994).
[76]S. Grimme, J. Antony, T. Schwabe, and C. Mück-Lichtenfeld, Org. Biomol. Chem. **5**, 741 (2007).
[77]S. Nosé, J. Chem. Phys. **81**, 511 (1984).
[78]I. Mayer and P. R. Surján, Int. J. Quantum Chem. **36**, 225 (1989).
[79]I. Mayer and Á. Vibók, Int. J. Quantum Chem. **40**, 139 (1991).
[80]J. Noga and Á. Vibók, Chem. Phys. Lett. **180**, 114 (1991).
[81]P. Valiron, Á. Vibók, and I. Mayer, J. Comput. Chem. **14**, 401 (1993).
[82]Á. Vibók and I. Mayer, Int. J. Quantum Chem. **43**, 801 (1992).
[83]S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).
[84]D. Feller, J. Chem. Phys. **98**, 7059 (1993).
[85]B. Paizs, P. Salvador, A. G. Császár, M. Duran, and S. Suhai, J. Comput. Chem. **22**, 196 (2001).
[86]I. Alkorta, C. Trujillo, J. Elguero, and M. Solimannejad, Comput. Theor. Chem. **967**, 147 (2011).
[87]T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).
[88]S. K. Min, E. C. Lee, H. M. Lee, D. Y. Kim, D. Kim, and K. S. Kim, J. Comput. Chem. **29**, 1208 (2008).
[89]T. Verstraelen, M. Van Houteghem, V. Van Speybroeck, and M. Waroquier, J. Chem. Inf. Mod. **48**, 2414 (2008).
[90]See http://molmod.ugent.be/software for more information on MD-TRACKS, a trajectory analysis toolkit for MD and Monte Carlo simulations, and other software for preparing and post-processing data from molecular simulations.
[91]L. Vanduyfhuys, T. Verstraelen, M. Vandichel, M. Waroquier, and V. Van Speybroeck, "*Ab initio* parametrized force field for the flexible metal-organic framework MIL-53(Al)," J. Chem. Theory Comput. (in press).
[92]K. Levenberg, Quart. Appl. Math. **2**, 164 (1944).
[93]D. Marquardt, SIAM J. Appl. Math. **11**, 431 (1963).
[94]E. Artacho, E. Anglada, O. Diéguez, J. D. Gale, A. García, J. Junquera, R. M. Martin, P. Ordejón, J. M. Pruneda, D. Sánchez-Portal, and J. M. Soler, J. Phys.: Condens. Matter **20**, 064208 (2008).
[95]J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).
[96]E. Anglada and J. M. Soler, Phys. Rev. B **73**, 115122 (2006).
[97]See supplementary material at http://dx.doi.org/10.1063/1.4749929 for O–O RDF in Fig. S2; dispersion contributions to the interaction potentials for oxygen-hydrogen pair in Fig. S3; and an extended potential energy scan ranging from 0.6 Å to 7.5 Å in Fig. S4.
[98]A. K. Soper, Chem. Phys. **202**, 295 (1996).
[99]A. K. Soper, J. Phys. Condens. Matter **9**, 2717 (1997).
[100]A. K. Soper, Mol. Phys. **99**, 1503 (2001).
[101]D. T. Bowron, J. L. Finney, and A. K. Soper, J. Phys. Chem. B **102**, 3551 (1998).
[102]A. K. Soper, Phys. Rev. B **72**, 104204 (2005).
[103]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[104]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **78**, 1396 (1997).

Paper III

Critical analysis on the accuracy of models predicting or extracting information on liquid structure

M. Van Houteghem, A. Ghysels, T. Verstraelen, W. Poelmans, M. Waroquier, V. Van Speybroeck

submitted

Critical analysis on the accuracy of models predicting or extracting information on liquid structure.

Marc Van Houteghem,[1] An Ghysels,[1] Toon Verstraelen,[1] Ward Poelmans,[1] Michel Waroquier,[1] and Veronique Van Speybroeck[1,*]

[1] *Center for Molecular Modeling, QCMM Alliance Ghent-Brussels, Ghent University, Technologiepark 903, B-9052 Zwijnaarde, Belgium*

Abstract

This work aims at a critical assessment of properties predicting or extracting information on the density and structure of liquids. State of the art NVT and NpT molecular dynamics simulations have been performed on five liquids: methanol, chloroform, acetonitrile, tetrahydrofuran and ethanol. These simulations allow the computation of properties based on first principles, including the equilibrium density and radial distribution functions (RDFs), characterizing the liquid structure. Refinements have been incorporated in the MD simulations by taking into account Basis Set Superposition Errors (BSSE). An extended BSSE model for an instantaneous evaluation of the BSSE corrections has been proposed, and their impact on the liquid properties has been assessed. If available, the theoretical RDFs have been compared with the experimentally derived RDFs. For some liquids significant discrepancies have been observed and a profound but critical investigation is presented to unravel the origin of these deficiencies. This discussion is focused on THF where the experiment reveals some prominent peaks completely missing in any MD simulation. Experiments providing information on liquid structure consist mainly of neutron diffraction measurements offering total structure factors as the primary observables. The splitting of these factors in reciprocal space into intra- and intermolecular contributions is extensively discussed, together with their sensitivity in reproducing correct RDFs in coordinate space.

*veronique.vanspeybroeck@ugent.be

I. INTRODUCTION

The structure of liquids is a recurring theme in the scientific literature. Liquids do not have a structure in the sense that crystals have a structure. The forces between liquid molecules are not as strong as those holding solids together, but they show some attractive character arising from dipole-dipole interactions such as hydrogen bonding. The intermolecular forces lead to characteristic correlations, which determine the structure of the liquid and which affect many physical (thermodynamical) properties of the liquid. It is important to know the microscopic structure of liquids in as much detail as possible because it determines their macroscopic properties, e.g. thermodynamic functions.

Characteristics of liquid structure are thus mainly governed by intermolecular interactions, of which hydrogen bonding constitutes the main class if present. These special forces are not as strong as the electrostatic ionic forces, but they have a characteristic that makes them unique among the repertoire of forces at the disposal of molecules-direction. Water is by far the most popular among the hydrogen-bonded molecules.[1–8] The liquid phase of methanol, consisting of a hydrogen-bonding hydroxyl group and a hydrophobic methyl group, is also extensively studied as it is the smallest alcohol which can be studied to characterize the hydrogen bonding in alcohols.[9–12]

Neutron diffraction measurements with isotopic substitution (NDIS) are one of the most powerful methods to extract experimental information on liquid structure. The differences in the experimental scattering patterns by replacing a particular chemical element by their isotope (e.g. hydrogen by deuterium) in the solvent molecule enable the extraction of structure factors determining the differential cross section of the neutron scattering process. However, the interpretation of the experimentally obtained structure factors is a difficult task as the measured scattering intensities are the result of a weighted summation of different pairwise spatial correlation functions formed by all atoms in the system. Disentangling the various correlations remains a very difficult and challenging task. In particular the separation of the experimental total structure factor into intramolecular and intermolecular contributions is extremely complex even within the NDIS methodology. The total structure factor is composed of partial structure factors reflecting the pairwise spatial correlations between different atom types with multiplicative prefactors containing concentrations and scattering lengths of each atom type. In case of a solute and a solvent, e.g. aqueous solution of a polyatomic

organic molecule, the atomic concentrations in the solution can be handled as an additional degree of freedom. While the intermolecular contribution varies with concentration, the intramolecular contribution does not change. Exploiting this concentration invariance, NDIS experiments with different concentrations should in principle be able to extract intermolecular partial structure factors [13]. Undoubtedly this method has its benefits and has large potential to become a standard procedure, but it can not be applied on liquids composed of one type of molecules.

In general there are not enough different scattering experiments available to extract all information needed to completely characterize the liquid. Even with the help of three-dimensional models of the liquid structure with empirical parameters which can be adjusted to reproduce the experimental structure factors, this method is prone to large inaccuracies/errors. The numerical instabilities become even strengthened when inverting the structure factors to real space distribution functions, as the inverse Fourier transform is error-prone and ill-conditioned [14–17]. These are serious deficiencies and the results are absolutely not free from any bias.

By the group of Soper [18] the incompleteness of the experimental data to extract the partial structure factors is circumvented by combining the experimental measurements with additional Monte Carlo simulations based on intermolecular site-site potentials. Structure refinements are proposed such to reproduce experimental data as best as possible. The method is pure empirical in the sense that perturbations are induced iteratively to the various potentials till satisfactory agreement with the experimental data is reached. This methodology is called Empirical Potential Structure Refinement (EPSR) model and is closely related to the original reverse Monte Carlo simulation technique (RMC) as originally introduced by McGreevy and Pusztai [16, 19]. RMC is a method of modeling a structure and only produces a result that is consistent with the experimental data. Both RMC and EPSR are inverse methods of structural modeling, in which building and refining of the particle configurations during the course of the simulations form an essential ingredient.

Related methods to reverse MC modeling combined with MD computer simulations based on interatomic potentials have also been applied by other authors [10, 11] on alcohols and haloforms.

In the above RMC-EPSR iterative approaches the radial distribution function (RDF) plays a central role. It is a quantity determined from intermolecular forces and forms a

bridge between the microscopic structure of the liquid and macroscopic properties such as (thermodynamic) energy (E), entropy (S) and pressure (p). Appropriate theoretical tools to compute RDFs are offered by molecular dynamics simulations. They can be performed with force fields [10, 11, 13, 20] or based on density functional theory [21, 22]. MD modeling has proven to be very effective in helping to guide assignments of peaks in the experimentally derived radial distribution functions, though these last quantities are not direct observables and obtained after some error-prone manipulations, as already mentioned. We need to be very cautious in using them as reference quantities to judge the quality of the theoretical predictions. A striking example are the intermolecular partial radial distribution functions for tetrahydrofuran (THF) [18] showing sharp peaks at short distances, which are completely missing in all MD simulations published in literature so far. This apparent discrepancy between theory and the empirical EPSR model requires a profound and thorough investigation to find out the origin of the appearance of this sharp peak in the experimental RDF of THF, and why theory does not succeed in reproducing it. This discussion should be extended in a broader context as similar, however less pronounced, features are also observed in other liquids. A discussion on this item forms one of the principal focuses of this work.

In order to unravel this apparent mismatch between theory and experiment it is better to focus on the structure factors and to perform the assessment in reciprocal space instead of in real coordinate space. A Fourier transform to convert RDFs to SFs requires some mathematical manipulations but the accuracy of these computations is better controlled. In this work we will not only compute (partial) RDFs and their counterparts, the (partial) SFs, but we will also try to investigate the impact of intramolecular contributions on the SFs, which experimentally poses a lot of complications. Theoretically, separation of the intra- and intermolecular contributions of the various properties does not cause any problem, and may assist in finding a correct interpretation of the features observed experimentally in the structure factors. In addition we will examine the sensitivities of small perturbations in the structure factors on the radial distributions.

Apart from the above discussion, the structure of five liquids will be investigated with DFT-based molecular dynamics simulations: methanol (MeOH), chloroform (TCM), acetonitrile (MeCN), tetrahydrofuran (THF) and ethanol (EtOH). The availability of first principles MD simulation studies on liquids to derive their liquid structure is less abundant in literature. The choice of the five solvents were based on (i) their importance for the

industry, and (ii) their classification : non-polar (TCM), polar aprotic (MeCN, THF), and polar protic (MeOH and EtOH). The liquid structure of methanol has already been studied earlier [22] .

In DFT-MD simulations, it has been shown that (i) dispersion and (ii) the basis set superposition error (BSSE) have a significant effect on the RDFs of water and methanol.

(i) Standard DFT methods do not account for dispersion interactions.[23, 24] Dispersion is a long-range attractive van der Waals interaction between electron clouds arising from instantaneous transition dipoles. Schmidt *et al.*[25] have found that adding the Grimme dispersion correction[26, 27] improves the predicted densities of water in the isothermal-isobaric ensemble (NpT), for both BLYP[28, 29] and PBE[30] functionals, compared to experiment. In addition, the first and second maxima of the oxygen-oxygen RDF of water using the BLYP-D functional instead of BLYP lead to a much better agreement with the experimental data, although the RDF for PBE-D still remains overstructured[25].

(ii) The BSSE results from an inconsistent description of the molecular orbitals as a result of using a limited number of localized basis functions (incomplete basis set). When two molecules are far apart, they can only "use" their own basis functions. When the intermolecular distance decreases, they may use each other's basis, causing an artificial strengthening of the intermolecular interactions and artificial shortening of intermolecular distances due to the overlap of basis functions. In ref. [22] a systematic correction of the BSSE with an exponentially decaying parameterized pairwise classical force-field energy term has been proposed by the authors that can be added in subsequent MD simulations. The model is applied to methanol using MD in a canonical ensemble (NVT) with the BLYP-D functional. It is found that upon introduction of this correction, the errors on the first peak height of the relevant partial RDFs of the atoms that participate in hydrogen-bonding are reduced from 12%-16% to 0.4%-1%, compared to experiment. Another conclusion is that, for methanol, the Grimme-dispersion correction and the BSSE have the same order of magnitude but an opposite sign.

This paper extends the work in ref. [22] in the sense that (i) MD simulations in the NpT ensemble have been performed on the five liquids under consideration, (ii) the BSSE correc-

tion scheme as proposed in ref. [22] has been extended to all intermolecular interactions.

(i) The isothermal-isobaric ensemble (NpT) is considered to allow density fluctuations of the liquid. An RDF is directly related to the density of the molecular system, so it is vital that the density in MD simulations is properly reproduced. In ref. [22] a canonical NVT ensemble was considered.

(ii) The BSSE-correction model proposed in ref. [22] is an exponentially decaying parametrized classical energy term for all intermolecular interactions. In ref. [22] one assumed a dominant BSSE contribution from the intermolecular oxygen-hydrogen interaction in methanol. This assumption is no longer applicable for the set of liquids under study. Therefore, in this work *all* the intermolecular interaction terms are taken into account in the parametrisation of the correction model for BSSE.

The following section outlines the technical details of the quantum MD simulations and the computational methodology to calculate the RDF, SF and BSSE. Section III reviews shortly the BSSE correction model and its extension, i.e. its mathematical form and the parameter calibration. The BSSE-correction model with calibrated parameters is then applied to a new series of NpT MD-simulations. Section IV presents the results of the BSSE computations, MD-runs, RDFs and SFs with the necessary discussion. Where experimental data is available, comparison with the experiment is made. In the last section the most relevant conclusions are summarized.

II. COMPUTATIONAL SECTION

A. Ab initio MD

The same computational settings are used as in ref. [22] apart from some convergence criteria. All simulations are performed with the CP2K code.[31] For the ab initio MD, the Quickstep module of CP2K[32] is used, employing the hybrid Gaussian and plane-wave (GPW) density functional method with a BLYP gradient-corrected functional. The core electron states are represented by the norm conserving Goedecker-Teter-Hutter (GTH) pseudopotential.[5, 33, 34] The GPW basis set consists of a triple zeta TZVP Gaussian type orbital basis for the real space representation and an auxiliary plane wave basis to compute

TABLE I. Experimental densities (at 298 K) and cell parameters for the NVT simulations for all studied solvents.

solvent	ρ_{exp} [g/cm^3]	cell parameter [Å]
MeOH	0.787[d]	20.583
TCM	1.48[b,c]	25.865
MeCN	0.777[a]	22.445
THF	0.881[e]	20.882
EtOH	0.785[d]	23.248

[a]ref. [36],[b]ref. [37] [c]ref. [38],[d]ref. [39],[e]ref. [40]

the long-range periodic electrostatic interactions in the reciprocal space.[14] It is known from the work of McGrath *et al.* [15] that the convergence of the volume requires a significantly higher auxiliary basis set cut-off compared to constant-volume simulations.[15, 35]. Therefore, the NpT and NVT MD runs are performed with the 800 Ry plane wave cut-off. An exception is the NVT simulation of methanol, where we use the same settings of ref. [22] with a 400 Ry cut-off (with Grimme-dispersion, without correction for BSSE).

Liquid MeOH, TCM, MeCN, and EtOH are modeled at room temperature in a periodic cubic simulation cell containing 129 molecules. Periodic boundary conditions are employed using the minimum image convention. Given the size of THF molecules, the unit cell for this molecular system contains 67 molecules. The simulation cell sizes for the NVT-simulations are chosen to correspond with the experimental densities ρ_{exp} of the solvents, as summarized in Table I. The NVT/DFT-D3 MD production runs of MeOH, TCM, MeCN, THF and EtOH amount to 8.1 ps. The NpT/DFT-D3 and NpT/DFT-D3 + CP production runs are carried out for at least 12.5 – 15.0 ps and 16.0 – 19.9 ps respectively. The exact lengths of the MD runs done for this work are given in Table S.1 of the supporting information (SI). The simulations are carried out at a temperature of 300 K with a Nosé thermostat[41] with time constant 1 ps. The isobaric-isothermal ensemble is implemented in CP2K making use of the Martyna-Tuckerman-Tobias-Klein equations algorithm[42, 43], and computing the stress tensor is implemented as described by Schmidt *et al.*[25] The external pressure is 1 bar and the time constant of the barostat is 1 ps. The Verlet integrator is used with an integration time step of 1 fs.

Once calibrated parameters are obtained for the classical BSSE correction energy term (see Section III B), a new MD simulation is performed with a combination of the Quickstep and the molecular mechanics module of CP2K (FIST). The total energy is calculated at each timestep as the sum of the DFT-energy and the classical force-field energy.

As such we perform three types of MD simulations per solvent:

- NVT/DFT-D3: dispersion-corrected DFT using the DFT-D3 scheme in the NVT-ensemble
- NpT/DFT-D3: dispersion-corrected DFT using the DFT-D3 scheme in the NpT-ensemble
- NpT/DFT-D3 + CP: dispersion- and BSSE-corrected DFT using the DFT-D3 scheme and the counterpoise (CP) correction scheme, as explained in Section II C

To reduce the amount of equilibration time, three measures have been taken: (i) all initial structures are generated by Packmol[44] (a software package especially developed for building initial configurations for MD simulations), (ii) all MD runs are preceded by equilibration NVT-simulations, (iii) the positions at the start of the production runs of the NpT/DFT-D3 simulations are taken as the starting geometry for the NpT/DFT-D3 + CP runs.

B. Radial distribution function and structure factor

The RDF is the number of particle pairs $\mathrm{d}n(r)$ in the spherical shell with radius r and $r + \mathrm{d}r$, relative to the number of particle pairs $\mathrm{d}n_0(r)$ in an uncorrelated ideal gas:

$$g(r) = \frac{\mathrm{d}n(r)}{\mathrm{d}n_0(r)}, \tag{1}$$

where r is the interparticle distance. In practice, the space around a given atom is discretized in concentric spherical shells with width δr. The number of atoms n_i in each shell is counted during the MD run, such that a histogram of the n_i values can be built.

The RDF and SF are correlated through a three-dimensional Fourier transform[45]. If $g_{\alpha\beta}(r)$ is the partial pair correlation function of atom types α and β, then the partial SF

$S_{\alpha\beta}$ may be expressed as:

$$\begin{aligned} S_{\alpha\beta}(k) - 1 &= \rho \int \exp(i\boldsymbol{k}\cdot\boldsymbol{r})(g_{\alpha\beta}(r)-1)d\boldsymbol{r} \\ &= 4\pi\rho \int_0^\infty \mathrm{d}r r^2 \frac{\sin(kr)}{kr}(g_{\alpha\beta}(r)-1), \end{aligned} \tag{2}$$

Here, ρ is the average particle number density, while in the second equality the angle dependence is integrated out in the understanding that $g_{\alpha\beta}(r)$ is isotropic in space. The total structure factor $S_{\mathrm{tot}}(k)$ is a weighted sum of the partial SFs, which for neutron diffraction reads as:

$$S_{\mathrm{tot}}(k) = \sum_{\alpha\leq\beta} (2-\delta_{\alpha\beta})\, c_\alpha b_\alpha c_\beta b_\beta \left[S_{\alpha\beta}(k)-1\right]. \tag{3}$$

k gets a physical significance in the sense that it represents the momentum transfer in a neutron scattering experiment. The weighting factors are composed of the coherent scattering lengths b_α which characterize the interaction strength between atoms and incident radiation and which can vary from isotope to isotope, and the atomic fractions $c_\alpha = N_\alpha/N$ with N_α the number of type α atoms, and N the total number of atoms in the system. The Kronecker $\delta_{\alpha\beta}$ avoids double counting of pairs of atoms of the same type. $S_{\mathrm{tot}}(k)$ is a direct experimental observable and with the use of Eqs. (2) and (3) directly accessible by theoretical simulations, in particular molecular dynamics. All other quantities, including the partial SFs, are not directly available from experiment. In an indirect way, however, they may be derived from the measured total structure factors obtained by changing the isotopic composition of the sample [18], but still the procedure is ill-conditioned.
By means of an inverse Fourier transform,

$$\begin{aligned} f(r) &= \frac{1}{\rho}\left(\frac{1}{2\pi}\right)^3 \int \exp(-i\boldsymbol{k}\cdot\boldsymbol{r}) S_{\mathrm{tot}}(k) d\boldsymbol{k} \\ &= \frac{1}{\rho}\left(\frac{1}{2\pi^2}\right) \int_0^\infty k^2 \mathrm{d}k \frac{\sin(kr)}{kr} S_{\mathrm{tot}}(k) \\ &= \sum_{\alpha\leq\beta} (2-\delta_{\alpha\beta}) c_\alpha b_\alpha c_\beta b_\beta (g_{\alpha\beta}(r)-1), \end{aligned} \tag{4}$$

$S_{\mathrm{tot}}(k)$ is converted to some total pair correlation function $f(r)$, which has no real physical significance but whose last expression in Eqs. (4) can serve as a test for the numerical integrations needed for the computation of the various Fourier transforms. Note that $\lim_{r\to\infty} f(r) = 0$ since $\lim_{r\to\infty} g_{\alpha\beta}(r) = 1$.

Experimental total structure factors are not free from contributions arising from intramolecular scattering (neutrons can scatter from atoms belonging to the same molecule). On the contrary, they even dominate the entire behavior of the measured total SF. From physical point of view, in unraveling the liquid structure, intramolecular contributions should be projected out as only intermolecular terms are of physical interest. This is from experimental viewpoint not an easy task. Theory can help in estimating the impact of intramolecular parts on the various quantities, and in particular those which are accessible by experiment. Within this context, we present a schematic model wherein the intramolecular contribution to the partial pair correlation function $g_{\alpha\beta}(r)$ can be modeled as a sum of one or more Dirac delta functions:

$$g_{\alpha\beta}(r) = g_{\alpha\beta}^{inter}(r) + \sum_i A_{\alpha\beta,i}\delta(r - R_{\alpha\beta,i}) \tag{5}$$

It is not excluded that inside a molecule multiple intramolecular distances are present between atom types α and β; they are discriminated by means of the index i. Eq. (5) is valid for rigid liquid molecules, but may be regarded as a realistic approximation in molecules like THF where one may expect nearly constant intra-atomic distances $R_{\alpha\beta,i}$ in the same molecule during the simulations. A better approximation is the use of an alternative representation for the Dirac delta function with some width or standard deviation for the interatomic distance in the molecule. These "finite width" effects on the structure factors will not differ much from the limiting case where the distribution goes over to the Dirac delta function. Summarizing, the intramolecular contributions to the partial structure factors read as:

$$S_{\alpha\beta}^{\text{intra}}(k) = 4\pi\rho \sum_i A_{\alpha\beta,i} R_{\alpha\beta,i}^2 \frac{\sin(kR_{\alpha\beta,i})}{kR_{\alpha\beta,i}} \tag{6}$$

Their behavior is determined by the sinc function or a superposition of sinc functions, with a large central peak at $k = 0$ and damping oscillation terms with a period of $2\pi/R_{\alpha\beta,i}$. These aspects will be further discussed in the section IV.

C. Counterpoise correction for BSSE

The hybrid Gaussian and plane-wave DFT method uses localized Gaussian basis sets that are centered at the nuclei of the atoms, of which relatively few basis functions are required to describe molecular orbitals. A downside of the method however is the BSSE that arises due to this limited number of basis functions, and thus incomplete basis sets.

The most widely employed method to correct for BSSE is the counterpoise (CP) method as introduced by Boys and Bernardi.[46] If a complex AB is considered, they estimated the artificial stabilization for monomer A as the energy shift ΔE_{A} when the basis for chemical system A is augmented with basis set functions of monomer B (which are located at the monomer B position), i.e. if A approaches monomer B close enough so that A can utilize the basis functions of B, and similarly for ΔE_{B}.[47] These estimated errors are then subtracted from the uncorrected interaction energy $\Delta E^{int}(\mathrm{AB})$ of a complex AB to obtain the *counterpoise-corrected* interaction energy:

$$\Delta E_{\mathrm{CP}}^{\mathrm{int}}(\mathrm{AB}) = \Delta E^{\mathrm{int}}(\mathrm{AB}) - \Delta E_{\mathrm{A}} - \Delta E_{\mathrm{B}} = \Delta E^{\mathrm{int}}(\mathrm{AB}) + \Delta E_{\mathrm{CP}}, \tag{7}$$

where ΔE_{CP} is the so-called counterpoise correction. The CP-computations are carried out on pairs of molecules.

III. EXTENDED CORRECTION CP MODEL

In this section, we develop the CP correction model for ΔE_{CP} and the calibration method for its parameters. The model of ref. [22] is briefly reviewed and extensions, new in this work, are clarified.

A. CP model

The proposed CP model is an additive pairwise interaction between certain atom types in two different molecules[22]. With each pair of atom type t, a unique set of parameters (A_t, B_t) is associated:

$$\Delta E_{\mathrm{CP}}^{\mathrm{mod}} = \sum_{\mu=1}^{N-1} \sum_{\nu=\mu+1}^{N} \sum_{\substack{i \,\epsilon\, M_\nu \\ j \,\epsilon\, M_\mu}} A_{t(ij)} \exp\left(-B_{t(ij)} r_{ij}\right), \tag{8}$$

where r_{ij} is the distance between the atoms i and j, M_ν is a set of atomic indices of molecule ν, μ and ν are molecule indices and N is the total number of molecules. The functional form of each term is an exponential, decaying with the separation distance r_{ij} of the atoms and is necessarily of repulsive nature ($A_t > 0$). To reduce the number of parameters, the $A_{t(ij)}$-parameters in this paper are computed according to the mixing-rule:

$$A_{t(ij)} = \sqrt{A_{t(ii)} \times A_{t(jj)}}, \tag{9}$$

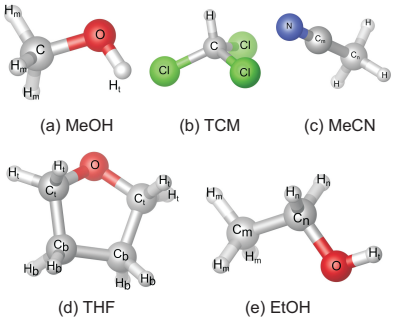


FIG. 1. Atom types for the solvents MeOH, TCM, MeCN, THF, EtOH.

such that only the $A_{t(ii)}$ parameters need to be determined (section III B). The parameter $B_{t(ij)}$ is here calculated as

$$B_{t(ij)} = \frac{1}{\tau(R_i + R_j)}, \tag{10}$$

where R_i and R_j are the VdW-radii of the atoms that make up the atom-pair $t(ij)$. We assume that τ is a general dimensionless factor independent of the atom types. The R_i values are taken from ref. [48].

Fig. 1 gives the nomenclature of the atoms types of the molecules as used in this work. Each pair of atom types is denoted following these labels throughout the remainder of the paper.

B. Parameter calibration

The training data is obtained from CP computations on snapshots from the NpT/DFT-D3 MD simulations (Section II C), and is used to fit the parameters of the model (Eq. (8)). The procedure, schematically presented in Fig. 2, consists of the following steps:

(i) A snapshot of the liquid structure is selected every 800 steps (= 800 fs) of the NpT/DFT-D3 MD simulation, corresponding to a time greater than the velocity auto

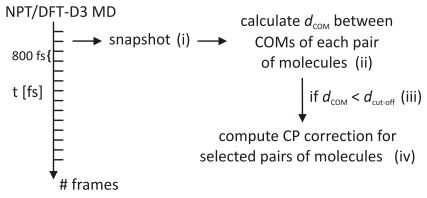


FIG. 2. Flowscheme for obtaining training data from MD simulations.

correlation time of the center of mass (COM) of the molecules to ensure statistically independent samples. The correlation times are computed with the MD-TRACKS program[49, 50], and are all below 270 fs (see Table S.2 of the SI). In this way a respectable number of frames ($15 - 20$) of the MD simulation are selected. The exact number depends on the type of the solvent.

(ii) At each snapshot, the distance d_{COM} between the COMs of every possible pair of molecules is computed with the minimum image convention.

(iii) The pairs of molecules for which this distance is smaller than a cut-off distance $d_{\text{cut-off}}$, are selected for the computation of training data.

(iv) We then calculate the CP-correction $\Delta E_{\mathrm{CP}}^{\mathrm{ref}}$ with CP2K for each of the selected pairs, which serves as training data.

The same specifications (basis set, electronic structure method, 800 Ry cut-off, ...) are used for the CP-computations as those for the MD simulations of the solvents (Section II A). The box size has been chosen such that there is a margin of minimal 5 Å between the dimer and the edge of the simulation box. The sizes of the unit cells for the five solvents are given in Table S.3 of the SI.

We use the COM RDF as a criterium to determine a proper $d_{\text{cut-off}}$ for selecting molecule pairs for the training CP-computations, i.e. $d_{\text{cut-off}}$ is chosen as the distance that contains the whole first peak of the COM RDF determined from the NpT/DFT-D3 simulations. Using this approach, it is ensured that at least the first layer of intermolecular interactions is represented in the training set. The COM RDF of methanol is shown in Fig. 3 as an example. Based on this RDF the $d_{\text{cut-off}}$ for methanol is set at 6.2 Å. Following this criterium,

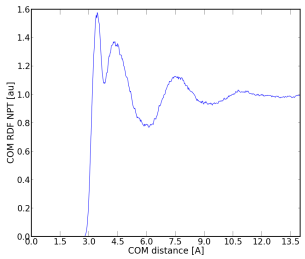


FIG. 3. The COM RDF of methanol derived from the NpT/DFT-D3 MD simulation.

the $d_{\text{cut-off}}$ of 4.0 Å chosen in ref. [22] is improved. This latter value for $d_{\text{cut-off}}$ resulted in 3850 molecule pairs, while the higher cut-off distance in this paper results in a total number of 15714 pairs. The $d_{\text{cut-off}}$ values and the COM RDFs for the other solvents are given in the SI (Table S.3 and Fig. S.1, respectively).

The parameters (A_i, τ) in Eq. (8) are estimated from the training data set of CP-computations ($\Delta E_{\text{CP}}^{\text{ref}}$) by a non-linear least-squares fit. In this paper, the CMA-evolution strategy [51] (CMA-ES) is applied to minimize the cost function. CMA is a population-based stochastic search algorithm for parameter optimization. It neither approximates nor uses gradients, making it an efficient optimization method for problems where derivative-based methods may fail due to a rugged search landscape presenting multiple discontinuities, sharp bends, noise and local optima.

The considered cost function is the root-mean-square deviation (RMSD) between the natural logarithms of the model $\Delta E_{\text{CP}}^{\text{mod}}$ and the training $\Delta E_{\text{CP}}^{\text{ref}}$ corrections, and is further discussed in the SI.

Once average calibrated parameters (A_i, τ) are obtained, a new MD simulation is performed with the Quickstep and FIST modules of CP2K. The energy in the MD run is computed as the DFT-energy (Quickstep module) plus the CP model term (Eq. (8)) (FIST module).

IV. RESULTS AND DISCUSSION

A. CP-computations on NpT/DFT-D3 MD

From the procedure as outlined in Section III B, 16 – 19 snapshots from the NpT/DFT-D3 MD simulations are used to select 7982 – 15111 molecule pairs, with values for $d_{\text{cut-off}}$ ranging between 6.2 and 7.9 Å. CP-computations with CP2K on these selected pairs serve as training data for the five considered liquids. To avoid negative parameters, $(\ln A_i, \ln \tau)$ are varied as parameters in the algorithm and to assess the impact of the stochastic factor inherent to the CMA-algorithm, the procedure is conducted 200 times for each solvent. This results in 200 estimates for each of the parameters. For instance, MeOH has 4 parameters ($\ln \tau$, $\ln A_{\text{H-H}}$, $\ln A_{\text{C-C}}$, $\ln A_{\text{O-O}}$) that are varied in the CMA algorithm, and other parameters (B parameters, other A parameters) are calculated with Eq. (9) and Eq. (10). The final parameters are then averages over the 200 parameter estimates, and are shown in Table II.

TABLE II: Final A_i, τ and B_i parameters for MeOH, TCM, MeCN, THF and EtOH.

solvent	interaction	A [kJ/mol]	τ [/]	B [Å^{-1}]
MeOH	H-H	3.1689	0.1796	1.3514
	C-C	0.0779		0.8416
	O-O	7.9499		0.6626
	H-C	0.4968		1.0372
	H-O	5.0192		1.1116
	C-O	0.7869		0.8899
TCM	H-H	0.0491	0.1949	1.2455
	C-C	0.3867		0.7757
	Cl-Cl	1.6284		0.7802
	H-C	0.1377		0.9560
	H-Cl	0.2827		0.9594
	C-Cl	0.7935		0.7780
MeCN	H-H	0.3212	0.2387	1.0169

Continued on next page

TABLE II – *Continued from previous page*

solvent	interaction	A [kJ/mol]	τ [/]	B [Å^{-1}]
	C-C	0.0084		0.6334
	N-N	0.6938		0.6884
	H-C	0.0519		0.7805
	H-N	0.4720		0.8210
	C-N	0.0763		0.6597
THF	H-H	0.8028	0.1927	1.2596
	C-C	0.4101		0.7846
	O-O	3.9314		0.8801
	H-C	0.5737		0.9669
	H-O	1.7764		1.0362
	C-O	1.2696		0.8296
EtOH	H-H	1.7398	0.1757	1.3816
	C-C	0.7422		0.8605
	O-O	9.0476		0.9653
	H-C	1.1363		1.0606
	H-O	3.9674		1.366
	C-O	2.5913		0.9099

Fig. 4 shows the correlation between training $\Delta E_{\mathrm{CP}}^{\mathrm{ref}}$ and model $\Delta E_{\mathrm{CP}}^{\mathrm{mod}}$ data for methanol. The CP model-energies are computed with Eq. (8), where the final parameters (average over 200 parameter estimates, Table II) are inserted. The plots for the other four solvents look similar (Fig. S.2 of the SI).

The condition number of the normalized sampling covariance matrix (see Table S.4 of the SI) is high for all solvents, implying that the calibrated parameters are not transferable. One indeed notices no significant trend in the final parameters (A_i,τ,B_i) in Table II. The maximum values for the CP-correction energy ΔE_{CP} for the solvents are: MeOH: 4.5 kJ/mol, TCM: 1.1 kJ/mol, MeCN: 0.4 kJ/mol, THF: 3.0 kJ/mol, EtOH: 4.5 kJ/mol. These results

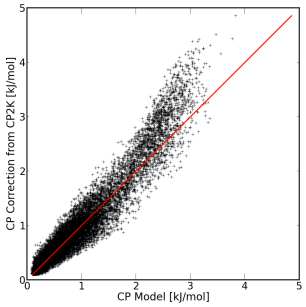


FIG. 4. CP-energies computed with CP2K ($\Delta E_{\mathrm{CP}}^{\mathrm{ref}}$) versus model CP-energies ($\Delta E_{\mathrm{CP}}^{\mathrm{mod}}$) in methanol.

are confirmed by relaxed PES-scans on a complex consisting of two dimers of each considered solvent. The training CP-corrections for the solvents MeOH , EtOH and THF are relatively large. Those for TCM and MeCN are small.

B. Molecular dynamics

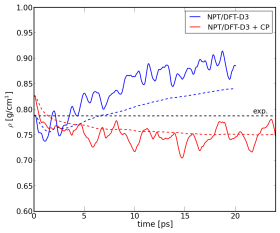
This work encompasses in total 15 MD simulations (NVT, NpT and NpT + CP for five solvents). Production run trajectories have been carried out and their duration depends on the type of the solvent but especially on the type of the dynamics. NpT dynamics require run times of at least 20 ps or even more. The instantaneous densities of the five considered solvents found in our NpT MD simulations, including the equilibration run, are presented in Fig. 5.

It should be stressed that NpT dynamics of molecular liquids within the DFT methodology poses much more problems and has a much larger complexity than standard NVT. Their outcome largely depends on the accuracy of the DFT interaction potentials. It is well-known that the convergence of the pressure requires a significantly higher basis set cutoff than normally used for standard NVT MD simulations. Another problem is that the equilibration

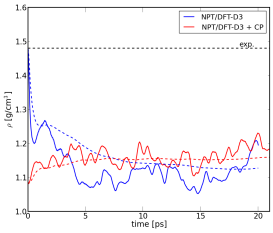
time for NpT dynamics can be significantly larger, as the volume fluctuations are typically low frequency. A comprehensive DFT study on the performance of two popular gradient-corrected exchange correlation functionals on the structure and density of liquid water is presented in [25] in the NpT ensemble. Both PBE and BLYP functionals underestimate the density by about 25 % and 12 % respectively. Adding Grimme corrections of the type D1 [26] the authors of ref. [25] found a significant improvement of the densities. Our results are in line with the findings of the previous work on water. The densities computed from the NpT runs for the five liquids under study show large fluctuations. When plotting the running average (average over all previous steps) a more fluent behavior is obtained (dotted line in Figure 5), as in ref. [25]. These curves have a tendency to converge but they mask the fluctuations which are still present. Even when enlarging the simulation times ($>$ 20 ps) the fluctuations will not disappear, are even not damping and the standard convergence criteria will probably never be reached. The results for the five solvents are different. Overall, the agreement with experiment becomes better when including the BSSE-CP corrections. This is remarkably the case in methanol and ethanol. In only one case – chloroform – the predicted density is largely underestimated. In the absence of hydrogen bonds the remaining intermolecular forces – which are weaker – become now decisive in determining the density of the liquid. However, they are prone to the interaction potentials, involved in the DFT functionals by construction, and therefore one could expect a large dependence of the predicted density of liquid chloroform on the employed density functional. Within this context we also examine the geometry of each separate liquid molecule (the monomer). When averaging the C-H and C-Cl bond lengths of all chloroform molecules in the snapshots taken during the NpT simulations, an average bond length of 1.82 Å is found for C-Cl and 1.09 Å for C-H. A geometry optimization of the monomer in gas phase within the B3LYP/311+g(d,p) level of theory (LOT), yields 1.79 Å and 1.08 Å respectively. The average "volume" taken by a single chloroform molecule (pyramidal structure) in the simulation is thus 7.5 % larger than that in gas phase. It partly explains the underestimation by 20 % for the density measured in the NpT simulation with BLYP. It also implies that in this specific case of chloroform the BLYP functional is not adequate enough to predict the right geometry and the remaining weak intermolecular forces in the absence of hydrogen bond interactions. The interplay of all these small effects has direct consequences in the reproduction of the correct density.

For the convenience of the reader, the average densities ρ_{av} for the five liquids under study

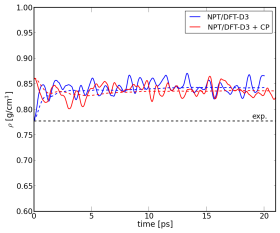
FIG. 5. Evolution of the density as a function of time for the NpT/DFT-D3 and NpT/DFT-D3 + CP MD simulations. The production run starts at a time indicated by an arrow. The horizontal lines give the experimental density, which has been imposed in the NVT/DFT-D3 runs.



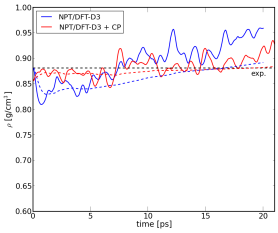
(a)MeOH



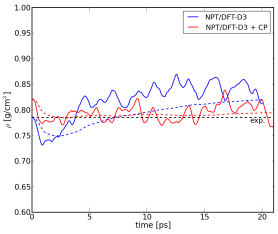
(b)TCM



(c)MeCN



(d)THF



(e)EtOH

TABLE III. Average densities ρ_{av} for the NpT/DFT-D3 and NpT/DFT-D3 + CP MD simulations, compared to experiment (NVT/DFT-D3, see Table I), for the five considered solvents. The values between brackets show the deviation compared to experiment.

solvent	ρ_{av} [g/cm^3]		
	exp NVT/DFT-D3	NpT/DFT-D3	NpT/DFT-D3 + CP
MeOH	0.787	0.872 (0.085)	0.746 (-0.041)
TCM	1.479	1.107 (-0.372)	1.162 (-0.317)
MeCN	0.777	0.845 (0.068)	0.839 (0.062)
THF	0.881	0.909 (0.028)	0.889 (0.008)
EtOH	0.785	0.837 (0.052)	0.795 (0.01)

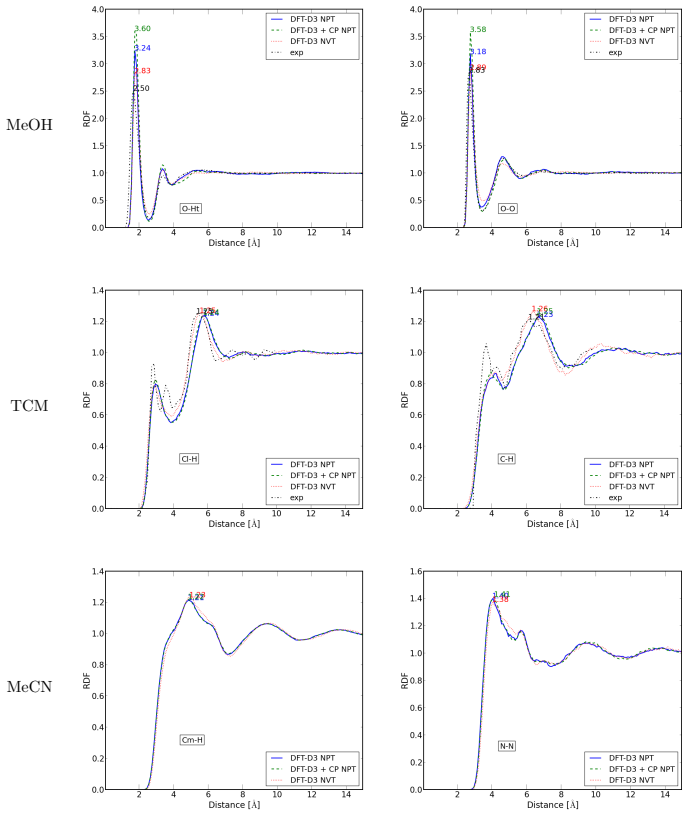
are tabulated in Table III. These averages are taken over the production runs only, starting from 5 ps unless otherwise stated (arrow in Fig. 5).

C. Radial distribution functions

A good reproduction of the liquid density does not necessarily lead to a correct description of the structure of the liquid. This is best illustrated in the case of water[25] where PBE has the tendency toward overstructuring, whereas the obtained density in NpT dynamics is closer to the experimental value than with BLYP.

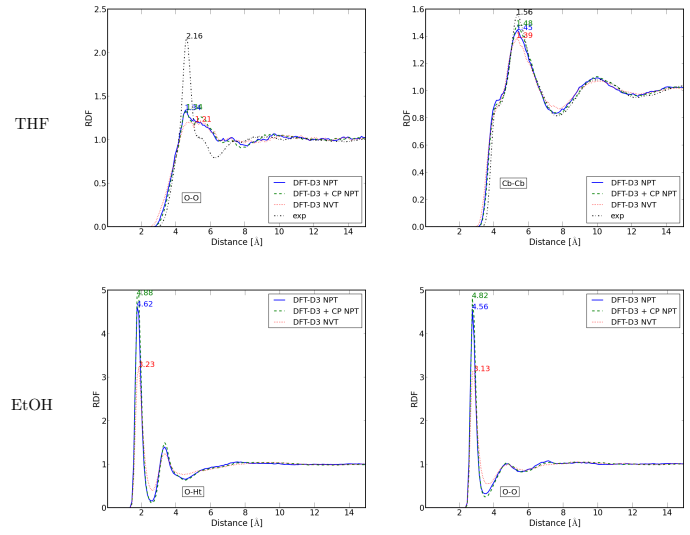
We calculated the intermolecular partial radial distribution functions for all five liquids from our NpT simulations. We only display a selection of them in Fig. 6; a complete list of the RDFs is given in the SI (Figure S.3-S.7). The most characteristic features of the RDFs, determining the structure of the liquid, are given by the height and position of the first peak. They are tabulated in Table IV. Structure properties found in constant volume NVT simulations are also taken up, where the density is set to the experimental value. They offer additional information in the discussion.

FIG. 6. Radial distribution functions. Experimental data from refs. [52, 53] (MeOH), ref. [54] (TCM), and ref. [18] (THF).



Continued on next page

FIG. 6 – *Continued from previous page*



(i) NpT simulations clearly give rise to an overstructured **methanol** liquid. The heights for the first oxygen-hydrogen (O-H_t) and oxygen-oxygen (O-O) peak are largely overestimated. In contrast to NVT simulations [22], the BSSE-CP corrections even strengthen the discrepancy. We have no explanation for it, as the BSSE switches the density to a slightly underestimated value (0.75), but closer to the experimental value of 0.79 (see Fig. 5). On the other hand in similar work on water [25], it is observed that an underestimation of the density is generally accompanied by a locally overstructured liquid. In many other studies in the literature [7, 8, 22, 55–59] on water theory systematically overestimates the first intermolecular peak. Experiments indeed do not measure pair interactions directly, as indicated in the introduction. The

various manipulations in extracting "experimental" partial RDFs are not free from inaccuracies what we will discuss in detail in the remainder of this paper. We should take care in drawing final conclusions.

(ii) Despite the serious underestimation of the density in **chloroform**, we now observe a lower first peak than experiment in most of the partial RDFs (see also Fig. S.4 of the SI). BSSE corrections on the RDFs are small. The experiment reveals a second peak at around 3.1 Å in the partial RDF of Cl-H, but this peak is in no way reproduced by the simulations. Similar observations are noticed in other partial RDFs, and in particular in C-H and H-H. These prominent peaks are completely missing in our simulations and other MD studies [60–62] are in line with our results. RMC calculations performed by Pothoczki *et al.*[54] without additional constraints imposed by diffraction data also reveal no additional peak in the region 3.5 Å – 4.0 Å.

(iii) No experimental data are available for **acetonitrile**. All simulations predict the same RDFs. BSSE corrections are negligible.

(iv) The RDFs of **ethanol** show large similarities with those of methanol. The first peak raises enormously in height when switching from NVT to NpT, and BSSE makes it even stronger.

(v) The situation is dramatic in **THF** where the experiment predicts sharp prominent peaks at short distances, and which in no way are reproduced by any theoretical molecular dynamics simulation, neither in this work nor elsewhere in literature. This strongly deviant behavior, also noticed in the case of TCM, requires special attention and a serious in depth discussion to reconcile theory and experiment. As the primary observable is the total structure factor, the discussion should be held in reciprocal space. What characteristic in the (partial) structure factor in momentum space can cause such a prominent peak at short distances in coordinate space? This should be unraveled in the next subsection.

Table IV shows the amplitude and position of the first peaks of the RDFs from Fig. 6.

TABLE IV. Amplitudes and positions of the first intermolecular peak of the RDFs shown in Fig. 6. The positions are expressed in Å.

solvent	interaction	property	exp	NpT/DFT-D3	NpT/DFT-D3 + CP	NVT/DFT-D3
MeOH	O-H_t	amplitude	2.50	3.24	3.60	2.83
		position	1.68	1.75	1.75	1.75
	O-O	amplitude	2.83	3.18	3.58	2.89
		position	2.67	2.75	2.75	2.75
TCM	Cl-H	amplitude	0.93/0.79/1.25	0.79/1.24	0.83/1.24	0.79/1.25
		position	2.87/3.80/5.36	2.95/5.75	2.95/5.75	2.85/5.55
	C-H	amplitude	1.06/0.90/1.21	0.87/1.23	0.87/1.25	0.87/1.26
		position	3.68/4.02/6.15	4.25/6.65	3.95/6.65	4.25/6.35
MeCN	C_m-H	amplitude		1.21	1.22	1.23
		position		4.85	4.85	4.95
	N-N	amplitude		1.40	1.41	1.38
		position		4.05	4.15	4.05
THF	O-O	amplitude	2.16	1.34	1.34	1.21
		position	4.62	4.55	4.65	5.15
	C_b-C_b	amplitude	1.56	1.45	1.48	1.39
		position	5.34	5.35	5.35	5.35
EtOH	O-H_t	amplitude		4.62	4.88	3.23
		position		1.75	1.75	1.85
	O-O	amplitude		4.56	4.82	3.13
		position		2.75	2.75	2.85

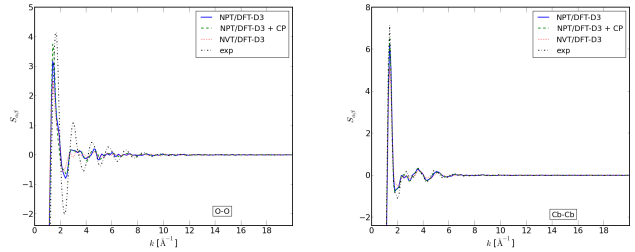
D. Structure factors of THF

Intermolecular partial structure factors $S_{\alpha\beta}(k)$ are computed from the partial RDFs by a Fourier transform as expressed in Eq. (2). This transformation from r-space to k-space poses no problem as each RDF is theoretically a well-behaved function. In the reciprocal

space all structure factors show the behavior of a typical spectral function with damped oscillations at increasing k. They are displayed in Fig. S.8 of the SI. We focus on the partial $S_{\text{O-O}}$ structure factor to explore the possible ingredients lying on the origin of the strong peak at 4.6 Å in real space. For completeness, the factor ρ representing the average particle number density $\langle N/V \rangle$ in Eq. (2) is computed over the whole MD production run and amounts to 0.0987 Å^{-3}, 0.0962 Å^{-3}, 0.0956 Å^{-3} for the NpT/DFT-D3, NpT/DFT-D3 + CP and NVT/DFT-D3 simulations respectively. The last number naturally coincides with the experimental estimate. The theoretical prediction of $S_{\text{O-O}}$ is displayed in Fig. 7 for the three models under consideration. Also displayed is the Fourier transform resulting from the experimental RDF $g_{\text{O-O}}(r)$ (with the strong first peak). We realize that this property has been constructed after a lot of detours, and thus not merit the label "exp" in the strict sense, but we could not dispose of the experimentally deduced partial $S_{\text{O-O}}(k)$ structure factor, which has been used by Soper *et al.*[18] to extract the corresponding RDF after applying an inverse Fourier transform. A first source of errors lies in the sequence of Fourier and inverse Fourier transforms. Final conclusions cannot be drawn from Fig. 7: the indirectly derived experimental structure factor shows a quasi-perfect damped harmonic oscillator behavior, while the theoretical curves are dominated by anharmonicities. In the paper of Bowron *et al.*[18] only total structure factors are reported as measured in neutron scattering experiments, and not partial structure factors. Since we dispose of all theoretical partial RDFs belonging to THF, we are able to compute all partial structure factors by performing a Fourier transform following Eq. (2). A weighted summation of these structure factors, as given in Eq. (3), yields the theoretical prediction of the total structure factor, which can directly be compared with the experimentally measured $S_{\text{exp}}(k)$. Both curves are compared in Fig. 8 for THF. The coherent neutron scattering lengths are tabulated in Table V (values taken from ref. [63]).

A new ingredient enters the discussion: the discrimination between intermolecular and intramolecular contributions to the total structure factor. Theoretically this poses no problem, while experimentally separation of the intramolecular contributions is not a trivial task and requires adapted experimental procedures such like neutron diffraction measurements with isotopic substitution (NDIS) as applied in ref. [18]. However, for complex systems, with more than two atom types, it is generally not possible to perform enough isotope variation experiments to extract all the partial structure factors required to completely characterize

FIG. 7. Partial SFs $S_{O\text{-}O}(k)$ and $S_{C_b\text{-}C_b}(k)$ for THF. The dashed-dotted line labeled with "exp" is not derived directly from experiment, but computed from the empirically derived RDF [18].



(a)$S_{O\text{-}O}$ (b)$S_{C_b\text{-}C_b}$

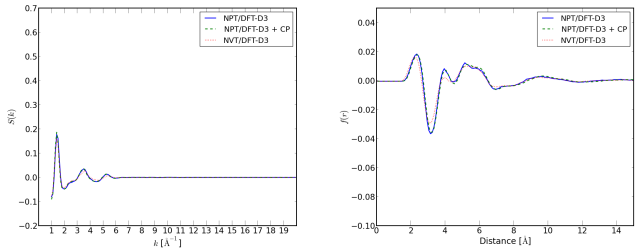
TABLE V. Coherent neutron scattering lengths[63].

isotope	b_α [fm]
C	6.646
O	5.803
H	-3.739
D	6.671

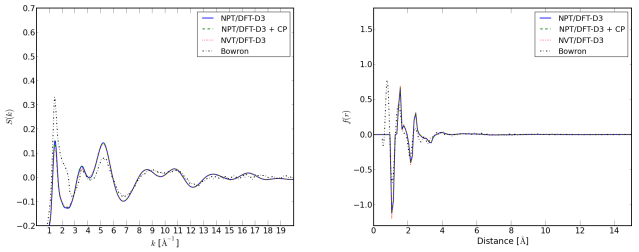
the system. Experimentalists introduce a three-dimensional model of the structure of the disordered liquid and, by means of RMC or EPSR, structures [17, 64] are modeled that are finally consistent with the available experimental data. In this particular case of THF only three measured structure factors for C_4H_8O, C_4D_8O and the 1:1 mixture [18] are at the disposal of the experimentalists. It remains an ill-conditioned routine and there are multiple possibilities of interatomic potentials which are able to generate structures yielding RDFs and SFs which succeed in reproducing the available total SFs. In addition the separation of the total SF into inter- and intramolecular contributions remains a challenging and almost unfeasible task.

To illustrate the difference of the total structure factor between the two samples, we include a similar figure for the deuterium sample (Fig. 9). There are indeed some fundamental

FIG. 8. Total structure factor for THF C_4H_8O obtained as a weighted sum of the partial SFs limited to intermolecular contributions (upper left) and to inter- and intramolecular contributions (bottom left). Also given are the corresponding total pair correlation functions $f(r)$ as defined in Eq. (4) (right column). Units are barn/sr/atom.



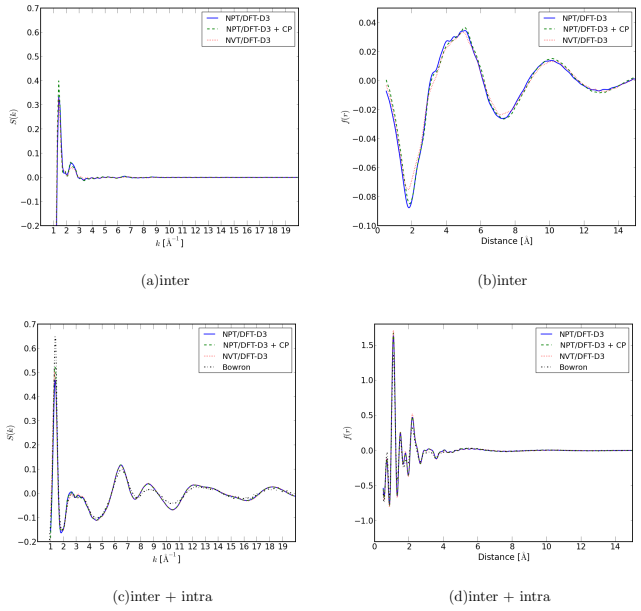
(a)inter (b)inter



(c)inter + intra (d)inter + intra

differences between the two liquids. Experimentalists make use of it to split the total S_{tot} into the partial structure factors. The observed differences are contributed to the negative scattering length of the proton (Table V). Note, however, that the experimentally measured total structure factors, displayed in the Figs. 8 and 9, are those as resulting directly from the diffraction experiments including both intra- en intermolecular contributions. We don't dispose of these spectra without intramolecular contributions. The agreement with

FIG. 9. Total structure factor for C_4D_8O obtained as a weighted sum of the partial SFs limited to intermolecular contributions (upper left) and to inter- and intramolecular contributions (bottom left). Also given are the corresponding total pair correlation functions $f(r)$ as defined in Eq. (4) (right column). Units are barn/sr/atom.



(a)inter (b)inter

(c)inter + intra (d)inter + intra

the theoretical predictions seems to be better for C_4D_8O. The peak positions and the minima are correctly reproduced. The amplitude of the first peak in $S^{\mathrm{exp}}_{\mathrm{tot}}$ amounts to 0.65 barn/sr/atom, much higher than the height predicted by the theory. It is interesting to note that BSSE corrections succeed in slightly increasing the peak height in better agreement with experiment.

In C_4H_8O the theoretical simulations predict the first peak at the correct position, but its

amplitude is barely half the experimental estimate. More important is the appearance of a shoulder structure at around $2.0\,\text{Å}^{-1}$ in $S_{\text{tot}}^{\text{exp}}$ which is not present in our MD simulations. It should be emphasized that the total structure factor including intra- and intermolecular contributions is the only experimental quantity which is measured directly without any model dependence and thus without any bias, and that the spectra displayed in the bottom left of Figs. 8 and 9 are the sole cases where theory can be directly compared with experiment. In principle, an assessment of the theoretical results should be based solely on this comparison. To estimate the impact of this shoulder structure in $S_{\text{tot}}^{\text{exp}}$ on other derived properties, we perform the inverse Fourier transform to evaluate the total pair correlation function $f(r)$ (Eq. 4) which we display in the bottom right of Fig. 8. A prominent peak raises up at a distance lower than 1 Å. As this peak is completely missing in the deuterated spectra, it is evident that this "ghost" peak at 1 Å is correlated with the shoulder structure in $S_{\text{tot}}^{\text{exp}}$. The key issue arising at present is now whether this shoulder structure arises from pure intramolecular interactions. The generation of a peak in $f(r)$ at intramolecular distances fully supports this conclusion.

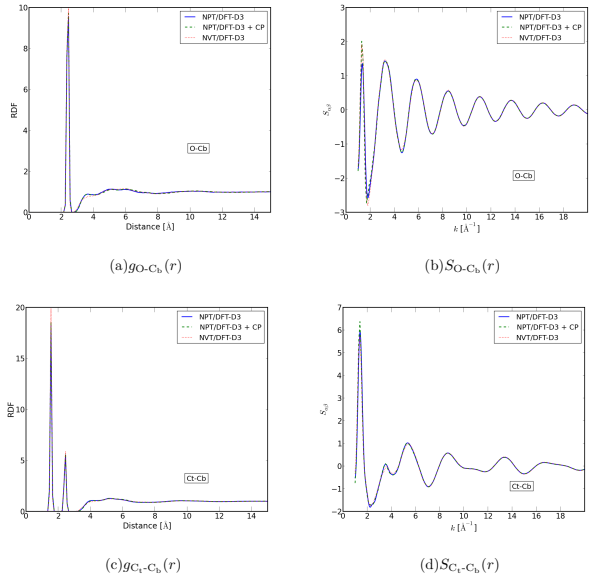
The influence of subtracting intramolecular contributions from the total structure factor is immense. The typical diffraction patterns fade out at relatively low values for the scattering vector ($k = 6.0\,\text{Å}^{-1}$ for C_4H_8O and $k = 3.0\,\text{Å}^{-1}$ for C_4H_8O). The major fluctuations at larger k-values are completely disappeared, and may be fully attributed to intramolecular contributions. This is conform the schematic model we presented in section II B where the strong peaks at short distances (≈ 1 Å) in the partial pair correlation functions $g_{\alpha\beta}(r)$ (we refer to Fig. 10 and Fig. S.11 of the SI) give rise to large damping oscillations in the whole k-region ranging from $1\,\text{Å}^{-1}$ to more than $20\,\text{Å}^{-1}$. We don't dispose of an experimental estimate of the total structure factor where the intramolecular contributions have been subtracted as its behavior has not been taken up in the paper of Bowron *et al.* [18].

A second observation after inspecting the inter and inter + intra S_{tot} spectra is that the position of the first peak is maintained at about $k = 1.5\,\text{Å}^{-1}$ and also its strength is mainly preserved, but this does not imply that intramolecular interactions do not contribute to this low-k peak. This forms an important and prominent feature with serious consequences, and is fully supported when comparing S_{tot} for both the cases. These considerations make clear that from experimental point of view the subtraction of the dominant intramolecular fluctuations from the total spectrum is not a trivial task to perform with the highest accuracy.

The experimental data are incomplete to extract a correct separation. The EPSR method, as developed by Soper[2, 17, 64], assists in disentangling the various partial structure factors decomposed in inter and intra terms, by means of an iterative refinement procedure of the used empirical interatomic potentials. However, EPSR explores a broad range of structural models that are consistent with the available scattering data. The method does not ensure an error free and unique partial structure factor. In addition the functional form of the potential can bias the model, and may prevent certain atomic configurations. Another feature strengthening our statement that the extraction of "experimental" derived properties is prone to large inaccuracies, is the observation that total SFs are very insensitive to changes in partial intermolecular RDFs. A small relative error in the computed intramolecular SF will not be visible from a relative large error in the intermolecular SF. Such small errors in the intramolecular SF are surely present, which can be noticed in Fig. 8(c) and Fig. 9(c), from the difference between theoretical and experimental SF at large k. This is best illustrated in Fig. 7 where the partial structure factors $S_{\text{O-O}}$ and $S_{\text{C}_\text{b}\text{-C}_\text{b}}$ are hardly affected by the type of the MD simulation, while the corresponding RDFs exhibit clear structural differences (Fig. 6). This explicitly shows that the derivation of partial intermolecular RDFs from a small number of structure factors is ill-conditioned: there are multiple possibilities for the RDF that match a corresponding SF. It is an artefact that is well-known by the experimentalists doing neutron scattering to reveal the structure of liquids, but which has not yet been emphasized so strongly as in the case of THF. An inverse Fourier transform of the the total structure factor leads to the total pair correlation function $f(r)$ (Eq. (4)). Although this quantity has no real physical significance, it gives an indication how sensitive the correlation actually is between a function in reciprocal and coordinate space. The peaks of the intermolecular $f(r)$ are one order of magnitude smaller than those of the combined intra- and intermolecular pair correlation function (Fig. 8). Also given is the $f(r)$ reported by Bowron *et al.*[18], which nicely agree with ours apart from a large peak at 0.8 Å as already discussed before. Its behavior is completely determined by the fluctuations in the large k-region of the total structure factor dominated by the intramolecular contributions. It is not likely that termination effects could alter this conclusion.

The same findings are found when investigating the Fourier transforms of the inter- and intramolecular partial RDFs, as given in Fig. 10, Fig. S.9 and Fig. S.10 of the SI, and comparing them with the Fourier transforms of the intermolecular partial RDFs (Fig. 8 and

FIG. 10. Total (intra + inter) partial RDFs $g_{\mathrm{O\text{-}C_b}}(r)$ and $g_{\mathrm{C_t\text{-}C_b}}(r)$ (left column) in THF. Also given are the corresponding partial structure factors (right column).



(a) $g_{\mathrm{O\text{-}C_b}}(r)$ (b) $S_{\mathrm{O\text{-}C_b}}(r)$

(c) $g_{\mathrm{C_t\text{-}C_b}}(r)$ (d) $S_{\mathrm{C_t\text{-}C_b}}(r)$

Fig. S.8). The figures confirm manifestly an intramolecular contribution at the first peak at small k values. This is not really surprising. The typical high peaks at very short distances in the RDFs can be easily regarded as Dirac delta peaks generating a behaviour like a sinc curve $(\sin(x)/x)$ as proposed in section II B. In the presence of one single peak the partial structure factor clearly shows a one-component damped oscillation with a central peak at about $1\,\mathrm{Å}^{-1}$. Concluding, the theoretical simulations give evidence that the extraction of intermolecular partial RDFs and even partial structure factors from experimental data should be regarded as extremely error-prone.

A similar profound analysis can be performed on chloroform where visible spurious spikes

are also observed in the region of short hydrogen-hydrogen distances (below 3.5 Å).[54] This maximum is reproduced neither by any of our MD simulations nor the RMC study of Pothoczki.[54] Despite H-H partial structure factors only contain intermolecular terms, the appearance of the spurious peak is very probably the result of some noise on the measured total scattering structure factor for $CHCl_3$ and $CDCl_3$, and the smallest perturbation may give rise to serious changes of the radial distributions in real space, as clearly demonstrated in the case of THF.

V. CONCLUSIONS

This work comprises the study of the structure of five liquids : methanol, chloroform, acetonitrile, tetrahydrofuran and ethanol by means of DFT-based molecular dynamics simulations in both a NVT and NpT ensemble. First the influence of BSSE on the various liquid properties has been investigated. For this purpose the BSSE correction scheme as proposed in [22] has been extended to all intermolecular interactions. MD simulations in a NpT ensemble show that they have a beneficial effect on the reproduction of the density in all five liquids under study. In only one case – chloroform – a serious underestimation of the density is observed. We ascribe this deficiency to the inappropriate description of the employed DFT functional BLYP-D3 in describing the intermolecular interactions in the absence of hydrogen bonds, as was confirmed by Schmidt *et al.* [25] in NpT dynamics on water. Second, properties have been computed from the simulations for a structural determination of the five liquids. In first instance all partial RDFs for the three methods under consideration (NpT/DFT-D3, NpT/DFT-D3+CP and NVT/DFT-D3) have been calculated, and compared with experiment if available. BSSE corrections turn out to overstructure the liquid in NpT simulations, but for the majority of the liquids under study they remain minor and of no real importance in the further discussion on the overall performance of the RDFs. In the liquids TCM and THF – where hydrogen bonds are not occurring – the experimental agreement is dramatic. Peaks appear at short distances, which are completely missing in any MD simulation. As experimental partial RDFs are not directly measurable but derived after an inverse Fourier transform of the partial structure factors, which in turn are determined from an iterative protocol combining neutron scattering experiments and MD runs with empirical interatomic potentials, these properties are not free from cumulating inaccuracies

generated in the various steps in the procedure. A fair assessment can only be performed in reciprocal space, where the total structure factor, as a direct experimental observable, can be compared with theory, as do the empirical RMC and EPSR methods. However, the separation of the intermolecular contribution from the more pronounced intramolecular terms, that would allow the characterization of the liquid structure, has proven to be difficult. It is well known that intramolecular interactions largely dominate the total structure factors. This is completely confirmed by our theoretical results, but what has not yet been reported in literature is that this dominance is complete in the sense that the intramolecular terms cover the whole momentum range, including the low k region where we expected that it is mainly controlled by intermolecular contributions. Even the first large peak in the spectrum is composed of both contributions. Another aspect revealing from our computation, is that the signal from intermolecular scattering fades out quite fast with increasing momentum transfer. These facts suggest that the first and probably the most crucial source for generating errors is the extraction of the intermolecular total structure factor from the total measured scattered amplitude S_{tot}. The isotopic substitution technique is very powerful, but if the isotopic composition of the sample can only be changed by replacing hydrogens by deuteriums, the number of diffraction measurements generating independent information remains still small. To overcome this incompleteness, EPSR-refined models are constructed determining all the partial structure factors, of which a weighted sum leads to the $S_{\mathrm{tot}}^{\mathrm{inter}}$. What we really need is an unbiased model-independent intermolecular total structure factor without the help of interatomic potentials. The lack of this quantity hinders a fully reliable comparison with theory. In the work of Bowron *et al.*[18] no information on $S_{\mathrm{tot}}^{\mathrm{inter}}$ is given. In the present case we only dispose of an experimental estimate of the total structure factor including both intra and intermolecular contributions. Our MD simulations succeed in a very satisfactory reproduction of S_{tot} in the deuterated sample. On the contrary the proton sample displays a shoulder at about $2\,\mathrm{Å}^{-1}$, which is not present in the MD simulations but which causes a prominent peak at $0.8\,\mathrm{Å}$ in the total pair correlation function $f(r)$. This could be an indication that the shoulder structure in the structure factor originates from intramolecular interactions. If this structure is not filtered out in the derivation of the intermolecular total structure factor, it affects the partial SFs with undesirable consequences in the RDFs. The experimentally derived partial structure factor $S_{\mathrm{O\text{-}O}}$ in THF (Fig. 7) also displays large damping oscillations in the low k region, which are signatures of an insufficient

filtering of the intramolecular contributions, resulting to spurious and prominent peaks in the RDFs.

ACKNOWLEDGMENTS

This work is supported by the Fund for Scientific Research - Flanders (FWO), the Research Board of Ghent University (BOF) and BELSPO in the frame of IAP/6/27, the Belgian Prodex office and ESA. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by Ghent University, the Hercules Foundation and the Flemish Government department EWI. VVS acknowledges the European Research Council under the European Community's Seventh Framework Programme (FP7(2007-2013) ERC grant agreement number 240483). AG and TV are post-doctoral researchers funded by the Foundation of Scientific Research - Flanders.

[1] G. N. I. Clark, C. D. Cappa, J. D. Smith, R. J. Saykally, and T. Head-Gordon. *Mol. Phys.*, 108(11):1415, 2010.

[2] A. K. Soper. *ISRN Physical Chemistry*, 2013:279463, 2013.

[3] A. K. Soper. *J. Phys.: Condens. Matter*, 9(13):2717, 1997.

[4] A. K. Soper. *Chem. Phys.*, 258(2-3):121, 2000.

[5] J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello. *J. Chem. Phys.*, 122(1):014545, 2005.

[6] T. Todorova, A. P. Seitsonen, J. Hutter, I-F. W. Kuo, and C. J. Mundy. *J. Phys. Chem. B*, 110(8):3685, 2006.

[7] I-C. Lin, A. P. Seitsonen, M. D. Coutinho-Neto, I. Tavernelli, and U. Rothlisberger. *J. Phys. Chem. B*, 113(4):1127, 2009.

[8] M. J. McGrath, I.-F. W. Kuo, and J. I. Siepmann. *Phys. Chem. Chem. Phys.*, 13(44):19943, 2011.

[9] V. K. Yadav and A. Chandra. *Chem. Phys.*, 415:1, 2013.

[10] A. Vrhovšek, O. Gereben, S. Pothoczki, M. Tomšič, A. Jamnik, S. Kohara, and L. Pusztai. *J. Phys.: Condens. Matter*, 22(40):404214, 2010.

[11] A. Vrhovšek, O. Gereben, A. Jamnik, and L. Pusztai. *J. Phys. Chem. B*, 115(46):13473, 2011.

[12] L. Zoranić, F. Sokolić, and A. Perera. *J. Chem. Phys.*, 127(2):024502, 2007.

[13] P. E. Mason, G. W. Neilson, C. E. Dempsey, D. L. Price, M.-L. Saboungi, and J. W. Brady. *J. Phys. Chem. B*, 114(16):5412, 2010.

[14] G. Lippert, J. Hutter, P. Ballone, and M. Parrinello. *J. Phys. Chem.*, 100(15):6231–6235, 1996.

[15] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, C. J. Mundy, J. VandeVondele, J. Hutter, F. Mohamed, and M. Krack. *J. Phys. Chem. A*, 110(2):640, 2006.

[16] R. L. McGreevy. *J. Phys.: Condens. Matter*, 13(46):R877, 2001.

[17] A. K. Soper. *Phys. Rev. B*, 72(10):104204, 2005.

[18] D. T. Bowron, J. L. Finney, and A. K. Soper. *J. Am. Chem. Soc.*, 128(15):5119, 2006.

[19] R. L. McGreevy and L. Pusztai. *Mol. Simulat.*, 1(6):359, 1988.

[20] J. L. Thomas, D. J. Tobias, and A. D. MacKerell Jr. *J. Phys. Chem. B*, 111(45):12941, 2007.

[21] T. D. Kühne, M. Krack, and M. Parrinello. *J. Chem. Theory Comput.*, 5(2):235, 2009.

[22] M. Van Houteghem, T. Verstraelen, A. Ghysels, L. Vanduyfhuys, M. Waroquier, and V. Van Speybroeck. *J. Chem. Phys.*, 137(10):104506, 2012.

[23] S. Kristyán and P. Pulay. *Chem. Phys. Lett.*, 229(3):175–180, 1994.

[24] P. Hobza, J. Šponer, and T. Reschel. *J. Comput. Chem.*, 16(11):1315, 1995.

[25] J. Schmidt, J. VandeVondele, I.-F. William Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy. *J. Phys. Chem. B*, 113(35):11959, 2009.

[26] S. Grimme. *J. Comput. Chem.*, 25(12):1463, 2004.

[27] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. *J. Chem. Phys.*, 132(15):154104, 2010.

[28] A. D. Becke. *Phys. Rev. A*, 38(6):3098, 1988.

[29] C. Lee, W. Yang, and R. G. Parr. *Phys. Rev. B*, 37(2):785, 1988.

[30] J. P. Perdew, K. Burke, and M. Ernzerhof. *Phys. Rev. Lett.*, 77(18):3865, 1996.

[31] CP2K, 2011. `http://cp2k.berlios.de`.

[32] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. *Comput. Phys. Commun.*, 167(2):103, 2005.

[33] S. Goedecker, M. Teter, and J. Hutter. *Phys. Rev. B*, 54(3):1703, 1996.

[34] C. Harwigsen, S. Goedecker, and J. Hutter. *Phys. Rev. B*, 58(7):3641, 1998.

[35] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, C. J. Mundy, J. VandeVondele, J. Hutter, F. Mohamed, and M. Krack. *Chem. Phys. Chem.*, 6(9):1894, 2005.

[36] G. P. Cunningham, G. A. Vidulich, and R. L. Kay. *J. Chem. Eng. Data*, 12(3):336, 1967.

[37] J. A. Riddick, W. B. Bunger, and T. K. Sakano. *Organic Solvents: Physical Properties and Methods of Purification*. Wiley-Interscience, third edition, 1970.

[38] C. L. Yaws. *Thermodynamic and Physical Property Data*. Golf Publishing, 1992.

[39] J. Gao, D. Habibollazadeh, and L. Shao. *J. Phys. Chem.*, 99(44):16460, 1995.

[40] F. Franks, M. A. J. Quickenden, D. S. Reid, and B. Watson. *Trans. Faraday Soc.*, 66:582, 1970.

[41] S. Nosé. *J. Chem. Phys.*, 81(1):511, 1984.

[42] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein. *Mol. Phys.*, 87(5):1117, 1996.

[43] M. E. Tuckerman, J. Alejandre, R. López-Rendón, A. L. Jochim, and G. J. Martyna. *J. Phys. A: Math. Gen.*, 39(19):5629, 2006.

[44] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. *J. Comput. Chem.*, 30(13):2157, 2009.

[45] D. McQuarrie. *Statistical mechanics*. Harper & Row, 1976.

[46] S. F. Boys and F. Bernardi. *Mol. Phys.*, 19(4):553, 1970.

[47] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt - van de Rijdt, and J. H. van Lenthe. *Chem. Rev.*, 94(7):1873, 1994.

[48] R. S. Rowland and R. Taylor. *J. Phys. Chem.*, 100(18):7384, 1996.

[49] T. Verstraelen, M. Van Houteghem, V. Van Speybroeck, and M. Waroquier. *J. Chem. Inf. Mod.*, 48(12):2414, 2008.

[50] Center for Molecular Modeling, 2013. `http://molmod.ugent.be/software`.

[51] N. Hansen and A. Ostermeier. *Evol. Comput.*, 9(2):159, 2001.

[52] T. Yamaguchi, K. Hidaka, and A. Soper. *Mol. Phys.*, 96(8):1159, 1999.

[53] T. Yamaguchi, K. Hidaka, and A. Soper. *Mol. Phys.*, 97(4):603, 1999.

[54] S. Pothoczki, L. Temleitner, S. Kohara, P. Jóvári, and L. Pusztai. *J. Phys.: Condens. Matter*, 22(40):404211, 2010.

[55] J.-W. Handgraaf, E. J. Meijer, and M.-P. Gaigeot. *J. Chem. Phys.*, 121(20):10111, 2004.

[56] M. E. Martín, M. L. Sánchez, F. J. del Valle, and M. A. Aguilar. *J. Chem. Phys.*, 116(4):1613, 2002.

[57] T. Ishiyama, V. V. Sokolov, and A. Morita. *J. Chem. Phys.*, 134(2):024509, 2011.

[58] H. Yu, D. P. Geerke, H. Liu, and W. F. Van Gunsteren. *J. Comput. Chem.*, 27(13):1494, 2006.

[59] S. Patel and C. L. Brooks. *J. Chem. Phys.*, 122(2):024508, 2005.

[60] W. Dietz and K. Heinzinger. *Ber. Bunsenges. Phys. Chem.*, 89(9):968, 1985.

[61] H. J. Böhm and R. Ahlrichs. *Mol. Phys.*, 54(6):1261, 1985.

[62] H. Torii. *J. Mol. Liq.*, 119(1–3):31, 2005.

[63] V. F. Sears. *Neutron News*, 3(3):26, 1992.

[64] A. K. Soper. *Chem. Phys.*, 202(2–3):295, 1996.

Paper IV

MD-TRACKS: A Productive Solution for the Advanced Analysis of Molecular Dynamics and Monte Carlo simulations

T. Verstraelen, M. Van Houteghem, V. Van Speybroeck, M. Waroquier

Journal of Chemical Information and Modeling, 48 (12), 2414-2424, 2008

2414 J. Chem. Inf. Model. 2008, 48, 2414–2424

MD-TRACKS: A Productive Solution for the Advanced Analysis of Molecular Dynamics and Monte Carlo simulations

Toon Verstraelen, Marc Van Houteghem, Veronique Van Speybroeck, and Michel Waroquier*

Center for Molecular Modeling, Ghent University, Proeftuinstraat 86, 9000 Ghent, Belgium

Received July 11, 2008

In this paper, we present MD-TRACKS, an advanced statistical analysis toolkit for Molecular Dynamics and Monte Carlo simulations. The program is compatible with different molecular simulation codes, and the analysis results can be loaded into spreadsheet software and plotting tools. The analysis is performed with commands that operate on a binary trajectory database. These commands process not only plain trajectory data but also the output of other MD-TRACKS commands, which enables complex analysis work flows that are easily programmed in shell scripts. The applicability, capabilities, and ease of use of MD-TRACKS are illustrated by means of examples, that is, the construction of vibrational spectra and radial distribution functions from a molecular dynamics run is discussed in the case of tetrahydrofuran. These properties are compared with the experimental data available in the literature. MD-TRACKS is open-source software distributed at http://molmod.ugent.be/code/.

1. INTRODUCTION

Molecular dynamics (MD) and Monte Carlo (MC) simulations are well-established modeling techniques in diverse research fields, ranging from catalysis over biochemistry to solid-state physics.[1−5] Historically, most MD/MC simulation software relied on molecular mechanics models to cope with the cost of computing the potential energy and the interatomic forces.[6] However, during the past decade, the development of specialized algorithms and continuously progressing computer technology have made ab initio molecular dynamics[7−11] a viable and attractive alternative to the conventional molecular mechanics methods. Despite the ever-increasing computing power, molecular mechanics remains relevant for its scalability toward large systems, longer time scales, and the applications in hybrid QM/MM methods.[12] Today, a myriad of MD/MC simulation software is available (CPMD,[7,13] CP2K,[8,14] LAMMPS,[15,16] DL_POLY,[17] CHARMM,[18] NAMD,[19] GULP,[20] GROMACS,[21] GROMOS,[22] CERIUS2,[23]...) that enables one to generate a vast amount of trajectory data by integrating the equations of motion of the system under study. The trajectory data does not only contain the time-dependent atomic coordinates, but also many other quantities as a function of time, for example, atomic velocities, forces, cell parameters, pressure, temperature, partial charges, dipole moments, polarizability, different types of energies, orbitals, and so on. To translate the raw trajectory data into relevant scientific results, a proper statistical analysis is indispensable.

Several trajectory analysis solutions are readily available, for example, Visual Molecular Dynamics[24] (VMD), GROMACS,[21] PTRAJ,[25] SIRIUS,[26] and SMILYS.[27] They are certainly of high quality, and some of them offer impressive visualization functions for molecular dynamics simulations. Despite the presence of these valuable analysis tools, we felt the need for a more generic software program that enables complex analysis that goes beyond the standard functions present in the current tools. Ideally, a trajectory analysis program should also be flexible enough to be adapted for many different applications. In this paper, we present MD-TRACKS, a versatile, user-friendly, and freely available toolkit that addresses this challenge. We have tested the existing analysis software and prototypes of MD-TRACKS in foregoing studies[28−30] which has strongly influenced the software design. MD-TRACKS has some distinctive characteristics that make it suitable for a wide range of applications:

(1) MD-TRACKS is compatible with multiple MD/MC simulation codes. Prior to the actual analysis, the trajectory data is converted into a simple, manageable, fast, and cross-platform binary database. The current version of MD-TRACKS (0.003) has an interface to CP2K,[14] CPMD,[13] LAMMPS,[15,16] DL_POLY,[17] and CERIUS2,[23] and our software is easily extended to process trajectory data from other simulation codes.

(2) An analysis task is solved with a series of consecutive MD-TRACKS commands in a solution work flow. Each command loads only the actively used parts of the database in memory to analyze huge amounts of data without memory limitations.

(3) MD-TRACKS commands are orthogonal, that is, the output of each command is written to the database and can be used as input for any other command. MD-TRACKS has a built-in plotting function, but it is also possible to convert the analysis results from the binary database into plain text format that is supported by most spreadsheet applications and plotting tools.

(4) An MD-TRACKS programming library is provided to create custom Python scripts that can access the binary database. Once parts of the database are loaded in memory, efficient numerical operations are possible through NumPy.[31] If one can perform a very specific analysis task only partially

* To whom correspondence should be addressed. E-mail: michel.waroquier@ugent.be.

10.1021/ci800233y CCC: $40.75
Published on Web 11/18/2008

Table 1. Overview of the Most Important Modules in the MD-TRACKS Programming Library

module	functionality
`tracks.core`	low-level classes and auxiliary functions that efficiently read from or write to the binary MD-TRACKS database
`tracks.convert`	routines that convert the output of molecular dynamics and Monte Carlo programs into the binary format
`tracks.parse`	a set of functions that facilitate the interpretation of command line arguments
`tracks.vector tracks.cell`	tools for the manipulation of a collection of track files that represent time-dependent three-dimensional vectors or matrices

with the current MD-TRACKS commands, one can easily implement the remainder of the analysis in a specialized script using the MD-TRACKS programming library.

This paper describes both the implementation and the usage of MD-TRACKS. It is assumed that the reader has a basic knowledge of UNIX systems. The following section discusses the database format, the structure of a typical MD-TRACKS command, and an overview of the commands in the current MD-TRACKS version (0.003). In the third section, we give an impression of the capabilities of MD-TRACKS by showing how different types of vibrational spectra and radial distribution functions are easily computed from conventional trajectory data. In section four, we describe how MD-TRACKS can be obtained.

2. IMPLEMENTATION

2.1. MD-TRACKS Database Format. The MD-TRACKS toolkit stores all trajectory data and analysis output in a subdirectory `tracks` of the working directory where the molecular dynamics or Monte Carlo simulation software has written its output files. The commands that start with `tr-from-` convert the simulation output into binary files in the tracks directory. For example, `tr-from-cpmd-traj` converts a CPMD[13] trajectory file into binary data.

Each file in the directory `tracks` corresponds to a single time-dependent scalar, for example, `tracks/atom.pos.0000000.y` contains the values of the *Y* coordinate of the first atom for all iterations of the MD/MC simulation. This approach has the advantage that the database can be managed with simple UNIX commands like `ls`, `cp`, `rm`, and `mv`.

A single file in the database is called *a track*. Each track file consists of a fixed length header, followed by a binary data stream. The header contains a unique fingerprint and a format description of the binary data stream. The information in the header makes it possible to interpret the binary stream correctly on all computer architectures. The binary format has two major advantages, compared to conventional text files: (i) binary data is more efficient, both in terms of disk space and input/output performance, and (ii) binary data support random access, that is, one can read and write in track files at arbitrary positions without any overhead, while text files have to be read or written line by line.

MD-TRACKS follows these database-related conventions that facilitate the analysis process:

• All values in the MD-TRACKS database are stored in atomic units. The output files from MD/MC simulations are converted into atomic units by the `tr-from-*` scripts. Each MD-TRACKS command that produces human readable output, will present numerical data in the units specified by the user. All unit conversions are based on the CODATA basic constants and conversion factors.[32]

• Filenames in the MD-TRACKS database are reserved for a specific purpose. The most important reserved names are given below.

◦ `tracks/time`: This track contains the time axis, that is, the time in the simulated molecular system at each iteration step. In most cases, this is the integration time step multiplied by the step counter.

◦ `tracks/potential_energy`: This track stores the potential energies felt by the nuclei. It includes all interatomic interactions, that is, it is not limited to the Coulomb repulsion between the positively charged nuclei. In a molecular mechanics simulation, this also includes the valence interactions, Van Der Waals interactions, and so on. In the case of an ab initio simulation, this file contains the sum of the nuclei−nuclei, nuclei−electron and electron−electron interactions.

◦ `tracks/kinetic_energy`: This track contains the kinetic energy of the nuclei.

◦ `tracks/atom.pos.${index}.${c}`: These tracks hold the atomic positions, where `${index}` identifies the atom by an integer of seven characters and `${c}` is *x*, *y*, or *z*.

◦ `tracks/atom.vel.${index}.${c}`: These tracks contain the atom velocities, using the same conventions as in the previous item.

The proposed filename conventions are not strictly imposed but are strongly recommended when working with MD-TRACKS.

2.2. MD-TRACKS Commands. An overview of the MD-TRACKS commands (in version 0.003) is given in the Appendix. To execute a command, one enters its name at the UNIX command line shell, followed by options and arguments. The names of the commands are chosen to interact well with the tab-completion function in most popular UNIX shells. When the command line option `--help` is given as an argument to an MD-TRACKS command, the documentation for that command is printed on screen.

The MD-TRACKS commands are not meant to be entered manually, unless one explores new types of simulations or analysis techniques. In all other cases, it is much more efficient to collect a series of MD-TRACKS commands in a shell script that can be reused for the analysis of many MD/MC trajectories. This script-based approach automates the analysis process, but one still has the possibility to tune all parameters in this automated procedure.

Each command in the MD-TRACKS toolkit is based on the MD-TRACKS programming library which is a Python package that can be reused to write new MD-TRACKS commands for specialized applications. The most relevant modules in the MD-TRACKS programming library are listed in Table 1.

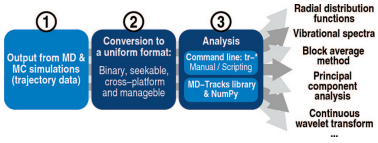


Figure 1. Workflow of an MD-TRACKS application.

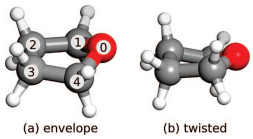


Figure 2. Two stable conformers of the THF molecule. On the basis of the molecular mechanics model for linear and cyclic ethers of Vorobyov et al.,[36] the twisted conformer is 2.5 kJ mol^{-1} lower in energy than the envelope conformer. The pseudorotation phase (as defined by Cremer and Pople[51]), is ±90° for the twisted and 0° or 180° for the envelope conformer.

3. EXAMPLE APPLICATION: VIBRATIONAL AND STRUCTURAL PROPERTIES OF TETRAHYDROFURAN

In this section, we present a selection of the various possibilities of MD-TRACKS by demonstrating how the vibrational and structural properties of tetrahydrofuran (THF) can be derived from a molecular dynamics simulation. This gives an idea of the usage pattern and the advantages of the MD-TRACKS software design. The typical MD-TRACKS workflow is illustrated in Figure 1. THF was chosen as a case study in view of its important role as structural unit in carbohydrates and biological systems and in the context of a general interest in the conformations and the ring-puckering of small-membered rings.[33,34]

A general consensus about the minimum energy structure of THF has not been achieved yet.[35,33] It is not the intention of this paper to resolve the question of the equilibrium geometry. Therefore, we will use a recently developed molecular mechanics force field for linear and cyclic ethers.[36] All conformational energies, vibrational frequencies and other analysis results mentioned below are obtained with this force field. The potential energy surface of THF in this molecular mechanics model has two local minima, corresponding to the twisted and envelope (or bended) conformer,[37,38] illustrated in Figure 2. The envelope conformer is 2.5 kJ mol^{-1} lower in energy. The vibrational frequencies of both conformers based on the harmonic oscillator approximation are listed in Table 2.

3.1. Vibrational Spectra of THF in the Gas Phase. In this example, we will apply MD-TRACKS to compute the infrared and inelastic neutron scattering (INS) spectrum of THF. In addition, we analyze the parts of the spectrum that are inherent to the pseudorotation motion, which is present in many puckered cyclic organic molecules.[39,40] The concept of pseudorotation was originally proposed by Kilpatrick et al. to explain the exceptionally high entropy of cyclopentane.[41] The IUPAC definition defines a psueodortation as a conformational change resulting in a structure that appears to have been produced by rotation of the entire initial molecule and is superimposable on the initial one, unless different positions are distinguished by substitution or isotopic labeling. No angular momentum is generated by this motion; this is the reason for the term. In the case of ring molecules, the conformational changes consist of puckering modes. The pseudorotation of tetrahydrofuran (THF) has been extensively investigated during the past decades, both in experimental and theoretical studies.[33,34,35,37,42,43]

Table 2. Vibrational Frequencies of the two THF Conformers (in cm^{-1}) Based on the Harmonic Oscillator Approximation Applied to the Molecular Mechanics Model for Linear and Cyclic Ethers by Vorobyov et al[36]

envelope	twisted
262	252
573	558
574	585
756	790
827	846
856	860
936	936
950	950
982	971
1052	1049
1067	1080
1120	1135
1166	1163
1193	1183
1287	1286
1321	1319
1334	1331
1350	1354
1451	1450
1465	1472
1472	1489
1487	1514
1580	1571
1595	1606
2851	2850
2851	2851
2885	2884
2888	2887
2897	2897
2903	2905
2920	2923
2931	2932

We performed an NVE molecular dynamics simulation of a single THF molecule at an average temperature of 300 Kelvin in the gas phase. The simulation time is 1 ns, and the integration time step is 1 fs. The simulation has been carried out with CP2K.[14] In the following paragraphs, we give step-by-step instructions for the analysis of the molecular dynamics simulation. As the analysis progresses, we unravel the relation between the harmonic frequencies of the two conformers, the vibrational density of states from the molecular dynamics simulation and the experimentally measured infrared spectrum.

The remainder of the text contains a transcript from the command line terminal and can be used by a potential user to reproduce this specific example. The output of the commands is hidden or reduced for reasons of clarity. All commands are preceded by a ">" sign and are printed in bold. The screen output generated by the MD-TRACKS commands is printed using a normal font weight. Long

commands are split over multiple lines because of the limited column width, but in practice, they should be entered without line breaks.

3.1.1. Setup of the MD-TRACKS Database. The output of the molecular dynamics simulation, as generated by CP2K (development version of June 27, 2008), is distributed over several files. The file `md-1.ener` contains the elementary energy terms as a function of time. The files `md-pos-1.xyz` and `md-vel-1.xyz` contain the atomic coordinates and velocities at each integration time step. The file `md-MM_DIPOLE-1.data` contains the time derivative of the dipole moment. The general output file md.out will not be used in this example. We load the relevant files in the binary MD-TRACKS database with the commands `tr-from-cp2k-ener`, `tr-from-xyz`, and `tr-from-txt`. If an MD-TRACKS database does not yet exist, it is automatically created. The conversion is done as follows. (The CP2K input files are not shown.)

```
> ls
md-1.ener
md-MM_DIPOLE-1.data
md-pos-1.xyz
md-vel-1.xyz
md.out
> tr-from-cp2k-ener md-1.ener
> tr-from-xyz md-pos-1.xyz pos
> tr-from-xyz md-vel-1.xyz vel -u au
> ls
...
tracks/
> ls tracks/
atom.pos.0000000.x
atom.pos.0000000.y
atom.pos.0000000.z
atom.pos.0000001.x
...
atom.pos.0000012.z
atom.vel.0000000.x
...
atom.vel.0000012.z
conserved_quantity
kinetic_energy
potential_energy
time
```

The command `tr-from-xyz` has two mandatory arguments: the XYZ trajectory filename and a suffix (e.g., pos or vel) that is used to generate the filenames in the binary tracks database. By default `tr-from-xyz` assumes that the text based XYZ file contains atom coordinates in angstroms. When the velocity file is converted, the option `-u au` is used to indicate that the XYZ file contains data in atomic units.

In the remainder of this example, we also need the time derivative of the dipole moment generated by the THF molecule. The conversion of the file `md-MM_DIPOLE-1.data` demonstrates how the generic `tr-from-txt` command converts ASCII data to the binary format when a specific `tr-from-*` command is not available. The records of interest in the file `md-MM_DIPOLE-1.data` have the following format (one line):

```
MM DIPOLE [NON-PERIODIC] DERIVATIVE
(A.U.)| 0.000226 −0.000289 −0.00070
```

The three numbers correspond to the *x*-, *y*-, and *z*-components of the time derivative of the dipole moment in atomic units. The file contains such a record for each time step. We can use the ubiquitous UNIX tools `grep` and `cut` to filter out the data of interest:

```
> grep DERIVATIVE md-MM_DIPOLE-1.data |
  cut -c 50-
0.000226 -0.000289 -0.000070
0.000226 -0.000289 -0.000070
0.000220 -0.000262 -0.000066
0.000186 -0.000186 -0.000060
0.000135 -0.000079 -0.000053
...
```

The `grep` command prints only the records from the file `md-MM_DIPOLE-1.data` that contain the word `DERIVATIVE`. The pipe symbol, `|`, prevents that the output of grep is printed on screen. Instead, the filtered records are redirected as input to the cut command, which discards the first 50 characters from each line. The command `tr-from-txt` reads text data formatted in columns from the standard input and writes this information into the binary database. We can use a second pipe symbol to redirect the output of the cut command to `tr-from-txt`.

```
> grep DERIVATIVE md-MM_DIPOLE-1.data |
  cut -c 50- | tr-from-txt
tracks/dipole.derivt.x
tracks/dipole.derivt.y
tracks/dipole.derivt.z
> ls tracks
...
dipole.derivt.x
dipole.derivt.y
dipole.derivt.z
...
```

So far, all given MD-TRACKS commands were specific for CP2K. From now on, the analysis is completely generic. One can replace the setup of the tracks database by specific commands for another molecular dynamics program and then continue with the instructions below to perform a similar analysis.

3.1.2. Standard Spectral Analysis. The infrared adsorption spectrum can be derived from a molecular dynamics simulation based on linear response theory. The classical approximation of the infrared adsorption spectrum is given by[44,45]

$$\alpha(\nu) \approx \lim_{\tau\to\infty} \frac{1}{\tau} \sum_{j=x,y,z} \left| \int_0^\tau \mathrm{d}t \exp(-i\nu t) \frac{\mathrm{d}\mu_j}{\mathrm{d}t} \right|^2 \qquad (1)$$

where ν is the frequency and μ_j are the Cartesian components of the dipole moment. This expression represents the power spectrum of the time derivative of the dipole moment. The command `tr-spectrum` is a generic tool to compute power spectra based on the numerical FFT algorithm.[46] When the time derivative of the dipole moment is given as input, the infrared spectrum is generated. When the atomic velocities are given as input, the inelastic neutron scattering (INS) spectrum or velocity power spectrum is computed. One could also compute the power spectrum of the time derivative of the polarizability, which leads to the Raman spectrum. The velocity power spectrum can be used as a classical approximation[47−49] of the vibrational density of states.[50]

Note that formula 1 is entirely equivalent to the Fourier transform of the autocorrelation function of the time-derivative of the dipole moment. The following MDTRACKS commands compute the infrared adsorption and the INS spectrum:

```
> tr-spectrum tracks/dipole.derivt.*
  tracks/time tracks/spectrum.ir
  --blocks=250
> ls tracks/
...
spectrum.ir.frequencies
spectrum.ir.wavenumbers
spectrum.ir.amplitudes
...
> tr-spectrum tracks/atom.vel.*
  tracks/time tracks/spectrum.vib
  --blocks=50
> ls tracks/
...
spectrum.vib.frequencies
spectrum.vib.wavenumbers
spectrum.vib.amplitudes
...
```

In the first command line, the three components of the time derivative of the dipole moment are given as input. The time axis is used to create a proper frequency and wavenumber axis. The output is written to files that start with `tracks/spectrum.ir`.

The option `--blocks=250` instructs the `tr-spectrum` command to divide the input data in 250 blocks of the same size. The spectrum is computed for each block and finally the average over all spectra is computed. A higher number of blocks improves the statistical accuracy of the final spectrum, but reduces the resolution on the wavenumber axis. For the infrared spectrum, the resolution on the wavenumber scale (*X*-axis) is 8 cm^{-1}, and the relative statistical error on the infrared adsorption (*Y*-axis) is 8%. The INS spectrum is obtained in a similar way. The latter is obtained with a resolution of 1.5 cm^{-1} and a relative statistical error of 9%. The INS spectrum is less sensitive to statistical noise because it is based on more input data: for each atom, there are three Cartesian velocity components, while the dipole moment has only three components in total.

The output files `tracks/spectrum.ir.frequencies` or `tracks/spectrum.ir.wavenumbers` can be used as the *X*-axis when plotting the spectrum. The corresponding amplitudes of the spectrum are stored in `tracks/spectrum.ir.amplitudes`.

In Figure 3, the INS spectrum is compared with the frequencies obtained within the harmonic oscillator approximation. A strict one-to-one correspondence between harmonic frequencies and peaks in the INS spectrum cannot be made. The main reason is that the molecular dynamics simulation takes into account finite temperature effects, while the harmonic oscillator approximation is only valid close to zero Kelvin. In the remainder of the text, we will show that at room temperature the THF molecule does not oscillate very long in one of the local minima of the potential energy surface. Instead THF continuously alters from the twisted to the envelope state and visits all intermediate structures. It is therefore impossible to assign peaks in the vibrational spectra to one of the conformers. It is clear however that

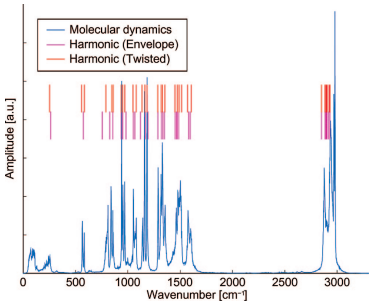


Figure 3. Vibrational frequencies in the THF molecule. The harmonic frequencies of the envelope (purple) and twisted (red) conformer are plotted as vertical lines. The vibrational density of states based on the molecular dynamics simulation is plotted in blue.

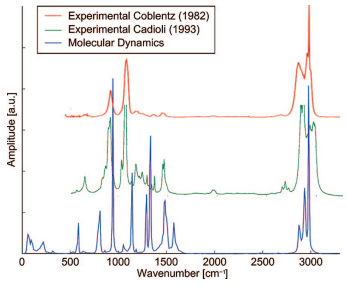


Figure 4. Infrared spectrum. The blue line represents the simulated spectrum. The green[58] and red[34] curves are experimental spectra from the literature.

clusters of harmonic frequencies correlate with clusters of peaks in the spectrum. The correlation fails visibly in two cases:

(1) In the region below 150 cm^{-1}, a band in the INS spectrum appears, which is completely absent in the harmonic oscillator approximation. This band represents a genuine vibration of the THF molecule and can not be attributed to coupling with rotational degrees of freedom. (The angular momentum is set to zero at the beginning of the molecular dynamics simulation and remains negligible.)

(2) The peaks around 3000 cm^{-1} are blue-shifted in the velocity power spectrum when compared to the harmonic frequencies. Figure 4 compares the simulated infrared spectrum with the experimental result. The molecular mechanics model approximates the experimental peak positions, but it fails to predict the infrared activity. In the remainder of this section, we will study the origin of adsorption band below 150 cm^{-1}.

3.1.3. Transitions between the THF Conformers. The distinct conformers of the THF molecule are well characterized by the ring puckering coordinates that can be computed

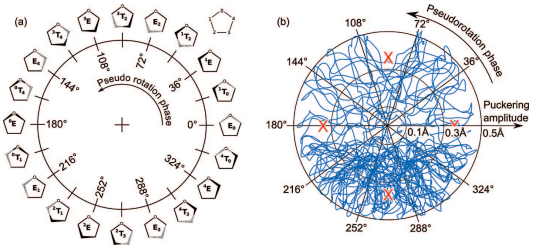


Figure 5. (a) Illustration of the pseudorotation phase, based on Figure 3 in the work of Altona and Sundaralingam.[59] (b) A polar plot of the puckering coordinates during the first 10 ps of the MD simulation. The radius is the puckering amplitude; the angle corresponds to the pseudorotation phase. The coordinates of the stable THF conformers are indicated with crosses.

with `tr-ic-puckering`. Our implementation relies on the general definition of puckering coordinates by Cremer and Pople.[51] In the case of a five-membered ring structure, there are two puckering coordinates: the puckering amplitude, q, which expresses the deviation from the planar ring structure and the pseudorotation phase, ϕ, which discriminates between all possible envelope and twisted geometries (see Figure 5a).

The time evolution of the ring puckering coordinates are computed with the following command:

```
> tr-ic-puckering 5
tracks/atom.pos.0000000
tracks/atom.pos.0000001
tracks/atom.pos.0000002
tracks/atom.pos.0000003
tracks/atom.pos.0000004 tracks/puck.pos
> ls tracks/
...
puck.pos.amplitude.0000002
puck.pos.phase.0000002
...
```

The first argument (`5`) stands for the number of atoms in the ring structure. The following five arguments are file prefixes that correspond to the five ring atoms (in consecutive order). The last argument is a filename prefix for the output files. The number of ring puckering coordinates is N-3 where N is the number of atoms in the ring. Conventionally, these coordinates are labeled with an integer index that starts from two.[51] For an eight-membered ring, one would obtain the following output files:

```
> ls tracks/
...
puck.pos.amplitude.0000002
puck.pos.amplitude.0000003
puck.pos.amplitude.0000004
puck.pos.phase.0000002
puck.pos.phase.0000003
...
```

During the molecular dynamics simulation, the THF molecule passes through all possible envelope and twisted conformers. This is demonstrated in Figure 5b where the time-dependent puckering coordinates during the first 10 ps are plotted as a solid line. The coordinates of the stable conformers are drawn as crosses. From this picture, it is clear the ring puckering motion can not be considered as a harmonic oscillation around a minimum on the potential energy surface. This suggests that the vibrational spectra from the MD simulation will exhibit features that are not present in the frequencies from the harmonic oscillator approximation. In other words, the band below 150 cm^{-1} is most likely related with the pseudorotation of the THF molecule. In the next part of this example application, this correspondence is unambiguously demonstrated.

3.1.4. Peak Assignment in Simulated Vibrational Spectra. For a proper understanding of the relation between the pseudorotation phase and the vibrational spectra, we should compute the contribution of the pseudorotation to the spectrum. A straightforward power spectrum of the time derivative of the pseudorotation phase can be misleading. The unit of the amplitude of this spectrum differs from the original INS spectrum, which disturbs a strict comparison. Alternatively, one can project the Cartesian velocity vector on the tangent of the internal coordinate of interest at each time step. The spectrum of these projected velocities only includes contributions of the motion along the selected internal coordinate and has the same unit as the original INS spectrum. The projected velocities are always smaller than or equal to the original velocities. The total spectrum will therefore be an (approximate) upper limit for the spectrum of the original velocities. When a peak in the *projected* spectrum coincides with its counterpart in the total spectrum, one can assume that there is no motion along other (orthogonal) coordinates that contributes to this peak. Such a peak is then completely resolved.

The following command will compute the ring puckering coordinates of the THF molecule, their time derivatives, and the projection of the Cartesian velocity vector on the tangent of each puckering coordinate:

```
> tr-ic-puckering 5
tracks/atom.pos.0000000
tracks/atom.pos.0000001
tracks/atom.pos.0000002
tracks/atom.pos.0000003
tracks/atom.pos.0000004
tracks/atom.vel.0000000
tracks/atom.vel.0000001
tracks/atom.vel.0000002
tracks/atom.vel.0000003
```

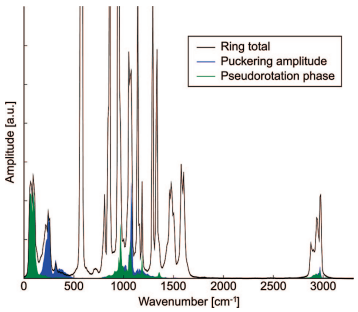


Figure 6. Analysis of the low-frequency bands in the velocity power spectrum of THF. The black line is the total velocity power spectrum of the five ring atoms. The blue curve represents the power spectrum of the velocity vector projected on the tangent of the puckering amplitude coordinate. The green curve is a similar spectrum that corresponds to the pseudorotation phase.

```
tracks/atom.vel.0000004
tracks/puck.pos tracks/puck.vel --project
> ls tracks/
...
atom.vel.0000000.x.proj.puck.vel.amplitude.0000002
atom.vel.0000000.x.proj.puck.vel.phase. 0000002
atom.vel.0000000.y.proj.puck.vel. amplitude.0000002
...
atom.vel.0000004.y.proj.puck.vel.amplitude.0000002
atom.vel.0000004.z.proj.puck.vel.amplitude.0000002
atom.vel.0000004.z.proj.puck.vel.phase.0000002
...
puck.pos.amplitude.0000002
puck.pos.phase.0000002
puck.vel.amplitude.0000002
puck.vel.phase.0000002...
```

Then we compute the power spectra of the projected velocities.

```
>tr-spectrum
tracks/atom.vel.*.proj.puck.vel.amplitude.0000002 tracks/time
tracks/spectrum.puck.amplitude
--blocks=250 > tr-spectrum
tracks/atom.vel.*.proj.puck.vel.phase.0000002
tracks/time tracks/spectrum.puck.phase --blocks=250
> ls tracks/
...spectrum.puck.amplitude.amplitudes
spectrum.puck.amplitude.wavenumbers
spectrum.puck.phase.amplitudes
spectrum.puck.phase.wavenumbers
...
```

Figure 6 gives an overview of the results. The black curve represents the total velocity power spectrum of the five ring atoms. The blue curve is the spectrum of the velocity vector projected on the tangent of the puckering amplitude coordinate, and the green curve corresponds to the pseudorotation phase. The contribution of the hydrogen atoms is not included in this figure because the ring puckering coordinates are not influenced by the hydrogen positions. The plot clearly reveals the origin of the lowest frequency band (from 20 to 150 cm^{-1}). It is entirely the result of the pseudorotation motion. The second-lowest band (from 150 to 280 cm^{-1}) is mainly caused by the puckering amplitude vibration. Apparently the projected velocities also correlate with bond stretch and bending angle modes at higher frequencies, mainly, because the tangents of the bond stretch, bending angle, and puckering coordinates are not orthogonal.

3.1.5. Final Remarks. This example stresses an important technical advantage of MD-TRACKS. The orthogonal design of the MD-TRACKS commands does not impose a predefined work flow during the analysis. In this case, the output of the command `tr-ic-puckering` is used as input for the command `tr-spectrum`. The puckering coordinates could have been used as an input for other commands, such as `tr-hist` or `tr-blav`. As demonstrated above, `tr-spectrum` also processes the output of many commands, for example, `tr-from-xyz`, `tr-from-txt`, `tr-ic-*`, etc. More advanced analysis tasks are carried out by a whole series of MD-TRACKS commands, grouped in a shell script. Each line in such a script processes the results from the previous commands. The final analysis results are converted to ASCII format with `tr-to-txt`, or they can be directly plotted with `tr-plot`.

3.2. Structural Properties of Liquid THF. In this section, we study radial distribution functions (RDF's) or pair distribution functions[52] of THF in the liquid phase. We consider both the center-of-mass RDF and the atom−atom RDF, which can be compared to neutron diffraction experiments.[53] The comparison between an experimental and a simulated RDF is a stringent test for the validity of the nonbonding interactions in the molecular dynamics simulation.

All radial distribution functions below are derived from an NVT molecular dynamics simulations of 64 THF molecules in a cubic box of 20.5 Å, using periodic boundary conditions. A Nosé−Hoover thermostat[54] with a relaxation time of 0.1 ps was applied to control the temperature of the system. The integration time step is 1 fs, and the total simulation time is 1 ns. The simulation has been carried out with CP2K.[14] The following subsections are organized in the same style as in the previous example.

3.2.1. Setup of the MD-TRACKS Database. This part is very similar to the previous example. We extract the time dependent atom positions with the following commands:

```
> ls
init.psf
md-1.ener
md-MM_DIPOLE-1.data
md-pos-1.xyz
md-vel-1.xyz
md.out
> tr-from-xyz md-pos-1.xyz pos
--slice=::1000
> ls tracks/
atom.pos.0000000.x
atom.pos.0000000.y
atom.pos.0000000.z
atom.pos.0000001.x
...
atom.pos.0000831.z
```

The option `--slice=::1000` instructs the script `tr-from-xyz` to read a frame from the trajectory file `md-pos-1.xyz` every 1000 time steps. The `file init.psf` is a CHARMM[18] topology file that will be used below to determine which atoms belong to the same molecule or to identify the chemical environment of an atom.

3.2.2. Center-of-Mass Radial Distribution Function. A center-of-mass RDF expresses the probability of finding two THF molecules at a certain distance apart in the liquid, relative to the probability of finding a pair of molecules that are homogeneously distributed at the same density. We will first derive the centers of mass of each molecule:

```
> tr-split-com tracks/atoms.pos pos
init.psf
> ls tracks/
...
com.pos.0000000.x
com.pos.0000000.y
...
com.pos.0000063.z
...
```

The arguments of `tr-split-com` are interpreted as follows: (i) the prefix for the track files that contain the atom coordinates, (ii) a tag for the output files, and (iii) the CHARMM topology file to identify the individual molecules. Consequently, we run the command that computes the radial distribution function based on these centers of mass:

```
> all_mol_prefixes=$(tr-select init.psf mol
  True --prefix=tracks/com.pos)
> echo $all_mol_prefixes
tracks/com.pos.0000000
tracks/com.pos.0000001...
> tr-rdf $all_mol_prefixes 20.5*A, 15*A 100
tracks/com.rdf
> ls
...
tracks/com.rdf.bins
tracks/com.rdf.hist
...
```

The first command `tr-select` lists molecules or atoms based on a filter expression. The first argument is the CHARMM topology file for the system under study. The second argument can be `at` or `mol` to indicate which objects one would like to select (atoms or molecules). The last argument is a filter expression that must evaluate to true for the atoms or molecules of interest. When `True` is literally given as filter expression, all atoms, or in this case molecules, are listed. The option `--prefix=tracks/com.pos` determines the format in which the list is printed. The construction `all_mol_prefixes=$(...)` assigns the output of the command `tr-select` to the variable `all_mol_prefixes`. The command `echo $all_mol_prefixes` prints the contents of this variable on screen. The command `tr-rdf` computes the actual radial distribution function for a list of time-dependent Cartesian coordinates. In this example, the centers of mass are used to compute the radial distribution function. The second argument specifies the box dimension so that the periodicity is properly taken into account when computing pair distances. The third argument is the maximum distance for which the RDF is computed. The fourth argument defines the number of bins in the histogram and the last argument is a prefix used for the output files. In this example, the file `tracks/com.rdf.hist` contains the *y*-values of the radial distribution function.

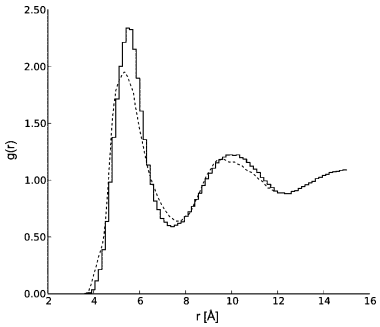


Figure 7. Center-of-mass radial distribution function of liquid THF at room temperature. The solid line represents the histogram derived from the molecular dynamics simulation. The dashed line is the experimentally observed radial distribution function by Bowron et al.[53]

The results are depicted in Figure 7, together with the experimental center-of-mass RDF by Bowron et al.[53] The overall correspondence is satisfactory, except for the maximum of the first peak, which is slightly overestimated in the simulated distribution. The area under the first peak of the simulated pair distribution reveals that (on average) there are 12.9 $\pm$ 0.5 molecules in the first shell that surrounds a given THF molecule. This compares very well to the experimental value of 12.6 $\pm$ 0.3.[53]

3.2.3. Atom−Atom Radial Distribution Functions. A more fine-grained picture of the relative position and orientation of THF molecules in the liquid phase is given by the atom−atom RDF, which expresses the probability of finding an atom of type A and B at a certain distance in the liquid, relative to the probability of finding these atoms at the same distance when they are homogeneously distributed. Analogously to the center-of-mass RDF, we first select the atoms for which we want to compute the RDF. In a second step, the actual RDF's are computed:

```
> O_prefixes=$(tr-select init.psf at
'a.symbol=="O"' --prefix=tracks/atom.pos)
> C1_prefixes=$(tr-select init.psf at
'a.symbol=="C" and a.nsymbols="O,C,H_2"
--prefix=tracks/atom.pos)
> C2_prefixes=$(tr-select init.psf at
'a.symbol=="C" and a.nsymbols="C_2,H_2"
--prefix=tracks/atom.pos)
> echo
tracks/atom.pos.0000001
tracks/atom.pos.0000004...
> tr-rdf $O_prefixes 25*A, 15*A 100
tracks/O.rdf
> tr-rdf $O_prefixes - $C1_prefixes 25*A,
15*A 100 tracks/OC1.rdf
> tr-rdf $O_prefixes - $C2_prefixes 25*A,
```

```
15*A 100 tracks/OC2.rdf
> ls tracks/
...
O.rdf.bins
O.rdf.hist
OC1.rdf.bins
OC1.rdf.hist
OC2.rdf.bins
OC2.rdf.hist
...
```

In the first three lines in the transcript above, three groups of atoms are defined with the command `tr-select`: the oxygen atoms (O), the carbon atoms that are directly bonded to an oxygen atom (group C1), and the carbons atoms that are not directly bonded to an oxygen atom (group C2). For a detailed description of the filter expressions, we refer to the documentation of `tr-select`, which can by consulted with the command `tr-select --help`. Consequently, the RDF's are computed with `tr-rdf`. The first example is based on distances between atoms in a single set, in this example, the set of oxygen atoms. The latter two RDF's consider the distances between the atoms in set A and B but not the distances within each set A or B. In this case, A is the set of oxygen atoms, and B is the set of C1 or C2 atoms.

The three radial distribution functions are plotted in Figure 8. The simulated distributions do not match perfectly with the experimental data. Mainly at short distances, the experimental RDF's show sharp peaks that are not present in their simulated counterparts. The experimental data suggest that our simulations underestimate the liguid structure in terms of relative orientation of neighboring THF molecules.

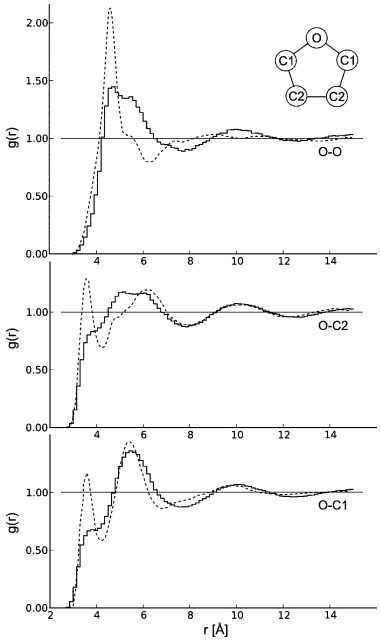


Figure 8. Atom−atom radial distribution functions for the pairs O−O, O−C1, and O−C2. C1 is the set of carbon atoms directly bonded to oxygen, and the C2 set contains to the remaining carbon atoms. The simulated RDF is plotted as a solid line, while the experimental curves are plotted as dashed lines.

4. PROGRAM AVAILABILITY

The MD-TRACKS toolkit is distributed as open source software under the conditions of the GNU General Public License, version 3. The software can be downloaded from the Code Web site of the Center for Molecular Modeling: http://molmod.ugent.be/code/. Documentation, installation instructions, and technical support are also available on this web site. In addition, there is a web-interface to the revision control systems that logs all changes in the source code. MD-TRACKS is released together with ZEOBUILDER,[55] which is a highly suitable GUI toolkit for the construction of initial molecular geometries for molecular dynamics simulations.

5. CONCLUSIONS

MD-TRACKS is a powerful and free molecular trajectory analysis toolkit. The MD-TRACKS toolkit consists of many commands that operate on an efficient and cross-platform binary trajectory database. There are three levels at which the MD-TRACKS toolkit can be used: one enters the individual MD-TRACKS commands at a UNIX command line shell, or one collects these commands in specialized shell scripts that automate the analysis work, or one creates new MD-TRACKS commands based on the MD-TRACKS programming library. Currently, the MD-TRACKS program has an interface CP2K,[14] CPMD[13] LAMMPS,[15,16] DL_POLY,[17] and Cerius2.[23]

The analysis of a molecular dynamics simulation of tetrahydrofuran with MD-TRACKS gives nontrivial insights in the vibrational and structural properties of this solvent. The transition of THF between the two (symmetric) twisted conformers is an anharmonic oscillation. It results in a broad band in the vibrational spectra between 20 and 150 cm^{-1}. The molecular mechanics model of Vorobyov et al.[36] leads in general to analysis results that correlate well with experimental observations. It should be clear that these examples merely cover a small part of the functionality of the MD-TRACKS toolkit. Even a full listing of the current MD-TRACKS commands does not reflect the continuous development of new features and improvements.

ACKNOWLEDGMENT

This work is supported by the Fund for Scientific Research, Flanders, and the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT Vlaanderen).

APPENDIX

The algorithms implemented in the commands below are discussed in standard text books on molecular dynamics and Monte Carlo simulations.[52,56]

- `tr-ac`: Computes the autocorrelation function.
- `tr-ac-error`: Computes the error on an autocorrelation function.
- `tr-angular-momentum`: Determines the angular momenta of one or more molecules.
- `tr-blav`: Applies the block-average method.
- `tr-calc`: Evaluates arbitrary mathematical functions on the data in track files
- `tr-corr`: Computes correlation coefficients
- `tr-cwt`: Computes the continuous wavelet transform.
- `tr-derive`: Numerically differentiates a function.
- `tr-fit-peaks`: Fits peaks in a spectrum.
- `tr-fluct`: Computes fluctuations.
- `tr-from-atrj`: Converts Cerius2 trajectory files (into the binary track files).
- `tr-from-cp2k-cell`: Converts CP2K unit cell data.
- `tr-from-cp2k-ener`: Converts CP2K energy files.
- `tr-from-cp2k-stress`: Converts CP2K stress tensor files.
- `tr-from-cpmd-ener`: Converts CPMD energy files.
- `tr-from-cpmd-traj`: Converts CPMD atom trajectory files.
- `tr-from-dlpoly hist`: Converts DL_POLY history files.
- `tr-from-dlpoly output`: Converts DL_POLY output files.
- `tr-from-lammps-dump`: Converts LAMMPS dump files.
- `tr-from-txt`: Reads data from a column-based plain text file and writes the data to binary track files.
- `tr-from-xyz`: Converts XYZ trajectory files.
- `tr-hist`: Computes histograms.
- `tr-ic-bend`: Computes a (time-dependent) bending angle.
- `tr-ic-dihed`: Computes a dihedral angle.
- `tr-ic-dist`: Computes an interatomic distance.
- `tr-ic-dtl`: Computes a distance to a line.
- `tr-ic-oop`: Computes an out-of-plane distance.
- `tr-ic-psf`: Enumerates and computes (a subset of) the internal coordinates based on the molecular topology.
- `tr-ic-puckering`: Computes the generalized puckering coordinates for an n-membered ring.
- `tr-integrate`: Numerically integrates a function.
- `tr-irfft`: Computes the inverse real Fourier transform.
- `tr-length`: Prints the length of a track file.
- `tr-mean-std`: Computes the (time-dependent) mean and the standard deviation.
- `tr-msd`: Derives the mean square displacement of a set of coordinates as a function of the time interval.
- `tr-msd-fit`: Derives the diffusion coefficient from the data obtained with tr-msd.
- `tr-norm`: Computes the time-dependent norm of a vector.
- `tr-pca`: Applies the principal component analysis method.
- `tr-plot`: Generates charts directly from data in the binary MD-TRACKS database.
- `tr-qh-entropy`: Computes the vibrational entropy, using the quasi-harmonic approximation.
- `tr-rdf`: Computes different types of radial distribution functions:
- `tr-reduce`: Reduces a data set with block averages.
- `tr-rfft`: Computes the forward real Fourier transform.
- `tr-select`: Prints atom or molecule indexes that fulfill a given filter expression.
- `tr-select-rings`: Prints atom indexes that belong to n-membered strong rings.[57]
- `tr-shortest-distance`: Computes the (time-dependent) shortest distance between two sets of atoms.
- `tr-slice`: Reduces a data set with subsampling.
- `tr-spectrum`: Computes various types vibrational spectra.
- `tr-split-com`: Computes the time dependent centers of mass of the molecules in the trajectory.
- `tr-to-txt`: Reads data from the binary database and convert it into plain text format.
- `tr-to-xyz`: Converts the trajectory data in the database to the XYZ format.

REFERENCES AND NOTES

(1) Rapaport, D.D.C. Introduction. In *The Art of Molecular Dynamics Simulationo*, 1st ed.; Cambridge University Press: Cambridge, U.K., 1997; pp 1−10.
(2) Tornroth-Horsefield, S.; Wang, Y.; Hedfalk, K.; Johanson, U.; Karlsson, M.; Tajkhorshid, E.; Neutze, R.; Kjellbom, P. Structural mechanism of plant aquaporin gating. *Nature*. **2006**, *439*, 688–694.
(3) Beckstein, O.; Tai, K.; Sansom, M. Not ions alone: Barriers to ion permeation in nanopores and channels. *J. Am. Chem. Soc.* **2004**, *126*, 14694–14695.
(4) Caleman, C.; van der Spoel, D. Picosecond melting of ice by an infrared laser pulse: A simulation study. *Angew. Chem. Int Ed.* **2008**, *47*, 1417–1420.
(5) Berendsen, H. J. C. *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*, 1st ed.; Cambridge University Press: Cambridge, U.K., 2007; p 624.
(6) Allen, M. P.; Tildesley, D. J. Introduction. In *Computer Simulation of Liquids*; Oxford University Press: New York, 1989; pp 1−32.
(7) Car, R.; Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **1985**, *55*, 2471.
(8) Vandevondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.
(9) McGrath, M. J.; Siepmann, J. I.; Kuo, I. W.; Mundy, C. J.; VandeVondele, J.; Hutter, J.; Mohamed, F.; Krack, M. Isobaric−isothermal Monte Carlo simulations from first principles: Application to liquid water at ambient conditions. *Chem. Phys. Chem.* **2005**, *6*, 1894–1901.
(10) Kuhne, T. D.; Krack, M.; Mohamed, F. R.; Parrinello, M. Efficient and accurate Car−Parrinello-like approach to Born−Oppenheimer molecular dynamics. *Phys. Rev. Lett.* **2007**, *98*, 066401.
(11) Dal Peraro, M.; Ruggerone, P.; Raugei, S.; Gervasio, F. L.; Carloni, P. Investigating biological systems using first principles Car−Parrinello molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2007**, *17*, 149–156.
(12) Lin, H.; Truhlar, D. QM/MM: What have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185–199.
(13) Parrinello, M.; Hutter, J.; Marx, D.; Focher, P.; Tuckerman, M.; Andreoni, W.; Curioni, A.; Fois, E.; Roetlisberger, U.; Gianozzi, P.; Deutsch, T.; Alavi, A.; Sebastiani, D.; Laio, A.; Vandevondele, J.; Seitsonen, A.; Billeter, S. *CPMD*, version 3.11.1; IBM Corp, MPI für Festkörperforschung Stuttgart: Stuttgart, Germany, 1990−2006, 1997−2007.
(14) CP2K. http://cp2k.berlios.de (accessed Mar 24, 2008).
(15) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
(16) LAMMPS. http://lammps.sandia.gov/cite.html (accessed Jul 9, 2008).
(17) Smith, W.; Forester, T. R. DL_POLY_2.0: A general-purpose parallel molecular dynamics simulation package. *J. Mol. Graph.* **1996**, *14*, 136–141.
(18) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
(19) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable

molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
(20) Gale, J. GULP—A computer program for the symmetry adapted simulation of solids. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 629.
(21) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306–317.
(22) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; Gunsteren, W. F. V. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719–1751.
(23) *Cerius2 Modelling Environment*; Accelrys Software Inc.: San Diego, CA, 2006.
(24) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.
(25) Case, D.; Darden, T.; T. E. Cheatham, I. I. I.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Merz, K.; Pearlman, M.; Crowley, M.; Walker, R.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D.; Schafmeister, C.; Ross, W.; Kollman, P.; *AMBER 9*; University of California: San Francisco, CA, 2006.
(26) Sirius. http://sirius.sdsc.edu/ (accessed Jul 10, 2008).
(27) Krüger, P.; Szameit, A. SIMLYS version 2.0. *Comput. Phys. Commun.* **1992**, *72*, 265–268.
(28) Pauwels, E.; Verstraelen, T.; Waroquier, M. Effect of temperature on the EPR properties of a rhamnose alkoxy radical: A DFT molecular dynamics study. *Spectrochim. Acta, Part A* **2008**, *69*, 1388–1394.
(29) Pauwels, E.; Verstraelen, T.; De Cooman, H.; Van Speybroeck, V.; Waroquier, M. Temperature study of a glycine radical in the solid state adopting a DFT periodic approach: Vibrational analysis and comparison with EPR experiments. *J. Phys. Chem. B.* **2008**, *112*, 7618–7630.
(30) Lesthaeghe, D.; Vansteenkiste, P.; Verstraelen, T.; Ghysels, A.; Kirschhock, C. E. A.; Martens, J. A.; Speybroeck, V. V.; Waroquier, M. MFI fingerprint: How pentasil-induced IR bands shift during zeolite nanogrowth. *J. Phys. Chem. C.* **2008**, *112*, 9186–9191.
(31) Oliphant, T. Numpy. http://www.numpy.org/ (accessed Jun 15, 2008).
(32) Mohr, P. J.; Taylor, B. N. CODATA recommended values of the fundamental physical constants: 1998. *Rev. Mod. Phys.* **2000**, *72*, 351.
(33) Legon, A. C. Equilibrium conformations of four- and five-membered cyclic molecules in the gas phase: determination and classification. *Chem. Rev.* **1980**, *80*, 231–262.
(34) Cadioli, B.; Gallinella, E.; Coulombeau, C.; Jobic, H.; Berthier, G. Geometric structure and vibrational spectrum of tetrahydrofuran. *J. Phys. Chem.* **1993**, *97*, 7844–7856.
(35) Duffy, P.; Sordo, J. A.; Wang, F. Valence orbital response to pseudorotation of tetrahydrofuran: A snapshot using dual space analysis. *J. Chem. Phys.* **2008**, *128*, 125102.
(36) Vorobyov, I.; Anisimov, V.; Greene, S.; Venable, R.; Moser, A.; Pastor, R.; Mackerell, A. Additive and classical drude polarizable force fields for linear and cyclic ethers. *J. Chem. Theory Comput.* **2007**, *3*, 1120–1133.
(37) Rayon, V. M.; Sordo, J. A. Pseudorotation motion in tetrahydrofuran: An ab initio study. *J. Chem. Phys.* **2005**, *122*, 204303.
(38) Harris, D. O.; Engerholm, G. G.; Tolman, C. A.; Luntz, A. C.; Keller, R. A.; Kim, H.; Gwinn, W. D. Ring Puckering in five-membered rings. I. General theory. *J. Chem. Phys.* **1969**, *50*, 2438–2445.
(39) Vansteenkiste, P.; VanSpeybroeck, V.; Verniest, G.; DeKimpe, N.; Waroquier, M. Applicability of the hindered rotor scheme to the puckering mode in four-membered rings. *J. Phys. Chem. A.* **2006**, *110*, 3838–3844.
(40) Vansteenkiste, P.; VanSpeybroeck, V.; Verniest, G.; DeKimpe, N.; Waroquier, M. Four-membered heterocycles with a carbon−heteroatom exocyclic double bond at the 3-position: puckering potential and thermodynamic properties. *J. Phys. Chem. A.* **2007**, *111*, 2797–2803.
(41) Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, R. The thermodynamics and molecular structure of cyclopentane. *J. Am. Chem. Soc.* **1947**, *69*, 2483–2488.
(42) Coulombeau, C.; Jobic, H. Contribution to the vibrational normal modes analysis of d8-THF by neutron inelastic scattering. *THEOCHEM.* **1995**, *330*, 127–130.
(43) Yang, T.; Su, G.; Ning, C.; Deng, J.; Wang, F.; Zhang, S.; Ren, X.; Huang, Y. New diagnostic of the most populated conformer of tetrahydrofuran in the gas phase. *J. Phys. Chem. A.* **2007**, *111*, 4927–4933.
(44) Berens, P.; Wilson, K. Molecular dynamics and spectra. I. Diatomic rotation and vibration. *J. Chem. Phys.* **1981**, *74*, 4872–4882.
(45) McQuarrie, D. A. The time-correlation function formalism, I. In *Statistical Mechanics*; Rice, S., Ed.; Harper & Row: New York, 1976; pp 467−542.
(46) Cooley, J. W.; Tukey, J. W. An algorithm for the machine calculation of complex fourier series. *Math. Comput.* **1965**, *19*, 297–301.
(47) Dickey, J. M.; Paskin, A. Computer simulation of the lattice dynamics of solids. *Phys. Rev.* **1969**, *188*, 1407.
(48) Loong, C.; Vashishta, P.; Kalia, R. K.; Jin, W.; Degani, M. H.; Hinks, D. G.; Price, D. L.; Jorgensen, J. D.; Dabrowski, B.; Mitchell, A. W.; Richards, D. R.; Zheng, Y. Phonon density of states and oxygen-isotope effect in $Ba1\text{-}xK_xBiO_3$. *Phys. Rev. B.* **1992**, *45*, 8052.
(49) Rahman, A.; Mandell, M. J.; McTague, J. P. Molecular dynamics study of an amorphous Lennard-Jones system at low temperature. *J. Chem. Phys.* **1976**, *64*, 1564–1568.
(50) McQuarrie, D. A. Crystals. In *Statistical Mechanics*; Rice, S., Ed.; Harper & Row: New York, 1976; pp 194−221.
(51) Cremer, D.; Pople, J. A. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.
(52) Allen, M. P.; Tildesley, D. J. Statistical Mechanics. In *Computer Simulation of Liquids*; Oxford University Press: New York, 1989; pp 33−70.
(53) Bowron, D.; Finney, J.; Soper, A. The structure of liquid tetrahydrofuran. *J. Am. Chem. Soc.* **2006**, *128*, 5119–5126.
(54) Nose, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
(55) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. ZEO-BUILDER: A GUI toolkit for the construction of complex molecular structures on the nanoscale with building blocks. *J. Chem. Inf. Model.* **2008**, *48*, 1530–1541.
(56) Frenkel, D.; Smit, B. "Monte Carlo Simulations" and "Molecular Dynamics Simulations". In *Understanding Molecular Simulation*, 2nd ed.; Academic Press: San Diego, CA, 2002; pp 23−108.
(57) O'Keeffe, M.; Hyde, B. Nets and infinite polyhedra. In *Crystal Structures I. Patterns and Symmetry*; Mineralogical Society of America: Washington, DC, 1996; pp 289−375.
(58) Smith, A. *The Coblentz Society Desk Book of Infrared Spectra*, 2nd ed.; Carver, C., Ed.; The Coblentz Society: Kirkwood, MO, 1982; pp 1−24.
(59) Altona, C.; Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.

CI800233Y

Paper V

UV-Raman and ^{29}Si NMR Spectroscopy Investigation of the Nature of Silicate Oligomers Formed by Acid Catalyzed Hydrolysis and Polycondensation of Tetramethylorthosilicate

A. Depla, E. Verheyen, A. Verfeyken, M. Van Houteghem, K. Houthoofd, V. Van Speybroeck, M. Waroquier, C. Kirschhock, J.A. Martens

Journal of Physical Chemisty C, 115 (22), 11077-11088, 2011

THE JOURNAL OF PHYSICAL CHEMISTRY C

ARTICLE

pubs.acs.org/JPCC

UV-Raman and ^{29}Si NMR Spectroscopy Investigation of the Nature of Silicate Oligomers Formed by Acid Catalyzed Hydrolysis and Polycondensation of Tetramethylorthosilicate

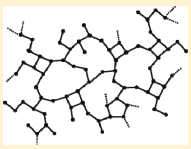
Anouschka Depla,[†] Elke Verheyen,[†] An Veyfeyken,[†] Marc Van Houteghem,[‡] Kristof Houthoofd,[†] Veronique Van Speybroeck,[‡] Michel Waroquier,[‡] Christine E. A. Kirschhock,[†] and Johan A. Martens*[,†]

[†]Center for Surface Chemistry and Catalysis, K.U. Leuven, Heverlee, Belgium

[‡]Center for Molecular Modeling, Ghent University, Zwijnaarde, Belgium

Supporting Information

ABSTRACT: Tetramethylorthosilicate (TMOS) was hydrolyzed and polymerized under strongly acidic conditions in the presence of substoichiometric quantities of water. The polymerization reaction was monitored during 64 h using ^{29}Si NMR and UV-Raman spectroscopy. The nature of the oligomers and the condensation reaction pathways were unraveled using this combination of experimental techniques together with molecular modeling. ^{29}Si NMR and UV-Raman signals which previously were not documented in literature could be assigned. TMOS rapidly was converted into short straight methoxylated silicate chains. Subsequently the growth of oligomers proceeded by condensations between a hydrolyzed middle group of a chain with an end-group of another chain. Larger oligomers were attached to each other via condensations between middle groups generating multiply branched structures. Rings were formed late in the reaction scheme through internal condensations of sizable silicate molecules. Oligomers that were characteristic of the different stages of the polymerization process were proposed. Oligomerization pathways starting from tetramethylorthosilicate and tetraethylorthosilicate (TEOS) are significantly different. While with TMOS rings are formed only late in the oligomerization scheme, with TEOS rings are formed at early stages through cyclo-dimerization. This insight into the different nature of the oligomers obtained from TMOS and TEOS will assist the design of new silica sol−gel materials.



INTRODUCTION

Silica sol−gel chemistry refers to the hydrolysis and condensation of a monomeric silicon alkoxide source in the presence of solvent and homogeneous catalyst into sol and gel. Many industrial manufacturing schemes of specialty silica materials including fibers, coatings, films, gels, glasses, and ceramics are based on silica sol−gel chemistry.[1−7] In sol−gel processes the properties of the final silica material are largely dependent on the nature of the silicate polymer and, especially, the formation of linear versus branched chains and ring structures. Crucial parameters of the silica sol−gel processes are the water:alkoxide molar hydrolysis ratio designated as the *r* value, the pH, the nature of the alkoxide, and the addition of organic solvent.[8−23] In general terms, high pH, and high *r* values favor the formation of highly branched silicate networks suited for manufacturing of colloidal silica or silica gel.[24−30] The use of a homogeneous acid catalyst leads to more linear polymer chains and is recommended when materials with low fractal dimension such as films and fibers are desired.[12]

Recent studies have shown that this traditional vision on the impact of the catalyst selection on branched versus linear polymer growth may be too simplistic. Insight into vibrational fingerprints revealed with UV-Raman spectroscopy of silicate oligomers is steadily growing.[31−46] Using ^{29}Si NMR and small angle X-ray scattering, Brinker and Assink observed significantly branched silicate chains obtained under acidic conditions.[47] Further evidence for the presence of branched oligomers was provided by ^{29}Si NMR and Raman spectroscopy.[12,17,48−51] In support of these observations, occurrence of cyclization reactions in acid catalyzed silica sol−gel processes has been detected.[12,17,48,49,52−59]

In an earlier publication we reported a similar investigation of acid catalyzed silica polymerization starting from tetraethylorthosilicate (TEOS) using ^{29}Si NMR, UV-Raman, and molecular modeling.[60] The identity of oligomers formed at *r* values of 2 and lower could be identified and the reaction scheme projected. In the polymerization process departing from TEOS, linear chains no longer than hexamer are formed next to rings composed of three to six silicate units and branched specimen. The main reaction pathways in acid catalyzed silica oligomerization departing from TEOS are chain extension through monomer addition, formation of rings via cyclo-dimerization, and branchings through attachment of dimer and trimer.

Previously we noticed significant differences in the properties of silica gel obtained from TMOS and TEOS under almost

Received: January 19, 2011
Revised: May 5, 2011
Published: May 16, 2011

The Journal of Physical Chemistry C ARTICLE

Table 1. Overview of Samples

sample code	*r* value	TMOS (mL)	MeOH (mL)	H_2SO_4 (mL)	H_2O (mL)	TMOS:H_2O:MeOH:H_2SO_4 molar ratio
R-0.1	∼0.1	10.442	8.783	0.673	0.103	1:0.1:3.07:0.35
R-0.9	0.9	10.183	8.565	0.656	0.597	1:0.9:3.07:0.35
R-1.3	1.3	9.876	8.307	0.636	1.180	1:1.3:3.07:0.35

identical reaction conditions.[61,62] While the use of TEOS led to the formation of microporous silica gel, its substitution with TMOS led to the formation of mesoporous material.[61] A standing question was how the choice of alkoxide affected the oligomerization reaction scheme. In this work we unraveled the molecular aspects and underlying mechanism of the silica oligomerization process departing from TMOS. We applied the same methodology of combined UV-Raman and ^{29}Si NMR, successfully applied on the TEOS sol−gel chemistry.[60] Raman frequencies and ^{29}Si NMR chemical shifts have been computed for the different species using ab initio molecular modeling techniques. Significant differences in silicate speciation were observed, increasing the insight into the nature of the final silica gel materials that are obtained from TMOS versus TEOS.

■ EXPERIMENTAL SECTION

In a typical sample preparation TMOS (Acros, 99%) and methanol (VWR, absolute grade, 99.9%) were mixed. Concentrated H_2SO_4 (Fisher, 98%) and optional water were added dropwise in this order under continuous stirring. The molar ratio of TMOS:MeOH:H_2SO_4 was 1:3.07:0.35. Water was added to reach a molar H_2O:TMOS ratio of 0.5 or 1.0 initially. After preparation in open recipients the samples were either sealed in polypropylene bottles (20 mL) for sample aging or transferred into NMR sample tubes, sealed, and introduced in the ^{29}Si NMR instrument. The samples were hygroscopic and rapidly absorbed some water from ambient air. The actual *r* value was determined based on the hydrolysis and condensation state of the Si atoms determined by ^{29}Si NMR sequences following a method explained in a previous publication.[60] In short, the *r* value was calculated from the number of hydrolyzed groups, assuming quantitative reaction of all water present. The *r* value for the samples with initial molar H_2O:TMOS ratio of 0.5 and 1.0 increased to 0.9 and 1.3, respectively. Specifically for the UV-Raman investigation a third sample with minimum water content was prepared (R-0.1). The chemical compositions of the samples R-0.1, R-0.9, and R-1.3 are given in Table 1.

Raman spectra were recorded using a Raman instrument combining of a Coherent Innova 300C MotoFred laser, a Roper Scientific Trivista TR557 A&S triple-stage spectrometer, and a CCD camera (Princeton Instruments). The spectrometer stages were equipped with 900−900−3600 grooves/mm grating combination. An excitation beamline at 244 nm wavelength with an output power of 10 mW was used. The silicate samples were analyzed in open shallow cylindrical vessels rotating horizontally. To minimize local heating, a short exposure time to the laser beam of 80 s was applied. Each spectrum was an average accumulation of three recordings, each on a fresh sample. Possible instrumental frequency shifts were corrected using PTFE as a shift reference. Spectra were normalized for intensity based on the 1455 cm^{-1} vibration, assigned to the asymmetrical C−H bending of methoxide groups of both TMOS and MeOH. A Savitzky−Golay algorithm with a filter coefficient of 5 was applied to smooth the spectra.

^{29}Si NMR spectra of samples R-0.9 and R-1.3 were recorded on a Bruker Avance II+ 600 spectrometer (14.1 T). Samples (3 mL) were loaded in 10 mm quartz tubes. The spinning frequency was 16 Hz. Experimental conditions comprised a recycle delay of 80 s and pulse lengths of 12.0 ms (R-0.9) or 10.0 ms (R-1.3). These pulse lengths correspond to a tip angle of 90°. Recordings of samples R-0.9 and R-1.3 contained 65536 and 16384 time domain data points (td), respectively. Tetramethylsilane served as external chemical shift reference. Measurements were performed without lock, under isothermal conditions at 298 K. Shimming was performed on a solution of TMS-salt (trimethylsilylpropane sulfonic acid, Merck) in 3 mL of D_2O. Each acquisition in the time sequence represented an accumulation of 28 scans.

■ THEORETICAL SECTION

All theoretical predictions for Raman vibrational frequencies Raman[63] and ^{29}Si NMR chemical shifts[64] were derived from static quantum chemical methods using the Gaussian03 program.[65] Ab initio molecular dynamics simulations are the most natural way to reproduce IR and Raman spectra, but for larger oligomers they require a high computational cost as long simulation times are needed. Alternatively, classical molecular dynamics methods make calculations feasible but require adequate force fields. For Raman spectra the Fourier transform of the autocorrelation function of the time variation of the electric polarizability matrix should be evaluated from the trajectories. This is very sensitive to the applied force field, and for the specific purpose of this paper we have preferred the static approach.[66] All chain structures from monomer up to octamer and some cyclosilicates (three- and four-rings) were optimized in vacuo at the B3LYP/6-31+g(d) level of theory.[67] This level of theory has been recognized as one of the best performing functionals for geometry optimizations and has already been used in previous studies on silica oligomerization.[29,60] DFT based harmonic vibrational frequencies are normally scaled to compare with experimental values. The scale factor is uniform for a given level of theory[68] and has been determined via an optimal correlation between theoretical and experimental values for all structures ranging from monomer up to pentamer (see Supporting Information). This scale factor amounts to 1.0335 for Raman frequencies and is in good agreement with theoretically determined scale factors proposed by Radom et al.[68] for low frequencies. The model assumes all species fully saturated with methoxide groups, justified at low molar hydrolysis ratio. The configurational dependency of the spectra was verified by considering different conformers for each species. For all species a large number of conformers was investigated ranging from linear to curled up conformers and various internal rotations were applied to scan for a large variety of structures. From these structures we have selected the most stable ones differing in binding energy by no more than 5 kJ mol^{-1}. The coordinates of all selected conformers are given in the Supporting Information. Vibrational frequencies for all conformers have been computed, yielding mean values and widths. Maximal variation between conformers appeared to be ∼20 cm^{-1}.

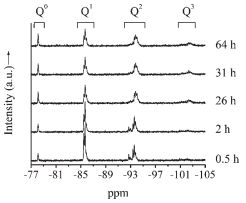


Figure 1. ^{29}Si NMR spectra of the polycondensation process of R-0.9 sample.

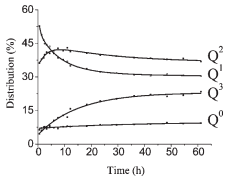


Figure 2. Q^n distribution in R-0.9 sample according to ^{29}Si NMR against time.

For ^{29}Si NMR, chemical shifts were calculated for all silicon atoms of the structures mentioned above and averaged out over Q^0, Q^1, Q^2, and Q^3 sites. A distinction was made between Q^2 species in chains and Q^2 species in rings. Single-point calculations with a triple-ζ basis set 6-311++g(d,p) were performed on B3LYP/6-31+g(d) optimized geometries for the ^{29}Si NMR signals.[69] Solvent effects on NMR shifts are well-known and have been investigated to some extent on nitrogen shielding constants.[70] Most of the previous ab initio calculations of chemical shifts on small silica species have ignored the effect of the molecular environment.[71] Recently some studies included these effects either by using an implicit solvent model[72] or by adding some discrete solvent molecules to the silica clusters.[73] The silica clusters under consideration in this paper are relatively large, ranging from 20 to about 120 atoms. Explicitly solvating these clusters with solvent molecules is not feasible as it would require a lot of explicit solvent molecules.[74] Moreover, the position of each of these solvent species would have to be equilibrated by means of molecular dynamics simulations.[75] To account for the solvent environment, we used a scale factor for the NMR shifts by determining an optimal correlation between theoretical and experimental values for the Q^0, Q^1, Q^2, and Q^3 sites (See Supporting Information). The scale factor was determined as 0.9596.

RESULTS AND DISCUSSION

Identification of Silicate Oligomers Using ^{29}Si NMR, UV-Raman, and Molecular Modeling. Our samples prepared from TMOS at low r values in strongly acidic conditions were homogeneous and suitable for ^{29}Si NMR spectroscopy. The R-0.9 sample was prepared and loaded into the NMR apparatus. About 70 ^{29}Si NMR spectra were taken in the period from 0.5 to 64 h. Chemical shifts of specific signals were determined in the 70 spectra and averaged. For the majority of signals the accuracy of the chemical shift was within 0.01 ppm. Examples of spectra recorded 0.5, 2, 26, 31, and 64 h after preparing the sample are shown in Figure 1. The spectra comprised signals of Q^0, Q^1, Q^2, and Q^3 silicate connectivities, the subscript in the Q symbol referring to the number of siloxane bonds of the Si atom. The evolution of the distribution of Q^n environments in R-0.9 sample according to ^{29}Si NMR is presented in Figure 2. The Q^0 signal was constant and represented less than 10% of all silicon atoms. End groups (Q^1) were dominant in the first hours (Figure 2). A formation of dimers and short chains can account for such high content of end groups. The Q^1 concentration decreased from over 50% initially to a final level of ca. 30% of the Si atoms. Si atoms in chains linked to two Si atoms via siloxane bridges (Q^2) were present all the time. Their concentration reached a maximum after ca. 6 h. The content of branchings (Q^3) originally was very low but steadily increased all along the investigated 64 h period. The concentration of fully condensed Si atoms (Q^4) remained below the detection limit. At this low water content and high acidity the network cannot reach a high level of connectivity within the monitored time frame.

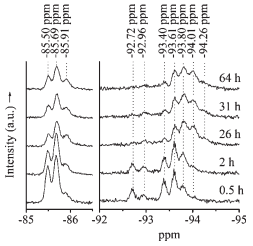


Figure 3. Detail of Q^1 (left) and Q^2 (right) range of ^{29}Si NMR spectra of Figure 1.

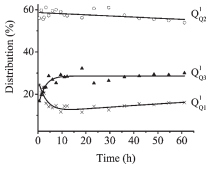


Figure 4. Distribution of Q^1 ^{29}Si NMR signals in R-0.9 sample against aging time. Lines are meant to guide the eye.

Details of the ^{29}Si NMR spectra of Figure 1 in the Q^1 and Q^2 regions are provided in Figure 3. One signal was observed in the Q^0 range, assigned to TMOS (Q^0, $\delta = -78.14$ ppm, Figure 1). The Q^1 region showed three well-resolved signals (Figure 3). Next to the Q^1 signal assigned to dimer (Q^1_{Q1}, $\delta = -85.50$ ppm),[48−51,54] two additional Q^1 signals were observed at $\delta = -85.69$ and -85.91 ppm. Si atoms terminating chains of trimers and longer chains

typically show a resonance around −85.7 ppm,[48,54,76] leading to the assignment of the signal at −85.69 ppm to Q^1 end groups connected to a Q^2 and denoted Q^1_{Q2}. The third Q^1 signal at $\delta = -85.91$ ppm has hardly been discussed in the literature.[48] From the evolution of the internal distribution of Q^1 signals (Figure 4) it appeared that the species associated with the −85.91 ppm signal were formed later than dimer and chains. The $\delta = -85.91$ ppm signal grew in parallel with the signals in the Q^3 range. This led to a tentative assignment of the −85.91 ppm resonance to Q^1 end groups connected to a Q^3 silicon, denoted Q^1_{Q3}. On the basis of their chemical shift all Q^1 groups were methoxylated.

The ^{29}Si NMR spectra in the Q^2 range showed eight resolved peaks (Figure 3). The assignment of the signals at −93.40, −93.61, and −93.80 ppm to Q^2 Si atoms of linear fully methoxylated trimer, tetramer and pentamer was based on the literature.[48−51,54,76] These three signals dominated the Q^2 range after 0.5 and 2 h. After 26 h and later two additional upfield signals gained intensity with maximum intensity around −94.01 and −94.26 ppm, respectively (Figure 3). These signals have not been documented before. Their position and gain in intensity at later stages of the experiment led to a tentative assignment to Q^2 middle atoms of chains containing more than five Si atoms, i.e., chains of six Si (−94.01 ppm) and seven Si units (−94.26 ppm), respectively. The presence of a shoulder upfield of the −94.26 ppm signal hints at the formation of even longer chains.

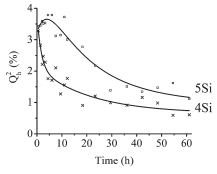


Figure 5. Content of hydrolyzed Q^2_n groups in chains with 4Si, 5Si, and 6 or more Si atoms (6Si+) according to ^{29}Si NMR against time during the polycondensation of sample R-0.9. Lines are meant to guide the eye.

The five already discussed Q^2 signals were assigned to fully methoxylated molecules. In the Q^2 range the first two ^{29}Si NMR spectra (0.5 and 2 h) showed additional signals downfield of the signals of fully methoxylated molecules at $\delta = -92.72$, −92.96, and −93.2 ppm, respectively (Figure 3). Due to the low intensity the last signal will not be quantitatively interpreted. Signals in this chemical shift region have been assigned in literature either to Q^2 signals of cyclosilicates or to chains containing one hydrolyzed Q^2 silicon.[48−50,54,76] Brunet et al.[48] observed ^{29}Si NMR signals at chemical shifts similar to those presently observed and assigned them to Q^2 silicons of branched 4-rings. Kelts and Armstrong observed similar spectra in an acid catalyzed TMOS oligomerization and originally assigned them to hydrolyzed linear groups.[50] In a later study in which TMOS and TEOS hydrolysis and condensation were compared, those authors modified their interpretation of the discussed signals in favor of cyclosilicates.[49] In the present study the intensity of the signals at $\delta = -92.72$ and −92.96 decreased strongly in the period from 2 to 26 h (Figure 3). Provided the assignments to cylcosilicates were correct, these Q^2 silicons either would have been converted to Q^3 via the formation of branchings or else to Q^2 and Q^1 groups of chains via a break-up of these rings. According to Brunet et al.[48] the Q^3 signature of ring Si atoms carrying a branching are to be found at $\delta \sim -101$ ppm. In our experiment, Q^3 signals in the region around −101 ppm were weak in early NMR spectra and decreased rather than increased in intensity when the discussed Q^2 signals weakened. Only weak signals assigned to Q^3 environments were observed in the region −100 to −103 ppm (Figure 1) indicating the existence of silicate rings early in the synthesis was unlikely.

An alternative interpretation of the Q^2 signals at $\delta = -92.72$, −92.96, and −93.2 ppm was assumption of the presence of a hydrolyzed group on Q^2 Si atoms in chains. The presence of a hydrolyzed group in a chain has been reported to cause a downfield shift of ca. 0.9 ppm[50] or ca. 1.3 ppm[48,54,76] compared to the nonhydrolyzed oligomer. Compared with the chemical shifts of the Q^2 signals of fully methoxy terminated tetramer ($\delta = -93.61$ ppm), pentamer ($\delta = -93.80$ ppm), and hexamer (Q^2, $\delta = -94.01$) each of

Table 2. Experimental and Theoretical ^{29}Si-NMR Chemical Shifts

		experimental ^{29}Si NMR signal (ppm)		
Q^i region	species	R-0.9[a]	R-1.3	theoretical region (ppm)[b]
Q^0	monomer	−78.14		−74.97; −77.16
Q^1				−82.78; −88.36
Q^1_{Q1}	dimer	−85.50	−85.5	
Q^1_{Q2}	end-group of Q^2	−85.69	−85.7	
Q^1_{Q3}	end-group of Q^3	−85.91	−85.9	
Q^2				−90.96; −99.65
	3Si	−93.40	−93.4	
	4Si	−93.61	−93.6	
	5Si	−93.80	−93.8	
	6Si$^+$	−94.01; −94.26	−94.0; −94.3	
Q^2 hydr.		−92.72; −92.96; −93.2	−92.4; −92.7; −93; −93.2	
Q^3	branches	[−100; −103]	[−99.5; −103.5]	−92.84; −105.38

[a] Chemical shifts of specific signals have been determined and averaged over about 70 ^{29}Si NMR spectra. The error margin on the experimental ^{29}Si NMR signal (ppm) for R-0.9 was 0.01 for all fully methoxide terminated species and 0.03 for the hydrolyzed specimen. [b] Theoretical values were scaled with a factor of 0.9596, determined by the optimal correlation between the theoretical and experimental values for the Q^0, Q^1, Q^2, and Q^3 environments (see Supporting Information).

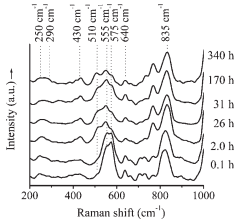


Figure 6. UV-Raman spectra of the condensation process of R-0.9 sample.

the signals at $\delta = -92.72$, -92.96, and -93.20 ppm Q^2 exactly present this downfield shift by ~0.9 ppm. According to this assignment, in the ^{29}Si NMR spectra after 0.5 and 2 h, Q^2 groups in tetramers were most abundant (signals at -92.72 ppm for tetramer with hydrolyzed group and -93.61 ppm for unhydrolyzed tetramer, respectively). The content of hydrolyzed Q^2 groups in chains with four and five Si atoms was plotted against time in Figure 5. In the first hours, the content of hydrolyzed Q^2 groups in 4Si chains rapidly decreased, while later hydrolyzed groups in chains with 5Si atoms dominated among the hydrolyzed Q^2 groups.

Linear trimer hydrolyzed at Q^2 was expected to cause a Q^2 ^{29}Si NMR signal around -92.50 ppm considering a downfield shifting by 0.9 ppm of the -93.40 ppm signal assigned to fully methoxylated trimer. However, at the expected chemical shift no peak was observed (Figure 3). Similarly, in the Q^0 and Q^1 range the ^{29}Si NMR spectra did not show any signal that could be assigned to hydrolyzed species (Figure 1). The concentration of hydrolyzed monomer, dimer, trimer, and of end groups of chains was low in this TMOS system. These groups apparently were short-lived and their concentration was therefore below the detection limit during the first ^{29}Si NMR recording. Apparently, hydrolyzed groups persisted only on Q^2 silicons in chains with four and more Si atoms.

In the time sequence of ^{29}NMR spectra (Figure 1) relatively weak Q^3 signals were observed. Systematic investigation of the 70 collected spectra revealed intensity maxima occurred at $\delta = -100.9$, -101.15, -101.45, and -101.8 ppm at early times. With time, Q^3 signals at more negative chemical shifts of -102.08, -102.35, and -102.6 ppm were detected. A listing of ^{29}Si NMR signals and their assignment in the R-0.9 sample is provided in Table 2.

Raman spectra of sample R-0.9 during its evolution are presented in Figure 6. The TMOS monomer was responsible for a Raman signal at ~640 cm^{-1}. Raman signals of dimer and end groups of chains coincided around 575 cm^{-1}.[1,16,17,51,60,77] Molecular modeling confirmed these assignments (Table 3). The representative silica structures are shown in Figure 7. The Raman active band at 575 cm^{-1} corresponds to a global stretching vibration. The frequency of the maximum of the band depended on the relative concentration of the dimer versus end groups. In UV-Raman spectra of sample R-0.9 the signal at ~575 cm^{-1} was in the lower range of the wavenumbers reported for this signal in literature pointing at dominating dimers. In the time series of Raman spectra, the ~575 cm^{-1} band decreased in intensity in favor of

Table 3. Experimental and Theoretical Raman Shifts

species	experimental shift (cm^{-1})	theoretical shift region (cm^{-1})
monomer	640	[639−649]
dimer (+end groups)	575	[568−578]
3Si	555	[543−553]
4Si	~ 540	[522−542]
5Si	510	[515−535]
6Si		[515−535]
7Si		[515−535]

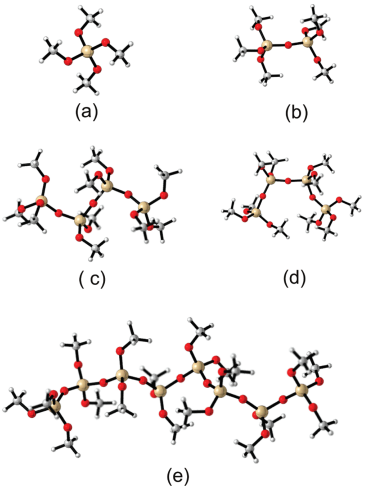


Figure 7. Some of the lowest energy conformers of monomer (a), dimer (b), tetramer (open (c) and closed (d) configuration), and octamer (e).

signals at ~555 and ~510 cm^{-1}. The latter two vibrations could not be readily assigned. From ^{29}Si NMR spectra (Figure 3) it was known that linear trimers, tetramers, and pentamers were dominant. Literature reports the vibrations of linear trimer and tetramer at wavenumbers of 525 and 484 cm^{-1}, respectively.[16,51] Absorptions at such wavenumbers were absent in the Raman spectra of the R-0.9 sample. According to ab initio modeling, trimer, tetramer, and pentamer vibrations were to be expected in the regions of 543−553, 522−542, and 515−535 cm^{-1}, respectively. On the basis of these models, the experimentally observed Raman band around 510 cm^{-1} was attributed to pentamer. The assignment of the Raman band at 550−555 cm^{-1} was less obvious and required

The Journal of Physical Chemistry C ARTICLE

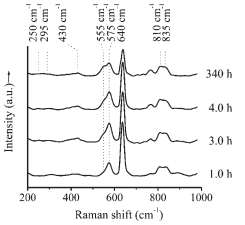


Figure 8. UV-Raman spectra of the condensation process of R-0.1 sample.

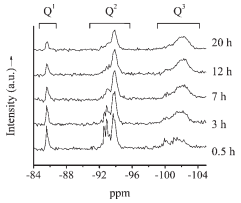


Figure 9. ^{29}Si NMR spectra of the condensation process of R-1.3 sample.

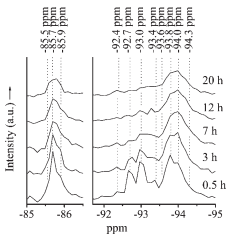


Figure 10. Detail of Q^1 range (left) and Q^2 range (right) of the ^{29}Si NMR spectra of Figure 9.

further investigation. A sample with very low r value (R-0.1) was prepared in order to further limit the extent of oligomerization. In this sample the Raman signatures of monomer (∼640 cm^{-1}) and dimer (∼575 cm^{-1}) were prominent (Figure 8). A shoulder was present on the peak at ∼575 cm^{-1} which developed into a distinct vibration at ∼555 cm^{-1}. The band at ∼510 cm^{-1} observed in the R-0.9 sample was absent. On the basis of these spectra and the molecular modeling locating trimer vibrations to the 543−553 cm^{-1} range, the ∼555 cm^{-1} vibration was assigned to linear trimers. The Raman spectra of sample R-0.9 (Figure 6) revealed a gradual transformation of dimer (∼575 cm^{-1}) and linear trimer (∼ 555 cm^{-1}) into linear pentamer (∼510 cm^{-1}) in agreement with NMR. The Raman signal of linear tetramers according to NMR also present in this sample presumably was contained in the envelope of ∼510 and ∼555 cm^{-1} peaks (Figure 6). According to ab initio modeling, tetramers indeed were to be expected in the region of 522−542 cm^{-1}. A closer inspection revealed the maximum of the peak at ∼555 cm^{-1} shifted to 550 cm^{-1} hinting tetramer contribution increased with time. Tetramer vibration has tentatively been assigned to be located around ∼540 cm^{-1}. ^{29}Si NMR of R-0.9 sample revealed the formation of 6Si and longer chains. According to our theoretical calculations, discrimination between Raman signatures of pentamers and longer linear chains was expected to be difficult (Table 3). The Raman bands of linear chains containing more than five Si atoms were expected in the region of 515−535 cm^{-1}, identical to the region of pentamers (Table 3). Thus Raman spectroscopy was less convenient for discriminating between longer chains. Hence, the broad band observed at ∼510 cm^{-1} was assumed to contain contributions from pentamers as well as from longer chains.

The Raman signals at ∼810 and ∼835 cm^{-1} comprised the asymmetrical Si−O stretch (810 cm^{-1}) and the CH_3-rocking mode (835 cm^{-1}).[17,78] Raman bands at ∼250, ∼ 295, and ∼430 cm^{-1} in TMOS oligomerization hardly have been addressed in literature. The presence of these bands has been found to be highly dependent on the type of mineral acid, and they were estimated to be less relevant to the present analysis of oligomerization pathways. Vibrations at ∼665, ∼710, ∼730, and ∼770 cm^{-1} observed in the R-0.9 sample occurred in a region usually assigned to hydrolyzed species.[13,16,17,24,51] Hydrolysis of methoxy groups lowers the molecular mass and was expected to cause the vibrations to shift toward higher wavenumbers. However, the R-0.1 sample with minimum water content displayed similar signals (Figure 8) making assignment of the vibrations at ∼665, ∼710, ∼730, and ∼770 cm^{-1} to hydrolyzed species unlikely. Most probably is the vibration at ∼770 cm^{-1} associated with −OCH_3 and the evolution of a network with increasing connectivity. This is supported by the absence of this vibration in TEOS systems[60] and the strength of the signal observed in a fresh gel of the R-1.3 sample (vide infra) which contains still a large amount of Q^2 and Q^3 (see Supporting Information). Assignment of this signal to monomeric silica appears highly unlikely as ^{29}Si NMR clearly indicates no such species are present. However the assignment of the vibrations at ∼665, ∼710, ∼730, and ∼770 cm^{-1} remains elusive.

Further confirmation of the assignments of ^{29}Si NMR and Raman signals in the R-0.9 sample was obtained by comparison to a sample with slightly higher molar hydrolysis ratio. ^{29}Si NMR spectra of the R-1.3 sample are shown in Figure 9 with details in Figure 10. These ^{29}Si NMR spectra were less resolved compared to the R-0.9 sample due to a lower number of data points in the time domain (cf. Experimental Section). Furthermore, in this sample with more advanced hydrolysis a distribution of chemical shifts has to be expected which also contributed to the observed line broadening. In this sample no ^{29}Si NMR signal in the Q^0 range was discernible, revealing all TMOS had reacted. The ^{29}Si NMR spectra recorded in the period 0.5−20 h indicated dominance of Q^2 and Q^3 environments (Figure 11). The Q^3 content steadily increased at the expense of Q^2. The number of Q^1 species remained almost constant.

The spectral features in the Q^1 and Q^2 range observed in the R-0.9 sample were also encountered in R-1.3. In the Q^1 range an

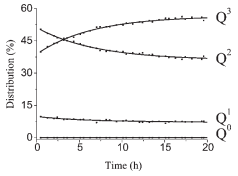


Figure 11. Evolution of Q distribution according to ^{29}Si NMR in R-1.3 sample.

unresolved signal covered the signals of dimer (−85.5 ppm) and end groups of chains with more than 2Si atoms (−85.7 and −85.9 ppm). In the Q^2 range the ^{29}Si NMR spectrum comprised signals of Si atoms of linear chains with one hydrolyzed group and fully methoxylated chains. Signals that could be ascribed to cyclosilicates were absent. In the Q^2 range the homologous series of methoxylated chains containing 3Si (−93.4 ppm), 4Si (−93.6 ppm), 5Si (−93.8 ppm), 6Si (−94.0 ppm), and even more Si (a signal assigned to −94.3 ppm) was observed. The Q^2 signal of 5Si (−93.8 ppm) dominated after 0.5 h, while after 20 h, the most intense Q^2 signal was assigned to 6Si (−94.0 ppm). Thus, the linear oligomers grew on average about one silicate unit longer than in R-0.9, where the pentamer signal (−93,8 ppm) dominated in the spectra after 26, 31, and 64 h (Figure 3). As before, chains with hydrolyzed methoxy groups at Q^2 positions were present in R-1.3 sample. The signals around −92.7 and −93.0 ppm were assigned to hydrolyzed tetramer and pentamer, respectively. In the R-1.3 sample the signal of pentamer, hydrolyzed at Q^2, dominated the signals of hydrolyzed species after 0.5 h (Figure 10), while in the R-0.9 sample the tetramer was dominant (Figure 3). The signal at −92.4 ppm was assigned to hydrolyzed Q^2 in trimer. Whereas this signal was absent in R-0.9, in the R-1.3 sample with slightly higher water content the trimer was partially hydrolyzed at Q^2. Similar to R-0.9 the ^{29}Si NMR spectra of R-1.3 indicated no hydrolysis of end groups. The most significant difference between R-0.9 and R-1.3 samples was the higher relative intensity of Q^3 signals (region δ = −99.5 to −103.5 ppm) in R-1.3 (Figure 1 compared to Figure 9). A listing of ^{29}Si NMR signals in R-1.3, their assignment, and a comparison with R-0.9 are provided in Table 2.

In the UV-Raman spectra of sample R-1.3 the most prominent signal relevant to silicate speciation was at ∼510 cm^{-1} (Figure 12) confirming the abundant presence of pentamers and longer chains, in agreement with ^{29}Si NMR. Increasing cross-linking and network formation causes a change in Si−O−Si bond length and angle. These alterations may lead to a shift of the vibrational frequency. The maximum of the band at ∼510 cm^{-1} gradually shifted after 0.5 h to ∼500 cm^{-1} after 3 h and to 490 cm^{-1} after 20 h. In the R-1.3 sample the dimer signal (∼575 cm^{-1}) and the trimer and tetramer signals (∼550 cm^{-1}) were rather weak.

Reaction Pathways of Acid Catalyzed TMOS Polymerization at Low *r* Value. The detailed assignment of the silicate species obtained in acid catalyzed silica polymerization at low *r* value presented in the previous section allowed a detailed description of the reaction pathways. The first ^{29}Si NMR spectrum was recorded 30 min after preparing the R-0.9 sample.

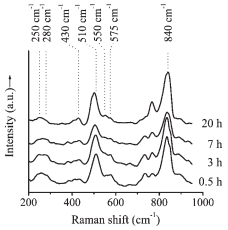


Figure 12. UV-Raman spectra of the condensation process of R-1.3.

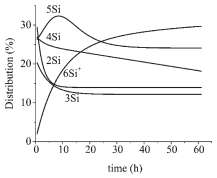


Figure 13. Evolution of the molar distribution (%) of silicate chain lengths in R-0.9 sample. Experimental data points were left out for clarity.

Less than 10% of all the added TMOS remained after 30 min. Apparently, in the first minutes, which for practical reasons could not be investigated, the reactions went rapidly. Hydrolysis is an acid catalyzed reaction. In acid catalyzed silica oligomerization departing from TMOS with substoichiometric quantities of water is a nucleophilic substitution reaction with a hydrolyzed Si−OH group acting as nucleophile and methoxy as a leaving group

$$\mathrm{Si{-}OH + CH_3{-}O{-}Si \rightarrow Si{-}O{-}Si + MeOH}$$

Under the described conditions the methoxy group will be protonated, facilitating elimination. The *r* value of 0.9 estimated from the content of siloxane and hydrolyzed groups (Figure 2 and Figure S) was in agreement with a conversion level of TMOS of around 90%. This consistency meant that very few water molecules were left unreacted in the sample.

With the recorded ^{29}Si NMR spectra Q^2 environments in differently sized chains were separately detected and quantified (Figure 10 and Table 2). Dimers could be quantified based on their Q^1 signal. With this information, the evolution of the distribution of differently sized silicate chains was calculated and is shown in Figure 13. The figure with experimental data points is given in the Supporting Information. Note that in this estimation the contribution of Q^3 silicons to chains was neglected. Information on the distribution of Q^3 silicons over differently sized chains could not be unambiguously separated.

In the first hours, the content of dimers (2Si) in R-0.9 decreased rapidly from ca. 30 mol % after 30 min to ca. 12% after 20 h. Initially present 3Si and 4Si chains disappeared in favor

The Journal of Physical Chemistry C ARTICLE

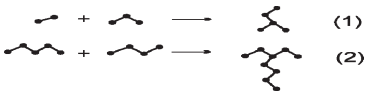


Figure 14. Examples of chain extension via condensation reactions between Q^1 and Q^2 groups. In reaction 1, dimer reacts with trimer into branched Si_5 oligomer; in reaction 2 tetramer and pentamer react together into branched Si_9 oligomer.

Table 4. Representative Silicate Oligomers of Acid Catalyzed TMOS Polycondensation

		Q_n ratio[a]				
sample	time (h)	Q^1_{Q3}	Q^1_{Q2}	Q^2	Q^3	representative structure[b]
R-0.9	0.5	2	6.4	6.8	1	model A
	6	1	2.3	4	1	model B
	57	0.4	0.7	1.7	1	model D
R-1.3	0.5		0.24[c]	1.26	1	model E
	20		0.13[c]	0.7	1	model F

[a] Q^1_{Q2} and Q^1_{Q3} represent Q^1 signals next to Q^2 and Q^3, respectively. [b] Presented in Figure 16. [c] $Q^1_{Q2} + Q^1_{Q3}$.

of chains with five and more Si atoms. The content of 5Si chains reached a maximum around 10 h. Later, chains with six and more Si atoms became dominant. Chain growth was not fed by monomer, since the monomer concentration remained constant. Although end-on condensation of short chains could account for part of the growth, the steady increase of the Q^3 content (Figure 2) revealed the importance of another type of growth proceeding via condensation reactions between Q^1 and Q^2 groups.

According to NMR, after 30 min reaction, hydrolyzed groups at Q^2 in silicon chains with four or more Si atoms were available for condensation reactions (Figure 3 and Figure 5). The majority of hydrolyzed groups at Q^1 silicons must have been more reactive and converted before the recording of the first ^{29}Si NMR spectrum.

Two examples of condensation reactions involving Q^1 and Q^2 are visualized in Figure 14. Addition of a dimer to the central Si atom of a trimer leads to the formation of a branched Si_5 molecule with a main chain composed of 4Si atoms (Figure 14). The condensation of a pentamer via a Q^2 group with a tetramer via a Q^1 group leads to a Si_9 oligomer with a main chain composed of 7Si atoms and a 2Si side chain. Such processes can explain the observed chain extension (Figure 13) in parallel with an increase of branchiness (Figure 2). An important consequence of this interpretation is that the chain lengths to which the ^{29}Si NMR and UV-Raman signals were assigned (Table 2 and Table 3) represent linear chains as well as main chains of branched oligomers.

The quantification of Si groups composing oligomers based on ^{29}Si NMR led us to propose representative oligomers at different stages of the aging of sample R-0.9 (Table 4 and Figure 16). A plot of experimental Q^1_{Q3} versus Q^3 concentration (Figure 15) was a convenient guide for construction of these representative species. For instance, in the ^{29}Si NMR spectrum of R-0.9 sample recorded after 30 min, the Q^1_{Q3}/Q^3 ratio equaled 2. The smallest

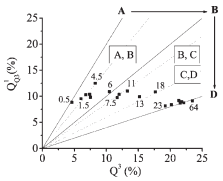


Figure 15. Q^1_{Q3} vs Q^3 content of R-0.9 sample during its evolution. The figures next to the data points refer to the time (h). Q^1_{Q3}/Q^3 ratio's characteristic of model structures A−D shown in Figure 16 and Table 4 are represented with lines.

branched oligomer that can be formed involves the condensation between a dimer and the central Si atom of a trimer (molecule A in Figure 16, Table 4). This pentamer has two Q^1_{Q3} groups and one Q^3. The 3D structure of this species is shown in Figure 17.

In the period between 0.5 and 20 h, hydrolyzed silicon atoms were observed at central positions of 4Si and 5Si chains (Figure 5). This period was characterized by the decrease of the content of dimer and trimer (Figure 13). Molecules B1 and B2 (Figure 16) are the result of condensation reactions between the end groups of trimer and the middle groups of tetramer and pentamer, respectively. In B1 and B2, the longest silica chain count six and seven Si atoms, respectively, explaining the experimental observation of an increasing number of such molecules (Figure 13). Similarly, reactions of end groups of dimers with middle groups of tetramers and pentamers will also lead to a branched specimen. A dominance of B type molecules in combination with some linear molecules can account for the Q^n distribution observed at 6 h in R-0.9, viz., Q^1_{Q3}:Q^1_{Q2}:Q^2: Q^3 of 1:2.3:4:1 (Table 4). Molecules of type B presenting 1 branching have experimental Q^1_{Q3} to Q^3 ratios of 1 (Figure 16), in excellent agreement with experimental observation (Figure 15 and Figure 13).

After 6 h, the content of dimer and trimer did not change any more (Figure 13). The content of 4Si and 5Si chains with a hydrolyzed Q^2 group continuously decreased in the period up to 20 h. Condensation reactions between these tetramers and pentamers and Q^1_{Q2} and Q^1_{Q3} groups of abundantly present molecules of type B led to type C molecules represented by molecules C1 and C2, respectively (Figure 16). The formation of molecules C presenting two branchings readily explained the observed decrease of the content of 4Si and 5Si chains in favor of 6Si and longer ($6Si^+$) chains (Figure 13).

After 20 h, almost no hydrolyzed Q^2 groups remained. ^{29}Si NMR revealed a further gradual decrease of the content of Q^1 and Q^2 silicons in favor of Q^3 (Figure 2). These changes could be interpreted as a result of condensation reactions between type C molecules leading to heavier, even more branched structures D, examples of which are shown in Figure 16 (molecules D1 and D2). At the end of the experiment of R-0.9 (~57 h), most of the silicon was incorporated into structures of type D, whereof the Q^1_{Q3}:Q^1_{Q2}:Q^2:Q^3 ratio matches the experimental ratio of 0.4:0.7:1.7:1 (Table 4). Molecules of D are characterized by a Q^1_{Q3} to Q^3 ratio of 0.5 (Figure 15). Model molecules A, B, and D represented key stages of the silica polycondensation obtained around 0.5, 6, and 57 h, respectively. In the plot of experimental

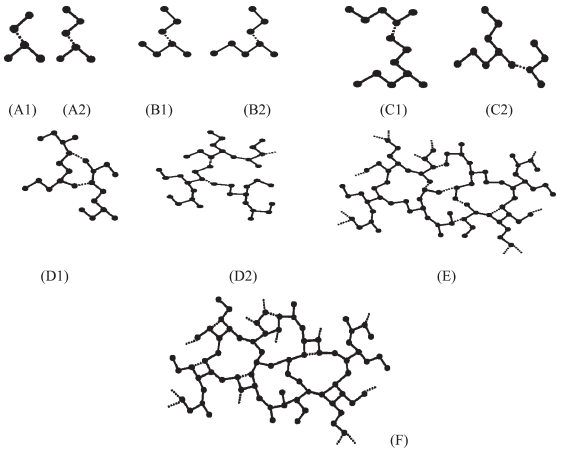


Figure 16. Condensation reactions (dotted lines) leading to the formation of representative molecules of the oligomerization process starting from TMOS under acid conditions and under stoichiometric water content (cf. Table 4).

Q^1_{Q3} versus Q^3 content (Figure 15) the actual data points clearly highlighted the gradual transition from molecules A to D.

In the periods 0.5 and 6 h the Q^1_{Q3}/Q^3 ratio shifted from 2, characteristic of molecules of A, to 1, characteristic of molecules of B. Between 6 and 23 h, the experimental Q^1_{Q3}/Q^3 ratios corresponded to a silica speciation containing B and C. An almost pure C population was reached around 13 h, and around 23 h NMR indicated sole presence of species D. According to the Q^1_{Q3}/Q^3 ratio, little evolution in the molecular structure was observed later, between 23 and 64 h.

On the basis of this insight into the development of silicate network reached in the R-0.9 sample, the evolution of R-1.3 sample could also be understood. According to the Q^n distribution sample R-1.3 was already at a further stage after 30 min compared to R-0.9 sample after 57 h (Table 4). With respect to Q^3 silicons, the content of Q^1 and Q^2 was already significantly lower, viz., 0.24 instead of 1.1 for Q^1 and 1.26 instead of 1.7 for Q^2 (Table 4). The Q^n ratio of the R-1.3 sample could be simulated by subjecting molecules D to inter- and intramolecular reactions and generation of molecules of type E (Figure 16) having a Q^1:Q^2:Q^3 ratio of 0.3:1.2:1 in excellent agreement with the ratios (Table 4). The evolution of the Q^n distribution of this sample (Figure 11) revealed that the evolution of the polymer involved mainly a transformation of Q^2 into Q^3 silicons. Intramolecular condensation of two Q^2 silicons of molecule E leads to molecule F, having Q^1:Q^2:Q^3 ratio of 0.2:0.6:1 matching with an experimental distribution observed at the final stage of this sample (Table 4).

These molecular models provided further insight into the Raman spectra. The Raman signal observed in R-1.3 sample around 550 cm^{-1}, a region originally assigned to isolated tetramers and trimers, originates from side chains of the polymer network represented in molecules E and F.

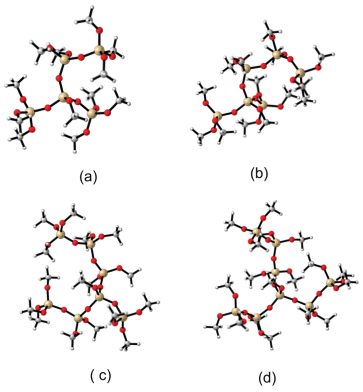
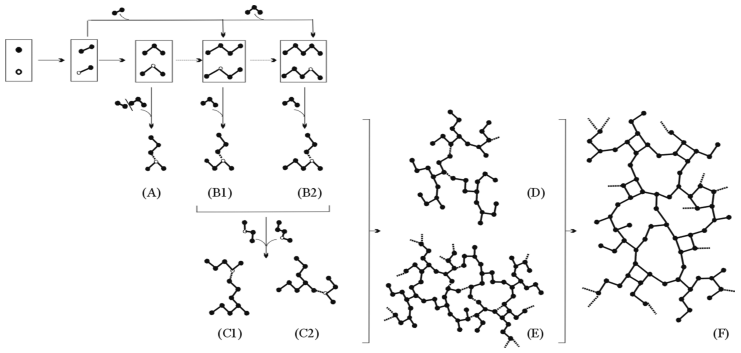


Figure 17. The lowest energy conformers of molecules A and B presented in Figure 16: A1 (a), A2 (b), B1 (c), and B2 (d).

Scheme 1. Acid Catalyzed Silica Oligomerization Scheme Departing from TMOS[a]



[a] Fully methoxide terminated and hydrolyzed silicon atoms are presented by solid (●) and open (○) dots, respectively.

Another intriguing aspect of this TMOS based sol−gel chemistry is the low tendency for formation of rings. According to the presented molecular structures A−F rings with four to six Si atoms are formed very late in the reaction network upon condensation of Q^2 silicons. The absence of Q^2 signals characteristic of cyclosilicates in the detailed NMR spectra of sample R-0.9 precluded the formation of rings in molecules A−D. Most likely, rings are formed with increasing number of intramolecular condensation reactions exemplified in molecules E and F. Note that these structures do not contain Q^2 silicon in their four-, five- and six-rings. In this model rings are almost entirely composed of Q^3 silicon.

The investigated samples did not evolve to the gel stage. Molecule E can already be thought of as a representative of the approached gel network. The presumed nature of the final gel network should then consist of a highly branched polymer, containing a limited amount of branched four-, five-, and six-rings. On the basis of our advanced analysis of the polymerization process, the here proposed final gel network shows striking similarities to the gel network proposed by Kelts and Armstrong for TMOS polymer grown at low pH.[49] The main difference between this work and that of Kelts and Armstrong is the absence of cylcosilicates as precursor units (Scheme 1). According to the presented observations, attachment of side chains is the main reaction pathway in TMOS based silica oligomerization. Surprisingly, this mechanism is significantly different from the reaction pathways observed in TEOS based silica oligomerization, where ring formation through cyclo-dimerization was recognized as important reaction pathway.[60] Steric and electronic factors most probably are at the origin of these differences.

The differences in polymerization pathways using TMOS and TEOS are expected to have an effect on the properties of the obtained silica gel. Departing from TMOS the gel network is characterized by a much higher branching degree and lower content of rings. The formation of microporous versus mesoporous silica gel depending on the use of TEOS versus TMOS[61] likely is due to these differences, but further research is needed to elucidate the details.

■ CONCLUSIONS

A detailed analysis of NMR spectra and interpretation of silica speciation supported by UV-Raman spectroscopy and molecular modeling enabled us to propose a detailed reaction scheme (Scheme 1) of the conversion of TMOS into silica gel network under conditions of acid catalysis and low molar hydrolysis ratio. At an r value of 0.9 TMOS readily hydrolyzed and oligomerized into short chains with three up to five Si atoms. Once most of the water was consumed, the condensations went slower. Residual hydrolyzed silicate groups were present almost exclusively at central positions of four and five Si chains. Characteristic stages of the building of the silica gel network could be identified and characteristic molecular structures were proposed (Figure 16). The condensations corresponded to an attachment of dimer and trimer to middle sections of short chains (molecules A and B, Figure 16). At the next stage tetramer and pentamer were attached via hydrolyzed middle groups to end groups (molecules C) until they were depleted. Intermolecular condensations via end and middle groups led to still larger structures D characterized by the absence of rings. A sample with r value of 1.3 evolved much more rapidly and reached a further developed polymer structure already after 30 min. This sample had a highly branched silicate network lacking fully condensed silicon atoms and containing only a limited amount of four-, five-, and six-membered rings.

This paper adds to the assignment of both ^{29}Si NMR and Raman signals. ^{29}Si NMR signals at $\delta = -92.4$, -92.72, -92.96, and -93.2 ppm were assigned to hydrolyzed Q^2 groups in chains of three, four, five, and six or more Si atoms, respectively. Signals at -93.61, -93.80, -94.01, and -94.26 ppm were assigned to fully methoxylated chains of four, five, and six or more Si atoms either isolated or part of a larger network. Distinction between ^{29}Si NMR signals of Q^1 end groups next to Q^2 and Q^3 is revealed in NMR signatures at $\delta = -85.69$ and -85.91, respectively. Raman signatures at ∼555, ∼540, and ∼510 cm^{-1} were assigned to chains of three, four, five, and more Si atoms.

■ ASSOCIATED CONTENT

Supporting Information. Full *XYZ* coordinates of all optimized structures for the molecular modeling of Raman intensities, Raman frequencies, and ^{29}Si NMR chemical shifts, correlation graphs for Raman and NMR scale factor determination, and complete ref 65. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: Johan.Martens@biw.kuleuven.be.

■ ACKNOWLEDGMENT

A.D. acknowledges the Research Foundation Flanders (FWO) for a research fellowship. Theoretical work was supported by the Research Board of Ghent University (BOF) and BELSPO, through Interuniversity Attraction Poles (IAP-PAI P6/27). We greatly acknowledge Professor Tatjana Parac-Vogt and Karel Duerinckx for ^{29}Si NMR measurements on a Bruker Avance II^{+} 600 spectrometer (Center for Molecular Design and Synthesis at K.U. Leuven, Belgium). J.A.M. and C.E.A.K. acknowledge the Flemish Government for long-term structural funding (Methusalem), the Belgian Prodex office and ESA. V.V.S acknowledges the European Research Council under the European Community's Seventh Framework Programme (FP7 (2007-2013) ERC Grant Agreement Number 240483).

■ REFERENCES

(1) Rouse, J. H.; Ferguson, G. S. *J. Am. Chem. Soc.* **2003**, *125*, 15529.
(2) Shin, J. H.; Metzger, S. K.; Schoenfisch, M. H. *J. Am. Chem. Soc.* **2007**, *129*, 4612.
(3) Uemura, T.; Hiramatsu, D.; Yoshida, K.; Isoda, S.; Kitagawa, S. *J. Am. Chem. Soc.* **2008**, *130*, 9216.
(4) Wang, J.; Pamidi, P. V. A.; Zanette, D. R. *J. Am. Chem. Soc.* **1998**, *120*, 5852.
(5) Livage, J. *Chem. Mag. (Den Haag)* **1990**, 67.
(6) Livage, J.; Babonneau, F.; Sanchez, C. *Sol-gel optics*; Kluwer Pub., 1994.
(7) Livage, J.; Ganguli, D. *Sol. Energy Mater. Sol. Cells* **2001**, *68*, 365.
(8) Aelion, R.; Loebel, A.; Eirich, F. *J. Am. Chem. Soc.* **1950**, *72*, 5705.
(9) Assink, R. A.; Kay, B. D. *J. Non-Cryst. Solids* **1988**, *99*, 359.
(10) Bernards, T. N. M.; Vanbommel, M. J.; Boonstra, A. H. *J. Non-Cryst. Solids* **1991**, *134*, 1.
(11) Bertoluzza, A.; Fagnano, C.; Antonietta Morelli, M.; Gottardi, V.; Guglielmi, M. *J. Non-Cryst. Solids* **1982**, *48*, 117.
(12) Brinker, C.; Scherer, G. *Sol-Gel Science: The Physics and Chemistry of Sol-Gel Processing*; Academic Press: Boston, MA, 1990.
(13) Colomban, P. *J. Raman Spectrosc.* **1996**, *27*, 747.
(14) James, P. F. *J. Non-Cryst. Solids* **1988**, *100*, 93.
(15) Lin, C. C.; Basil, J. D. *Mater. Res. Soc. Symp. Proc.* **1986**, *73*, 585.
(16) Matsuyama, I.; Satoh, S.; Katsumoto, M.; Susa, K. *J. Non-Cryst. Solids* **1991**, *135*, 22.
(17) Mulder, C. A. M.; Damen, A. *J. Non-Cryst. Solids* **1987**, *93*, 169.
(18) Pouxviel, J. C.; Boilot, J. P.; Beloeil, J. C.; Lallemand, J. Y. *J. Non-Cryst. Solids* **1987**, *89*, 345.
(19) Livage, J. *Stud. Surf. Sci. Catal.* **1994**, *85*, 1.
(20) Aegerter, M. A.; Mennig, M. *Sol-gel technologies for glass producers and users*; Kluwer Academic Publishers: Boston, MA, 2004.
(21) Livage, J.; Beteille, F.; Roux, C.; Chatry, M.; Davidson, P. *Acta Mater.* **1998**, *46*, 743.
(22) Livage, J.; Sanchez, C. *J. Non-Cryst. Solids* **1992**, *145*, 11.
(23) Nabavi, M.; Doeuff, S.; Sanchez, C.; Livage, J. *J. Non-Cryst. Solids* **1990**, *121*, 31.
(24) Artaki, I.; Bradley, M.; Zerda, T. W.; Jonas, J. *J. Phys. Chem.* **1985**, *89*, 4399.
(25) Brinker, C. J.; Keefer, K. D.; Schaefer, D. W.; Ashley, C. S. *J. Non-Cryst. Solids* **1982**, *48*, 47.
(26) Petry, D. P.; Haouas, M.; Wong, S. C. C.; Aerts, A.; Kirschhock, C. E. A.; Martens, J. A.; Gaskell, S. J.; Anderson, M. W.; Taulelle, F. *J. Phys. Chem. C* **2006**, *113*, 20827–20836.
(27) Marino, I.-G.; Lottici, P. P.; Bersani, D.; Raschellà, R.; Lorenzi, A.; Montenero, A. *J. Non-Cryst. Solids* **2005**, *351*, 495.
(28) Mora-Fonz, M. J.; Catlow, C. R. A.; Lewis, D. W. *J. Phys. Chem. C* **2007**, *111*, 18155.
(29) Trinh, T. T.; Jansen, A. P. J.; van Santen, R. A. *J. Phys. Chem. B* **2006**, *110*, 23099.
(30) Zerda, T. W.; Artaki, I.; Jonas, J. *J. Non-Cryst. Solids* **1986**, *81*, 365.
(31) Yu, Y.; Xiong, G.; Li, C.; Xiao, F.-S. *J. Catal.* **2000**, *194*, 487.
(32) Dutta, P. K.; Shieh, D. C.; Puri, M. *Zeolites* **1988**, *8*, 306.
(33) Knops-Gerrits, P.-P.; De Vos, D. E.; Feijen, E. J. P.; Jacobs, P. A. *Microporous Mater.* **1997**, *8*, 3.
(34) Dutta, P. K.; Rao, K. M.; Park, J. Y. *J. Phys. Chem.* **1991**, *95*, 6654.
(35) De Man, A. J. M.; Van Santen, R. A. *Zeolites* **1992**, *12*, 269.
(36) Russel, P. E. *Microbeam analysis*; San Fransisco Press: San Fransisco, CA, 1989.
(37) Bornhauser, P.; Bougeard, D. *J. Raman Spectrosc.* **2001**, *32*, 279.
(38) Mozgawa, W. *J. Mol. Struct.* **2001**, *596*, 129.
(39) Dutta, P. K.; Del Barco, B. *J. Phys. Chem.* **1985**, *89*, 1861.
(40) Dutta, P. K.; Del Barco, B. *J. Phys. Chem.* **1988**, *92*, 354.
(41) Dutta, P. K.; Twu, J. *J. Phys. Chem.* **1991**, *95*, 2498.
(42) Bremard, C.; Le Maire, M. *J. Phys. Chem.* **1993**, *97*, 9695.
(43) Mozgawa, W.; Bajda, T. *J. Mol. Struct.* **2006**, *792*−*793*, 170.
(44) Mozgawa, W.; Jastrzebski, W.; Handke, M. *J. Mol. Struct.* **2006**, *792*−*793*, 163.
(45) Krause, K.; Geidel, E.; Kindler, J.; Förster, H.; Smirnov, K. S. *Vib. Spectrosc.* **1996**, *12*, 45.
(46) Pavel, C. C.; Zibrowius, B.; Loffler, E.; Schmidt, W. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3440.
(47) Brinker, C. J.; Assink, R. A. *J. Non-Cryst. Solids* **1989**, *111*, 48.
(48) Brunet, F.; Cabane, B. *J. Non-Cryst. Solids* **1993**, *163*, 211.
(49) Kelts, L. W.; Armstrong, N. J. *J. Mater. Res.* **1989**, *4*, 423.
(50) Kelts, L. W.; Effinger, N. J.; Melpolder, S. M. *J. Non-Cryst. Solids* **1986**, *83*, 353.
(51) Lippert, J. L.; Melpolder, S. B.; Kelts, L. M. *J. Non-Cryst. Solids* **1988**, *104*, 139.
(52) Bailey, J. K.; Macosko, C. W.; Mecartney, M. L. *J. Non-Cryst. Solids* **1990**, *125*, 208.
(53) Brus, J.; Karhan, J.; Kotlik, P. *Collect. Czech. Chem. Commun.* **1996**, *61*, 691.
(54) Balfe, C. A.; Martinez, S. L. *Mater. Res. Soc. Symp. Proc.* **1986**, *73*, 27.
(55) Engelhardt, G.; Michel, D. *High-resolution solid-state NMR of silicates and zeolites*; John Wiley and Sons Ltd.: Chichester and New York, 1987.
(56) Klemperer, W. G.; Ramamurthi, S. D. *J. Non-Cryst. Solids* **1990**, *121*, 16.
(57) Ng, L. V.; Thompson, P.; Sanchez, J.; Macosko, C. W.; McCormick, A. V. *Macromolecules* **1995**, *28*, 6471.
(58) Sanchez, J.; McCormick, A. *J. Phys. Chem.* **1992**, *96*, 8973.
(59) Pienaar, C. L.; Chiffoleau, G. J. A.; Follens, L. R. A.; Martens, J. A.; Kirschhock, C. E. A.; Steinberg, T. A. *Chem. Mater.* **2007**, *19*, 660.
(60) Depla, A.; Lesthaeghe, D.; van Erp, T. S.; Aerts, A.; Houthoofd, K.; Fan, F.; Li, C.; Van Speybroeck, V.; Waroquier, M.; Kirschhock, C. E. A.; Martens, J. A. *J. Phys. Chem. C* **2011**accepted.
(61) Depla, A.; Kirschhock, C.; Martens, J. *Stud. Surf. Sci. Catal.* **2010**, *175*, 801.

(62) Maier, W. F.; Tilgner, I. C.; Wiedorn, M.; Ko, H. C. *Adv. Mater.* **1993**, *5*, 726.
(63) Lazzeri, M.; Mauri, F. *Phys. Rev. Lett.* **2003**, 90.
(64) Casanovas, J.; Illas, F.; Pacchioni, G. *Chem. Phys. Lett.* **2000**, *326*, 523.
(65) Frisch, M. J.; et al. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2003 (For complete list of authors, see Supporting Information).
(66) Van Houteghem, M.; Verstraelen, T.; Van Neck, D.; Kirschhock, C. E. A.; Martens, J. A.; Waroquier, M.; Van Speybroeck, V. *J. Chem. Theory Comput.* **2011**, *7*, 1045.
(67) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(68) Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683.
(69) Sykes, D.; Kubicki, J. D.; Farrar, T. C. *J. Phys. Chem. A* **1997**, *101*, 2715.
(70) Mennucci, B.; Martinez, J. M.; Tomasi, J. *J. Phys. Chem. A* **2001**, *105*, 7287.
(71) Cho, H.; Felmy, A. R.; Craciun, R.; Keenum, J. P.; Shah, N.; Dixon, D. A. *J. Am. Chem. Soc.* **2006**, *128*, 2324.
(72) Butler, K.; Slater, B.; Lewis, D. W. *Zeolites and Related Materials: Trends, Targets and Challenges: Proceedings of the 4th International FEZA Conference*; Studies in surface science and catalysis 174; Elsevier: Amsterdam and London, 2008; p 725.
(73) Moravetski, V.; Hill, J. R.; Eichler, U.; Cheetham, A. K.; Sauer, J. *J. Am. Chem. Soc.* **1996**, *118*, 13015.
(74) Van Speybroeck, V.; Moonen, K.; Hemelsoet, K.; Stevens, C. V.; Waroquier, M. *J. Am. Chem. Soc.* **2006**, *128*, 8468.
(75) Verstraelen, T.; Szyja, B. M.; Lesthaeghe, D.; Declerck, R.; Van Speybroeck, V.; Waroquier, M.; Jansen, A. P. J.; Aerts, A.; Follens, L. R. A.; Martens, J. A.; Kirschhock, C. E. A.; van Santen, R. A. *Top. Catal.* **2009**, *52*, 1261.
(76) Brunet, F.; Cabane, B.; Dubois, M.; Perly, B. *J. Phys. Chem.* **1991**, *95*, 945.
(77) Gnado, J.; Dhamelincourt, P.; Pelegris, C.; Traisnel, M.; Mayot, A. L. *J. Non-Cryst. Solids* **1996**, *208*, 247.
(78) Matos, M. C.; Ilharco, L. M.; Almeida, R. M. *J. Non-Cryst. Solids* **1992**, *147−148*, 232.

List of publications

Updated October 2013

Publications in international peer-reviewed journals

I. M. Van Houteghem, T. Verstraelen, D. Van Neck, C. Kirschhock, J.A. Martens, M. Waroquier, V. Van Speybroeck
Atomic Velocity Projection Method: A New Analysis Method for Vibrational Spectra in Terms of Internal Coordinates for a Better Understanding of Zeolite Nanogrowth
J. Chem. Theory Comput., 7 (4), 1045-1061, 2011

II. M. Van Houteghem, T. Verstraelen, A. Ghysels, L. Vanduyfhuys, M. Waroquier, V. Van Speybroeck
Analysis of the basis set superposition error in molecular dynamics of hydrogen-bonded liquids: application to methanol
J. Chem. Phys., 137 , 104506, 2012

III. M. Van Houteghem, A. Ghysels, T. Verstraelen, W. Poelmans, M. Waroquier, V. Van Speybroeck
Critical analysis on the accuracy of models predicting or extracting information on liquid structure
submitted

IV. T. Verstraelen, M. Van Houteghem, V. Van Speybroeck, M. Waroquier
MD-TRACKS: A Productive Solution for the Advanced Analysis of Molecular Dynamics and Monte Carlo simulations
J. Chem. Inf. Model., 48 (12), 2414-2424, 2008

V. A. Depla, E. Verheyen, A. Verfeyken, M. Van Houteghem, K. Houthoofd, V. Van Speybroeck, M. Waroquier, C. Kirschhock, J.A. Martens
UV-Raman and ^{29}Si NMR Spectroscopy Investigation of the Nature of Silicate Oligomers Formed by Acid Catalyzed Hydrolysis and Polycondensation of Tetramethylorthosilicate
J. Phys. Chem. C, 115 (22), 11077-11088, 2011