From the inhomogeneous electron gas to classical force fields: a multi-scale model for Ionic Liquids

Von der Fakultät Mathematik und Physik der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von
Florian Dommert
aus Dachau

Hauptberichter: Prof. Dr. Christian Holm
Mitberichter: JP Dr. Johannes Kästner

Tag der mündlichen Prüfung: 30.1.2013

Institut für Computerphysik der Universität Stuttgart

2013

Contents

List of Figures

List of Tables

List of Publications

1. F. Dommert, J. Schmidt, B. Qiao, Y. Zhao, C. Krekeler, L. Delle Site, R. Berger and C. Holm, *J. Chem. Phys.*, 2008, **129**, 224501
2. C. Krekeler, F. Dommert, J. Schmidt, Y. Y. Zhao, C. Holm, R. Berger and L. Delle Site, *Phys. Chem. Chem. Phys.*, 2010, **12**, 1817–1821
3. J. Schmidt, C. Krekeler, F. Dommert, Y. Zhao, R. Berger, L. Delle Site and C. Holm, *J. Phys. Chem. B*, 2010, **114**, 6150–6155
4. F. Dommert, J. Schmidt, C. Krekeler, Y. Y. Zhao, R. Berger, L. Delle Site and C. Holm, *J. Mol. Liq.*, 2010, **152**, 2–8
5. H. Wang, F. Dommert and C. Holm, *J. Chem. Phys.*, 2010, **133**, 034117
6. K. Wendler, S. Zahn, F. Dommert, R. Berger, C. Holm, B. Kirchner and L. Delle Site, *J. Chem. Theory Comput.*, 2011, **7**, 3040–3044
7. K. Wendler, F. Dommert, Y. Y. Zhao, R. Berger, C. Holm and L. Delle Site, *Faraday Discuss.*, 2012, **154**, 111–132
8. F. Dommert, K. Wendler, R. Berger, L. Delle Site and C. Holm, *ChemPhysChem*, 2012, **13**, 1625–1637
9. F. Dommert and C. Holm, *Phys. Chem. Chem. Phys.*, 2013, **15**, 2037–2049

1. Zusammenfassung

Im letzten Jahrzehnt sind ionische Flüssigkeiten, wie etwa Imidazolium-basierte Salze, immer stärker in den Blickpunkt der Forschung geraten, vor allem jene, deren Schmelztemperatur unterhalb der Raumtemperatur liegt. Um diese Klasse von Substanzen besser verstehen zu können, wurde im Rahmen des DFG SPP 1191 ein Ansatz verfolgt, der die verschiedenen Orts- und Zeitskalen miteinander verknüpft. Dabei wurde speziell darauf geachtet möglichst unabhängig von experimentellen Resultaten bei der Modellierung der Salze zu bleiben. Denn das finale Ziel ist es, ionischen Flüssigkeiten mit dem Computer zu *designen*. Während in der Arbeitsgruppe von Robert Berger die Systeme mit quantenchemischen Methoden charakterisiert wurde, schufen diese Resultate die Grundlage für eine semi-quantenmechanische dynamische Beschreibung kleiner Systeme in der Arbeitsgruppe von Luigi Delle Site.

In dieser Arbeit kommt es nun zu einer Verknüpfung der semi-quantenmechanischen mit der klassischen Skala, die es schließlich zulässt thermodynamische Eigenschaften zu berechnen, wie zum Beispiel die Leitfähigkeit oder Viskosität. Hiefür werden Methoden entwickelt, um die Energien und Kräfte in Potentialen zu parametrisieren, damit die Koordinaten und Geschwindigkeiten der Atome mittels Newton'scher Mechanik mit dem Computer effizient propagiert werden können.

Dafür werden zunächst die Konzepte der klassischen Molekulardynamik vorgestellt und Methoden zu deren Effizienzsteigerung weiterentwickelt. Danach wird der Prozess der Kraftfeldoptimierung, der vor allem die Partialladungen der Atome und deren Parameter für kurzreichweitige Wechselwirkungen betrifft, detailliert diskutiert. Schließlich wird die Methodik zur Kraftfeldvalidierung erläutert, die sicher stellen soll, dass ein akkurates und zuverlässiges Kraftfeld vorliegt. Die dafür notwendigen Protokolle waren entweder schon teilweise in Programmen vorhanden oder wurden, wenn nötig, implementiert, womit der Weg für eine Bestimmung der Parameter für beliebige Systeme für die Zukunft geebnet ist.

Im Rahmen dieses Projektes wurde schon eine Reihe von Artikeln veröffentlicht,[1,7,9–17] die diese Verfahren vorstellen und Anwendungen aufzeigen. In dieser Arbeit wird jedoch noch konkreter auf die Problematik eingegangen und weitere Resultate werden vorgestellt. Zudem wird dem Leser alle nötige Information, von der Theorie der Molekulardynamik-Simulation bis zur Validierung der Ergebnisse, präsentiert, damit die Vorgehensweise leicht verständlich wird und das nötige Wissen vermittelt wird, um Kraftfelder für klassische Molekulardynamik zu optimieren.

1.1. Die Prinzipien der klassischen Molekulardynamik

In der klassische Molekulardynamik (MD) wird ein Ensemble von Teilchen mit Hilfe von Newtons Bewegungsgleichungen unter verschiedenen Annahmen in Raum und Zeit propagiert. Eine davon ist die Born-Oppenheimer-Näherung, die vorhersagt, dass die Koordinaten des Atoms durch die Kernkoordinaten beschrieben werden können. Darüberhinaus ist ein Kraftfeld notwendig, das die interatomaren Energien adäquat beschreibt, woraus letztendlich Kräfte berechnet werden, die für die Integration der Bewegungsgleichungen nötig sind. Um den iterativen Prozess der Integration effizient auf Computern zu ermöglichen, wurden verschiedene Ansätze entwickelt, deren Verständnis für die Durchführung von MD-Simulationen notwendig ist. Im Folgenden werden die für diese Arbeit relevanten Methoden vorgestellt. Desweiteren wird eine Technik zur Berechnung von elektrostatischen Kräften präsentiert und weiterentwick-

elt, um die Simulationen in Hinblick auf Genauigkeit und Effizienz zu optimieren.[13]

Simulation verschiedener thermodynamischer Ensembles Im Laufe der Zeit wurden verschiedene Schemata vorgestellt, die eine Integration der Bewegungsgleichungen erlauben. Dabei wurden diese sogenannten Integratoren durch verschiedene Ansätze hergeleitet, doch in den letzten Jahren hat sich der Trotter-Formalismus etabliert, der eine einfache Verallgemeinerung der verschiedenen Integrationskonzepte zulässt. Dieser basiert auf dem Theorem von Trotter, das aussagt, dass die Propagation der Koordination im Phasenraum durch eine Operation innerhalb einer Semigruppe dargestellt werden kann und die schrittweise Zeitevolution verdeutlicht. Um diese Evolution der Koordinaten auch zeitreversibel durchzuführen, wurde eine symmetrische Zerlegung der Hamiltonfunktion von Kreutz und Gocksch vorgeschlagen. Mit diesem Formalismus lässt sich jedes Integrationsschema leicht darstellen, da verschiedene Schritte der Integration Teilen der Hamiltonfunktion entsprechen. Darüber hinaus eröffnet sich die Möglichkeit zu einer Fehleranalyse des Integrationsschemas und Beurteilung des Verhaltens des Integrators bezüglich chaotischen Verhaltens. Das ist von großer Bedeutung, denn die Analyse einer MD Simulation stützt sich auf das ergodische Theorem, welches ein dichtes Sampling des Phasenraums voraussetzt, um einen Ensemblemittelwert über eine Zeitmittelung zu erhalten. Ein chaotisches Verhalten ist nötig, um zeitabhängige Größen aus einer langen Trajektorie zu berechnen, weil nach einer bestimmten Anzahl von Integrationsschritten die Konfiguration unabhängig von der Ausgangskonfiguration ist.

Da eine Vielzahl von Integratoren existiert, werden hier nur der Verlet-, Leap-Frog- und Velocity-Verlet-Algorithmus diskutiert, da jene am weitesten verbreitet und für diese Arbeit relevant sind. Der Velocity-Verlet-Integrator korrespondiert komplett zur symmetrischen Zerlegung der Hamiltonfunktion und ist somit zeitreversibel. Zudem ist dieser symplektisch, was bedeutet, dass das Volumen des Phasenraums bei der Integration erhalten bleibt, womit ein ergodisches Sampling ermöglicht wird. Bei einem Vergleich der erwähnten Integratoren untereinander stellt sich heraus, dass alle eine äquivalente Propagation der Ortskoordinaten liefern. Damit eignen sich alle für die Simulation eines mikrokanonischen Ensembles.

Unterschiede tauchen jedoch bei der Berechnung der Teilchengeschwindigkeiten auf. Während keine explizite Berechnung im Verlet Algorithmus auftaucht, werden die Geschwindigkeiten der Teilchen im Leap-Frog- und Velocity-Verlet-Schema benutzt, um die Bewegungsgleichungen zu integrieren. Das erlaubt die Simulation eines kanonischen NVT-Ensembles, da die Temperatur des Systems durch die Geschwindigkeiten der Teilchen bekannt ist. Um die Temperatur konstant zu halten, wurden verschiedene Methoden vorgeschlagen, die alle auf eine Skalierung der Geschwindigkeiten abzielen. Aber es hat sich gezeigt, dass eine einfache Skalierung der Geschwindigkeiten gemäß einer vorgegebenen Relaxationszeit kein kanonisches Ensemble generiert, sondern ein weiterer stochastischer Term addiert werden muss, um die erwartete Geschwindigkeitsverteilung zu erhalten. Ein anderer Ansatz, genannt extended-Lagrangian, beruht auf einer Skalierung der Geschwindigkeiten, wofür Bewegungsgleichungen durch eine Pseudo-Lagrangefunktion, die durch bestimmte Erhaltungssätze definiert ist, gegeben sind. Im Fall eines kanonischen NVT-Ensembles gilt es verschiedene Virialtheoreme zu erfüllen, was durch weitere generalisierte Koordinaten und entsprechenden Lagrange-Multiplikator festgelegt wird, die letztendlich den Skalierungsfaktor für die Geschwindigkeit bestimmen.

Geht man einen Schritt weiter und zielt auf die Simulation eines NpT-Ensembles ab, wird

die Gibbsche Freie Energie erhalten, womit durch zusätzliche Lagrange-Multiplikatoren weitere Bewegungsgleichungen für die Vektoren der Simulationsbox konstituiert werden. Da bei der Integration durch Velocity-Verlet die Orte und Geschwindigkeiten zum gleichen Zeitpunkt bekannt sind und nicht um einen halben Zeitschritt verschoben, wie im Falle von Leap-Frog, eignet sich dieser als einziger, die vorgegebenen Bewegungsgleichungen für das NVT- und NPT-Ensemble exakt zu integrieren. Jedoch werden, im Vergleich zu Leap-Frog, mehr Kommunikationsschritte bei einer parallelen Version auf einem Supercomputer zwischen dem verteiltem Speicher benötigt. Da auf einem Computer immer numerische Fehler auftreten und die Abweichung der Integration durch Leap-Frog verglichen zu Velocity-Verlet mit der reziproken Teilchenzahl abfallen, eignet sich somit auch Leap-Frog für Systeme mit großen Teilchenzahlen und kann Velocity-Verlet vorgezogen werden. Nachdem nun Leap-Frog als geeigneter Integrator identifiziert wurde, wird eine Methode vorgestellt und weiterentwickelt, die es erlaubt die Berechnung der elektrostatischen Kräfte unter periodischen Randbedingung in Hinblick auf Genauigkeit und Geschwindigkeit zu optimieren.

Optimierung der Simulationsparameter Da die Anpassung des Kraftfeldes und dessen Verifizierung eine sehr große Anzahl an MD-Simulationen benötigt, spielt Effizienz und Genauigkeit eine entscheidende Rolle. Während die Berechnung der kurzreichweitigen und gebunden Wechselwirkungen nicht viel Spielraum zur Optimierung läßt, existieren Möglichkeiten den Rechenaufwand für die langreichweitigen Elektrostatik effizient zu gestalten.

In dieser Arbeit werden fast ausschließlich Systeme unter periodischen Randbedingungen untersucht, um Randeffekte so klein wie möglich zu halten. Daher bieten sich sogenannte Ewald-Methoden zur Berechnung der Elektrostatik an. Dabei wird bei der Summation der Coulombenergie im realen und reziproken Raum nur eine endliche Anzahl von Termen berücksichtigt. Zwar konvergiert diese Summe nur bedingt, aber bei geeigneter Wahl der Parameter sehr schnell. Jedoch ist der Raum der Parameter sehr groß, womit eine sorgfältige Auswahl derer in Hinblick auf Effizienz und Genauigkeit erschwert wird.

Erst kürzlich wurde eine Verallgemeinerung derartiger Methoden, die auf einer Interpolation der Ladung auf einem Gitter basieren, zugänglich. Diese erlaubt zudem für diese Art von Algorithmen eine a-priori Fehlerabschätzung abzuleiten. Das ermöglichte, die vorhandene Fehlerapproximation des Smooth-Particle-Mesh-Ewald-Algorithmus im Rahmen dieser Arbeit auf ungerade Interpolationsordnungen zu erweitern. Ein entsprechendes Programm, genannt `g_pme_error`, wurde für GROMACS implementiert, das die Fehlerberechnung erlaubt und die Parameter in Hinsicht auf Genauigkeit optimiert. Damit ist nun ein wichtiges Werkzeug für das Programmpaket GROMACS vorhanden, um ionische Flüssigkeiten auf atomarer Ebene effizient und akkurat zu simulieren, da elektrostatische Kräfte eine große Rolle spielen.

Das Kraftfeld Um die Bewegungsgleichungen eines Systems von Teilchen auf einem Computer zu lösen, ist nicht nur ein passender Integrationsalgorithmus notwendig, sondern auch eine adäquate Parametrisierung der Energie des System, die zur Berechnung der Kräfte dient. Leider gibt es keine eindeutige Lösung für das Problem, die interatomaren Wechselwirkungen auf klassische Näherungen abzubilden. Jedoch haben sich bestimmte Formen der Parametrisierungen etabliert, da sie nicht nur ein mathematisch praktisches, sondern auch physikalisch sinnvolles Modell konstatieren.

Im Allgemeinen unterscheidet man bei einem Kraftfeld zwischen zwei Arten von Wechselwirkungen. Auf der einen Seite stehen die Atombindungen, die das Molekül zusammenhalten, und auf der anderen Seite die intermolekularen Wechselwirkungen, die sich aus der Ladung der Atome, Dispersions- und Polarisationseffekten ergeben. Mit wachsender Molekülgröße spielen auch auf intermolekularer Ebene letztere eine Rolle und es muss darauf geachtet werden, dass die Parameter sorgfältig aufeinander abgestimmt sind. Zudem sollte sichergestellt werden, dass die Parameter auf verschiedene Bereiche des Phasenraums übertragbar und mit einer möglichst großen Anzahl von Molekülen kompatibel sind. Um dieser Vielzahl an Anforderungen gerecht zu werden, ist eine sorgfältige Konstruktion und Validierung des Kraftfelds immer notwendig.

Für die Bestimmung der intermolekularen Parametrisierung sind oft computergestützte quantenmechanische Berechnungen ausreichend und es ist eine relativ eindeutige Methodik vorhanden. Jedoch lassen die Methoden zur Bestimmung der restlichen Parameter sehr viel Spielraum. Allein die Berechnung der Partialladungen der Atome kann auf sehr viele verschiedene Arten erfolgen, wobei es im Voraus sehr schwer zu beurteilen ist, welche der Methoden den genauen Ladungszustand des Moleküls am besten widerspiegeln kann.

Noch schwieriger ist es, die kurzreichweitigen Wechselwirkungen zu parametrisieren, denn im Gegensatz zu den Partialladungen, die ausschließlich elektrostatische Wechselwirkung beschreiben, muss die kurzreichweitige Parametrisierung verschiedene Effekte beschreiben wie Polarisation, Dispersion und Pauli-Abstoßung. Um diese Effekte in einem effektiven Potential zu verallgemeinern, werden für gewöhnlich viele experimentelle Resultate benötigt, womit das Kraftfeld schließlich angepaßt wird. Daraus ergibt sich, dass bereits ein bestimmtes Maß an Informationen vorhanden sein muss, was prädiktive Aussagen erschwert und teils sogar unmöglich macht.

Aus diesem Grund stützt sich die Kraftfeldparametrisierung in dieser Arbeit allein auf die experimentelle Massendichte. Alle weiteren nötigen Daten werden mit Simulationen bestimmt, für die keine weiteren experimentellen Befunde nötig sind. Das erlaubt auch Systeme wie ionische Flüssigkeiten zu parametrisieren, über die oft nur ein geringer Schatz an Wissen vorhanden ist.

Tauglichkeitstest von Kraftfeldern für ionische Flüssigkeiten Obwohl ionische Flüssigkeiten schon seit einiger Zeit mit MD-Simulationen untersucht werden, sind nur wenige Kraftfelder verfügbar, die sowohl akkurat, als auch zwischen verschiedenen Kombinationen von Anionen und Kationen transferierbar sind. Da diese Kraftfelder oft nicht ausreichend getestet sind, wurden zwei verschiedene Kraftfelder verglichen, die ein relativ großes Repertoire an ionischen Flüssigkeiten modellieren, und deren Konstruktion auf verschiedenen Methoden basiert.

Es wurden verschiedene statische und dynamische Größen untersucht, wobei sich herausstellte, dass die dynamischen Eigenschaften wie Diffusion oder Leitfähigkeit stark unterschätzt wurden. Ein Grund dafür ist, dass beide Kraftfelder vollständig im Vakuum oder der Kristallphase parametrisiert wurden und keinerlei Information über die Flüssigphase in die Kraftfeldpotentiale eingegangen ist. Es hat sich aber gezeigt, dass der Polarisationzustand in der kondensierten Phase, welche bei den untersuchten Kraftfeldern komplett außen vor bleibt, vor allem für die Dynamik eine entscheidende Rolle spielt.

Jedoch diente die Untersuchung nicht allein zur Validierung der Kraftfelder, sondern es wurden auch verschiedene Methoden zur Berechnung der unterschiedlichen Größen angewandt, um daraus die effizientesten zu identifizieren. Dabei stellt sich heraus, dass die Einstein-Helfand-Methode eine genaue Bestimmung der Leitfähigkeit erlaubt, ohne den Weg über die Integration der Stromautokorrelationsfunktion zu gehen, der sehr viel Statistik und eine hochfrequente Speicherung der Geschwindigkeiten verlangt.

Da sich die Leitfähigkeit als kollektive dynamische Systemeigenschaft sehr gut zur Kraftfeldvalidierung eignet und ein sorgsamer Umgang mit Ressourcen aufgrund der Masse an anfallenden Simulationen nötig ist, wurde die genannten Methoden in das Programm `g_current` implementiert. Dieses wurde Teil des GROMACS-Programmpaketes, das in dieser Arbeit ausschließlich für die klassischen MD Simulationen verwendet wird.

1.2. Kraftfeldladungen von und für die flüssige Phase

Das Hauptziel dieser Arbeit ist, Techniken und Programme zu erarbeiten, die eine Konstruktion und Optimierung eines Kraftfeldes erlauben. Um mit minimaler experimenteller Information eine entsprechende Parametrisierung vorzunehmen, werden die Systemeigenschaften über verschiedenen Größenordnungen hinweg auf quantenmechanischer Basis untersucht, um die mit wachsender Systemgröße zunehmende Anzahl an Näherungen zu verifizieren. Einer der entscheidenden Schritte beim Übergang von der semi-quantenmechanischen zur klassischen MD ist schließlich die Reduktion der elektronischen Freiheitsgrade auf ein Punktladungsmodell. Meistens werden Punktladungen von isolierten Zuständen der Moleküle berechnet, die jedoch nicht zwangsläufig in der flüssigen Phase vorzufinden sind, obwohl gerade diese von Interesse ist. Im Folgenden wird näher erläutert, warum dies der Fall ist und wie in dieser Arbeit das Problem umgangen wird.

Schwierigkeiten in der Partialladungsberechnung Obwohl eine Vielzahl von Techniken zur Berechnung von Partialladungen vorhanden ist, sind nur wenige auf dicht gepackte Zustände wie die flüssige Phase anwendbar. Das resultiert aus dem Problem, das elektrostatische Potential einer kontinuierlichen Ladungsdichte auf ein Modell mit Punktladungen zu projizieren. Im isolierten Zustand kann dafür leicht eine passende Mannigfaltigkeit im Raum spezifiert werden, nicht jedoch im kondensierten Zustand.

Da sich gezeigt hat, dass die elektrostatischen Eigenschaften von ionischen Flüssigkeiten, wie das molekulare Dipolmoment, zwischen isolierten Ionen und Ionenpaarclustern stark variiert und vorhandene Kraftfelder nur den isolierten Zustand korrekt beschreiben, wird in dieser Arbeit die Methode von Blöchl angewendet. Diese macht sich die periodischen Randbedingungen des Systems und dessen Dichte zunutze und umgeht die Schwierigkeiten, indem sie im reziproken Raum operiert. Die Resultate sind sehr vielversprechend, da das elektrische Dipolmoment in der flüssigen Phase reproduziert wird und sich eine Reduktion der Ionenladungen ergibt, die auch in NMR-Experimenten gefunden wurde. Daher wird diese Methode der Konstruktion des Kraftfelds zugrunde gelegt.

CAB - Die Ladungszuordnungsmethode von Blöchl Im Gegensatz zum allgemeinen Ansatz, mit Punktladungen das elektrostatische Potential im realen Raum zu reproduzieren, basiert die Technik von Blöchl (CAB) auf dessen Multipolentwicklung und der Ausdehnung der Ladungsdichte. Dabei wird ein Ladungsmodell, das jedem Atom eine Partialladung zuordnet, die durch eine Überlagerung mehrerer Gaußglocken unterschiedlicher Varianz verschmiert wird, an die genannten Eigenschaften im reziproken Raum angepasst. Die Multipolentwicklung ist im Ursprung des reziproken Raums eindeutig durch die Ladungsdichte und die zugehörigen räumlichen Ableitungen bestimmt. Zudem wird die Gesamtanzahl der Elektronen erhalten, was Systemneutralität garantiert. Der Einfluss von Polarisation wird durch die Anpassung der Ausdehnung auch berücksichtigt. Dies wiederum spiegelt sich in einer Reduktion der Ionenladung wider.

Eine Anwendung der Methode auf einzelne Ionen und Ionenpaare zeigt, dass die erwartete Verbesserung des elektrostatischen Modells auftritt. Gegenüber quantenmechanischen Berechnungen wird die Polarisation zwar meist überschätzt, aber es wird eine gute Übereinstimmung mit Resultaten, die wie CAB auf der quantenmechanischen elektronischen Dichtefunktionaltheorie (DFT) basieren, beobachtet. Sobald jedoch mehr als ein Molekül involviert ist, werden auch die Abweichungen zwischen den verschiedenen DFT-basierten Methoden größer, was die Sensitivität gegenüber der Konfiguration unterstreicht. Letztendlich kann nur eine gemittelte Beschreibung dieser sensiblen Reaktion auf verschiedene molekulare Geometrien in einem Kraftfeld mit statischen Partialladungen erreicht werden. Hierfür wird jedoch eine bestimmte Anzahl an Konfigurationen für die CAB Methode benötigt, um die Partialladungen zu bestimmen, die Polarisation in der flüssigen Phase implizit beschreiben.

Berechnung der Partialladung aus der flüssigen Phase Zur Erzeugung einer Startkonfiguration für die DFT-gestützte ab-initio MD (AIMD), wurde zunächst ein System mit circa 30 Ionenpaaren (IP) und, um die Zulänglichkeit der Systemgröße zu verifizieren, ein weiteres System mit 240 Ionenpaaren mit klassischer MD equilibriert. Ein Vergleich der Resultate der klassischen MD mit 30 und 240 IP zeigt keine Effekte, die auf eine unzureichende Systemgröße schließen lassen. Damit war eine verläßliche Startkonfiguration für eine AIMD-Simulation verfügbar. Wie die klassische MD erfordert auch die semi-quantenmechanische AIMD eine genaue Kenntnis der Algorithmen und Erfahrung. Da dies den Rahmen dieser Arbeit sprengen würde, wurden die Trajektorien nicht selbst integriert. Vielmehr wurde dafür mit den Expertengruppen von Prof. Dr. Barbara Kirchner und Dr. Luigi Delle Site zusammengearbeitet, die letztendlich die notwendigen Konfigurationen für die Blöchl-Analyse, sowie die radialen Verteilungsfunktionen, die später für die Optimierung des Kraftfeldes nötig sind, zur Verfügung stellten.

Um eine Ladungsverteilung zu erhalten, die der flüssigen Phase entspricht, wurde darauf geachtet, dass keine künstliche Struktur durch eine zu kleine Systemgröße vorhanden ist. Zudem wurden 100 quasi-unabhängige Schnappschüsse untersucht, um die sensible Reaktion des Polarisationzustand auf die lokale Konfiguration zu berücksichtigen.

Eine Untersuchung des Systems [MMIM][Cl] zeigt, wie erwartet, eine Reduktion der Ionenladung auf $\pm 0.631\,\mathrm{e}$. Dabei weist die Verteilung der Partialladungen eine für den entsprechenden Atomtyp charakteristische Breite auf, die auf dessen Polarsierbarkeit zurückzuführen ist.

Vergleicht man die Ladungsverteilung der Gasphase und flüssigen Phase, kommt der unterschiedliche Polarisationszustand der Moleküle zum Vorschein. Beim Übergang vom Gas zur Flüssigkeit tritt die zu erwartetende Delokalisation der Elektronen über das ganze Molekül auf, die zu einem wachsenden Dipolmoment führt. Desweiteren begünstigt die in der Flüssigkeit berechnete Ladungsverteilung die Wasserstoffbrückenbindung an dem Wasserstoff des Imidazoliumrings, das am Kohlenstoff, der zwischen den Stickstoffen sitzt, gebunden ist. Diese spezifische Wechselwirkung ist schon seit längerem dafür bekannt, von großer Bedeutung für die Dynamik Imidazolium-basierter ionischer Flüssigkeiten zu sein. Dies ist ein weiterer Hinweis, dass die Blöchl-Ladungen eine sehr gute Repräsentation des Ladungszustands in der Flüssigkeit liefern.

Da die Blöchl-Technik jedoch auf einer Ladungsdichte beruht, die mit DFT berechnet wurde, muss der Einfluss des Austausch-Korrelation (XC) Funktionals geklärt werden. Obwohl ein Vergleich von Partialladungen für [EMIM][DCA], die mit verschiedenen Funktionalen berechnet wurden, zeigt, dass die Unterscheide meist innerhalb der Standardabweichung liegen, treten manchmal signifikante Variationen auf. Da eine Beurteilung der Genauigkeit eines XC-Funktionals ab-initio nicht möglich ist, wurden deshalb immer Resultate quantenmechanischer Rechnungen als Bezugspunkt ausgewählt, um ein passendes XC-Funktional zu finden.

Des weiteren gilt es zu klären, inwieweit kleine a-posteriori Modifikationen der Ladungen die Beschreibung der Systemeigenschaften beeinflussen. Diese Information ist von großer Bedeutung, denn es ist zu erwarten, dass für ein transferierbares Kraftfeld kleine Änderungen nötig sind. Dafür wurden zwei verschiedene Modelle mit Blöchl-Ladungen verglichen, die sich durch die Ladung mancher Ringatome unterscheiden. Während ein Modell direkt aus den Resultaten der Blöchl-Analyse hervorging, wurde bei dem anderen Modell die Ladung der Atome des gleichen Typs gemittelt, woraus eine symmetrische Ladung des Rings hervorging. Es stellte sich heraus, dass weder strukturelle oder dynamische, noch energetische Eigenschaften stark variieren. Das ist ein entscheidender Vorteil, denn es ebnet den Weg zur Konstruktion eines großen transferierbaren Kraftfelds.

Konstruktion von Partialladungen für ein Kraftfeld Das Ziel dieser Arbeit ist die Entwicklung einer Methode zur Kraftfeldparametrisierung, wobei einer der wichtigsten Aspekte die Transferierbarkeit ist. Aus diesem Grund müssen die Parameter sorgfältig aufeinander abgestimmt werden, wofür ein breites Spektrum an Kombinationen Imidazolium-basierter Kationen verschiedener Seitenkettenlänge mit den Anionen Chlorid, Dicyanamid und Thiocyanat, sowie die ionische Flüssigkeit Dimethylammoniumnitrat untersucht wurde. Ein Verständnis der Ladungsumverteilung und Reduktion liefert dabei die Theorie von Leontyev und Stuchebrukhov, die jedoch zuerst erweitert werden muss, um verschiedene Molekülspezies in einem Lösungsmittel zu beschreiben. Die erweiterte Theorie läßt es schließlich zu, einen generischen Satz von Partialladungen zu definieren, der leicht auf andere Kationen und Anionen erweiterbar ist, womit die Grundlage für ein großes transferierbares Kraftfeld gelegt ist.

Den Ausgangspunkt für die MDEC-Theorie von Leontyev und Stuchebrukhov bildet dabei die Zerlegung der Polarisierbarkeit des Mediums in einen nuklearen und elektronischen Anteil. Während der nukleare Anteil auf die Konfiguration der Atome und deren effektive Ladung zurückzuführen ist, ergibt sich der elektronische Anteil aus den delokalisierten Elektronen. Deren viel geringere Masse erlaubt es fast instantan und ohne Aufnahme von Impuls auf eine Veränderung des Zustands zu reagieren, solange das elektrische Feld schwach genug ist,

um die Näherung der linear-response Theorie zu erfüllen. Daraus ergibt sich das Bild, dass sich die Atome anstatt im Vakuum in einem Medium mit der dielektrischen Konstante ϵ_{el} aufhalten. Bei der Ladungsberechnung werden diese Effekte nicht entkoppelt, sodaß die Blöchl-Analyse den effektiven Ladungszustand wiedergibt, der ϵ_{el} implizit enthält. Jedoch ist ϵ_{el}, keine empirische Zahl, sondern materialabhängig und mit der Polarisation verknüpft. Für geladene Moleküle ergibt sich damit zwangsläufig eine Reduktion der ganzzahligen Ladung um den Faktor $1/\sqrt{\epsilon_{\mathrm{el}}}$. Zudem ist es auch möglich die effektive Ladungsreduktion experimentell zu verifizieren, da ϵ_{el} dem Quadrat des Brechnugsindex n_{D} entspricht, wobei für [EMIM][DCA] eine Abweichung im Bereich von nur 2 % zu beobachten ist.

Betrachtet man die Fluktuation der Ionenladung als Maß für die Polarisierbarkeit, erkennt man, dass diese relativ unabhängig von der Ionenkombination sind. Ein Vergleich der Ladungsverteilungen im polarisierten und isolierten Zustand verdeutlicht dabei die Delokalisation der Elektronen, falls Möglichkeiten zur Wechselwirkung gegeben sind. Letztendlich bildet all dies die Grundlage für die Konstruktion eine Satzes von Partialladungen, der auf eine große Anzahl von ionischen Flüssigkeiten anwendbar ist, sodass eine zeitaufwendige Blöchl-Analyse entfallen kann.

Die elektrostatischen Eigenschaften von ionische Flüssigkeiten weisen eine sehr starke Lokalität auf, die durch eine sehr hohe Ladungsdichte und Polarisierbarkeit hervorgerufen wird. Dies wurde durch die Partialladungen bestätigt, die aus Ladungsdichten periodischer Systeme mit acht beziehungsweise mindestens 30 Ionenpaaren hervorgingen. Die Unterschiede zwischen den einzelnen Ladungen lagen innerhalb der Standardabweichung, woraus sich die Möglichkeit ergibt, den Rechenaufwand für die Blöchl-Analyse massiv zu reduzieren.

Bei eine Betrachtung der polarisierten Ladungsverteilung kann kein entscheidender Einfluss der Ionenkombination festgestellt werden. Auch ein Vergleich der Dipolmomente, die mit den klassischen Partialladungen bestimmt wurden, mit den Resultaten einer Wannier-Analyse bestätigt die Verlässlichkeit der Blöchl-Ladungen, obwohl eine a-posteriori Symmetrisierung des Rings nötig ist, um eine angemessene Übereinstimmung des mittleren Dipolmoments zu erreichen. Zudem wird auch deutlich, dass eine mittlere Beschreibung des Polarisierbarkeit nicht ausreicht, um die Breite der Dipolmomentverteilung zu reproduzieren, die durch eine Wannier-Analyse gegeben ist, die auf der Elektronendichte basiert und im Gegensatz zu den statischen Partialladungen die Polarisierbarkeit explizit in Betracht zieht.

Um einen generischen Satz von Partialladungen zu konstruieren, wurden die Partialladungen von einer Serie Imidazolium-basierter ionischer Flüssigkeiten mit der Blöchl-Methode untersucht. Vergleicht man ϵ_{el} der verschiedenen Systeme, zeigt sich, dass die Variation von ϵ_{el} nur innerhalb verschiedener Kation-Anion Kombinationen signifikant ist. Jedoch ist bei Systemen mit dem gleichem Gegenion, aber wachsender Kettenlänge der Imidazolium-basierten Kationen, keine deutlichen Veränderungen der Ladungsverteilung des Rings beobachtbar. Auch die Ladung der Seitengruppen behält bestimmte Charakteristika, wie die Ausdehnung der Ladungsdichte auf die Wasserstoffe oder die Gesamtladung bestimmter Atomgruppen. Obwohl bei längeren Seitenketten eine weitere Ausdehnung der Ladungsverteilung möglich wäre, geschieht nichts derartiges, was darauf hindeutet, dass nur direkt an den Imidazoliumring gebundene Atomgruppen in der Seitenkette, sowie der Ring selbst, merklich polarisierbar sind. Dieser Mechanismus der intermolekularen Umverteilung der Ladung wird jedoch nur durch eine Gliederung des Kations in Domänen ersichtlich, die eine Gegenüberstellung von Systemen mit verschiedener Seitenkettenlängen erlaubt. Schließlich stimmen auch die polarisierten

Ladungsverteilung der Anionen überein. Diese Resultate zeigen, dass die Grundlage zur Konstruktion von generischen Kraftfeldladungen für Imidazolium-basierte ionische Flüssigkeiten gegeben ist.

Verglichen zu den Schwankungen, die durch Polarisation und dem Einfluß des XC-Funktionals hervorgerufen werden, ist die Variation der Partialladungen zwischen gleichen Kationen in Verbindung mit unterschiedlichen Anionen klein. Um diese weiter abzumildern, wird deshalb zur Konstruktion eines generischen Kraftfelds über die verschiedenen Systeme gemittelt. Leider sind noch nicht ausreichend Daten vorhanden, um dies mit Sicherheit auf beliebige Kettenlänge zu erweitern, da noch ein kleines Dipolmoment am Ende der Buthylkette zu finden ist. Jedoch kann am Beispiel von $[EMIM]^+$ das Vorgehen demonstriert werden und auch ein Ladungsmodell für beliebige Kettenlängen vorgeschlagen werden. Um die Anzahl unterschiedlicher Atomtypen zu minimieren, wurde untersucht, ob Gemeinsamkeiten zu finden sind, mit dem Ziel den weiteren Parametrisierungsaufwand so klein wie möglich zu halten. Leider war dies nicht der Fall, aber es wurde zunächst ein symmetrisch geladener Imidazoliumring durch Mittelung erzeugt, der, wie gezeigt, die Dipolmomente der AIMD in einem annehmbaren Rahmen widerspiegelt. Daraufhin wurden die Ladungen der $[EMIM]^+$-Kationen, die aus den verschiedenen Systemen hervorgingen, gemittelt und schließlich noch eine Korrektur von 0.02 e am Kohlenstoff der Methylgruppe, die direkt an den Ring gebunden ist, vorgenommen, um numerische Ungenauigkeiten auszugleichen und eine ganzzahlige Ladung des polarisierten Ladungszustands herzustellen. Zur ersten Verifizierung der generischen Partialladungen wurde das erzeugte Dipolmoment mit den AIMD Resultaten verglichen und, trotz einer kleinen Überschätzung, eine zufriedenstellende Übereinstimmung festgestellt. Hiermit wurde letztendlich der theoretische Hintergrund für die Entwicklung der elektrostatischen Kraftfeldparameter geliefert. Nun ist eine Methode notwendig, die es erlaubt, effizient die restlichen Parameter anzupassen, um schließlich ein zuverlässiges Kraftfeld zu generieren.

1.3. Techniken zur Anpassung und Konstruktion eines Kraftfelds

Bisher wurden ausschließlich Techniken vorgestellt, die sich mit der Bestimmung der Partialladungen befassen. Jedoch besteht ein Kraftfeld aus einem großen Satz an Parametern, die konsistent angepasst werden müssen. Um diesen Prozess so effizient wie möglich zu gestalten, wurde ein automatisiertes Konzept entwickelt, das eine Anpassung der Parameter an beliebige Referenzdaten erlaubt, und in ein Python Modul namens `PyPaTEGRO` implementiert. Im Folgenden wird dargestellt, auf welche Annahmen und Grundlagen sich die Optimierung stützt und schließlich die Methode am Beispiel [MMIM][Cl] vorgestellt.

Anpassung der Dihedralparameter Im Gegensatz zur Parametrisierung von kovalenten Bindungen, müssen die Dihedralpotentiale angepasst werden, sobald sich die elektrostatische oder kurzreichweitige Wechselwirkung ändert, da diese über die 1–4 Wechselwirkung miteinander verknüpft sind. Dieser Vorgang ist unproblematisch, wenn Referenzenergien eines anderen Kraftfelds vorliegen, wie in diesem Fall. Andernfalls ist es notwendig zuerst entsprechende Berechnungen durchzuführen. Liegen letztendlich die Torsionsenergieprofile der verschiedenen Dihedralwinkel vor, lassen sich die Parameter leicht mit einem nicht-linearen

Least-square-error Fit anpassen. Dieser Vorgang ist leicht automatisierbar und wird bei jeder Veränderung der Partialladungen oder kurzreichweitigen Wechselwirkung vorgenommen. Allein ein Vergleich der unmodifizierten mit den angepaßten Energien zeigt, dass ohne die Korrektur eine falsche Grundzustandskonfiguration energetisch favorisiert wäre.

Optimierung der kurzreichweitigen Wechselwirkungen Obwohl [MMIM][Cl] erst bei einer Temperatur von 398 K schmilzt, ist es abgesehen von der Symmetrie strukturell relativ gleichartig zu ionischen Flüssigkeiten, die bei Raumtemperatur flüssig sind. Aufgrund der geringen Größe eignet es sich deshalb zur Entwicklung einer Kraftfeldoptimierungsroutine und dient daher im Folgenden zur Entwicklung der Methodik als Beispiel. Obwohl nur eine Quelle für experimenelle Referenzdaten zugänglich ist, reicht diese aus um das Kraftfeld anzupassen und zu verifizieren. Dabei gilt es nun, die Parameter der Lennard-Jones (LJ) Potentiale zu optimieren, wofür zwei verschiedene Wege eingeschlagen werden können.

Eine Methode beruht auf einer iterativen Skalierung der Parameter, wodurch sich die Massendichte bei einer bestimmten Temperatur sehr leicht anpassen lässt. Dabei dienen die Referenz-RDFs der AIMD-Simulationen als Maß für die strukturelle Übereinstimmung. Die Problematik dieser Methode liegt jedoch in der gemeinsamen Optimierung der Parameter. Es ist sehr schwierig, die richtige Kombination zu finden, um gleichzeitig mehrere Referenzdaten anzupassen, wie etwa die Massendichte bei verschiedenen Temperaturen oder unterschiedliche RDFs. Jedoch erlaubt diese Methode, das Kraftfeld bezüglich eines sehr limitierten Gültigkeitsbereich schnell zu adaptieren.

Um diese Probleme aus dem Weg zu schaffen, wurde ein Verfahren entwickelt, dass eine Fehlerfunktion, die sich aus den verschiedensten Größen zusammensetzen lässt, minimiert, sodass letztendlich ein optimaler Satz von Parametern zur Verfügung steht. Diese Technik basiert auf der Entwicklung der Fehlerfunktion in eine Taylorreihe, wofür eine Reihe von Simulationen aufgesetzt und durchgeführt werden muss, um die notwendigen Ableitungen zuverlässig zu berechnen. Zu diesem Zweck wurde das Python-Modul `PyPaTEGRO` implementiert, das alle nötigen Dateien vorbereitet, um die Rechnungen massiv parallel auf einem Supercomputer ablaufen zu lassen. Desweiteren enthält es verschiedenste Routinen zur Simulationsanalyse und ermöglicht letztendlich die Berechnung eines Satzes von Parametern, der die Fehlerfunktion minimiert.

Diese Methode wurde wieder am Beispielsystem [MMIM][Cl] getestet. Um die Simulationszeiten so gering wie möglich zu halten, wurden nur statische Eigenschaften als Referenzdaten benutzt, wie die Massendichte bei unterschiedlichen Temperaturen und die RDFs zwischen dem Chlorid und den Wasserstoffen. Als Ausgangspunkt diente ein bekanntes Kraftfeld, das eine zu langsame Dynamik der Ionen wiedergibt. Bei diesem wurden die Partialladungen durch die Resultate einer Blöchl Analyse ersetzt und die Dihedralparameter angepasst. Hierfür wurde letztendlich die Fehlerfunktion minimiert, wobei verschiedene Kombinationen aus den LJ-Parametern der Wasserstoffe und des Chlorids am Optimierungsprozess beteiligt waren. Nach sieben Iterationsschritten wurde die Fehlerfunktion als optimiert betrachtet, da die Massendichten über einen Bereich von über 50 K innerhalb eines Fehlers von 1 % lagen und auch die Abweichung der RDFs von ihren Referenzen als annehmbar betrachtet wurden. Der letzte Schritt ist nun die Validierung des Kraftfelds, da bisher keine dynamischen oder energetischen Größen im Parametrisierungsprozess direkt involviert waren.

Validierung des Kraftfelds Zuletzt ist es nun nötig, die Verlässlichkeit des Kraftfelds hinsichtlich verschiedener Eigenschaften zu überprüfen. Wie erwähnt dienten ausschließlich statische Größen zur Parameteroptimierung, sodass keinerlei Information über die Reproduktion dynamischer oder energetischer Eigenschaften vorhanden ist. Diese Analyse erlaubt es zudem, Unzulänglichkeiten zu identifizieren, die für den weiteren Optimierungsprozess, auch in Hinblick auf ein transferierbares Kraftfeld, von großem Interesse sind. Daher wurden die Diffusionskonstanten, Leitfähigkeit und Verdampfungsenthalpie untersucht.

Obwohl das optimierte Kraftfeld BTFF gegenüber dem Ausgangskraftfeld CLaP die gewünschte Beschleunigung der Dynamik aufweist, kommt zum Vorschein, dass BTFF die Leitfähigkeit und damit auch Dynamik bei geringen Temperaturen überschätzt. Neben der schnelleren Dynamik weist BTFF gegenüber CLaP auch Unterschiede im Verhältnis der Diffusionskonstanten und der Korrelation der Ionen auf. Dieser Umstand kann leider nur mit Resultaten aus anderen Simulationen verglichen werden, wobei der Trend von BTFF gestützt wird. Diese ungewöhnlich schnelle Dynamik und Leitfähigkeit ist höchstwahrscheinlich der unzureichenden Übereinstimmung der RDFs zuzurechnen, da die ersten Maxima manchmal zu klein sind, was für eine zu geringe Wechselwirkung und dementsprechend Bindung führt. Dabei liegt vor allem die RDF zwischen Chlorid und dem Wasserstoffatom, das an das Kohlenstoffatom zwischen den Stickstoffen gebunden ist, im Blick, da bekannt ist, dass die Wechselwirkungen in diesem Bereich des Imidazoliumrings die Dynamik entscheidend beeinflussen.

Weiterhin spielt auch die Transferierbarkeit zwischen verschiedenen Temperaturen eine entscheidende Rolle. Die Resultate von entsprechenden Simulationen zeigen, dass BTFF ab etwa 440 K die experimentellen Resultate sehr gut reproduziert. Zudem wird der subtile Zusammenhang zwischen kurz- und langreichweitigen Wechselwirkung deutlich, was die Vermutung untermauert, dass eine zu geringe Wechselwirkung im entscheidenden Bereich des Imidazoliumrings besteht. Mit steigender Temperatur verliert diese Wechselwirkung seine Dominanz, und die unzulängliche Parametrisierung der kurzreichweitigen Parameter wird von der akkuraten Darstellung der elektrostatischen Wechselwirkung überspielt.

Ein entscheidender Einfluss des genannten Wasserstoffatoms zeigt sich jedoch in den Verdampfungsenthalpien. Gemäß den Erwartungen sind die Werte für BTFF viel kleiner als für CLaP, jedoch liegen sie auch unterhalb von Simulationsresultaten aus der Literatur, die zudem eine langsamere Diffusion als BTFF reproduzieren. Damit hat sich letztendlich gezeigt, dass zumindest bestimmte RDFs sehr genau angepaßt werden müssen, um über einen weiten Temperaturbereich eine verlässliche Beschreibung der Molekulardynamik zu liefern.

Mit diesen Ergebnissen kann nun der erste Zyklus zur Konstruktion eines großen transferierbaren Kraftfeld für ionische Flüssigkeiten abgeschlossen werden. Obwohl die Resultate nach dem erste Optimierungsprozess der kurzreichweitigen Wechselwirkungen bei niedrigen Temperaturen nicht zufriedenstellend waren, wird bei höheren Temperaturen die Dynamik sehr gut reproduziert. Durch Analyse der Defizite wurde zudem klar, dass einer akkuraten Modellierung der RDFs mehr Gewicht beizumessen ist, da diese ein Maß für die Balance zwischen kurz- und langreichweitiger Wechselwirkung darstellen.

Abschließend lässt sich zusammenfassen, dass nun eine Theorie und die Software vorhanden ist, um Kraftfelder zu parametrisieren. Obwohl etwas Erfahrung nötig ist, um den Vorgang möglichst effizient zu gestalten, sind die entscheidenden Grundpfeiler gelegt. Damit sollte eine weitere Hürde zur Parametrisierung von Kraftfeldern *in-silico* etwas geebnet worden sein, die es letztendlich erlauben soll, ionische Flüssigkeiten zu charakterisieren und verstehen.

2. Introduction

2.1. Room Temperature Ionic Liquids

The most simple example for a salt is sodium chloride consisting of two atomic ions. At room temperature this salt is present in the crystalline state, because its melting point is around 1000 K due to the strong ionic interaction. This is the case for all salts that consist of atomic ions, which are referred to as molten salts in their liquid state.

Another class of salts are ionic liquids (ILs), which have a much lower melting point than molten salts. They are constituted by ions, which are composed of several atoms such that a strong delocalization of the electrons occurs which weakens the electrostatic interaction. In addition the excluded volume and the configurational entropy increase which yields a significant reduction of the Gibb's Free energy of fusion and with that the melting temperature[18]. An example for this class of solvents is ethyl-methyl-imidazolium thiocyanate [EMIM][SCN] depicted in Figure 2.1 with a melting temperature of 327 K[19]. The charge of $[EMIM]^+$ is strongly delocalized around the imidazolium ring, which is one of the reasons for the very low melting point.

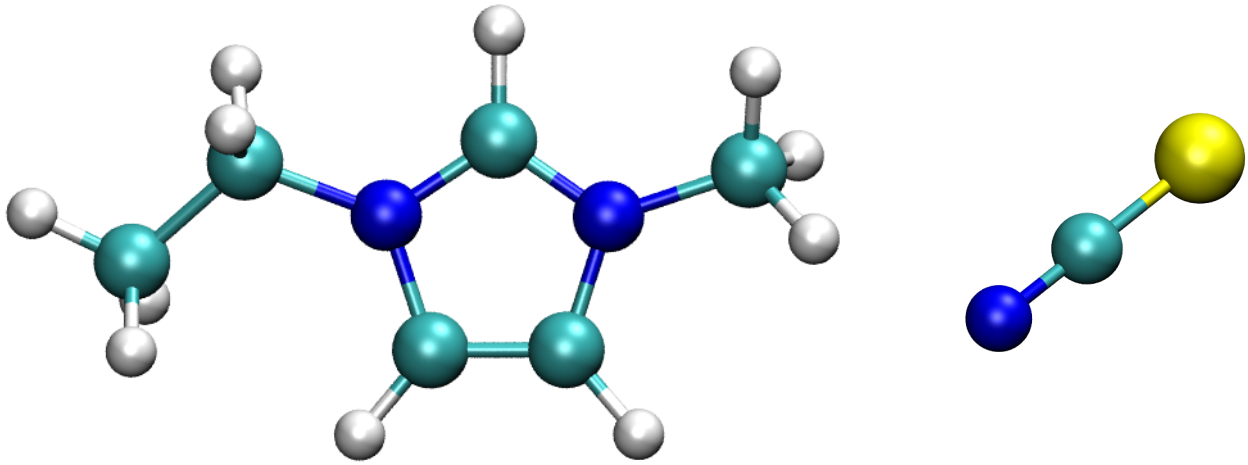


Figure 2.1.: A picture of an ethyl-methyl-imidazolium ($[EMIM]^+$) cation and a thiocyanate ($[SCN]^-$) anion.

If the size of the ions is further increased, a decrease of the melting temperature is expected and a further class of solvents is defined. Room temperature ionic liquids (RTILs) are per definition the class of salts, which are present in the liquid phase at room temperature. An example for this kind of compounds is [BMIM][DCA] melting at 267 K[20].

However, it can be shown that molten salts and ionic liquids are not completely different from each other. A comparison of the phase diagram for molten salts and ionic liquids in terms of reduced units reveals that the reduced melting temperature is almost constant, while only the reduced volume differs[21] highlighting the importance of the size of the ions.

Yet, several theories exist, but the class of ILs is not completely understood. Actually the first RTIL, Ethyl-ammonium nitrate (EAN), has been discovered by Walden[22] in 1914, but the real interest in this incredibly large class of compounds has been raised only some decades ago.[23–34] On the one hand, many RTILs have been identified with the gained knowledge as well as many potential applications. They have been "designed" for applications, which were inefficient or

impossible with common solvents yet. Some good examples are the processing of cellulose[35,36], the capture of flue gases in chemical reactions[37] and especially catalytic processes[30,38]. On the other hand, already during the synthesis of RTILs, their high sensitivity to impurities, especially water, has been recognized.[30] But apart from the synthesis, also many experimental techniques are not straightforwardly applicable, due to the high charge density, viscosity and impurity dependence[34]. The problem of inaccurate measurements is a large topic in the field of ionic liquids (ILs), because very small impurities in the range of a few ppm can affect the results significantly.

A good example is the disagreement of different density measurement for ethyl-methyl-imidazolium thiocyanate ([EMIM][SCN])[19,39,40], all depending on a linear relation between density ρ and temperature T:

$$\rho = aT + b \tag{2.1}$$

Comparing the different experimental results reveals the same slope for the different references up to precision $10^{-1}\ \mathrm{kg \cdot m^{-3} \cdot K^{-1}}$. But Domanska *et al.*[19] and Seki *et al.*[40] derive a value for the fitting parameter b around $1130\ \mathrm{kg \cdot m^{-3}}$, while Yokozeki and Shiflett[39] obtain $b = 1296\ \mathrm{kg \cdot m^{-3}}$. In order to avoid these inaccuracies, the sample has to be maintained pure over the whole chain from synthesis to the measurement, which is very difficult.

But in order to allow an efficient "design" of ILs, the molecular mechanisms responsible for their special behaviour have to be understood. A valuable tool for that kind of studies are computer simulations, whereby especially quantum-chemical post-Hartree-Fock (pHF), electronic density functional theory (DFT) and classical all-atom molecular dynamics (MD) simulations allow to obtain insight into the energies and forces on different scales providing the basis for a calculation of a vast number of properties. For RTIL systems, a new challenge in the field of computer simulations arises, because all levels of theory have to deal with different problems. Due to the relative large size of the solvent species, pHF and DFT calculations require an extremely large amount of computing time, while MD simulations suffer from a lack of detailled information about the inter- and intramolecular interactions. However MD simulations are a very suitable method to determine thermodynamic, structural and dynamic properties, but they rely on an accurate parameterisation of the interactions. In order to allow a "design" of RTILs *in-silico* avoiding potentially environmentally-unfriendly and expensive synthesis, such a parameterisation has to be developed, which is the main topic of this thesis.

2.2. From the inhomogeneous electron gas to classical molecular dynamics

From classical physics it is well known that an analytical solution of a many-body problem involving more than two bodies is not possible. The same situation arises for the Schrödinger equation. For this reason only the hydrogen atom allows an analytical treatment of the quantum mechanics, but as soon as more than two particles are involved, approximations are required. Finally systems of thousands of particles are of interest, which requires a consistent mapping of the information about the energetics and interactions via different scales shown in Figure 2.2.

In 1928 and 1930, Hartree[41] and Fock[42] established the basis for modern quantum chemical

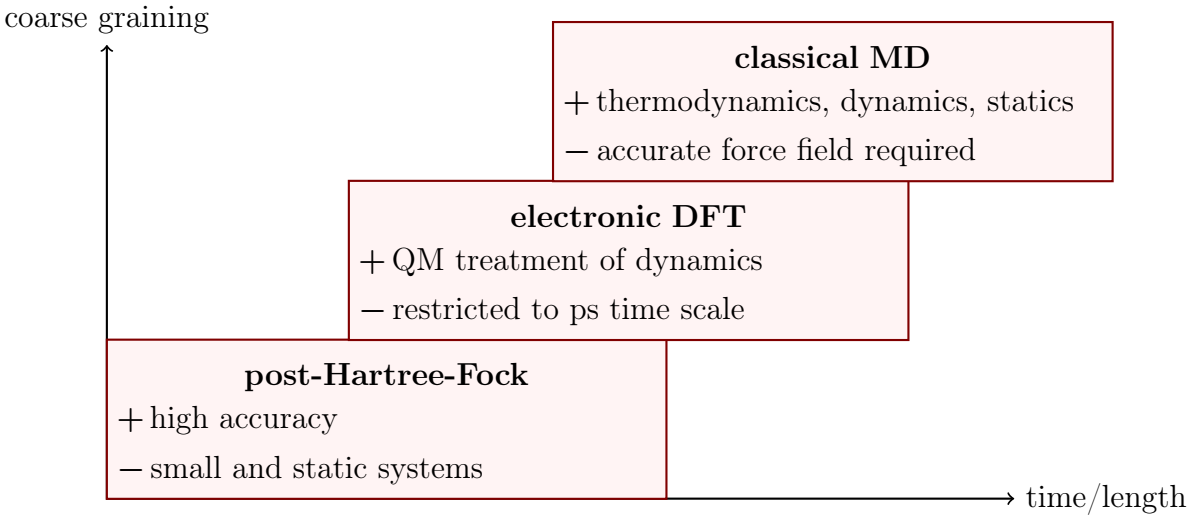


Figure 2.2.: Different computational methods with their advantages (+) and drawbacks (-) are applied on different scales to consistently map the system properties between each other.

calculations (HF), that allow a numerical solution of the Schrödinger equation for many-body systems. The method of Hartree[41] relies on a separation of the variables of the wave function Ψ, which is used to numerically integrate the Schrödinger equation. This description neglects the exchange interaction within the electrons as well as symmetry properties though they are required to describe degenerated states. These insufficiencies were outweighed by an extension of this description by Fock[42]. Therein a variational approach is applied to derive the solution of the many-electron Schrödinger equation, and the wave function Ψ is a linear combination of basis functions. It shows that the exchange interaction within the electrons is small, such that a treatment in terms of perturbation theory is appropriate. Hence to solve the complex set of equations, at first the Hartree approximation is derived by neglecting the cross-terms. Afterwards iterative application of perturbation theory allows to calculate the exchange interaction.

Nowadays many extensions of the Hartree-Fock method, called post-Hartree-Fock (pHF) schemes, exist, where one of the most famous is probably Møller-Plesset perturbation theory.[43] Møller and Plesset[43] have proven, that HF is an approximation correct up to first order in energy perturbation. Higher order correction terms are easily gained, as they rely only on single-electron wave functions, allowing to increase the accuracy of the calculations up to arbitrary order. Often the accuracy of the second-order Møller-Plesset approximation (MP2) is sufficient and hence it is widely used for studies on molecular compounds. However, the computational effort is increasing with N^5 with the number N of involved basis functions of the atomic wave functions. As depcitured in Figure 2.2, this method is very accurate method, but only applicable to small and static systems.

In order to allow one studies of larger systems in terms of quantum mechanics, the description of the system is more coarse grained, shown in the center of Figure 2.2. Therein, the electrons are considered as an inhomogeneous gas[44] of density $n(\boldsymbol{r})$ at the point in space $\boldsymbol{r}$. This method is called electronic density functional theory (DFT) and Hohenberg and Kohn[44] established a

theory that describes the energy of a system in terms of the electron density $n(\boldsymbol{r})$ given as a functional of an arbitrary external potential. Unfortunately the energy functional $F[n(\boldsymbol{r})]$ is rather complex and usually not analytically accessible. For a system of approximately constant electron density and in case of a slowly varying electron density, exact solutions are accessible, but more common cases require a subtle calculation of the functional. Finally a set of self-consistent equations has been derived by Kohn and Sham[45], that describes an inhomogeneous electron gas including exchange and correlation effects. These equations explicitly depend on the so-called exchange-correlation (XC) functional $E_{xc}[n]$, that is not known analytically. Many XC functionals have been proposed during the last decades, that mostly rely on a gradient approximation. Parameters involved in this approximation are obtained by a fit to HF or pHF calculations and adapted to achieve correct asymptotic behaviour, as for example the BLYP functional[46–48]. Another very popular functional has been introduced by Perdew *et al.*[49], known as the PBE functional. It completely relies on fundamental constants and is constructed in order to describe the analytically known local spin density correctly.

A choice of an XC functional from "scratch" does not necessarily yield accurate results, therefore careful benchmarking is necessary, to assure reasonable accuracy. Yet, DFT allows to study systems of hundreds and even thousands of atoms, such that "*bulk-like*" phases are accessible. While HF and pHF methods are only applicable to static calculations, DFT moreover can be combined with MD techniques to obtain the possibility to propagate the system in time with the XC functional as only empirical parameter. This is finally called "*ab-initio*" MD (AIMD). Mainly two schemes are applied in such kind of simulations, Born-Oppenheimer MD and Car-Parrinello MD, which differ in the derivation of the forces from the electron density.[50] Efficient algorithms, such as the programs CPMD[51] and CP2K[52] are available, but the simulation time is restricted to tens of picoseconds.

Hence, to simulate even larger systems on longer time scales, the computationally expensive explicit treatment of the electron density has to be incorporated into a parameterisation of the interaction into mathematically simple potentials. This kind of coarse-graining yields the scheme of classical molecular dynamics (MD) simulations, drawn in the upper right corner of Figure 2.2. The parameterisation process is a difficult task and not unique, because a careful mapping of the systems properties over different length scales into reasonable potentials is required. These are summarized in a so-called force field. Apart from the diversity of the applied interaction potentials, various methods to coarse-grain an MD simulation exist. It already starts on the level of an all-atom description of the system. At this point, it has to be decided, if static partial charges are applied or polarization is explicitly considered during the simulation. For this reason, different schemes have been developed to deal with even macroscopic length scales. An increasing length scale requires an increasing coarse graining of the simulation, starting with a condensation of several atoms into groups, over inclusion of explicit solvent molecules up to mean field descriptions of the interactions, such that only the actual object of interest is explicitly simulated.

Hence, to establish a reliable force field, the interactions given for the different scales have to be mapped consistently, but provide differing levels of accuracy. The important information has to be extracted on every level of theory in order to allow a proper description of the system in terms of classical physics. For it, a multi-scale approach has been suggested and this thesis takes the last step from the DFT to the classical level, that allows to parameterise and optimize classical force fields. It is based on an interconnection of all levels of theory to transfer almost

all important information straightforwardly from the QM to the classical level. Finally, the most challenging task comprises a construction of a force field, which is transferable within a certain set of compounds and points in the phase diagram. The main object inhere are all-atom force fields, because they usually provide the basis for further coarse-graining of the system. When the terms force field (FF) or molecular dynamics (MD) simulation are used in this work, they refer to an all-atom description of the system with static partial charges, if not stated otherwise.

2.3. Multiscale modeling of ionic liquids

A FF is the basis for every classical MD simulation and different levels of theory are involved in the parameterisation procedure. The advantages and disadvantages of the various methods have already been summarized in Figure 2.2. In order to allow one an efficient connection of the different scales experience on every level of theory is required. To this end a cooperation of the groups of Robert Berger, who has expertise in quantum chemical pHF studies, and Luigi Delle Site, experienced in the field of electronic DFT, had been established. Within the DFG SPP 1191 Ionic Liquids, we connected our knowledge on different levels of theory to establish a multiscale approach for classical force field refinement. This work deals with the transfer of the information from the DFT level to a classical force field. Herein partial charges are determined from the results of the DFT calculation and methods are developed to adapt the further parameters.

The multiscale approach begins at the pHF level of theory, which provides very accurate information about the electronic structure and geometry of small static systems. Though these calculations are restricted to the gas phase, this electronic structure is often used for the calculation of classical FF parameters, that should describe liquid phase behaviour. For some molecules this approximation works fine, but it fails for other systems, especially if strongly polarizable molecules are involved. To study the electronic properties of systems in the condensed phase electronic density functional theory methods are required, which sensitively depend on the exchange-correlation (XC) functional. Only careful benchmarking of the DFT with pHF results for small systems allows one to choose an XC functional, that provides reasonable accuracy. For this reason the first important connection of our multiscaling method is the connection between pHF and DFT depicted on the left side of the triangle in Figure 2.3. Calculation on small clusters consisting of a few ions with pHF and DFT allow one to compare a variety of properties, like structure, energy, forces or electrostatic moments. This in turn provides the basis for the choice of an accurate XC functional and is an inevitable connection between the different scales.

With that electronic structure, geometries, and even dynamics become accessible in the condensed phase based on quantum mechanical considerations because AIMD is applicable. Some liquid phase properties are already contained in the DFT or AIMD studies, such as information about structure and electostatics, but the system size and length of time propagation is still not large enough to predict every kind of bulk property, such as the mass density, viscosity or conductivity, which are of large interest in the field of ionic liquids.

For this reason, the DFT scale is bridged to classical MD simulations which allow one to simulate systems of thousands of atoms and reach time scales of tens of nanoseconds. These

time and length scales are finally sufficient to calculate the aforementioned properties as well as a vast number of other thermodynamic, static and dynamic properties, such as the heats of vaporization, melting temperatures or diffusion constants. Unfortunately, the accuracy of the results strongly depend on the applied FF. For many organic liquids and biomolecular systems reliable force fields such as OPLS-AA[53] or AMBER[54] are available, but also these FF are continuously refined. A very good example for the sensitivity of a parameterisation is the variety of available water models, which all have their advantages and disadvantages, such that a choice "*from scratch*" does not guarantee reasonable results. This problem results from the strong polarizability of water, that is not explicitly treated in a classical MD simulation. Different methods are available for an incorporation of these effects, but the computational effort of such polarizable models is much larger, restricting the accessible time and length scales drastically. For this reason, a reliable approach is necessary to parameterise all important properties of the system into a force field with a static charge distribution. As shown on the right side of the triangle in Figure 2.3, we connect the DFT and classical MD scale in order to include important condensed phase properties, which concerns especially the eletrostatics and structure. Thereby we extract the partial charges from bulk-like phase configurations and use the structure information given by radial distribution functions to adapt the short-range parameters of the force field.

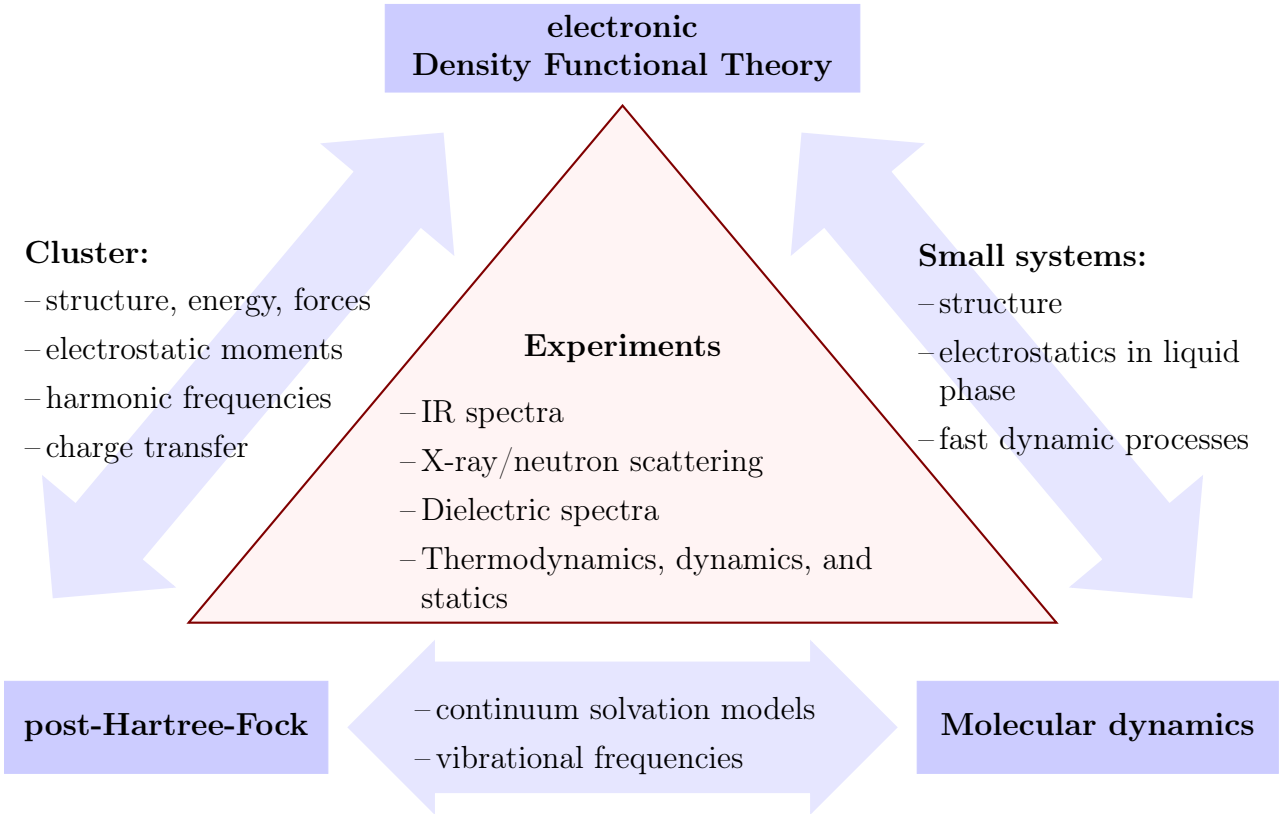


Figure 2.3.: An iterative cycle between the different levels of theory allows to benchmark the different methods and refine the FF parameterisation.

In order to transfer information between the different scales consistently, an iterative cycle involving all levels of theory is the long-term goal, such as depicted in Figure 2.3. As already

mentioned, the main object of this work is the connection of the DFT level with the classical part by an establishment of a procedure to refine or define classical FFs. Mostly the parameterisation of the Coulomb interaction is based on gas phase calculations, that completely neglect bulk effects. Afterwards this deficiency has to be removed in a demanding parameterisation process, which requires much knowledge about the system of interest. Studies on new compounds are difficult, because the parameterisation process requires quite long simulations to adapt the parameters to dynamic properties like the diffusion constant. To overcome this barrier the herein proposed method does not rely on the gas phase charge distribution, but the electronic properties are transferred from AIMD simulations capturing the properties of the condensed phase with the charge assignment method of Blöchl[55] (CAB). Only a single experimental property, the mass density, is involved in the setup of the AIMD simulations as well as in the FF parameterisation. Yet, this method is the first step towards a design of new chemical compounds *in-silico*.

So far, several articles have been published within this cooperation[1,7,9–12,14–17], which describe the development of the multiscale approach and its application. This also lead to a deeper understanding of the mechanisms dominating the behaviour of ionic liquids which in turn allows one to improve the already present methods. Additionally, an a–priori error estimate to optimize the calculation of electrostatic interactions in the classical MD simulations has been developed.[13]

3. The principles of classical molecular dynamics

Classical molecular dynamics (MD) simulations are a powerful tool to study condensed phases, because a system of hundreds of thousands of atoms can be propagated in time. Though one has to deal with atoms or molecules, their trajectory may be integrated following Newton's equation of motion. To render this possible, two assumption have to be valid.

The Born-Oppenheimer (BO) approximation[56] that shows that the motion of the atoms is given their nuclei. This arises from the large difference in mass between electrons and the nucleus resulting in a ratio of their velocities $v_{\text{electron}}/v_{\text{nucleus}} \approx 10^4$. Yet the electrons immediately react on a movement of the nucleus and the wave function of the systems can be separated into an electronic and a nuclear part. With that, the nuclei move in an effective potential where the coordinates of the electrons are treated as parameters. Hence the positions of the atoms can be integrated by considering them as point particles.

However, in order to derive the forces $\boldsymbol{F}$ which are required for the integration of the trajectories, a reliable parameterisation of them, a so-called force field (FF), is necessary. The FF consists of a set of parameters that describe the interactions in mathematically simple terms. These are often pair-wise additive and can be efficiently calculated on a computer. From the potential energy Φ given by the parameters, forces $\boldsymbol{F}$ are derived and applied to integrate Newton's equation of motion:

$$\boldsymbol{F} = -\boldsymbol{\nabla}\Phi. \tag{3.1}$$

This seems to be straightforward, but many aspects have to be considered, in order to accurately and efficiently propagate the coordinates and velocities of the particles. The next sections give a brief introduction in the "*art of molecular dynamics*" as Rapaport[57] stated. After a discussion of the mechanisms to simulate different thermodynamic ensembles, and an introduction to the principles of a FF, methods for the optimization of the simulations in terms of computational efficiency and accuracy are presented which mainly concerns the electrostatic interactions. An error estimate for the smooth-Particle-Mesh-Ewald method (SPME)[58] and the Particle-Particle-Particle mesh method with analytical differentiation (P^3M-AD)[59] is derived, which finally completes the set of accessible error estimates for Ewald-based methods. Yet an *a-priori* tuning of the simulations in terms of accuracy and speed becomes possible for all methods and also all sets of parameters. The chapter closes with a review about the development of IL FFs and a comparison of two of them, which are quite famous, where also the methods for the benchmarking are discussed in more detail.

3.1. Exploring different thermodynamic ensembles

To connect theory, experiment, and simulation correctly, it is necessary to assure that all branches discuss the same thermodynamic ensemble. Computer simulations have to be interpreted carefully due to finite size effects and numerical errors as well as require a suitable algorithm to realize a given situation. Finite-size effects arise from the limited number of particles that can be simulated and they are often minimized by the application of periodic boundary conditions. Apart from the limited machine accuracy, numerical errors result from unsuitable simulation parameters and can be significantly reduced by different optimization techniques. Finally the generation of the desired thermodynamic ensemble is a crucial part, because the ensemble averages for different kinds of ensemble only coincide for an infinite number of particles. Hence it has to be assured that the simulated ensemble is known.

This section sheds more light onto the relations between different kinds of simulations and ensembles. Therefore, canonical transformations, Liouville's law, and Trotter's theorems are discussed at first, in order to provide the basis for an easy understanding of integrators, thermostats and barostats that allow one to obtain different kinds of thermodynamic ensembles.

Canonical transformations, Liouville's Law, and the Trotter identity In classical mechanics the class of canonical transformations describes the time evolution of a system.[60] The basic idea is that a set of generalized coordinates and velocities exist determining the Lagrangian $\mathscr{L}$ uniquely as well as the corresponding Hamiltonian $\mathscr{H}$ which is derived by a Legendre transformation of $\mathscr{L}$ exchanging the generalized velocities $\dot{\boldsymbol{q}}$ with a generalized momentum $\boldsymbol{p}$. In terms of $\mathscr{H}$ the time evolution of the particles is given by the canonical equations:

$$\dot{\boldsymbol{q}} = \frac{\partial \mathscr{H}}{\partial \boldsymbol{p}}, \tag{3.2}$$

$$\dot{\boldsymbol{p}} = -\frac{\partial \mathscr{H}}{\partial \boldsymbol{q}}. \tag{3.3}$$

If this form is conserved under a transformation of $\boldsymbol{q} \to \boldsymbol{Q}$ and $\boldsymbol{p} \to \boldsymbol{P}$ to another set of generalized coordinates $\boldsymbol{Q}$ and momenta $\boldsymbol{P}$ the transformation is canonical. Yet a shift τ of the generalized coordinates $\boldsymbol{q}$ and momenta $\boldsymbol{p}$ in phase space:

$$\boldsymbol{q}(t+\tau) = \boldsymbol{Q}(\boldsymbol{q}(t), \boldsymbol{p}(t), \tau), \tag{3.4}$$

$$\boldsymbol{p}(t+\tau) = \boldsymbol{P}(\boldsymbol{q}(t), \boldsymbol{p}(t), \tau), \tag{3.5}$$

can be regarded as a canonical transformation, too. For this reason, an integration of the equations of motion for a certain time τ corresponds to a canonical transformation. Because Liouville's law states that a canonical transformation conserves the volume element $\mathrm{d}\Gamma = \mathrm{d}q_1\mathrm{d}p_1 \ldots \mathrm{d}q_n\mathrm{d}p_n$ of the phase space Γ, also a generation of a proper thermodynamic ensemble crucially depends on the conservation of phase space volume. Therefore, any scheme in an MD simulation involved in the integration of the canonical equations, such as barostats and thermostats as well as the integrator itself have to correspond to a canonical transformation. If the integration scheme conserves $\mathrm{d}\Gamma$ it is referred to as symplectic.

One possibility to express the equations of motions of the coordinates and momenta of the

particles is provided by the Liouville operator iL:[61–63]

$$\dot{f} = iLf, \text{ with} \tag{3.6}$$

$$iL = \dot{\boldsymbol{r}}\frac{\partial}{\partial \boldsymbol{r}} + \dot{\boldsymbol{p}}\frac{\partial}{\partial \boldsymbol{p}}. \tag{3.7}$$

The formal solution of equation (3.6) is simply derived by integration:

$$f(\boldsymbol{x}(t), \boldsymbol{p}(t)) = e^{iLt} f\left(\boldsymbol{x}(0), \boldsymbol{p}(0)\right), \tag{3.8}$$

but it is quite useless and does not offer much information about the time evolution of a system, which is determined by the canonical equations equations (3.2) and (3.3). In order to alleviate the interpretation of equation (3.8), some group theory is applied. From equation (3.8) it follows that the operator $T_t = e^{iLt}$ satisfies to the continuity condition: $\lim_{t\to 0} T_t f = f$, so its infinitesimal generator $\Omega = \lim_{t\to 0} t^{-1}(T_t - I)$ with the identity operation I is defined, and the semi-group condition: $T_{t+s} = T_t T_s$ for $s, t > 0$ is satisfied, because T_{t+s} has the same Ω as T_t and T_s, and $e^{iL(t+s)}f = f(t+s) = e^{iLt}f(s) = e^{iLt}e^{iLs}f(0)$. Hence, the Liouville operator iL is the infinitesimal generator Ω of a semi-group and the theorems of Trotter[64] guide a decomposition of Ω. One theorem proves that a decomposition of Ω in Ω_1 and Ω_2 such that $\Omega = \Omega_1 + a\Omega_2$ for $a > 0$ is only allowed if the operators T_1 and T_2 commute and therefore also their infinitesimal generators. In order to show that the Liouville operator only commutes in certain cases, the formalism of the Poisson bracket $\{\cdot, \cdot\}_{p,q}$[60] is introduced. Inserting equations (3.2) and (3.3) in equation (3.7) gives:

$$iLf = \frac{\partial \mathscr{H}}{\partial \boldsymbol{p}}\frac{\partial}{\partial \boldsymbol{r}}f - \frac{\partial \mathscr{H}}{\partial \boldsymbol{x}}\frac{\partial}{\partial \boldsymbol{p}}f, \tag{3.9}$$

$$iLf = \{\mathscr{H}, f\}_{p,q} = \dot{f} \tag{3.10}$$

Assuming the Hamiltonian $\mathscr{H}$ is split up in $\mathscr{H} = \mathscr{H}_1 + \mathscr{H}_2$, then the infinitesimal generators of the decomposition are given by $[\Omega_1, f]$ and $[\Omega_2, f]$. With the Jacobi identity of the Poisson brackets it is nicely illustrated that for commutative Ω_1 and Ω_2:

$$\left\{\Omega_1, \{\Omega_2, f\}_{p,q}\right\}_{p,q} - \left\{\Omega_2, \{\Omega_1, f\}_{p,q}\right\}_{p,q} = \left\{\{\Omega_1, \Omega_2\}_{p,q}, f\right\}_{p,q}, \tag{3.11}$$

$$\left\{\Omega_1, \dot{f}_2\right\}_{p,q} - \left\{\Omega_2, \dot{f}_1\right\}_{p,q} = 0, \tag{3.12}$$

$$\left\{\mathscr{H}_1, \dot{f}_2\right\}_{p,q} - \left\{\mathscr{H}_2, \dot{f}_1\right\}_{p,q} = 0, \tag{3.13}$$

equation (3.13) has to vanish. This is only the case, if f is a constant of motion in respect to $\mathscr{H}_1$ and $\mathscr{H}_2$ or $\mathscr{H}_2 = c\mathscr{H}_1$ for all $c \in \mathbb{R}$. For this reason Trotter's first theorem[64] usually does not apply. As a consequence the exponential in equation (3.8) must not be decomposed into products and applied one after each. In other words velocities and positions have to be known at the same point in time, to derive the acting forces and the resulting velocities and positions. The second theorem Trotter[64] proves that for any decomposition of iL a "product"

semi-group $S_{a,t}$ can be constructed:

$$S_{a,t} = \lim_{\tau \to 0} \left(T_\tau T'_{a\tau}\right)^{\left[\frac{t}{\tau}\right]}, \tag{3.14}$$

that conforms to the semi-group condition for all $a > 0$ and the infinitesimal generator of $S(a,t)$ becomes $\Omega_a = \Omega + a\Omega'$. With that, a decomposition of the Liouville operator is derived:

$$\exp(iLt) = \exp(iL_1 t + iL_2 t) = \lim_{\tau \to 0} \left[\exp(iL_1\tau)\exp(iL_2\tau)\right]^{\frac{t}{\tau}}. \tag{3.15}$$

Besides providing a possibility of decomposition, equation (3.15) justifies the stepwise integration of the equations motion, if the time steps τ are small enough:

$$f(t) = \lim_{\tau \to 0} \left[\exp(iL\tau)\right]^{\frac{t}{\tau}} f(0), \tag{3.16}$$

$$f(t) = \lim_{n \to \infty} \left[\exp(iL\tau)\right]^{n} f(0). \tag{3.17}$$

While equations (3.16) and (3.17) are exact for $\tau \to 0$, a small τ introduces an error of order $\mathcal{O}(\tau^3)$[61] for a single operation and an overall error of $\mathcal{O}(\tau^2)$ in the long time limit. As the way of a decomposition of iL is not unique only a suitable choice will help to interpret the physics behind this mathematical construction. If $iL = iL_p + iL_q$ is simply decomposed in two parts iL_p and iL_q depending only on the momenta and positions, respectively, equation (3.15) results in:

$$\exp(iLt) = \lim_{\tau \to 0} \left(\exp(iL_p\tau)\exp(iL_q\tau)\right)^n, \tag{3.18}$$

$$\Rightarrow \lim_{\tau \to 0} \exp(iL\tau) = \lim_{\tau \to 0} \exp(iL_p\tau)\exp(iL_q\tau). \tag{3.19}$$

However, the choice in equation (3.18) only yields a time-reversible scheme for the propagation of $f(t)$, if $\tau = 0$ or if $\boldsymbol{q}$ and $\boldsymbol{p}$ are both constants of motion, because:

$$f(\boldsymbol{q}(\tau), \boldsymbol{p}(\tau)) = \exp(iL_p\tau)\exp(iL_q\tau) f(\boldsymbol{q}(0), \boldsymbol{p}(0)), \tag{3.20}$$

$$= \exp(iL_p\tau) f(\boldsymbol{q}(0) + \dot{\boldsymbol{q}}(0)\tau, \boldsymbol{p}(0)), \tag{3.21}$$

$$= f(\boldsymbol{q}(0) + \dot{\boldsymbol{q}}(0)\tau, \boldsymbol{p}(0) + \dot{\boldsymbol{p}}(\tau)\tau), \tag{3.22}$$

and:

$$\exp(-iL\tau) f(\boldsymbol{q}(\tau), \boldsymbol{p}(\tau)) = \exp(-iL_p\tau) f(\boldsymbol{q}(0) + (\dot{\boldsymbol{q}}(0) - \dot{\boldsymbol{q}}(\tau))\tau, \boldsymbol{p}(0) + \dot{\boldsymbol{p}}(\tau)\tau), \tag{3.23}$$

$$= f(\boldsymbol{q}(0) + (\dot{\boldsymbol{q}}(0) - \dot{\boldsymbol{q}}(\tau))\tau, \boldsymbol{p}(0) + (\dot{\boldsymbol{p}}(\tau) - \dot{\boldsymbol{p}}(0))\tau). \tag{3.24}$$

Time-reversibility is an important aspect, as it is required for energy conservation, which is necessary for Newtonian or Hamiltonian dynamics. So equation (3.18) is unsuitable. The introduced Trotter decomposition is also applicable to Monte-Carlo simulations, which have to satisfy detailed-balance. A suitable decomposition was published by Creutz and Gocksch[65], who introduced a symmetric version of equation (3.18), such that an evolution of $f(t)$ in time

is given by:

$$f(\boldsymbol{q}(\tau),\boldsymbol{p}(\tau)) = \left(e^{iL_p\tau/2}e^{iL_q\tau}e^{iL_p\tau/2}\right) f(\boldsymbol{q}(0),\boldsymbol{p}(0)), \tag{3.25}$$
$$= e^{iL_p\tau/2}e^{iL_q\tau} f(\boldsymbol{q}(0),\boldsymbol{p}(0)+\dot{\boldsymbol{p}}(0)\tau/2), \tag{3.26}$$
$$= e^{iL_p\tau/2} f(\boldsymbol{q}(0)+\dot{\boldsymbol{q}}(\tau/2)\tau,\boldsymbol{p}(0)+\dot{\boldsymbol{p}}(0)\tau/2), \tag{3.27}$$
$$= f(\boldsymbol{q}(0)+\dot{\boldsymbol{q}}(\tau/2)\tau,\boldsymbol{p}(0)+\dot{\boldsymbol{p}}(0)\tau/2+\dot{\boldsymbol{p}}(\tau)\tau/2). \tag{3.28}$$

This provides a time-reversible integration scheme that is defined by the Hamiltonian $\mathscr{H}$ decomposed in $\mathscr{H} = \mathscr{H}_{p/2} + \mathscr{H}_q + \mathscr{H}_{p/2}$. As the transformation $f(0) \to f(\tau)$ is a solution of the canonical equations, it is a canonical transformation and hence the volume of the phase space element $\mathrm{d}\Gamma = \mathrm{d}q_1 \ldots \mathrm{d}q_n \mathrm{d}p_1 \ldots \mathrm{d}p_n$ is conserved. A small loop way to group theory was necessary to illustrate the step-wise time evolution of a system H propagated by different parts $\mathscr{H}_j$ of the corresponding Hamiltonian $\mathscr{H} = \sum \mathscr{H}_j$. Finally the respective Liouville operator $iL = \sum_j iL_j$ is obtained from the Hamiltonian $\mathscr{H} = \sum_i \mathscr{H}_i$ by a time-reversible decomposition. This formalism allows one to systematically include and understand various symplectic methods for the integration scheme in order to model different thermodynamic ensembles.

Ensemble averages and Ergodicity In order to derive ensemble averages $< P >_{\mathcal{E}}$ of an observable P from an MD simulation it is necessary, that the trajectory satisfies the ergodicity condition:

$$< P >_{\mathcal{E}} =< P >_t, \tag{3.29}$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} P(t_i) \tag{3.30}$$

Equation (3.29) is the ergodic theorem and the integration scheme has to conserve this system property. An analysis of this issue on the basis of the equations of motion is a difficult task, but the Trotter formalism provides an easy way to the solution of the problem.[62,63] If a pseudo Liouville operator $iL_\varepsilon = iL + \varepsilon/\tau$ is defined for a certain decomposition of iL ε acts as a measure of the introduced error. In order to supply a reasonable estimate for the error ε introduced by the stepwise integration, the Baker-Campbell-Hausdorff (BCH) formula[66] is applied to represent ε in terms of the $2k+1$-order commutators C_k of the elements of iL for a decomposition as suggested in equation (3.25):

$$\exp(iL_p\tau/2)\exp(iL_q\tau)\exp(iL_p\tau/2) = \exp[iL\tau + \sum_{k=1}^{\infty} C_k\tau^{2k+1}]. \tag{3.31}$$

Following Trotter[64] the left hand side of equation (3.31) is bound by construction for an infinitesimal small τ, but ε and iL_ε are only defined for a finite τ, if the series converges:

$$\varepsilon = \sum_{k=1}^{\infty} C_k\tau^{2k+1}. \tag{3.32}$$

In this case, the leading term in equation (3.32) will govern ε:

$$\varepsilon \approx C_1 \tau^3, \tag{3.33}$$

$$= -\frac{1}{24} \left\{ [2iL_p + iL_q, [iL_q, iL_p]] \right\} \tau^3. \tag{3.34}$$

A transition to infinitesimal small time steps τ yields an infinitesimal small error ε and the dynamics given by a Hamiltonian, that is described by its corresponding Liouville operator iL, are reproduced exactly. It should be noted, that if iL_ε exists it also defines a Hamiltonian $\mathscr{H}_\varepsilon$, that is exactly preserved by the time propagation scheme suggested in equation (3.25) and only differs from the true Hamiltonian $\mathscr{H}$ by a constant term in the order of $\varepsilon/\tau \sim \mathcal{O}(\tau^2)$. For this reason, no long term energy drift occurs and with that phase space of $\mathscr{H}_\varepsilon$ is sampled canonically and with that ergodicity is conserved. This means that if a trajectory, generated by any integrator conforming to the symmetric Trotter decomposition of $\mathscr{H}$, is long enough, the phase space is sampled dense enough and the ensemble average corresponds to the time average. Furthermore, it can be proven, that if a system is Lyapunov unstable[62], a long trajectory allows to calculate ensemble averages of time dependent properties. In case the momenta of the configurations only differ by a small amount δp, the deviation of the coordinates $\delta x(t)$ follows an exponential law dictated by the largest of the $6N$ Lyapunov exponents λ:

$$\delta q(t) \sim \Delta \exp(\lambda t). \tag{3.35}$$

For this reason, small differences in the initial conditions diverge quite fast resulting in independent configurations, which does not only allow to derive time-independent ensemble averages. Moreover, if a long trajectory is separated in blocks of sufficient length in time the resulting samples are independent from each other. Hence a large enough number of blocks also offers the possibility to derive ensemble averages of time-dependent properties.

Integrators for the microcanonical NVE ensemble If the total energy E given by a Hamiltonian $\mathscr{H}$ is conserved for a closed system of volume V consisting of N particles it corresponds to a microcanonical ensemble. The ensemble average $\langle \cdot \rangle$ of a quantity A is defined by the partition function $\Omega(\boldsymbol{q}, \boldsymbol{p})$ of the microstates $(\boldsymbol{q}, \boldsymbol{p})$ with energy E and Dirac's δ-function:

$$\langle A \rangle_{NVE} = \frac{\int\int A\Omega(\boldsymbol{q}, \boldsymbol{p}) \delta(\mathscr{H}(\boldsymbol{q}, \boldsymbol{p}) - E) \mathrm{d}\boldsymbol{q} \mathrm{d}\boldsymbol{p}}{\int\int \Omega(\boldsymbol{q}, \boldsymbol{p}) \mathrm{d}\boldsymbol{q} \mathrm{d}\boldsymbol{p}}. \tag{3.36}$$

Since not all microstates can be sampled with an MD simulations in a finite time span, the ergodic theorem (equation (3.29)) is used to derive $\langle A \rangle$ over n samples. If the samples are uncorrelated, the variance σ^2 is approximated by:

$$\sigma^2 = \frac{1}{n\,(n-1)} \sum_{i=1}^{n} \left(A_i - \langle A \rangle \right)^2, \tag{3.37}$$

which allows one to estimate the error $\Delta A = \sqrt{\sigma^2}$ of observable A introduced by the finite sampling of the phase space. If more than one observable A_j is involved in the calculation of a certain property $f(A_1, \ldots, A_N)$ Gauß's law of error propagation provides an estimate of the

variance of f, based on the variances σ_j of the different A_j:

$$\sigma_f^2 = \sum_{j=1}^{N} \left(\frac{\partial f}{\partial A_j}\bigg|_{A_j=\langle A_j\rangle} \right)^2 \sigma_j^2. \tag{3.38}$$

Because an efficient and symplectic integration scheme of an MD simulation is the basis for a successful generation of a thermodynamic ensemble, many integrators have been proposed during the development of computer simulations, but three of them: Verlet[67], Leap-Frog[59], and Velocity-Verlet[68], are very popular. The Verlet algorithm is derived from a sum of two Taylor expansions of the coordinates for $q(t+\tau)$ and $q(t-\tau)$ to ensure reversibility of the time evolution:

$$\boldsymbol{q}(t+\tau) = \boldsymbol{q}(t) + \dot{\boldsymbol{q}}(t)\tau + \frac{1}{2}\ddot{\boldsymbol{q}}\tau^2 + \ldots, \tag{3.39}$$

$$\boldsymbol{q}(t+\tau) + \boldsymbol{q}(t-\tau) = 2\boldsymbol{q}(t) + \frac{\boldsymbol{F}(t)}{2m}\tau^2, \tag{3.40}$$

$$\Rightarrow \boldsymbol{q}(t+\tau) = 2\boldsymbol{q}(t) + \frac{\boldsymbol{F}(t)}{2m}\tau^2 - \boldsymbol{q}(t-\tau). \tag{3.41}$$

Since all odd orders of the expansion cancel out each other, the error in the position is of order $\mathcal{O}(\tau^4)$, but from equation (3.41) a disadvantages become obvious. The velocities are not explicitly involved in the calculation. Obtaining them afterwards by averaging over $\boldsymbol{q}(t+\tau)$ and $\boldsymbol{q}(t-\tau)$ introduces an error in the order of $\mathcal{O}(\tau^2)$. If velocities are defined at every half time step $\tau/2$:

$$\boldsymbol{v}(t+\tau/2) = \frac{\boldsymbol{q}(t+\tau) - \boldsymbol{q}(t)}{\tau}, \tag{3.42}$$

$$\boldsymbol{v}(t-\tau/2) = \frac{\boldsymbol{q}(t) - \boldsymbol{q}(t-\tau)}{\tau}, \tag{3.43}$$

they can be included in the propagation scheme and equation (3.39) decomposes into a propagation scheme for the positions and velocities:

$$\boldsymbol{q}(t+\tau) = \boldsymbol{q}(t) + \boldsymbol{v}(t+\tau/2)\tau, \tag{3.44}$$

$$\boldsymbol{v}(t+\tau/2) = \boldsymbol{v}(t-\tau/2) + \frac{\boldsymbol{F}(t)}{m}\tau. \tag{3.45}$$

Though, velocities are explicitly derived during the simulation, they are not synchronously determined with the positions, but shifted by half a time step $\tau/2$ as depicted in Figure 3.1. The green lines demonstrate the propagation for a full time step τ of both velocities and positions. This is the update scheme of the so-called Leap-Frog algorithm. To access positions and velocities at the same point in time, the propagation of the velocities for a full time step is divided into two parts. In this case the well known scheme of the symmetric Trotter decomposition as given by equation (3.28) arises which is known as Velocity-Verlet algorithm. In this case the velocities $v(t)$ are integrated for $\tau/2$ before and after a full time step of the

integration of the positions, such that the forces at time t and $t+\tau$ are involved:

$$\boldsymbol{v}(t+\tau/2) = \boldsymbol{v}(t) + \frac{1}{m}\boldsymbol{F}(t)\tau/2, \tag{3.46}$$

$$\boldsymbol{q}(t+\tau) = \boldsymbol{q}(t) + \boldsymbol{v}(t+\tau/2)\tau, \text{ and,} \tag{3.47}$$

$$\boldsymbol{v}(t+\tau) = \boldsymbol{v}(t) + \frac{1}{2m}(\boldsymbol{F}(t) + \boldsymbol{F}(t+\tau))\tau. \tag{3.48}$$

It can be shown, that in terms of position all integrators are equivalent, because they are all connected by a reformulation and split of the calculations. The Velocity-Verlet algorithm introduces the smallest error of order $\mathcal{O}(\tau^3)$ to the current velocities compared to the Verlet and Leap-Frog method with an error in the order of $\mathcal{O}(\tau^2)$. However all of these methods conform to the symmetric decomposition scheme and its corresponding deviation from the true Hamiltonian given in equation (3.31) which is of order $\mathcal{O}(\tau^2)$ as discussed in the last paragraph. Moreover this shows that all introduced integrators are symplectic, which renders them suitable for a simulation of a microcanonical ensemble. Though the split of the propagation allows to access velocities and positions synchronously and increases the accuracy of the current velocity, it comes with the cost of a further calculation and communication step which matters for parallel computations on supercomputers.

Thermostats: a possibility for the simulation of a canonical NVT ensemble The introduced integrators allow a variety of studies, but they are restricted to closed systems. Often the situation occurs that a system R to study is in contact with a much larger one R', such that heat is transferred between R and R', while the temperature of the large system, called *reservoir* or *heat bath*, remains unchanged. Since R and R' are not isolated, energy is transferred and not strictly conserved for the single systems. The systems R and R' transfer heat until they achieve an equilibrium state, that maximizes the accessible phase space $\Omega(\boldsymbol{q}, \boldsymbol{p})$ of the microstates and the entropy S:

$$S = k_\mathrm{B} \ln \Omega. \tag{3.49}$$

If equilibrium is reached no heat is transferred and the system is characterised by the external parameters N, V, and T. With the Helmholtz free energy F:

$$F = (E - TS), \tag{3.50}$$

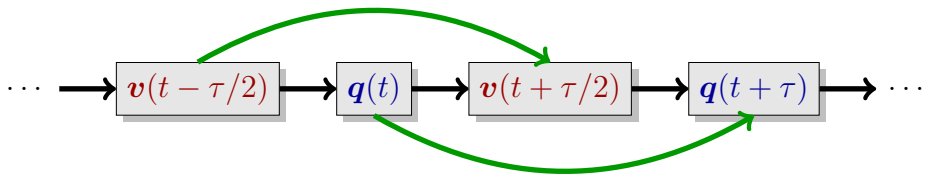


Figure 3.1.: Update scheme of the Leap-Frog algorithms for the velocities $\boldsymbol{v}$ and positions $\boldsymbol{q}$.

the equilibrium condition becomes:

$$\mathrm{d}F = \mathrm{d}E - T\mathrm{d}S = 0, \tag{3.51}$$

hence F is a constant of motion. While in the microcanonical ensemble the energy E is the conserved quantity, the Helmholtz free energy F is constant in the canonical NVT ensemble and equation (3.36) becomes:

$$\langle A \rangle_{NVT} = \frac{\iint A\Omega(\boldsymbol{q},\boldsymbol{p})\delta(E - TS - F_0)\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p}}{\iint \Omega(\boldsymbol{q},\boldsymbol{p})\delta(E - TS - F_0)\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p}}. \tag{3.52}$$

By some algebra and under the assumptions that $N = N_R \ll N_{R'}$ and $\ln\Omega$ is maximal, a partition function $\Omega_{NVT}(p,q)$ is derived for the small system R:

$$\ln\Omega_{NVT}(\boldsymbol{q},\boldsymbol{p}) = -\beta\mathscr{H} + \ln(\sum_i e^{-\beta\mathscr{H}_{RR',i}}), \tag{3.53}$$

where the sum involves all accessible microstates i of the total system and $\beta = \frac{1}{k_\mathrm{B}T}$ with the Boltzmann constant k_B. The second term of equation (3.53) is constant and is determined by the normalization condition, such that the integral of equation (3.52) becomes:

$$\langle A \rangle_{NVT} = \frac{\iint A\exp(-\beta\mathscr{H}(\boldsymbol{q},\boldsymbol{p}))\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p}}{\iint \exp(-\beta\mathscr{H}(\boldsymbol{q},\boldsymbol{p}))\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p}}. \tag{3.54}$$

In the microcanonical ensemble only the states with exactly same energy occur, but the energy of the states in the canonical ensemble is scattered around an average value defined by the external parameters:

$$\langle (E/N)^2 \rangle - \langle E/N \rangle^2 = k_B T^2 (C/N)/N, \tag{3.55}$$

where C is the extensive heat capacity. These fluctuations vanish in the thermodynamic limit $N \to \infty$, and the canonical and microcanonical ensemble become equivalent.

In order to achieve canonical conditions in an MD simulation, an obvious possibility is a scaling of the velocities, because the temperature is defined by the number of degrees of freedom n_dof and the kinetic energy E_kin:

$$\frac{n_\mathrm{dof}}{2} k_\mathrm{B} T = E_\mathrm{kin}. \tag{3.56}$$

Berendsen *et al.*[69] introduced such a thermostat that scales the velocities and respectively the kinetic energy E^kin such that they relax according to time constant τ_T:

$$\mathrm{d}E_\mathrm{kin} = \left(E_\mathrm{kin}^\mathrm{target} - E_\mathrm{kin}\right)\frac{\mathrm{d}t}{\tau_T}. \tag{3.57}$$

towards the target average kinetic energy $E_\mathrm{kin}^\mathrm{target}$ given by equation (3.56) This method allows to equilibrate a system in a reliable and quite fast manner, but it does not generate a canonical ensemble. A quite simple extension of this scheme was proposed by Bussi *et al.*[70] where a random term is added to equation (3.57) in order to obtain a canonical distribution of

velocities and according states:

$$dE_{\mathrm{kin}} = \left(E_{\mathrm{kin}}^{\mathrm{target}} - E_{\mathrm{kin}}\right)\frac{\mathrm{d}t}{\tau_t} + 2\sqrt{\frac{E_{\mathrm{kin}}^{\mathrm{target}}E_{\mathrm{kin}}}{n_{\mathrm{dof}}}}\frac{dW}{\sqrt{\tau_t}}. \tag{3.58}$$

The first term corresponds to the method suggested by Berendsen[69] and the second term influences the change dE_{kin} by the increment of a stochastic Wiener process dW that is specified by the Ornstein-Uhlenbeck process, an overdamped harmonic oscillator. The scaling factor α for the velocities:

$$\alpha^2 = e^{-\tau/\tau_T} + \frac{E_{\mathrm{kin}}^{\mathrm{target}}}{n_{\mathrm{dof}}E_{\mathrm{kin}}}(1-e^{-\tau/\tau_T})(R_1^2 + \sum_{i=2}^{n_{\mathrm{dof}}} R_i^2) + 2e^{-\tau/2\tau_T}\sqrt{\frac{E_{\mathrm{kin}}^{\mathrm{target}}}{n_{\mathrm{dof}}E_{\mathrm{kin}}}(1-e^{-\tau/\tau_T})}R_1, \tag{3.59}$$

is derived based on n_{dof} Gaussian distributed random numbers R_i and a parameter τ_T determining the strenght of coupling: For the limit $\tau_T \to 0$, an instantaneous adaption of the kinetic energy is achieved and an increasing τ_T smoothens the coupling towards longer time scales.

Another approach is the construction of a *pseudo* Lagrangian $\mathscr{L}_{pseudo}$, such that a constant temperature T^{ref} is introduced as a constraint. With that, equations of motion for the scaling factors of the velocities are derived by $\mathscr{L}_{pseudo}$. A famous thermostat of this kind was published by Nóse[71] and refined by Hoover[72]. It relies on the thermodynamic friction variable ξ which is coupled to the particle momenta and results in following equations of motions:

$$\dot{\boldsymbol{q}}_i = \frac{\boldsymbol{p}_i}{m_i}, \tag{3.60}$$

$$\dot{\boldsymbol{p}}_i = -\nabla\Phi - \xi\boldsymbol{p}_i, \text{ and} \tag{3.61}$$

$$\dot{\xi} = \left(\sum_{i=1}^{N}\frac{p_i^2}{m_i} - n_{\mathrm{dof}}k_{\mathrm{B}}T^{\mathrm{ref}}\right)/Q, \tag{3.62}$$

where an effective mass Q determines the coupling strength. From equation (3.60) it becomes obvious, that Q should be adapted to the desired temperature. However, since Q is not explicitly defined equation (3.62) can be reformulated such that the constant factor $n_{\mathrm{dof}}k_B$ is compensated by the effective mass Q':

$$\dot{\xi} = \left(T - T^{\mathrm{ref}}\right)/Q'. \tag{3.63}$$

Finally Q' is defined by a coupling time constant τ_T:

$$Q' = \frac{\tau_T^2 T}{4\pi^2} \tag{3.64}$$

as implemented in the MD program package GROMACS[73–76] and described in the corresponding user manual[77]. With that the coupling strength of the thermostat can be set independent of system size and reference temperature. This method is able to propagate the system canonically, but different conditions have to be satisfied, because only a single conservation law can be satisfied. External forces have to be absent and the center of mass has to remain fixed. However if the requirements are met, a Maxwell-Boltzmann distribution for the velocities and

energy fluctuations corresponding to equation (3.55) arise and ergodicity is conserved.

An extension to this method includes a coupling of the thermostat to one or more further thermostats. This method is named the *Nóse-Hoover-chain*(NHC) thermostat[78] and solves the mentioned problems. Since velocities are involved in the propagation of the thermostat, the Verlet integrator is obviously not applicable in this schemes. In terms of a Trotter decomposition, the final integrator for NHC is:

$$\exp(iL\tau) = \exp(iL_{\mathrm{NHC}}\tau/2)\exp(iL_{p,q}\tau)\exp(iL_{\mathrm{NHC}}\tau/2). \tag{3.65}$$

In this form equation (3.65) requires knowledge about the energy and with that positions and velocities at the same point in time, which is only provided by Velocity-Verlet. For Leap-Frog another decomposition can be derived[79]:

$$\exp(iL\tau) = \exp(iL_q\tau)\exp(iL_{\mathrm{NHC}}\tau/2)\exp(iL_p\tau)\exp(iL_{\mathrm{NHC}}\tau/2), \tag{3.66}$$

which is also time-reversible and symplectic. This is of importance, because Leap-Frog is computationally more efficient if communication is required in highly-parallel computations. Moreover, since for a chain length of one the Nose-Hoover thermostat is recovered from the NHC thermostat, equation (3.66) also includes this special case.

Barostats: discovering the canonical NpT ensemble In experiment the situation of an NpT instead of an NVT ensemble is often given, if a system is additionally able to perform mechanical work by a change of its volume. The equilibrium state for this ensemble is defined by a constant temperature and pressure. Instead of the Helmholtz free energy F, Gibb's free enthalpy G is a conserved quantity:

$$G = F + pV, \tag{3.67}$$
$$= E - TS + pV, \tag{3.68}$$

and the equilibrium condition becomes:

$$\mathrm{d}G = \mathrm{d}E - T\mathrm{d}S + p\mathrm{d}V, \tag{3.69}$$
$$= 0. \tag{3.70}$$

The accessible phase space is extended, because the coordinates q of the particles are not restricted to a certain volume anymore and the ensemble average corresponds to equation (3.52), but without limits for q. In the thermodynamic limit, the fluctuations of the temperature and volume tend to zero and all introduced ensembles become equivalent.

To simulate the effect of excerted pressure, similar methods as for thermostatting have been suggested. The pressure calculation is based on the tensor of the kinetic energy $\boldsymbol{E}_{\mathrm{kin}}$ and the

virial $\boldsymbol{\Xi}$ with:

$$\boldsymbol{E}_{\text{kin}} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i \otimes v_i, \text{ and,} \tag{3.71}$$

$$\boldsymbol{\Xi} = -\frac{1}{2}\sum_{i<j}^{N} \boldsymbol{q}_{ij} \otimes \boldsymbol{F}_{ij}, \tag{3.72}$$

where the direct vector product $\otimes$ is defined by:

$$(\boldsymbol{a} \otimes \boldsymbol{b})^{\alpha\beta} = a_\alpha b_\beta, \tag{3.73}$$

and the tensor notation for the indizes α and β is applied. Combining equations (3.71) and (3.72) finally determines the pressure tensor $\boldsymbol{P}$:

$$\boldsymbol{P} = \frac{2}{V}\left(\boldsymbol{E}_{\text{kin}} - \boldsymbol{\Xi}\right), \tag{3.74}$$

and the scalar pressure $p = trace(\boldsymbol{P})/3$.

The most simple approach to simulate constant pressure has been introduced by Berendsen *et al.*[69] which corresponds to their thermostatting scheme. The pressure is relaxed according to a coupling time τ_{P} towards a reference pressure P^{ref}:

$$\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} = \frac{\boldsymbol{P}^{\text{ref}} - \boldsymbol{P}}{\tau_P}, \tag{3.75}$$

$$= -\frac{1}{\beta_{ij} V}\frac{\mathrm{d}V}{\mathrm{d}t}. \tag{3.76}$$

which is connected to a change in volume by the isothermal compressibility β_{ij}. If the equations of motion for the positions q, given by the velocity v, are extended by a term proportional to q:

$$\dot{\boldsymbol{q}} = \boldsymbol{v} + a\boldsymbol{x}, \tag{3.77}$$

the change of the volume in time becomes:

$$\dot{V} = 3aV. \tag{3.78}$$

Combining equations (3.76) and (3.78) allows to determine a and the scaling matrix μ for the coordinates and box vectors during a time step τ becomes:

$$\mu_{ij} = \delta_{ij} - \frac{\tau}{3\tau_{\text{P}}}\beta_{ij}(P_{ij}^{\text{ref}} - P_{ij}). \tag{3.79}$$

It should be noted that only the positions and box vectors are scaled:

$$\boldsymbol{q} \rightarrow \mu_{ij}\boldsymbol{q}, \tag{3.80}$$

but not the velocities $\boldsymbol{v}$. This method is robust and suitable for the equilibration of a system, but the generated ensemble is unknown.

Another technique relies on equations of motions for the box vectors $\boldsymbol{b}_i$ that are derived from a *pseudo* Lagrangian. Based on the idea of Parrinello and Rahman[80], Nosé[71] introduced a set of generalized coordinates and momenta, that can be easily combined with their thermostat. The imbalance between P^{target} and P, which denote the external and internal pressure, respectively, drives the propagation of the barostat and a further effective mass W describes the coupling strength. If only a uniform deformation of the cell is allowed, the scalar pressure p^{target} is enough to define the equation of motion for the volume V:

$$W\ddot{V} = p - p^{\text{target}}. \tag{3.81}$$

The equations of motions are easily extended to the anisotropic case, where the box $b = b_{ij} = (\boldsymbol{b}_i)_j$ and the coupling parameter $W = W_{ij}$ are described by a matrix. Furthermore W can be reformulated such that the coupling strength is determined by a time constant τ_{P}, which is independent of the system size:[77]

$$\frac{\mathrm{d}^2 b}{\mathrm{d}t^2} = VW^{-1}(b^T)^{-1}(\boldsymbol{P} - \boldsymbol{P}^{\text{target}}), \tag{3.82}$$

and $(W^{-1})_{ij}$ is defined by:

$$(W^{-1})_{ij} = \frac{4\pi^2 \beta_{ij}}{3\tau_{\text{P}}^2 L}, \tag{3.83}$$

where L is the size of the largest box vector. Corresponding to the NH time constant τ_{T}, τ_{P} is representative for the oscillation period of the coupling bath and hence also for the maximal response to a change in $\boldsymbol{P}$. This method has been proven to be symplectic, but it does only approximate the true NpT ensembles, though static observables, that depend only on the positions of the atoms, are predicted correctly[71,81]. Only in the thermodynamic limit the pressure virial theorems:

$$\langle P \rangle = P^{\text{target}}, \text{ and,} \tag{3.84}$$

$$\langle PV \rangle = P^{\text{target}} \langle V \rangle - kT, \tag{3.85}$$

are valid, otherwise an error in the order of $\mathcal{O}(N^{-1})$ is introduced. As large systems with thousands of particles are considered in this work, the error is acceptable and the method will be applied for the generation of an NpT ensemble.

3.2. The force field

So far, methods for a proper propagation in time were discussed assuming that energies and the corresponding forces of the system are known. On the molecular scale quantum mechanics are expected to dominate the dynamics, but for a progagation in time, the atomics can be considered as point particles, in correspondence to the Born-Oppenheimer approximation. MD simulations of liquids are usually applied in the regime, where energy fluctuations are in the range of the thermal energy $k_{\text{B}}T$, which is larger than many QM contributions. This fact is used and a so-called force field (FF) is introduced that describes the atomic interactions by simple physical models. Much effort was spent in the development of FFs over the years. A certain kind of FFs are all–atom models, like AMBER[82] and OPLS-AA[53], which

provide a description of certain liquids and organic molecules on an atomic level. They are computationally expensive, but this detailed description is often required to implement all relevant degrees of freedom properly. For the establishment of a force field the parameters are usually fitted to experimentally or computationally derived properties in order to reproduce further experimental quantities. One should always bear in mind that altering only a single parameter can introduce unexpected artefacts and destroy the consistency of the whole FF. This section provides an introduction to the commonly applied terms, which are summarized in Figure 3.2, and where they originate from. In a classical MD simulation, the energies are usually separated into contributions from the bonded energy E_{bonded}, which determine the molecular structure, and non-bonded contributions $E_{\text{non-bonded}}$, which describe the interactions between the different atoms:

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{non-bonded}}. \tag{3.86}$$

The non–bonded terms $E_{\text{non-bonded}}$ are constituted by electrostatic interaction and a short-range interaction function, that describes Pauli repulsion, polarization and dispersive effects. The establishment of an accurate classical FF is a quite difficult task since n-body interactions have to be described by pairwise-additive potentials. This seems to be impossible, but an implicit description of the n-body interactions is often sufficient to describe the physics of the system on an appropriate level of accuracy. In the following paragraphs, these kinds of interactions and their connections are described in more detail and the standard potential forms, which are used in this work, are introduced.

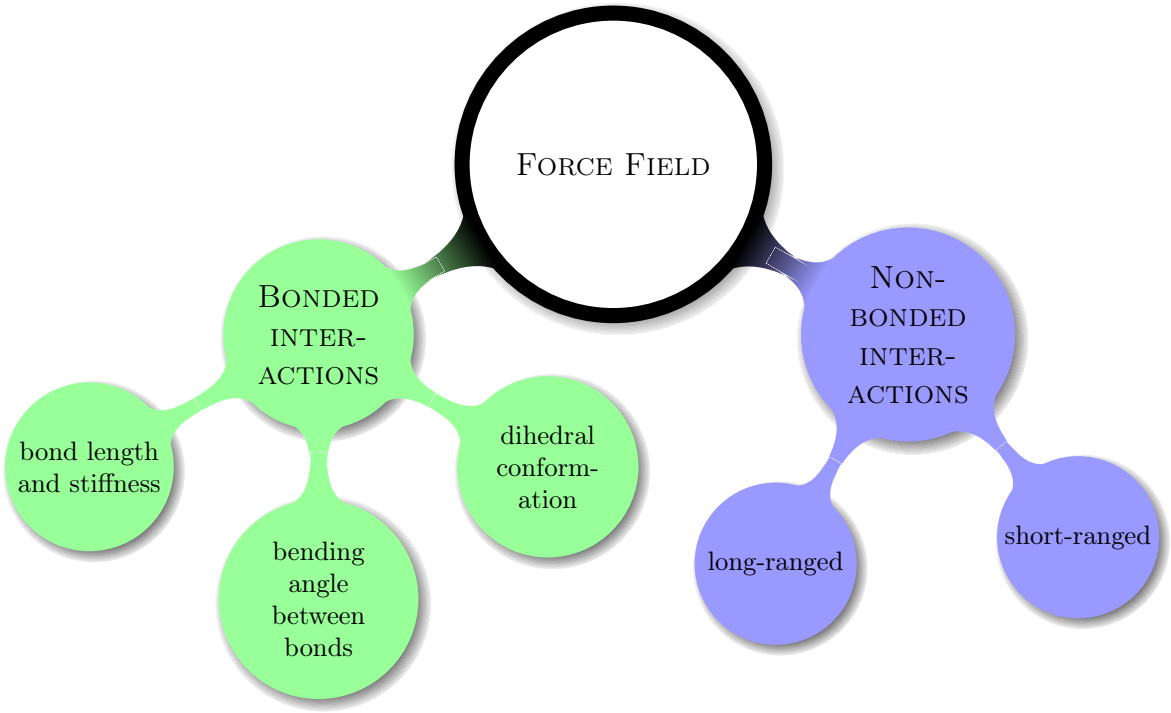


Figure 3.2.: A scheme of a classical force field.

Nonbonded interactions Usually, non-bonded (NB) interactions are characterized in respect to their interaction range: short-ranged and long-ranged. The short-ranged part is a superposition of various interactions. It should model dispersion, arising from the electron correlation, as well as Pauli repulsion, and polarization effects. There is no unique equation to combine this interactions in a simple pairwise-additive potential, but it has been shown that the Lennard-Jones (LJ) potential allows a good approximation:

$$E_{\mathrm{LJ}} = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right], \tag{3.87}$$

While the repulsive term $\sim r^{-12}$ is completely empirical, the attractive term $\sim r^{-6}$ should approximate the dispersion energy. Every kind of atom i has its own set of LJ parameters, σ_i and ϵ_j, and two methods are commonly employed to combine the parameters between different types of atoms i and j. One of these combination rules is based on the geometrical mean of the different types:

$$\sigma_{ij} = \sqrt{\sigma_i \sigma_j}, \quad \text{and} \tag{3.88}$$
$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}. \tag{3.89}$$

The other one is called Lorentz-Berthelot rule, where an arithmetic average for σ and a geometric average for ϵ is applied:

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_i + \sigma_j\right), \quad \text{and} \tag{3.90}$$
$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}. \tag{3.91}$$

Either of them have a physical background, but they have to be consistently applied during the force field construction. If the atoms carry a partial charge Coulomb interaction arises:

$$E_{\mathrm{coul}}(r) = \frac{1}{4\pi\epsilon_r} \frac{q_i q_j}{r}, \tag{3.92}$$

that also contributes to the short range part of the FF and constitutes the long-range part of the NB interactions. For a system that consists only of atoms, no bonds are involved and the force field is already given by the introduced terms. Unfortunately, they are not straightforward to derive, because they depend on the actual environment. It is possible to derive the parameters with the help of quantum mechanical calculations, but they are restricted to small and isolated or highly symmetric systems. The parameters derived from this calculations are often not valid in the regime of interest and optimization methods are necessary, which are described in section 4.1 and chapter 4. Partial charges are usually derived from a fit of the electrostatic potential (ESP) with methods like RESP[83] or CHELP[84]. Both rely on a minimization of the error between the true ESP and the ESP, that is given by point charges located on the atoms, on a suitable chosen surface. This is hard to achieve in a condensed state. In order to access the charge distribution of the liquid phase, an alternative method is introduced in this work, which relies on the electron density of bulk geometries. For the LJ parameters, straightforward methods are not present. Tosi and Fumi[85,86] proposed a technique to determine LJ parameters for crystals from QM calculations. Mostly these parameter can only be applied as an educated guess for a further parameterisation process, because the

parameters vary when the system crosses over from the crystalline to the liquid state. Many methods have been proposed to fit force fields to certain quantities as the heat of vaporization, diffusion constants, crystal configurations, or forces based on quantum-mechanics, which has become recently very popular since the increasing computer power allows to perform corresponding simulations. Finally they all have in common that an error function is minimized and therefore a larger number of appropriate simulations have to be setup, performed, and analyzed, tasks for which recently Hülsmann *et al.*[87] proposed a work flow to perform. The optimization process of a force field is computationally very demanding and an almost never ending story as new molecular compounds have to included. An efficient, massively parallel working technique is necessary to minimize the required amount of time, which is described in chapter 4. It is based on the idea to distribute the workload over a large number of processor allowing a linear scaling with the number of force field parameters, which have to tuned.

Bonded interactions The intramolecular forces, given by covalent bonds, describe the different stretching and bending modes of a molecule. Three different types of interactions are usually applied:

$$E_{\text{bond}} = \frac{1}{2} k_{\text{b}} \left(r - r_0\right)^2 , \tag{3.93}$$

$$E_{\text{angle}} = \frac{1}{2} k_{\text{a}} \left(\Theta - \Theta_0\right)^2 , \text{ and,} \tag{3.94}$$

$$E_{\text{dihedral}} = \sum_{i=1}^{4} \frac{V_i}{2} \left[1 + (-1)^{i-1} \cos\left(i\varphi\right)\right] . \tag{3.95}$$

Equation (3.93) and equation (3.94) describe the stretching and bending of the bonds, respectively, and equation (3.95) parameterises the energy due to intramolecular twisting. However, equation (3.95) can also be parameterised in a computationally more efficient form, the so called Ryckaert-Bellemans function[88]:

$$E_{\text{RB}} = \sum_{n=0}^{5} C_n \left(\cos\left(\Psi\right)\right)^n . \tag{3.96}$$

Apart from the different functional form, the corresponding dihedral angles are also defined in a different manner. Equation (3.95) utilizes the angles given by the IUPAC/IUB convention, where φ is the angle between the planes with $\varphi = 0$ corresponding to a *cis* configuration. The Ryckaert-Bellemans function uses the polymer convention resulting in $\Psi = \varphi - 180°$, which corresponds to $\Psi_{trans} = 0°$. In contrast to the nonbonded interactions, the parameterisation of the intermolecular energies is quite insensitive to the environment and straightforward. The bond and angle parameters are adapted to the ground state energies of the involved bonds and the corresponding eigenfrequencies, which are independent of the environment.

Usually the interactions between the atoms separated by one or two bonds are completely included in the bond and angle parameters. But the dihedral parameters describe an interaction of atoms that are separated by exactly three atoms, which is the definition of a 1–4 interaction. In many force fields the LJ and short-range electrostatic interaction partially contribute to the 1–4 interaction, but similar to the combination rules for the short range parameters, no physical background is provided. While OPLS uses a scaling factor ϵ^{1-4} of 0.5, the interactions

are scaled down by 0.8 in AMBER. Yet the total 1–4 energy E^{1-4} is:

$$E^{1-4} = E_{\text{dihedral}} + \epsilon^{1-4}(E_{\text{coul.}} + E_{\text{LJ}}). \tag{3.97}$$

For this reason, an optimization of the non-bonded parameters also involves a consistent refit of the dihedral parameters, because the overall torsional profile, given by equation (3.97), should be maintained. A corresponding technique is included in the parameter optimization routine (see Appendix A), and described in more detail in section 5.1.

Development of FFs is a difficult task, because most of the parameters depend on each other and a large number of issues has to addressed. A reasonable description of the molecular system over a wide range of temperature and pressure is required as well as compatibility with as much molecular compounds as possible. In order to allow one a straightforward addition of a further component to the FF, a strategy is proposed in this work covering all aspects, which are important to obtain a parameterisation as accurate and transferable as possible.

3.3. Optimization of simulations

MD simulations are a powerful tool to investigate macromolecular systems, but finite size effects and long-range interactions are of major concern. Finite size effects are usually circumvented by periodic boundary conditions (PBC), but this requires an efficient and appropriate treatment of the long-range interactions. In order to allow one an optimization of the corresponding simulation parameters in terms of accuracy and performance, the error introduced by the algorithm has to be estimated *a-priori*. In GROMACS an efficient treatment of the electrostatic interactions is only provided by the SPME algorithm[58]. For this kind of electrostatics solver an error estimate had to be developed,[13] which is only applicable to certain sets of parameters. For this reason, an error estimate for an arbitrary set of SPME parameters is developed in this section and a corresponding GROMACS tool, `g_pme_error`, which allows one to estimate the error and tune the parameters such that the error is below a certain limit. This tool also became part of the official GROMACS release.

Periodic boundary conditions and short ranged interactions On the molecular scale the major part of the atoms covers the surface region of the simulation box, which is mostly not desired. In order to get rid of this artefact, periodic boundary conditions (PBC) are applied. Instead of a single simulation box, also referred to as unit cell, an infinite lattice of unit cells is considered for the calculations. No surface effects arise, but the unit cell has to be large enough to avoid artefacts due to the artificial symmetry. So as to derive the respective forces all replicas have to be considered, which mainly concerns the long-range part of the forces.

The short range (SR) forces are characterized by a fast decay, such that they are are almost negligible beyond a certain cut–radius r_{c} and excluded from the force and energy calculation. This approximation is applied in different methods to decrease the number of required calculations. In the famous neighbourlist method introduced by Verlet[67], a neighbourlist cut–off $r_{\text{n}} > r_{\text{c}}$ is defined. If a list of neighbours is generated that is used for the force calculation. This method reduces the scaling of the calculations from an order $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$ as long as no particle diffusses from within $r < r_{\text{c}}$ to out of the neighbour search domain $r > r_{\text{n}}$.

However, an update of the neighbourlist involves $\mathcal{O}(N^2)$ operations. To circumvent the poorly scaling update of the neighbourlist, Hockney and Eastwood[59] proposed a method that assigns the particles to a grid, that can be constructed by an operation of order $\mathcal{O}(N)$. Yet also the force calculation is reduced to $\mathcal{O}(N)$.

Besides the computational load, also the numerical inaccuracies for the calculation of the SR interactions introduced by r_c for the can be reduced. A correction to the neglected dispersion energy can be derived under the assumption, that the system is homogeneous beyond r_c[89]. As forces between $r_c - \delta$ and $r_c + \delta$ are not smooth for all $\delta > 0$, because $E^\mathrm{SR} = 0$ for $r > r_\mathrm{c}$, energy and momentum are not strictly conserved. This artefacts are reduced to a minimum by a shift or a scaling of the short-range interaction, such that the interaction vanishes at r_c. A FF is always constructed in respect to a certain kind of potential with a respective r_c. For this reason, r_c has to be considered as a fixed force field parameter, that should not be changed arbitrarily, because artefacts may arise.

If large systems are regarded a single processor is often not capable to produce the required statistics in a reasonable amount of time. For this reason, the work is distributed over a network of processors, but this has to be done carefully to achieve an optimal load of the individual CPUs. In this scenario, the communication between the CPUs is always the bottleneck for the simulation and has to be reduced to a minimum. Since SR interactions are only considered within r_c, neighbor lists can be constructed, in order that no CPU time is wasted to find the small fraction of particles involved in the force calculations. The most obvious method is to divide particles on the different CPUs, but this is unsuitable for a large number of processors P, because the number of communications scale with P^2.[77] However the scaling can be improved, if the unit cell is decomposed in separate spatial domains and dedicated to the different CPUs communication is reduced strongly, because only information about particles crossing the domain border has to be exchanged. This scheme is called domain decomposition and allows to reduce the communication within the nodes to a minimum.[73,90]

Long-ranged electrostatic interactions in PBC As soon as the atoms are charged or gravitation is involved, long-range forces arise. A simple truncation of theses forces corresponding to the SR forces introduces serious artefacts, for which reason an appropriate treatment is necessary. Ewald[91] encountered this problem as he was interested in the calculation of the Madelung constant of a periodic crystal which involves the sum over an infinite number of unit cell replicas. He solved the problem by a split of the periodic potential in two parts $E = E_\mathrm{real} + E_\mathrm{recip}$, such that E_real and E_recip exponentially converge in real and reciprocal space, respectively. Though the sum only converges conditionally, for a cubic or cylindric summation over a few replicas the error is quite small.

This situation corresponds to a system with periodic boundary conditions (PBC). Yet many algorithms to solve the electrostatics are based on this idea. Though many other approaches have been suggested, like the Fast-Multipole method[92,93] (FMM), MMM techniques[94–96] or the Wolf summation[97] to mention just a few, only Ewald-based methods are described in this work, because they are relevant for the performed simulations. Moreover a method is proposed to optimize the parameters in terms of accuracy and speed.

A short summary of the Ewald technique is shown in Figure 3.3. If the calculations are splitted over real and reciprocal space, four different terms can be separated. Besides the

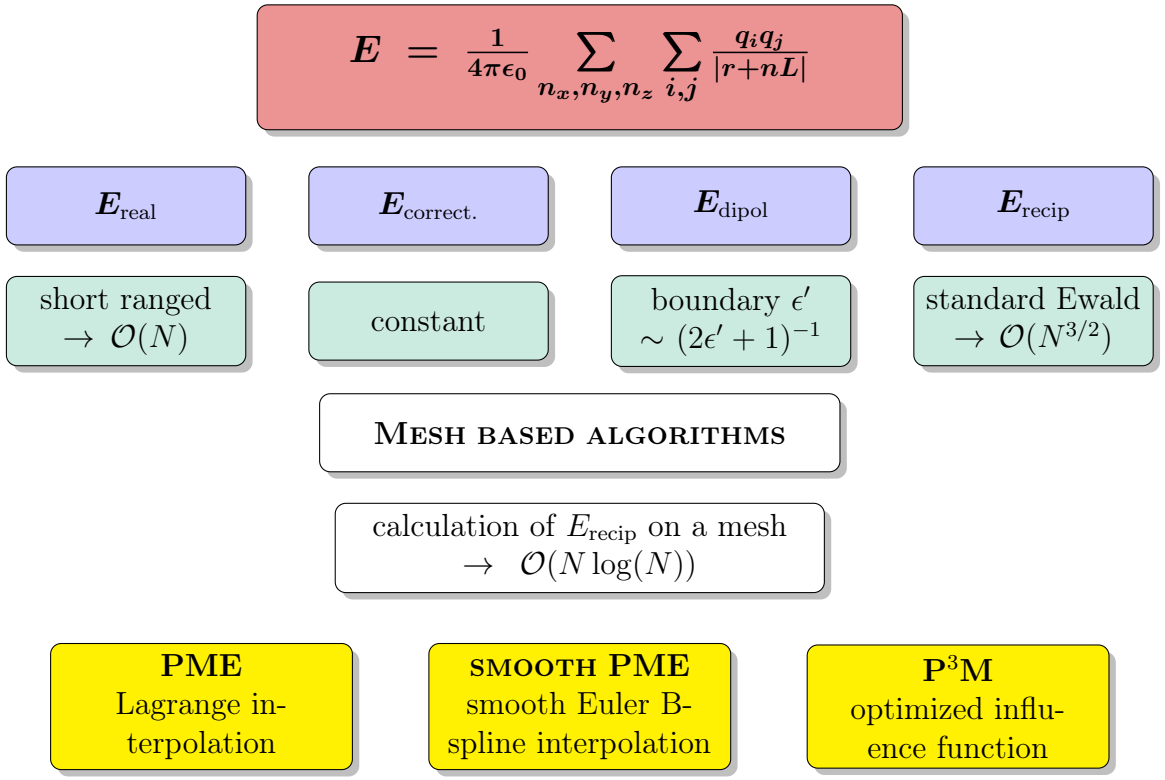


Figure 3.3.: Ewald-based methods for electrostatic interactions in PBCs.

constributions from the short-ranged term in real space E_{real} and the reciprocal term E_{recip}, the decomposition introduces a constant correction energy $E_{\text{correct.}}$. This term depends on the dipole moment $\boldsymbol{M}$ of the unit cell as well as the dielectric constant of the surrounding medium ϵ' and affects the fluctuations of $\boldsymbol{M}$. Neumann and Steinhauser[98] have shown that a conducting system like an ionic liquid, requires a perfectly conducting medium with $\epsilon' = \infty$. This boundary conditions are referred to as metallic. Yet the dipolar term E_{dipol} vanishes. As metallic boundaries are always applied in this work and E_{dipol} does not depend on the parameters of the Ewald techniques, it is not considered in the following discussion. The parameters required as input for the Ewald summation span a large space and only a suitable choice will yield an acceptable error and computation time. Kolafa and Perram[99] as well as Hummer[100] developed an *a-priori* error estimate that allows to determine an optimal set of Ewald parameters in terms of speed and accuracy, such that the distribution of the work load between real and reciprocal space provides a scaling of the summation of $\mathcal{O}(N^{3/2})$.

The standard Ewald summation is a superior method for systems containing hundreds of particles, but it is computationally too expensive for systems containing thousands or even more particles. This disadvantage arises from the scaling behaviour and the limited capabilities to run the computation on many processors in parallel. These problems were circumvented with the introduction of mesh based algorithms, which are summarized in Figure 3.3. Darden *et al.*[101] published the Particle Mesh Ewald (PME) method that achieves an $\mathcal{O}(N\log(N))$ scaling. PME uses fundamental Lagrange interpolation that has the advantage of error cancellation during charge assignment[102], but suffers from a discontinuous charge density for odd interpolation orders n. Furthermore, the charge density is not differentiable, independent of

n. Essmann *et al.*[58] suggested a charge assignment method with Euler B-splines. This results in a $(n-2)$-times continuous differentiable grid charge density, that has compact support.

Actually before the invention of PME and SPME, Hockney and Eastwood[59] already proposed the Particle-Particle-Particle mesh (P^3M) method. In contrast to PME or SPME, an optimized Green function $G^{\text{opt.}}$ is applied to derive energy and force from a mesh-based charge density. The term "optimized" regards to the ability to describe the *true* electrostatic interaction given by a continuous charge density as good as possible by the mesh-based charge density. In this context, it has to be mentioned that the optimized $G^{\text{opt.}}$ looks different for the energy and force, so one has to decide beforehand which one to use. Nevertheless, the differences are marginal. This results in a higher accuracy of P^3M compared to SPME for equal mesh sizes and corresponding n. Yet if available P^3M should always be the method of choice, but SPME is a widely applied method and implemented in many MD simulation packages, like GROMACS or AMBER. In contrast to P^3M, where the problem of an *a-priori* estimate of the error was already solved,[59,102,103] an approach for SPME had to be elaborated and has recently been published,[13] but the estimate is only applicable to even n.

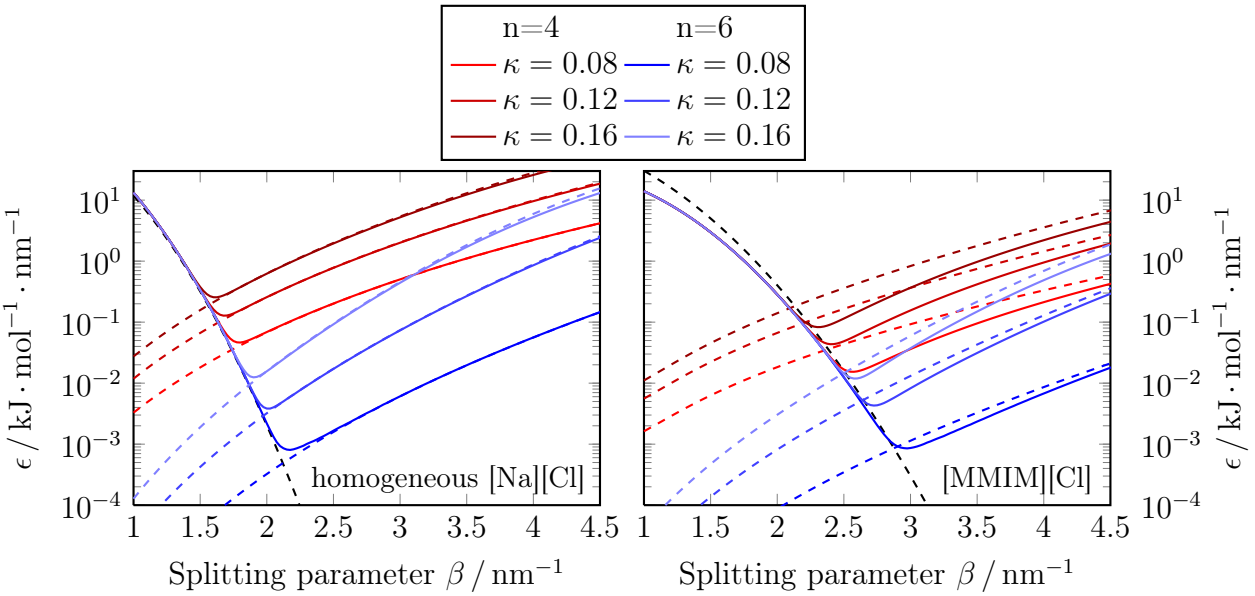


Figure 3.4.: Calculated (solid) and estimated (dashed) error introduced by the SPME algorithm for even n.

For a system of N atoms the estimated and calculated error is compared in terms of the root mean square force (RMSF) error $\epsilon = \sqrt{\frac{1}{N}\sum_N (\boldsymbol{F}_{\text{SPME}} - \boldsymbol{F}_{\text{ref}})}$ with reference forces $\boldsymbol{F}_{ref}$ derived from an Ewald sum of machine precision. Corresponding parameters for the Ewald sum have been obtain by the approach published by Hummer[100]. The error estimate assumes a homogeneous charge density and independent charges. On the left hand side of Figure 3.4 results are shown for 500 ion pairs in a cubic box of volume $V = (10\ \text{nm})^3$, where the ions were randomly placed within the box and overlapping is allowed. The solid and dashed lines depict the actual and estimated errors, respectively. A constant cutoff $r_c = 1.7$ nm is applied and the estimated error is split up in its contribution from real and reciprocal space. The calculated and estimated errors agree for different interpolation orders n and grid spacings $\kappa = L_\alpha / K_\alpha$

with box lengths L_α and grid points K_α for the spatial dimensions $\alpha = 1, \ldots, 3$.

In contrast, on the right hand side of Figure 3.4, the results for an inhomogeneous charge distribution are shown, where some of the atoms are connected via bonds and therefore not independent. The systems is equilibrated and no overlap of the atoms is possible. In this case, the error is overestimated. On the one hand side, this allows one to tune the parameters *a-priori*, such that an upper bound for the error is always guaranteed, but on the other hand side, this *overtuning* of the parameters causes a performance loss. Comparing the estimated error for $n = 4$ and $n = 6$ shows, that the estimate agrees much better for a larger n.

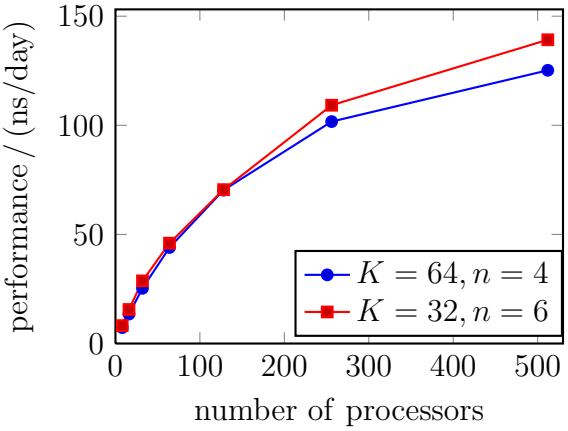


Figure 3.5.: Timing results for a system of 500 [MMIM][Cl] IPs, where up to 512 processors were involved. The SPME parameters were tuned to an error $\epsilon \approx 5 \cdot 10^{-2}\ \mathrm{kJ \cdot mol^{-1} \cdot nm^{-1}}$ for interpolation orders $n = 4$ and $n = 6$. In real space a cutoff $r_c = 1.1$ nm is applied.

In Figure 3.5 timing results for a system of 500 [MMIM][Cl] ion pairs are given with two sets of SPME parameters, tuned with a constant cutoff $r_c = 1.1\,\mathrm{nm}$. Interpolation orders $n = 4$ and $n = 6$ are applied to achieve an error $\epsilon \approx 5 \cdot 10^{-2}\ \mathrm{kJ \cdot mol^{-1} \cdot nm^{-1}}$, which results in $K = 64$ and $K = 32$ grid points in each spatial dimension, respectively. The system is simulated for 1100 steps with 8, 16, 32, 64, 128, 256, and 512 processors and the last 1000 steps are used for the timing. As GROMACS allows one to assign the workload of the reciprocal space on separate CPUs, the tool `g_tune_pme` has been used to perform the timings, because it tests different numbers of processors, exclusively dedicated to SPME. For the analysis, always the fastest settings were chosen.

Obviously, both sets of parameters perform equally, but sometimes the simulations with $n = 6$ are slightly faster. For this reason two points support to prefer $n = 6$ for the simulations. At first, a slightly higher performance is achieved over a wide range of applied processors compared to the tuned parameters with $n = 4$. At second, the behaviour of the error estimate for inhomogeneous systems agrees better with the actual error for $n = 6$, which allows a more precise tuning. The timing results may change significantly for a different system size, but the property of an improving estimate accuracy with increasing n remains. As the performance of the simulations for odd n is also of interest, an error estimate for SPME and $\mathrm{P^3M}$ is proposed that includes this condition.

An error estimate for an arbitrary set of SPME parameters Very recently, the mathematical framework to convert SPME and P^{3}M has been established by Ballenegger *et al.*[104]. It should be mentioned, that two different schemes to derive the forces from the mesh-bashed energy are available. The *ik*-scheme, which is a simple multiplication in Fourier space, is usually applied in the P^{3}M algorithm, because it conserves the momentum, but requires one Fast Fourier transformation (FFT) into reciprocal space and three FFTs back in real space, which requires a lot of global communication in parallel simulations. The other method is referred to as analytical differentiation (AD) scheme, which is also usually applied in combination with SPME. In the AD scheme the forces are derived from the analytically differentiated reciprocal energy. Though this method is not momentum conserving, it requires only a single FFT back in real space and conserves energy. For this reason, it is often preferred for large systems, which require a highly parallelized enviroment to achieve a reasonable performance. Actually, P^{3}M-AD and SPME only have different functions $\Omega(\boldsymbol{m})$ in their influence function $G(\boldsymbol{m})$:

$$G(\boldsymbol{m}) = \Omega(\boldsymbol{m})\varphi(\boldsymbol{m}), \tag{3.98}$$

with:

$$\varphi(\boldsymbol{m}) = \frac{4\pi}{\boldsymbol{m}^2} \exp\left(-\frac{\boldsymbol{m}^2}{4\beta^2}\right), \tag{3.99}$$

where β is the splitting parameter between real and reciprocal space and $\boldsymbol{m}$ is a grid vector in reciprocal space, which is spanned by $\boldsymbol{a}_i^\star = \varepsilon_{ijk}\frac{1}{\boldsymbol{a}_1 \cdot \boldsymbol{a}_2 \times \boldsymbol{a}_3}\boldsymbol{a}_j \times \boldsymbol{a}_k$ with the Levi-Civita tensor ε_{ijk}. The real space basis vectors $\boldsymbol{a}_i$ are orthonormal to $\boldsymbol{a}_j^\star$, $\boldsymbol{a}_i \cdot \boldsymbol{a}_j^\star = \delta_{ij}$. Every SPME algorithm is simply converted to a P^{3}M algorithm by a substitution of Ω. Only the routine which involves equation (3.98) has to be modified, such that the corresponding Ω of P^{3}M is calculated. This allows to use an optimized influence function and apply the error estimates derived for P^{3}M, which are valid for odd and even interpolation orders.

All details are provided for the implementation of P^{3}M, but a single piece of the error estimate for the AD scheme is missing, the estimate of the self-forces. In contrast to the *ik*-scheme, where the self-forces cancel out, the AD scheme generates self-forces, which have to be either subtracted during the simulation or estimated *a-priori*. Yet for studies of errors, the self-forces were subtracted in the simulation, but this is not standard case for all MD packages, such that a corresponding estimate is required. A formula for the self-force is present[105] and inhere the average is derived by integrating the square of the force over the volume V:

$$\boldsymbol{F}^{\text{self}}(\boldsymbol{r}) = \sum_{\boldsymbol{m}} \boldsymbol{b}^{(\boldsymbol{m})} \sin(2\pi \boldsymbol{m}\boldsymbol{s}), \tag{3.100}$$

with $s_\alpha = r_\beta / h_\alpha$ and $h_\alpha = L_\alpha / N_\alpha$ for $\alpha = 1, 2, 3$. The coefficient $\boldsymbol{b}^{(\boldsymbol{m})}$ is given by:

$$\boldsymbol{b}^{(\boldsymbol{m})} = \boldsymbol{k}_{\boldsymbol{N}\boldsymbol{m}} \frac{1}{2V} \sum_{\boldsymbol{k}_{\boldsymbol{n}}} G(\boldsymbol{k}_{\boldsymbol{n}}) f^{(\boldsymbol{m})}(\boldsymbol{k}_{\boldsymbol{n}}), \tag{3.101}$$

with the Green's function G and:

$$f^{(\boldsymbol{m})}(\boldsymbol{k}_{\boldsymbol{n}}) = \sum_{\boldsymbol{m}' \in \mathbb{Z}^3} \hat{\boldsymbol{U}}(\boldsymbol{k}_{\boldsymbol{n}+\boldsymbol{N}\boldsymbol{m}'}) \hat{\boldsymbol{U}}(\boldsymbol{k}_{\boldsymbol{n}+\boldsymbol{N}\boldsymbol{m}'+\boldsymbol{N}\boldsymbol{m}}). \tag{3.102}$$

The functions $\hat{\boldsymbol{U}}$ are proportional to the Fourier transform of the charges assignment function and given in Ballenegger *et al.*[104], (eq. 2.21). Integrating equation (3.100) over the volume of a unit cell provides the mean self-force:

$$\frac{1}{V}\int (F^{\text{self}}(\boldsymbol{r}))^2 d^3r = \frac{q_i^4}{V}\sum_{\boldsymbol{m},\boldsymbol{m}'}(\boldsymbol{b}^{(\boldsymbol{m})})(\boldsymbol{b}^{(\boldsymbol{m}')})\int \sin(2\pi\boldsymbol{m}\boldsymbol{s})\sin(2\pi\boldsymbol{m}'\boldsymbol{s})d^3r, \tag{3.103}$$

$$= q_i^4\sum_{\boldsymbol{m}}(\boldsymbol{b}^{(\boldsymbol{m})})^2. \tag{3.104}$$

The integral in equation (3.103) is $V/2$ in case of $\boldsymbol{m} = \boldsymbol{m}'$ and $\boldsymbol{m} = -\boldsymbol{m}'$. All other terms are identically zero, such that the average self-force for a particle reduces to the sum in equation (3.104). The average of the self-force also vanishes, because the integral is equal to zero. With that, only equation (3.104) contributes to the common error formula for P^{3}M for the reciprocal error:

$$(\Delta F^{\text{rec}})^2 = \frac{1}{V^2}\frac{Q}{N}\left(\sum_{i=1}^{N} q_i^2\right)^2 + \frac{1}{N}\Big(\sum_{m\in\mathbb{Z}^3}(\boldsymbol{b}^{(\boldsymbol{m})})^2\Big)\sum_{i=1}^{N} q_i^4, \tag{3.105}$$

$$= \frac{1}{V^2}\frac{Q}{N}\left(\sum_{i=1}^{N} q_i^2\right)^2 + \frac{S}{N}\sum_{i=1}^{N} q_i^4. \tag{3.106}$$

The factor Q is given in Ballenegger *et al.*[104], (eq. 4.8), and depends on the Green's or influence function G and the differentiation method. Therefore, Q and S can be calculated for both SPME and P3M. Thus an error estimate for all interpolation orders of SPME as well as an explicit estimate of the self-force F^{self} is available.

P^{3}M-AD is expected to achieve a higher accuracy for the same parameters compared to SPME, so it was implemented in the GROMACS 4.6 branch and is switched on with the `mdp` option `coulombtype=p3m-ad`. With it, an influence function G^{opt} optimized in respect to the energy is applied to calculate the reciprocal part of the electrostatic interaction. The GROMACS tool `g_pme_error` has been prepared in order to estimate the error and tune the parameters of SPME and P^{3}M-AD, in terms of accuracy, for a fixed cutoff r_c in real space. For the reciprocal part, the error estimate is based on equation (3.106) and the real space error is derived by the method of Kolafa and Perram[99]. An error estimate and tuning routine for the Ewald sum has also been included. It is based on the work of Hummer[100], that provides a formula for the reciprocal space cut-off, such that the error can easily is equally distributed over real and reciprocal space. Finally a combination of the GROMACS tools `g_tune_pme` and `g_pme_error` allows one to derive a set of SPME or P^{3}M paramters optimized in terms of accuracy and speed.

In order to numerically verify the error estimate, the homogeneous configuration of 500 [Na][Cl] IPs, discussed in the last paragraph, is chosen. A real space cutoff of 1.7 nm is applied to compare the errors for different interpolation orders and mesh sizes in respect to the splitting parameter β, shown in Figure 3.6. Meshes with $K = 32$ and $K = 48$ points in every spatial dimension are given on the left and right hand side, respectively. The estimated and calculated errors agrees well for both, SPME and P^{3}M-AD, for all n. For a large mesh size, the difference between SPME and P^{3}M-AD is negligible, but a significant difference arises for small mesh

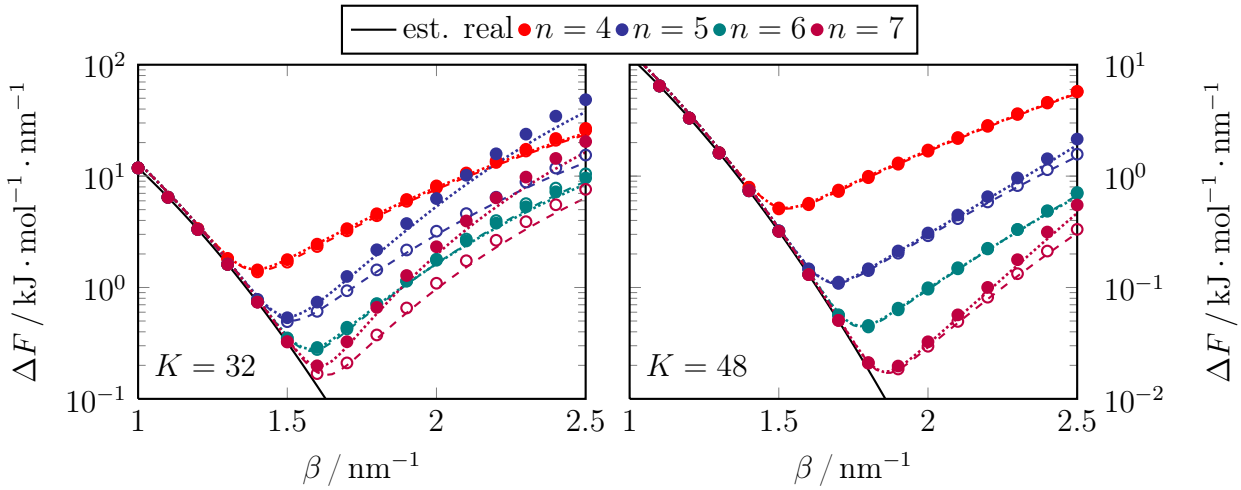


Figure 3.6.: RMSF errors for different n and K in every spatial dimension. Estimated errors are shown by markers. Dashed and densely dotted lines are the calculated errors for the SPME and P3M method, respectively.

sizes, especially if odd n are considered. It shows that a proper setting of β is important, because the minimum of the error is quite narrow, especially in the case of SPME. For small mesh sizes P^{3}M-AD should be the method of choice, while for larger mesh sizes the difference between the methods vanishes in the regime of β, where the minimal error is achieved.

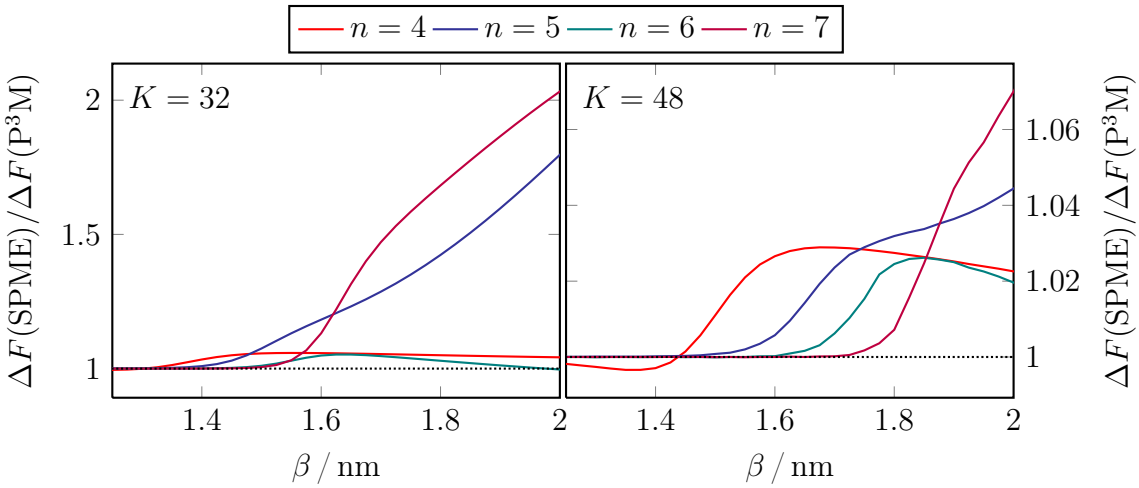


Figure 3.7.: Ratio of the RMS error of the forces ΔF of SPME and P3M-AD.

In Figure 3.7, the force error ratio $\Delta F(\text{SPME})/\Delta F(\text{P}^3\text{M})$ of SPME to P^{3}M-AD is shown, in order to study the difference in accuracy. P^{3}M-AD always shows a smaller or equal error in the minimal error regime, because the ratio always has a value larger than one there. So the choice of P^{3}M-AD will only bring advantages, but no drawbacks, compared to SPME. A consistent and comprehensive comparison of accuracy and speed for the different branches of the mesh-based Ewald techniques has been elaborated in the diploma thesis of Weik[106]. Besides the

introduced SPME and P^{3}M methods, also interlaced versions of these algorithms[107,108] are considered which rely on the average of two eletrostatic energies derived from grids that are shifted such that errors may cancel out. It has been shown that the interlaced version of P^{3}M-AD performs out all investigated Ewald-methods at a given method. This strongly suggest to convert and extend an SPME scheme to an interlaced P^{3}M-AD scheme, if energy conservation and performance are of interest.

Finally, the applicability of the new error estimate with equation (3.106) to inhomogeneous systems is studied. An equilibrated system of 500 [MMIM][Cl] ion pairs is chosen, with a unit cell given by a cubic box of $V = (4.62935\,\mathrm{nm})^3$. Instead of setting the number of mesh points K, the Fourier spacing $\kappa = L_\alpha/K_\alpha$ is used to determine the mesh size K_α by the length L_α of the three box vectors.

In Figure 3.8, errors for SPME and P^{3}M-AD with a cutoff $r_c = 1.5$ nm and Fourier spacings $\kappa = 0.10$, 0.12, and 0.14 nm are compared for $n = 4$ and $n = 6$. As expected, the error is overestimated for all cases, but the deviation of the estimate from the calculated error decreases with increasing interpolation order, shown in Figure 3.4. The overestimation seems not to be affected by the mesh size and no significant difference between SPME and P^{3}M-AD in terms of accuracy is observed.

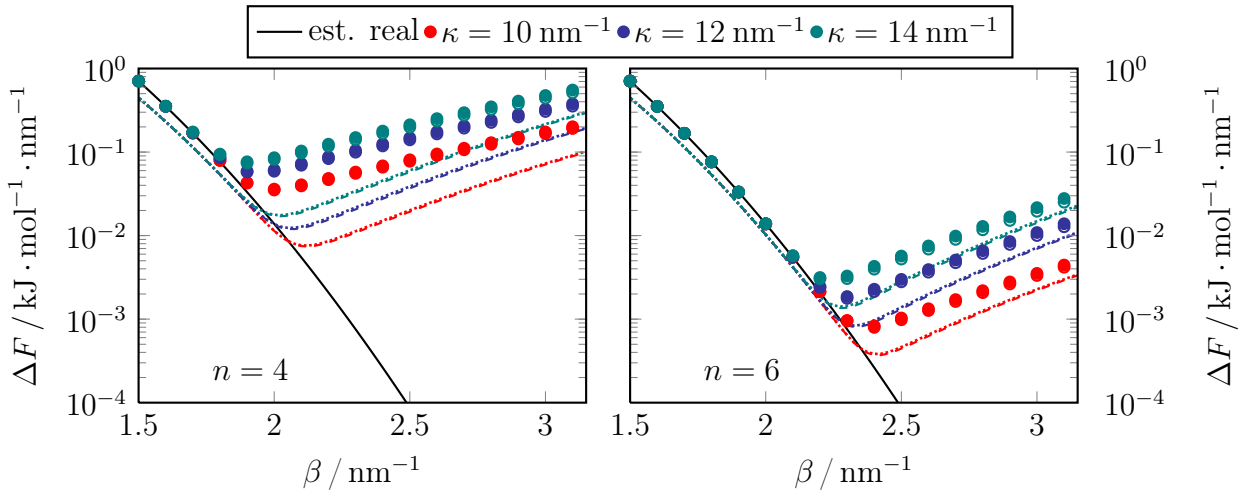


Figure 3.8.: RMSF errors of the electrostatic forces for an inhomogeneous [MMIM][Cl] system are compared for different n and κ. Estimated errors are given by markers. Dashed and densely dotted lines are the calculated errors for the SPME and P3M method, respectively.

Yet mesh-based algorithms and error estimates are easily transformed into each other. Due to an estimate of the mean self-force for P^{3}M-AD, provided here, and the conversion formulas to SPME,[104] an error estimate for SPME is available for all interpolation orders now as well as an implementation of the P^{3}M-AD scheme in GROMACS. The numerical test showed, that if the assumptions of distribution homogeneity and charge independence holds the estimates agree with the calculated errors. For more realistic system, such as the [MMIM][Cl] IL, the error is overestimated. This allows one to determine the upper bound of the error, but some performance is lost, because the same accuracy could be achieved with parameters that require

less computational effort. For this reason, the settings of the eletrostatic algorithm should be chosen carefully for every simulation regarding both, accuracy and speed.

3.4. Force fields for ionic liquids

Due to the increasing interest in RTILs in the last decades, also many computational techniques have been established and refined to study these compounds. This class of solvents brought new challenges to the fields of computational physics and chemistry, due to the size of the solvent molecules and their properties, such as a high viscosity, strong electrostatic interaction between the solvent molecules, and high polarizability. For quantum mechanical calculations, the size of the solvent species is the main problem. Though a variety of properties, such as electrostatics, energy, geometries or vibrational frequencies of single ion pairs or very small clusters can be studied, it is unclear how much their behaviour differs compared to bulk-like systems. Semi-quantum mechanical DFT calculations offer the possibility to study the dynamics of bulky systems, but suffer from a short simulation time and a lack of experience regarding the choice of a suitable XC functional. Finally, classical MD simulations aim at a generation of a thermodynamic ensembles that allows to calculate a variety of properties, such as density, viscosity, or conductivity, which are of main interest for IL systems. However, these MD simulations rely on an accurate force field which is often lacking for ILs. Fortunately, the computer power increases steadily, alleviating the QM calculations, which in turn provide the basis for the development of corresponding interaction potentials for the classical simulations. The construction of a FF is not a straightforward task and usually involves many iterative cycles. Nowadays, MD simulations are a common tool especially in biochemistry and biophysics. For this reason, a lot of effort has been put into the development of adequate parameterizations for biomolecular FFs to study proteins or DNA, where AMBER,[82,109] CHARMM,[110] and OPLS-AA[53,111] are some prominent examples. These FFs describe a large class of molecules over a certain temperature range, but their development is still ongoing, because new molecules are added and parameters are refined. However, an important property of force fields, which is often hard to achieve, is transferability over various molecules and temperatures, because a simple combination of parameters from different force fields for different molecules does not guarantee a valid parameterisation.

Hence, in order to study ILs with MD simulations, corresponding potentials are required. Reviewing the history of FF development for RTILs allows to identify the main methods which are commonly employed and introduced in the following paragraphs. As number of ionic liquids is rather large, the discussion is restricted to imidazolium-based ionic liquids, but the main conclusions are also valid for non-imidazolium ILs. For a more detailed overview, the reader is referred to literature.[16]

Transfer of established non-ionic liquid force fields In Figure 3.2 of section 3.2, a scheme of a classical FF is given which clearly shows the different types of interaction, usually present in an MD simulation. While the parameters for the bonded interactions can be straightforwardly calculated with QM methods, different methods are commonly used to obtain a suitable set of non-bonded parameters, which are often decomposed in long-range and short-range contributions. Usually electrostatics solely determine to the long range interactions,

but the short-range parameters have to describe various effects. Besides the contribution of the electrostatics to the short-range interactions, QM mechanisms, such as Pauli exclusion or electron-electron correlation, have to be included adequately in order to describe the interactions accurately. Different techniques are available to calculate a set of partial charges, but the parameterisation of the remaining short-range parameters is a tedious work. Hence, many FF for ILs are a combination of parameters from established FFs and partial charges derived for the particular ionic liquid of interest.

The first FF for an IL has been published by Hanke *et al.*[112] in 2001. While the bonded and short range parameters are chosen from different FFs,[113–116] the partial charges were derived by a second-order Møller-Plesset perturbation theory based distributed multipole analysis. This practice, to restrict the refinement of a FF to the recalculation of partial charges, continued and parameterisations for different ionic liquids were published,[28,117,118] based on different prominent FFs. The FFs were successfully validated against structural features obtained from neutron diffraction and diffusion constants determined by NMR experiments, but only few properties are studied, because further experimental knowledge was lacking, which is even nowadays often the case.

In 2002, an important concept for the charge calculation has been introduced by Morrow and Maginn[119]. While the short-range and bonded parameters were transferred from CHARMM, partial charges were obtained from ion pair instead of isolated ion calculations. This yields a reduction of the ionic net-charge, which sounds unphysical at first. Studying this effect in more detail (see section 4.4 and section 4.4) shows that it corresponds to an implicit description of polarization. Unfortunately, this concept has not been considered further at this point in time, but in recent years it became an essential part of FF parameterisation for ILs.

Unfortunately, the number of ILs is rather large, hence establishing a force field for every IL separately would consume a huge amount of time. For this reason, a FF transferable over many cations and ions is desirable. An important step towards this direction has been achieved by Canongia Lopes *et al.*[5], who established CLaP,[5,120–122] a transferable FF for a large number of ILs, by combining the short-range and bonded parameters from different established parameterisations with their own set of partial charges and small refinements on some bonded parameters. CLaP provides a proper description of the static properties, but fails in an accurate description of the dynamics, which is shown in section 3.5. However, due to its transferability, CLaP provided the starting point for many other IL FFs.

Hence, the technique to determine a set of IL specific partial charges and combine them with a present FF for a similar compound, allows one to obtain a more or less accurate FF for the specific IL. But the collection of the parameters may be tedious and a proper benchmark is required, in order to verify that all properties of the IL are described appropriately.

Explicit parametrization to the liquid phase All IL FFs, introduced so far, contain partial charges that have been parameterised in the gas phase with either a single ion or an ion pair. The parameters for the short-range interactions were simply chosen from established FFs, which were not explicitly parameterised for ILs. The bond parameters were either transferred from similar chemical complexes or explicitly derived. Thereby it has been assumed that the parameters do not significantly differ between gas and liquid phase as well as transferability between the FFs. Though some properties have been successfully validated, a benchmark

involving a large set of properties was not feasible unfortunately and a solid background for the assumed parameter transferability is lacking. However the results of Morrow and Maginn[119] have already shown that polarization plays an important role for ILs. For this reason, it should also be appropriately modelled by the force field parameters, but an explicit treatment is very time-consuming. Sometimes a proper adjustment of the short-range parameters is sufficient to model the polarization effects, but therefore corresponding experimetal data is required, such that the parameters can be fitted in order to match a certain set of adequately chosen properties, which are sensitive to the effects that have to be modelled. Since it has been observed, that polarization plays an important role for the dynamics of ILs[123], diffusion constants and rotational correlation times are suitable properties for an adaption of the force field parameters. Unfortunately, this is also not a straightforward task for ILs, because often data is lacking or inaccurate. Moreover, due to the high viscosity of ILs long simulations are required for a sampling of the phase space in order to derive well-converged ensemble averages. Finally, what's more, the parameters should be transferable to different states in the phase diagram. This difficult task is necessary to parameterise a generic FF for a large class of ILs.

An example for the breakdown of the transferability assumption can nicely be demonstrated. Liu *et al.*[124] introduced a parameterisation, denoted LHW in here, for imidazolium-based cations and the $[BF_4]^-$ and $[PF_6]^-$ anions. The bonded and short-range parameters are completely based on AMBER, apart for the acidic ring hydrogen H^1 of $[EMIM]^+$ (a scheme of the ions is shown in Figure 3.9). The short-range parameters of this hydrogen have been adapted in order to achieve agreement with molecular geometries obtained with accurate QM methods. The FF has been successfully validated against the mass density, but the phase space has been sampled only 100 ps, which does not allow to obtain a reliable averages of dynamic properties. Though, CLaP and LHW are quite similar, the most significant difference is the treatment of the bonded hydrogens. In CLaP, the hydrogen bond lengths are constraint, while LHW treats them as flexible. The effect of this difference and especially a benchmark of LHW against CLaP in terms of dynamics is discussed in the next section 3.5, but at this point the important message is that both FFs, CLaP and LHW, underestimate the dynamics of the IL. Finally this is not astonishing, because bulk properties were not involved in the parameterisation. It has already been shown that an adequate choice of the partial charges and short-range parameters allows one to describe an IL system in both aspects, statics and dynamics, properly, but an exploration of the huge amount of possible combinations of different FF parameters consumes very much time. For this reason, techniques are required to efficiently parametrize the bulk effects within ILs into the FF parameters.

The force-matching technique, which is becoming more and more prominent due to the increasing computer power, allows one to adapt the whole set of FF parameters to a provided set of coordinates and forces. In order to do so, reference data is obtained by an *ab-initio* method and the parameters of the FFs are fitted such that they match the reference forces. With that, information about the bulk properties is mediated by the forces. This method has been applied by Youngs *et al.*[125] to [MMIM][Cl], where all FF parameters were involved in the fitting process with the constraint of an integer ionic net-charge. The resulting FF parameters finally show the expected liquid character, given by an increase of the dipole moment, which is a result of an extended electron cloud around the ions, given by their mutual interaction. However, force-matching requires reference data, which is gathered by computationally very demanding calculations. Another possibility is a fit of the FF parameters to experimental liquid phase properties. With this method, Köddermann *et al.*[126] derived a force field for the

members of [C_nMIM][Tf_2N] family of ionic liquids. Diffusion and rotational correlation times were measured and the short-range parameters of CLaP were adapted to match these properties. Currently, this is the only parameterisation describing static, dynamic and energetic properties, accurately, while a certain transferability is provided. The reason might be that the fitting process requires accurate experimental data, which is often lacking, and a lot of computational and human effort.

An alternative is to apply a reduced net-charge of the ions, which obviously accelerates the system dynamics and was initially motivated by the screening of the electrostatic forces in the strongly charged IL system without considering the aspect of polarization explicitly. In this manner, many successful FFs for different ILs have been established.[127–132] This approach works quite well, but all sets of partial charges therein are either just uniformly scaled down from results of isolated ion calculations or based on calculations of ion pair (IP) clusters, which does not guarantee, that the charge distribution of the liquid state is given. Finally, the charge assigment method of Blöchl (CAB), described in detail in section 4.1, has been discovered and applied to [MMIM][Cl][11]. This method allows one to access the charge distribution in the liquid state, which naturally results in a reduced ionic net-charge. Combining all the pieces of the puzzle shows that CAB charges include the polarization of the electron cloud implicitly. Hence, CAB is a very suitable method to establish a set of partial charges for a FF. However, due to significant change of the partial charges compared to other calculations, an adaption of the short-range and some bonded parameters is required, too[12]. To this end, an approach and the corresponding computational framework is developed in here(see section 5.2 and Appendix A), which allows one to adapt the FF parameters in order to match a variety of properties.

FFs for ILs are still a heavily studied topic. Though much insight has already been gained, an accurate FF transferable between a large class of ionic liquids is required. However, this is a difficult task due to the variety of ILs as well as their properties. To this end, different strategies to establish an appropriate set of parameters are proposed in this thesis. The theory and computational framework is elaborated in chapter 4 and Appendix A, respectively. This computational framework minimizes the required amount of human input and is easily transferable to arbitrary systems. It should be applicable in all supercomputing centers that support the MD package GROMACS[73–76] and a Python interpreter[133].

3.5. Benchmarking force fields for ionic liquids

RTILs just recently became of major topic in research and industry, so available force fields have not always been validated sufficiently, to assess their reliability. The number of RTILs is very large and benchmarking is a quite time-demanding task, which is discussed in the following paragraphs. The FFs are not expected to be technically completely mature yet, compared to FFs like OPLS[53,111] or AMBER[54], which have been refined over decades. Many FFs are already available for ILs, but often they are restricted to a single cation/anion combination or only transferable between small modifications of the cation. In 2004, Canongia Lopes *et al.*[5] started publishing the FF CLaP that covers a large number of cation and anions.[5,121,122] The parametrization has been performed on isolated ions and only structural properties have been

chosen for a validation, but dynamics have neither been studied nor used for the FF setup. For this reason, the IL [EMIM][BF_4], depicted in Figure 3.9, has been studied to shed more light on the accuracy of CLaP in terms of statics and dynamics.[10] A further parametrization for [EMIM][BF_4] by Liu *et al.*[124], denoted LHW, is also available and investigated, that mainly differs from CLaP by the treatment of the covalently bonded cation hydrogens.

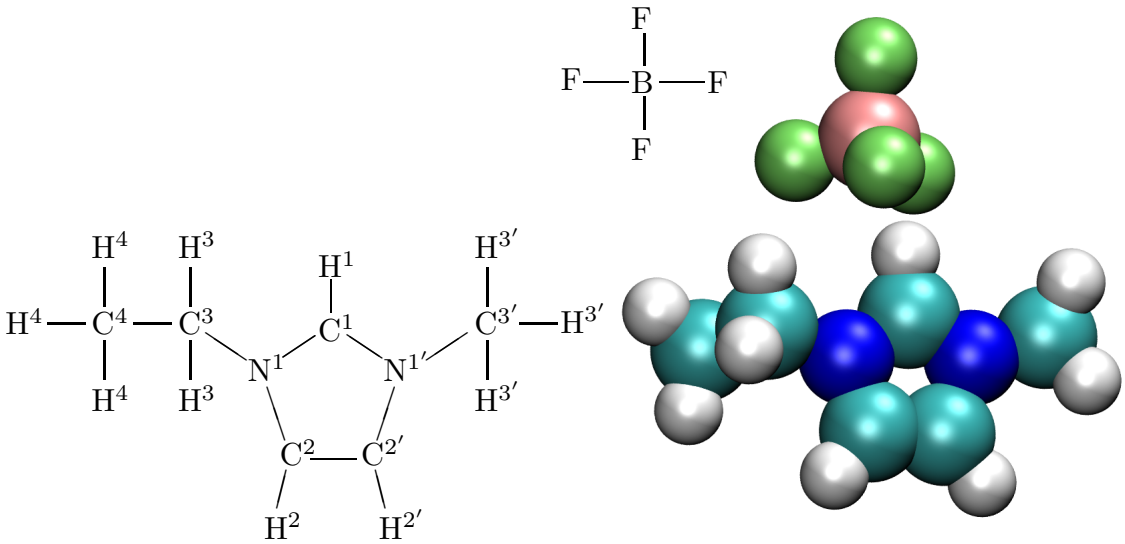


Figure 3.9.: A schematic description and picture of [EMIM][BF_4].

All MD simulations in this thesis have been performed with the package GROMACS 4,[73–76] which offers highly optimized and parallelized algorithms for an efficient treatment of an all-atom system on a supercomputer with many CPUs. A large number of suitable analysis tools is already present in GROMACS, but some properties, which are important for ILs, could not be assessed at the beginning of this thesis. GROMACS is Open-Source software, so the source code is accessible to everyone and freely modifiable. With that, the missing analysis routines have been implemented and finally became part of GROMACS distribution in the tool `g_current`. This tool allows a calculation of the current in a system, which is required to derive the static conductivity accurately, as well as the different contributions to the dipole moment, required in order to obtain the static dielectric constant. Although the GROMACS tool `g_dipoles` also allows a calculation of the dielectric constant, it is only applicable to systems, which do not contain free charges and since ILs do not belong to this kind of systems, it cannot be used.

For the visualization of molecules or simulation snapshots, the trajectory viewer VMD[134,135] is used exclusively, if not stated otherwise.

Simulation setup For a comparison of CLaP and LHW, a unit cell of 150 [EMIM][BF_4] ion pairs is simulated in an NpT ensemble at 1 bar pressure with periodic boundary conditions (PBC). The Leap-Frog integration scheme is employed with a time step of 1 fs. Nose-Hoover temperature and Parrinello-Rahman pressure coupling schemes are utilized with time constants of $\tau_T = 0.5$ ps and $\tau_P = 4$ ps, respectively. Starting configurations were chosen from

an earlier study.[9] ILs are know to be highly viscous and glassy systems, so long correlation times are expected. For this reason, simulations of 100 ns are performed at a temperature $T = 400$ K to access a sufficiently large volume of the phase space and accumulate enough statistics for collective properties. Apart from the single parameters, CLaP and LHW differ in their treatment of the covalently bonded hydrogens. In CLaP the bond length of the bond are fixed, while LHW allows fluctuations, so the parallel constraints solver P-LINCS[136] is involved in all simulations with CLaP. Usually a larger integration time step can be used, if the fast degrees of freedom are removed by constraints. Though this allows an immense acceleration of the simulations, in order to allow a fair comparison, the time step for both FFs is chosen equally. The further bonded potentials of the parameterisations are either obtained by *ab-initio* calculations or transferred from the FFs OPLS or AMBER. Besides the LJ parameters for H^1 in LHW, which has been adapted in order to match ground state structures, LJ parameters for both CLaP and LHW have also been transferred from OPLS or AMBER. Also the procedures for the partial charge assignment differ. LHW only relies on the RESP method[83] for the calculation of the partial charges, but CLaP has been constructed with the aspect of transferability in mind. For this reason, RESP is applied with the further constraints. In CLaP, the charging of hydrogens belonging to alkyl groups directly attached to the ring is symmetric and the charges of the first side chain atom is adapted to recover an integer net-charge. Partial charges for atoms separated more than three sites from the ring correspond in CLaP to the OPLS parametrization for alkanes. The electrostatic interaction is treated by the smooth particle mesh Ewald algorithm (SPME)[58] with a short range cut-off $r_c = 1.2$ nm and Fourier grid spacing of 0.12 nm at interpolation order 4, requesting a relative energy of real space and reciprocal space of 10^{-5}. Metallic boundary conditions are applied for the electrostatics, because an IL is a highly polar system and the fluctuation of the dipole moment is important and should not be correlated with the surrounding medium.[98] The LJ interactions are smoothly switched off between 0.9 and 1.0 nm and a neighbour list update is performed up to 1.2 nm every fifth step.

Structural properties ILs consist of bulky cations and anions, which are densely packed indicated by their high viscosity and mass density ρ. In Table 3.1, some properties of [EMIM][BF_4] are summarized. The simulated and experimental values for ρ agree and the dense packing is indicated by the relative large value around 1200 $\mathrm{kg \cdot m^{-3}}$, which is approximately 200 $\mathrm{kg \cdot m^{-3}}$ larger than the mass density of water at room temperature.[137]

More insight is obtained by the radial distribution functions (RDF), given in Figure 3.10, that nicely illustrate the layering structure of the positively and negatively charged ions in the liquid, which has also been observed for other ILs.[9,119,123,138]. The RDFs between the imidazolium ring and the boron atom B given by CLaP and LHW agree, but significant differences arise for the ring-ring and B-B RDFs. In CLaP, a more pronounced B-B RDF indicates stronger layering compared to LHW, which is supported by the less developed shoulder of the ring-ring RDF around 0.45 nm. Studies on other $[EMIM]^+$ based ILs[9] have shown, that this shoulder is attributed to an offset-stacking of the imidazolium rings. More detailed information about the local structuring of the liquid is gathered by the spatial distribution functions (SDF) that is the density distribution of a particle around a reference particle for

all spatial dimensions dimensions r, θ, and ϕ:

$$\gamma(r,\theta,\phi) = \frac{\mathrm{d}N(r,\theta,\phi)}{\mathrm{d}V}. \tag{3.107}$$

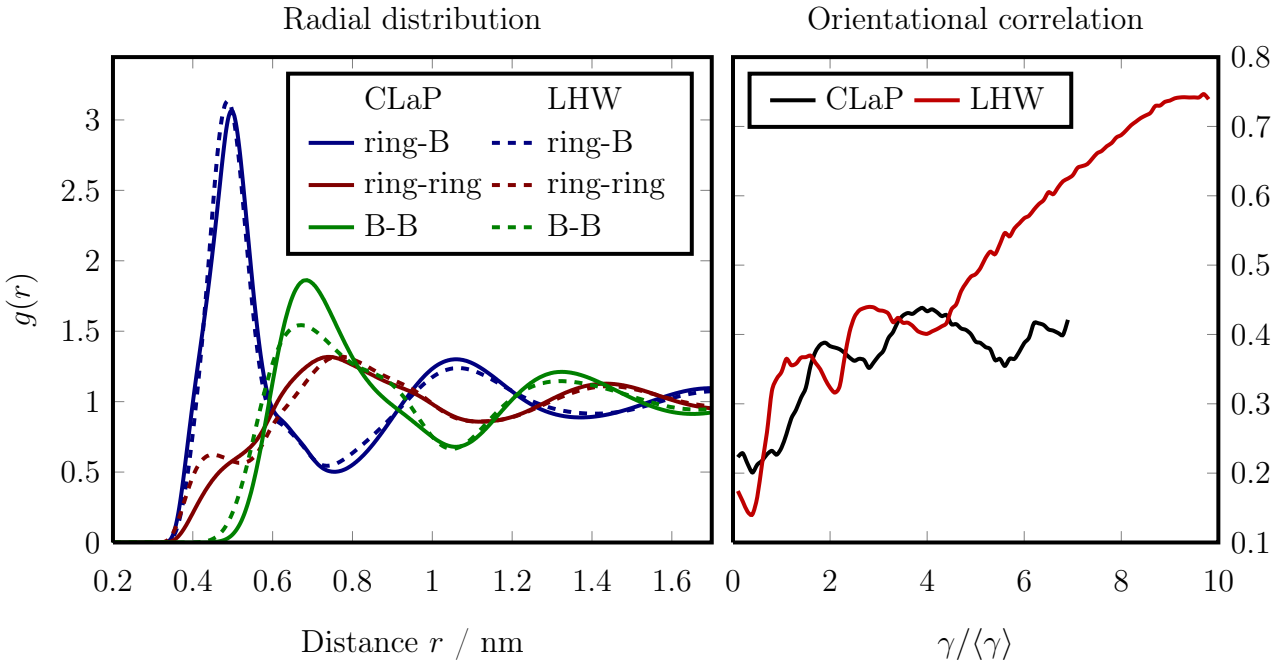


Figure 3.10.: Radial distribution and orientational correlation functions obtained by CLaP and LHW for the boron atom and the center of the imidazolium ring. The orientational correlation is determined by the angle between the normal vectors of the imidazolium rings and plotted in term of the SDF γ normalized by its average $\langle\gamma\rangle$.

In Figure 3.11, the SDF for B around the imidazolium ring as well as the SDFs of the rings and nearest B around an imidazolium ring are shown, plotted with the program GOPENMOL.[139] The ring-B SDF for CLaP and LHW significantly differ around the ring hydrogen H^1 between the methyl and ethyl group. For LHW, the SDF of B around H^1 is localized homogeneously, but the SDF, given by CLaP, is stretched towards the imidazolium ring. This indicates that LHW prefers hydrogen bonding at H^1 compared to CLaP. It has also been shown for [MMIM][Cl] by AIMD simulations[140] that CLaP underestimates the probability of hydrogen bonding at H^1. If the population of the nearest anions around the cations is considered in the bottom part of Figure 3.11 a larger spread is observed for LHW than for CLaP, which is reflected in the less pronounced layering. If the distribution of the rings around another ring is studied a first indicator for a strong offset-stacking in both FFs is given by the large population of rings above and below other rings, shown by the red and blue regions. It has to be noted that the center of this regions is not exactly above or below the respective ring, but shifted away from H^1 to the lower part of the ring. Also the domains above and below the ring are separated for LHW, which is not the case for CLaP.

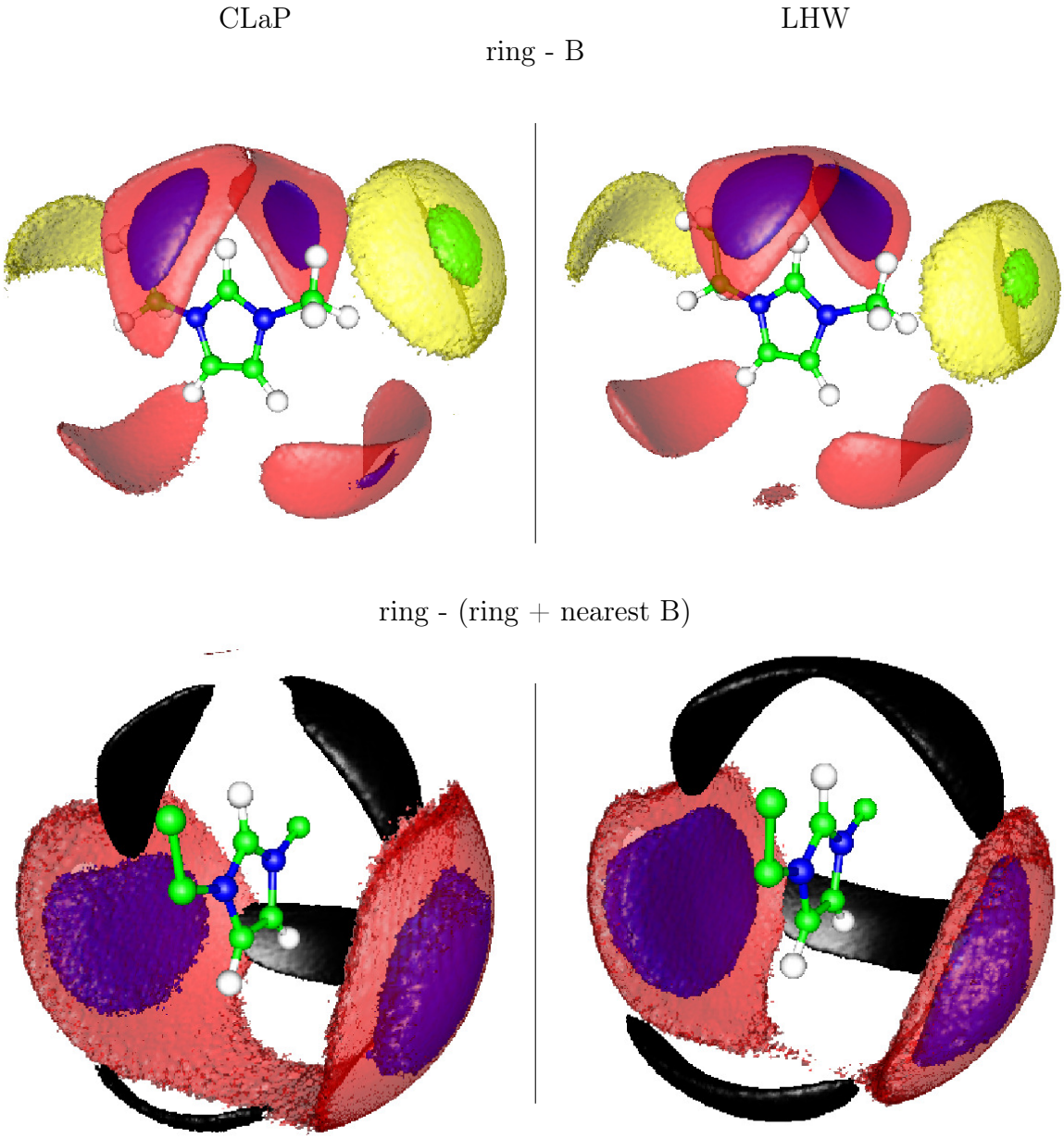


Figure 3.11.: Spatial distribution functions are plotted for the boron atom B and the imidazolium ring, both in respect to the imidazolium ring. The blue, green, black, red, and yellow isosurfaces depict the areas with ten, eight, six, five, and four times higher density than the average for the ring-B SDFs. The blue and red isosurfaces are plotted within a range from $0-0.59$ nm, while the green and yellow is given from 0.59 up to 0.73 nm and 0.75 nm for LHW and CLaP, respectively. For the SDF involving the nearest anions and cations around a ring, the color coding changed for blue and red providing the space with 5, and 2.5 times larger values for the density of the rings, respectively. For CLaP and LHW ranges of $0-0.54$ nm and $0-0.52$ nm are considered for the ring-ring SDFs, respectively. The color black depicts a six times higher average density of the anions around the ring in the range from $0-0.59$ nm for both FFs.

In order to allow one a final conclusion about the orientation, a correlation function $g_{\mathrm{SDF}}(\gamma)$:

$$g_{\mathrm{SDF}}(\gamma) = \frac{1}{N(\gamma)} \sum_{i=1}^{N(\gamma)} \frac{3\left[\cos(\alpha_i(\gamma))\right]^2 - 1}{2}, \tag{3.108}$$

is defined dependent on the SDF γ. It is derived from the $N(\gamma)$ angles $\alpha(\gamma)$ between the normal vectors of a reference imidazolium ring and the imidazolium rings around which correspond to the same γ. In combination with the SDF plots in Figure 3.11, this provides a good idea about the alignment of the imidazolium rings. If the results are compared, as shown in Figure 3.10, $g_{\mathrm{SDF}}(\gamma)$ provides are more parallel alignment of the rings in LHW starting at $\gamma = 4$, because $g_{\mathrm{SDF}}(\gamma) = 1$ corresponds to parallel orientation. Within the considered range, $g_{\mathrm{SDF}}(\gamma)$ given by LHW almost converges to 1. Yet due to the almost parallel alignment and the location of the populations, offset stacking is present in both FFs, but more pronounced by LHW.

	D / $10^{-10} \cdot \mathrm{m}^2 \cdot \mathrm{s}^{-1}$		σ / $\mathrm{S} \cdot \mathrm{m}^{-1}$					
Force field	cations	anions	σ_{NE}	σ_{GK}	σ_{EH}	ϵ / ϵ_0	μ_+ / D	ρ / $\mathrm{g} \cdot \mathrm{cm}^{-3}$
CLaP	1.39	0.65	3.6	3.5	3.5	12.6	1.79	1.183
LHW	0.66	0.42	2.2	2.1	2.2	11.3	1.74	1.218
Exp.	4.40[a]	3.94[a]	13.5[a]	8.3[a]		12.8[b]	1.56[c]	1.193[b]

[a] Obtained by complex impedance measurements[141].
[b] Experimental data at 298.15 K [18,142–146]
[c] μ_+ in [EMIM][TF_2N] [18,142–146]

Table 3.1.: Some properties of [EMIM][BF_4] at 400 K.

In summary, the detailed study elucidated some significant differences of CLaP and LHW. Although they seem to be small, if only the mass density is considered, a dissection of the structural properties shows that especially the region around the H^1 hydrogen influences the structure of the liquid. Supported by AIMD results, the more pronounced hydrogen bonding at this site is more properly modelled by LHW, resulting in a less layered liquid.

Dynamics CLaP and LHW differ in the modelling of the local liquid structure, but a more general static property, the mass density ρ, is predicted well by both of them. This is a result of the parameterisation techniques of the FFs, that rely only on static properties of the crystal or gas phase. The differences in the local liquid structure arise, because the liquid phase is not explicitly considered for the establishment of the force field, though it is of main interest. This lack of care is finally reflected in the poor modelling of the dynamics by both parameterisations.

A dynamic property, that is also accessible by experiment, is the diffusion of the ions. In Figure 3.12, the mean square displacement (MSD) $\Delta\boldsymbol{x}$ of the ions' center of mass $\boldsymbol{x}^{\mathrm{com}}$ is plotted:

$$\Delta\boldsymbol{x}(t) = \langle[\boldsymbol{x}^{\mathrm{com}}(t) - \boldsymbol{x}^{\mathrm{com}}(0)]^2\rangle, \tag{3.109}$$

which is related to the diffusion constant D for the ions at concentration $c(\boldsymbol{x}^{\mathrm{com}}, t)$ by Fick's

law and the assumption of conservation of total momentum[62]:

$$\frac{\partial c(\boldsymbol{x}^{\mathrm{com}}, t)}{\partial t} - D\boldsymbol{\nabla}^T \cdot \boldsymbol{\nabla} c(\boldsymbol{x}^{\mathrm{com}}, t) = 0. \tag{3.110}$$

Equation (3.110) is explicitly solvable and the solution allows one to calculate $\Delta\boldsymbol{x}(t)$ in terms of $c(\boldsymbol{x}^{\mathrm{com}}, t)$ known as the Einstein relation:

$$D = \frac{1}{2d} \lim_{t\to\infty} \frac{\partial}{\partial t} \Delta\boldsymbol{x}(t), \tag{3.111}$$

with the number of spatial dimensions d.

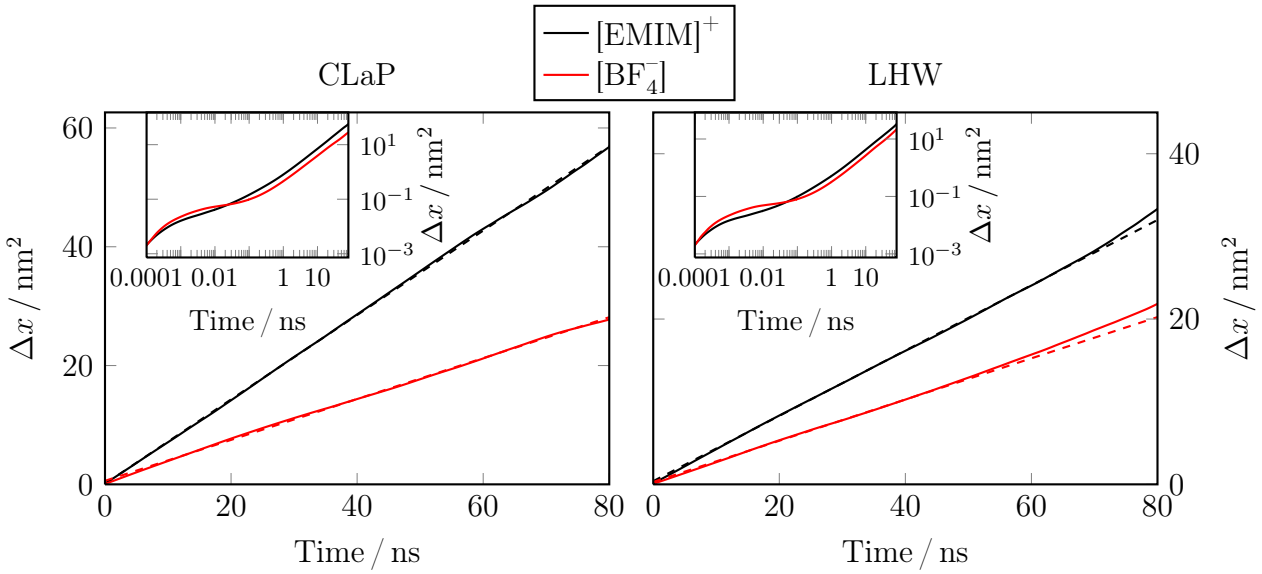


Figure 3.12.: Mean square displacement of the ions' center of mass.

In the insets of Figure 3.12, three different regimes are observed. Within the first tens of picoseconds the MSD is strongly increasing, which corresponds to the ballistic regime. The ions almost freely move, before the influence of the actual environment becomes significant. In the following sub-diffusive regime, the interactions with the other particles become dominant, such that further tens or even hundreds of picoseconds are required until the MSD converges to a linear behaviour, that reflects the average interactions and with that the diffusion of the system. If the diffusion constants of the cations and anions are derived by equation (3.111) and compared to experiments (see Table 3.1) both FFs underestimate the experimental value up to an order of magnitude. CLaP predicts a two times faster diffusion of the cations compared to LHW, but the anionic diffusion constant is only approximately 25% larger. For this reason, also collective dynamic properties like the conductivity are expected to be severly underestimated.

The diffusion is a particle-based property and averaging within a trajectory snapshot is possible, but statistics for collective properties like the dielectric constant or conductivity are only accumulated in the course of simulation time. However sometimes an single-particle properties allow an approximation. For the conductivity the diffusion coefficients D given an upper

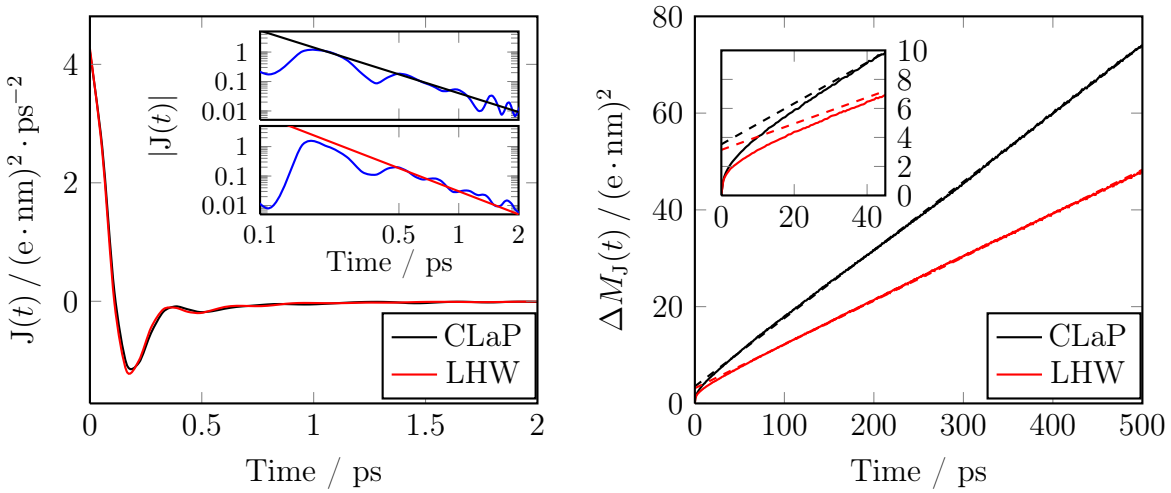


Figure 3.13.: Comparison of the current autocorrelation functions and their integrated versions, the mean square displacements of the $\boldsymbol{M}_\mathrm{J}$.

bound by the Nernst-Einstein method:

$$\sigma_\mathrm{NE} = \frac{N(q^\mathrm{net})^2}{V k_\mathrm{B} T}(D^+ + D^-), \tag{3.112}$$

involving the number of ion pairs N and the ion net charge q^net. This approach completely neglects the cross-correlation of the ions that reduces conductivity.

For an exact calculation of the conductivity including also the ion correlations a time span of 100 ns simulation time has been chosen to apply the seemingly straightforward Green-Kubo method. Therefore the current autocorrelation function $J(t) = \langle \boldsymbol{j}(t)\boldsymbol{j}(0)\rangle$[62] is integrated:

$$\sigma_\mathrm{GK} = \frac{1}{3V k_\mathrm{B} T}\int_0^\infty \langle \boldsymbol{j}(t)\boldsymbol{j}(0)\rangle \mathrm{d}t, \tag{3.113}$$

where the current $\boldsymbol{j}(t) = \sum_{i=1}^{N} q_i^\mathrm{net}\boldsymbol{v}_i^\mathrm{com}(t)$ is determined by the ionic charge q_i^net and the center of mass velocities $\boldsymbol{v}_i^\mathrm{com}$ of the ions i.

As shown in Figure 3.13 on the left hand side, $J(t)$ converges slowly and with the present statistics a direct integration is impossible. For this reason, the long time tail of $J(t)$ is fitted to a function $\sim -t^{-a}$ starting at $t_c = 1$ ps, that is also applied for an analytical integration of $J(t)$ starting at t_c. In the insets of Figure 3.13 on the left hand side, the agreement of the decay, given by the simulations and the fit, is clearly shown. The functions $J(t)$ converge in the long time limit $\sim t^{-2.12}$ and $\sim t^{-2.60}$ for CLaP and LHW, respectively. This limiting behaviour is an requirement for the existence of the dielectric susceptibility[citesega12a] which has to be at least $\sim t^{-2}$.

A further possibility to derive the conductivity is the Einstein-Helfand (EH) approach, that

relies on the translational dipole moment $\boldsymbol{M}_\mathrm{J} = \sum_{i=1}^{N} q_i^\mathrm{net} \boldsymbol{x}_i^\mathrm{com}$ of the ions:[147]

$$\lim_{t\to\infty} \Delta \boldsymbol{M}_\mathrm{J} = \lim_{t\to\infty} \langle [\boldsymbol{M}_\mathrm{J}(t) - \boldsymbol{M}_\mathrm{J}(0)]^2 \rangle = \lim_{t\to\infty} 6Vk_\mathrm{B}T\sigma_\mathrm{EH}t + 2\langle \boldsymbol{M}_\mathrm{J}^2 \rangle. \tag{3.114}$$

Similar to the relation of the velocity autocorrelation-function and the MSD of the ions, $\Delta \boldsymbol{M}_\mathrm{J}$ is the double integrated version of the current autocorrelation function.

In Table 3.1, the results for CLaP and LHW given by equation (3.112) are compared to the Green-Kubo approach and the Einstein-Helfand method. Both FFs predict an almost negligible correlation, because σ_NE is only slightly larger than the correlation-including values of σ_GK and σ_NE. With the Green-Kubo method CLaP and LHW predict conductivities of $3.5\ \mathrm{S}\cdot\mathrm{m}^{-1}$ and $2.1\ \mathrm{S}\cdot\mathrm{m}^{-1}$, respectively, which are too small compared to the experimental value of $8.3\ \mathrm{S}\cdot\mathrm{m}^{-1}$. The Einstein-Helfand method almost gives conductivities of $3.5\ \mathrm{S}\cdot\mathrm{m}^{-1}$ and $2.2\ \mathrm{S}\cdot\mathrm{m}^{-1}$ for CLaP and LHW, respectively. Yet almost the same conductivities are calculated with the equations (3.113) and (3.114), proving the consistency of the methods and the inaccurate modelling of the dynamics by the studied force fields. Though EH only relies on the particle position and does not require a frequent saving of the velocities, the value for σ is very sensitive to the chosen fit interval[citesega12a]. This sensitivity even increases for $\langle \boldsymbol{M}_\mathrm{J}^2 \rangle$, because a small error in the slope evolves to a large error in the intercept of the fit with axis. Unfortunately $\langle \boldsymbol{M}_J^2 \rangle$ is involved in the calculation of the static dielectric constant and is not directly accessible from MD simulations.

The static dielectric constant Dielectric properties of ionic liquids are of great interest, because the frequency spectrum allows to identify different kinds of interactions and motions. The static dielectric constant ϵ is a measure for the response of the medium to an external electric field. For uncharged species only the dipole of the particle is of interest, but charged ions also contribute to ϵ by their translational motion. Moreover if free charges are present the total dipole moment is ill-defined and dependent on the orientation of the unit cell. [citesega12a] dissected different methods that circumvent this problems. One of them, which has already been used in the field of ionic liquids[148] is based on a decomposition of the total dipole moment into a rotational part $\boldsymbol{M}_D = \sum_{i=1}^{N} \sum_{j=1}^{n_i} q_{j,i}(\boldsymbol{x}_{j,i} - \boldsymbol{x}_i^\mathrm{com})$, which is given by the positions $\boldsymbol{x}_{j,i}$ and partial charges $q_{j,i}$ of the n_j atoms j belonging to the ion i, and the translational part $\boldsymbol{M}_\mathrm{J}$. With this decomposition ϵ is calculated by:[10]

$$\epsilon = 1 + \frac{1}{3Vk_\mathrm{B}T\epsilon_0} \left(\langle \boldsymbol{M}_\mathrm{D}^2 \rangle - 2\langle \boldsymbol{M}_\mathrm{D}\boldsymbol{M}_\mathrm{J} \rangle + \langle \boldsymbol{M}_\mathrm{J}^2 \rangle \right). \tag{3.115}$$

The first term converges quite fast, but the cross-term involving $\boldsymbol{M}_\mathrm{D}$ and $\boldsymbol{M}_\mathrm{J}$ as well as the last term are not calculated straightforwardly. $\langle \boldsymbol{M}_J^2 \rangle$ is obtained from the Einstein-Helfand fit by equation (3.114). The cross-term is derived by an integration of the current:

$$\langle \boldsymbol{M}_\mathrm{J}\boldsymbol{M}_\mathrm{D} \rangle = \langle \int \boldsymbol{M}_\mathrm{D}(0) \cdot \boldsymbol{j}(t) \mathrm{d}t \rangle. \tag{3.116}$$

If the static dielectric constant is calculated from the simulations with CLaP and LHW the results are $\epsilon = 12.6$ and $\epsilon = 11.3$, respectively, which coincide well with the experimental value of 12.8. Also the dynamical property $\boldsymbol{j}(t)$, which is underestimated by CLaP and LHW,

is involved in equation (3.116), therefore the cross-term of equation (3.115) might also be underestimated. Additionally stronger correlations may occur within the ionic liquid than obtained from the simulations. This is also supported by recent simulations with polarizable force fields.[149,150] Therein the conductivity derived by EH is about 70% lower, than obtained with the correlation neglecting Nernst-Einstein (NE) approach. Yet only a reliable force field allows to shed more light on the influence on the cross-term.

Although both force fields describe the global structure and static properties, such as the mass density and dielectric constant similar, subtle differences arise for the local structure. However, both force fields fail in a proper description of dynamical properties. In order to obtain such a parametrization methods are developed in the following chapters that aim on an optimization and extension of partial charges and short range parameters of a given FF.

4. Force field partial charges from and for the liquid phase of ionic liquids

One of the main objectives of this work is the parameterisation and refinement of force fields (FFs) for ionic liquids (ILs). In this regard, one of the basic question is the calculation of the partial charges. Usually the set of charges for a force field are derived from QM calculations on isolated molecules with a Restrained Electrostatic Potential fit (RESP)[83] or CHarges from Electrostatic Potentials using a Grid based method (CHELPG)[84] are generated. However, the electronic properties of ILs change drastically when switching from an isolated molecule to small clusters[1,7], but these effects are also very localized as a result of the strong Coulomb screening in this kind of system[14,15,151]. In order to include liquid phase properties in the partial charges, an appropriate method is required. In the following sections, an approach suggested by Blöchl[55] is introduced and applied to ILs. Furthermore the MDEC theory of Leontyev and Stuchebrukhov[152] is introduced, which provides a consistent and implicit description of polarization with static partial charges. The final aim is the construction of a consistent set of partial charges. Thus different aspects are considered, such as the effect of the cation–anion combination, dipole moments or the influence of small changes in the electrostatic model. Finally, this allows us to suggest a reliable and straightforward method to derive a parameterisation of partial charges, which is transferable within different ILs and contains the characteristics of an IL under bulk conditions.

4.1. Difficulties in partial charge calculations

Mostly, partial charges of established force fields for liquids, ionic liquids or biomolecular molecules, such as OPLS-AA, CHARMM, CLaP or AMBER, were derived from QM calculations *in vacuo* or with implicit solvent models. The charge is often assigned with the methods CHELPG[84] or RESP.[83] These are based on a fit of the partial charges q_i to match the electrostatic potential (ESP), which is derived from a charge density $n(\boldsymbol{r})$ of single molecules or ion pairs, either *in–vacuo* or embedded in an implicit solvent. However, this approach has disadvantages, if the charge distribution for a liquid state is of interest. One aspect is the sensitivity of $n(\boldsymbol{r})$ to the configuration of the atoms. Hence, the partial charges should be derived by averaging the results for different configurations. However, often only the ground states of a molecule are considered. For this reason, the actual information about the liquid state, such as the correlation of the electrons from different molecules or the influence of temperature, is missing. While implicit solvent models help to include at least an average influence of the polarizing medium, the impact of the dense packing in the liquid state and the corresponding larger phase space is completely neglected. Though *ab–initio* molecular dynamics (AIMD) simulations allow one to model the bulk state of a system of tens of ion pairs in the ps regime, CHELPG and RESP are not applicable, because the choice of a reasonable manifold, which allows an accurate fit of the ESP, is very hard, due to the density of the atoms and periodicity of the charge density.

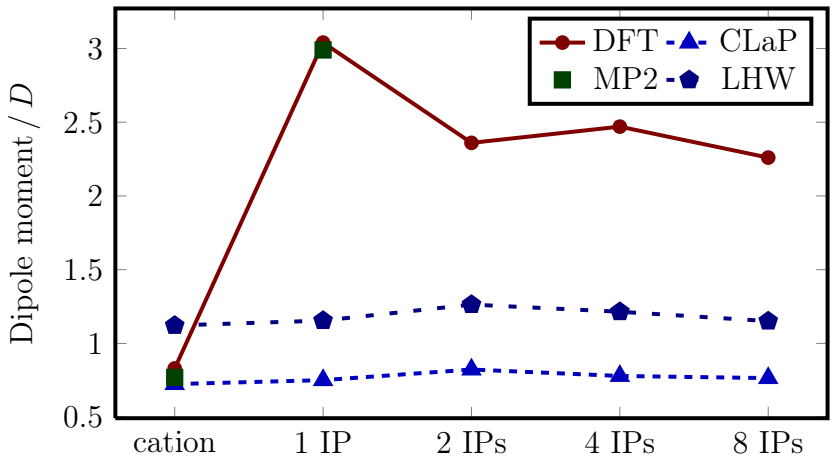


Figure 4.1.: The average cation dipole with respect to the center of mass[1] derived with quantum mechanical methods MP2 and DFT, is compared to calculations with the classical force fields CLaP and LHW.

However, if the electrostatic properties of ILs of the gas and bulk phase are compared significant differences arise,[1,2,11,12,15] as shown in Figure 4.1. While the electric dipole moment μ of the cation with respect to the center of mass is around 0.8 D for a single ion, μ increases to about 2.4 D in larger clusters of up to eight ion pairs, if *ab-initio* methods like MP2 or DFT are applied. This effect is a result of electron correlation, and therefore only observable if more than a single ion is present. The electron clouds of the ions extend, resulting in an increase of the dipole moment. A maximum of the cation dipole moment arises for single ion pairs (IPs), but with a growing number of ions, the space around a molecule is filled more and more homogeneously, such that a balance of the interactions between all molecules establishes and

a convergence of the dipole moment to $\approx 2.4\,\mathrm{D}$ is observed.[7]

If the dipole moments for the different systems are calculated via partial charges given by the FFs, CLaP[5] and LHW[124], μ remains almost constant with increasing system size at value of about 0.8 D and 1.2 D, respectively. Thus, they are not even close to the bulk value. This clearly shows that the charges of CLaP and LHW, are unable to model the electrostatic character of a bulk phase, because the FFs have been derived by single ion calculations. Static partial charges are not able to describe the increase in the dipole moment, because they do not model polarization effects. For this reason, a method is needed that allows the inclusion of bulk state characteristics of $n(\boldsymbol{r})$ in the partial charges.

Such a technique has been suggeted by Blöchl[55], that allows the derivation of the partial charges for periodic systems[55], which is perfectly suitable for a periodic *ab-initio* MD (AIMD) simulation. Applying this method to IL systems shows a significant reduction of the ion net-charge from ± 1 e to about ± 0.6 e, which has also been observed in QM gas phase calculations[11,119,153] and, more recently, in NMR measurements of ILs.[154] Furthermore, a change in the distribution of the charge within the ions is obtained, if results from gas and liquid state calculations are compared. Though different IL force fields with a reduced net-charge[127,128,130–132,153] have been suggested, the charge distribution has always been based on single ion or ion pair calculations. Only Youngs and Hardacre[129] include bulk effects, since the charges are derived by a force matching technique, which is applied to an AIMD trajectory under bulk conditions. However, in contrast to these methods, the partial charges derived with the Blöchl charge assignment method[55] (CAB) are directly connected to the charge density of the bulk phase. Yet, these charges are expected to be an improved electrostatic model for the characteristics of a liquid phase, compared to a model which relies on gas phase calculations.

Hence, the Blöchl method is a promising technique for the calculation of partial charges for IL force fields. In order to be able to apply it correctly, a precise description of the method itself is given in the following section.

4.2. CAB - The Charge Assignment method of Blöchl

While RESP and CHELPG rely on the electrostatic potential in real space, the charge assignment method of Blöchl[55] (CAB) acts on the charge density $n(\boldsymbol{k})$ in reciprocal space, which renders it perfectly suitable for systems with periodic boundary conditions. To derive partial charges q_i at atomic sites $\boldsymbol{R}_i$ a model charge density $\tilde{n}(\boldsymbol{r})$:

$$\tilde{n}(\boldsymbol{r}) = \sum q_i g_i(\boldsymbol{r}), \tag{4.1}$$

consisting of Gaussians g_i:

$$g_i(\boldsymbol{r}) = \frac{1}{\left(\sqrt{\pi} r_{\mathrm{c},i}\right)^3} \cdot \exp\left(-\frac{(\boldsymbol{r}-\boldsymbol{R}_i)^2}{r_{\mathrm{c},i}^2}\right), \tag{4.2}$$

with a decay length of $r_{c,i}$, is fitted to the expansion of the charge density in multipole moments Q_l:

$$Q_l = \int_V |\boldsymbol{r}|^l Y_l(\boldsymbol{r}) n(\boldsymbol{r}) d^3r, \tag{4.3}$$

in terms of a sum Y_l of spherical harmonics $Y_{lm}(\varphi, \theta)$:

$$Y_l = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_{lm}^*. \tag{4.4}$$

The guiding principle of this method is based on the property of the multipole expansion in reciprocal space $Q_l(\boldsymbol{k})$, that it is completely determined by the charge density and all its derivatives at $\boldsymbol{k} = 0$. It has been shown[11,55] that a few vectors in $\boldsymbol{k}$–space corresponding to an energy cutoff $K_{\text{cut}}^2 = 14\,\text{Ry}$ are enough to make the results converge. Due to to the translational invariance of the periodic lattice, the problem of a suitable choice for the origin of the multipole expansion does not need to be solved. Blöchl[55] showed that for a suitable construction of $\tilde{n}(\boldsymbol{r})$, a point charge model $\hat{n}(\boldsymbol{r}) = \sum_i q_i \delta(\boldsymbol{r} - \boldsymbol{R}_i)$ with partial charges equal to the partial charges of the Gaussians in $\tilde{n}(\boldsymbol{r})$ also matches the multipole moments given by $n(\boldsymbol{r})$, if the two following conditions are met. The difference of the multipole moments ΔQ_l given by the true and model charge densities is minimal:

$$\Delta Q_l = \left| \int_V |\boldsymbol{r}|^l Y_l \left(n(\boldsymbol{r}) - \tilde{n}(\boldsymbol{r}) \right) d^3r \right|, \tag{4.5}$$

as well as the width of the charge distribution ΔW is minimal:

$$\Delta W = \left| \int_V r^2 \left(n(\boldsymbol{r}) - \tilde{n}(\boldsymbol{r}) \right) d^3r \right|. \tag{4.6}$$

An extremal condition $F(q_i, \lambda)$ is formulated in terms of Langrangian multipliers

$$\begin{aligned} F(q_i, \lambda) = & \frac{V}{2} \sum_{\boldsymbol{k} \neq 0} w(\boldsymbol{k}) \left| n(\boldsymbol{k}) - \sum_i q_i g_i(\boldsymbol{k}) \right|^2 - \\ & - \lambda \left[n(\boldsymbol{k} = \boldsymbol{0}) V - \sum_i q_i g_i(\boldsymbol{k} = 0) V \right], \end{aligned} \tag{4.7}$$

with a weighting function $w(\boldsymbol{k})$:

$$w(\boldsymbol{k}) = \begin{cases} 4\pi \frac{\left(\boldsymbol{k}^2 - K_c^2\right)^2}{\boldsymbol{k}^2 K_c^2} & , \text{ if } |\boldsymbol{k}| < K_c^2, \\ 0 & , \text{ elsewhere,} \end{cases} \tag{4.8}$$

which is defined, in order to increase the importance of space around $\boldsymbol{k} = 0$. Since the weighting function is not defined for $\boldsymbol{k} = 0$, this case is treated with a Lagrangian multiplier

λ and the constraint for charge conservation separately. If $F(q_i, \lambda)$ is reformulated with:

$$b_i = V \sum_{\boldsymbol{k}\neq 0} w(\boldsymbol{k}) \Re\left[n^*(\boldsymbol{k}) g_i(\boldsymbol{k})\right], \tag{4.9}$$

$$A_{ij} = V \sum_{\boldsymbol{k}\neq 0} w(\boldsymbol{k}) g_i^*(\boldsymbol{k}) g_j(\boldsymbol{k}), \tag{4.10}$$

$$c_i = V g_i(\boldsymbol{k}=0) = 1, \tag{4.11}$$

$$N = V n(\boldsymbol{k}=0), \tag{4.12}$$

where $\Re(\,\cdot\,)$ is the real space part of a complex number, a system of linear equations arises:

$$\sum_j A_{ij} q_j + \lambda c_i = b_i, \quad \sum_i c_i q_i = N, \tag{4.13}$$

with the solution:

$$q_i = \sum_j \left(A^{-1}\right)_{ij} \left[b_j - c_j \frac{\sum_{k,l} c_k \left(A^{-1}\right)_{kl} b_l - N}{\sum_{m,n} c_m \left(A^{-1}\right)_{mn} c_n} \right]. \tag{4.14}$$

To match the width of the charge density a model density consisting of 3 to 4 Gaussians with increasing decay lengths $r_{c_\alpha,i}$ for every atomic site i is proposed[55]. It has been shown by Blöchl[55] that a minimal $r_{c_0,i}$ of $0.5 - 1$ Bohr radii a_0 and $r_{c_\alpha,i} = 1.5 \cdot r_{c_{(\alpha-1)},i}$ yields results accurate up to a few meV.

Partial charges and dipoles for single [MMIM][Cl] ion pairs derived by GPW and MP2 methods In order to benchmark the Blöchl method, hybrid Gaussian plane wave (GPW) calculations[155] with the QUICKSTEP[50] module implemented in the program package `CP2K`[52] have been performed on four ground state structures (see Figure 4.2). These structures were obtained by high level second order Møller-Plesset perturbation theory (MP2) and coupled cluster calculations (CC) and were kindly provided by Y.Y. Zhao.[2]

	charge q / e						
	DZVP		TZV2P		aug-QZV3P	MP2[2,11]	
geometry	BLYP-G	PBE	BLYP-G	PBE	PBE	CHELP	RESP
1	-0.69	-0.69	-0.65	-0.65	-0.63	-0.74	-0.74
2	-0.64	-0.64	-0.62	-0.62	-0.61	-0.71	-0.72
3	-0.79	-0.79	-0.77	-0.77	-0.76	-0.85	-0.85
4	-0.73	-0.73	-0.70	-0.69	-0.68	-0.78	-0.79

Table 4.1.: Charges of the chloride for the optimized structures, shown in Figure 4.2, derived by CAB and reference calculations on the MP2 level applying the established procedures CHELPG and RESP.

Partial charges for the chloride have been derived by the Blöchl method for different basis sets and exchange-correlation functionals and are compared in Table 4.1. For the GPW method, an energy cutoff $E_c = 280$ Ry and a cubic box size of $V = (21.618\ \text{Å})^3$ is applied. Further results,

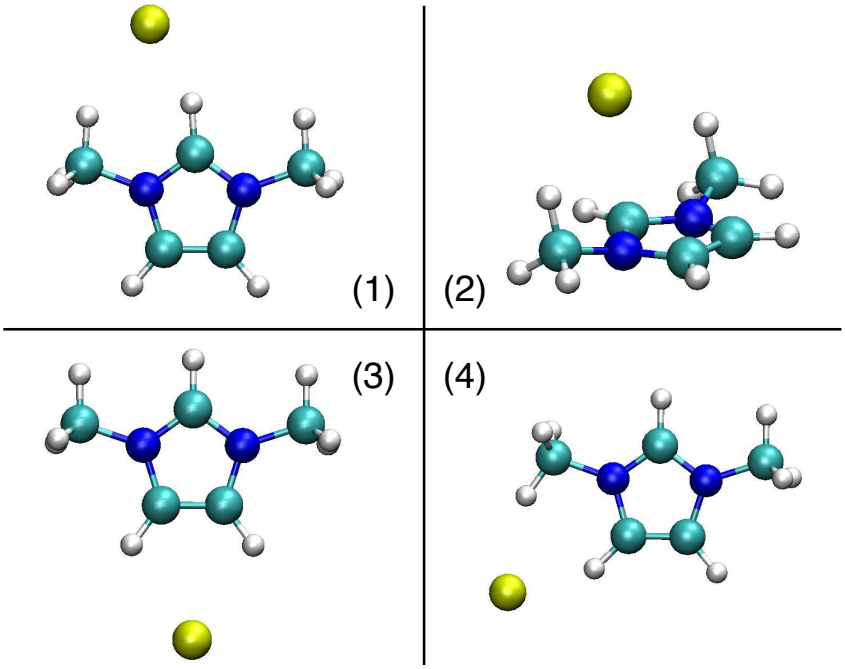


Figure 4.2.: Equilibrium structures derived by MP2 and coupled-cluster calculations[2] where structure 2 is the ground state.

which have been obtained by CHELPG or RESP by Schmidt *et al.*[11] with charge densities derived on the MP2 level of theory, are also compared. While a negligible dependence on the exchange-correlation functional is observed, an increase of the basis set size results in a less negative charge. Compared to the MP2 reference calculations, CAB tends to overestimate the charge reduction of the net-charge given by the RESP and CHELP procedure by about 10 %. This difference can be attributed to the insufficient description of the electron correlation by the exchange-correlation functionals, which might favour polarization. Since the results for the largest two basis sets do not differ significantly, the TZV2P basis set is expected to provide a good balance between accuracy and efficiency. However, the Blöchl method explicitly adapts the partial charges to the width of the charge distribution. This allows it to maintain a more precise description of the density than RESP or CHELPG, which rely on the electrostatic potential on a certain surface, significantly restricting the information about the density.

A comparison of the dipole moments of the four structures obtained from the Blöchl charges to MP2 calculations or the Maximally localized Wannier scheme (MLW) is given in Table 4.2. Consistently smaller dipole moments are obtained by CAB, apart from geometry 2. However this result is expected, because CAB charges provide an implicit description of polarization, which is discussed in detail in section 4.2. For the combination of TZV2P and PBE, the corrected results are shown, which fluctuate around the reference data. The large deviation for geometry 2 arises due to the location of the chloride, resulting in a structure with very small spatial extent. Yet, the sensitivity of the partial charges to the actual structure is highlighted.

Table 4.3 shows the consistency of the different schemes. A Blöchl analysis has been performed on single cations with an aug-QZV3P basis set and the PBE exchange-correlation functional

for the geometries 1–4. The dipoles have been calculated and averaged over the 4 structures. The results agree with the reference DFT calculations using the MLW scheme[1], but compared to the MP2 reference data[1], the Blöchl method predicts a 0.06 D larger dipole moment with respect to the center of mass (see Table 4.3), as well as the other DFT approach.

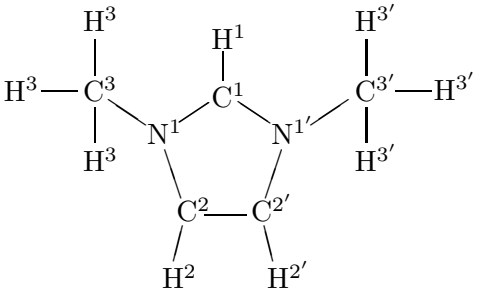


Figure 4.3.: Schematic drawing of the $[MMIM]^+$ cation.

In Table 4.4 the CAB charges averaged over the four geometries of isolated cations (for the nomenclature see Figure 4.3) are compared to the partial charges from the CLaP FF that were derived by CHELPG at the MP2 level of theory also using isolated ions. The largest differences are observed for the C^3, H^3, and N^1 atom types enhancing the dipolar character of a cation. Negative charge is transferred from the imidazolium ring to the carbon atoms of the methyl groups, if CAB is compared to CLaP. The charge of the C^2H^2 groups obtain with CAB is decreased compared to the CLaP FF and for C^1H^1 an increase of the charge is observed. This results in a quite polar region of the molecule around N^1, C^1, H^1, and $N^{1'}$, which is not that pronounced for the CLaP parameterisation. This finally proves that different charge assignment schemes can yield quite different partial charges as the mapping of the charge density is not unique. Especially if more than one molecule is involved, the results from CAB, CHELPG, RESP, and MLW differ significantly. The monopoles obtained with CAB charges are expected to allow a reliable classical description of the electrostatic interactions, because they also reproduce the dipole of a single cation. Thus the next step is the application of CAB to "*bulk-like*" systems, which allows the incorporation of the electrostatic character of a liquid

Totale dipole moment / D

	DZVP		TZV2P		aug-QZV3P	DFT (PBE)	
geometry	BLYP-G	PBE	BLYP-G	PBE	PBE	MP2[1]	MLW[1]
1	9.77	9.75	9.22	9.12 (14.03)	8.92	12.65	12.4
2	10.00	10.02	9.70	9.65 (15.57)	9.57	8.63	8.3
3	12.76	12.78	12.45	12.39 (16.09)	12.28	16.59	17.4
4	14.35	14.32	13.65	13.50 (19.57)	13.28	16.25	16.5

Table 4.2.: Total dipole moments of the ground state structures obtained with CAB from calculations with different basis sets and exchange-correlation functionals. The values in brackets include the effect of electronic polarization, disscussed in section 4.2. The results are compared to reference calculations on the MP2 level and a DFT approach using the Maximally localized Wannier (MLW) scheme[1].

	geometry				avg.	DFT[1]	MP2[1]
	1	2	3	4	$\langle\mu_i\rangle$	μ	μ
μ_{com} / D	0.79	1.17	0.49	0.88	0.83	0.83	0.77
μ_{cor} / D	2.13	2.68	1.72	2.20	2.18	2.19	n.a.

Table 4.3.: Dipole moments of a single cation in vacuum with respect to the center of mass μ_{com} and geometric center of the ring μ_{cor} are compared to DFT and MP2 reference data[1].

into the partial charges.

	partial charges q / e						
	H^1	H^2	H^3	C^1	C^2	C^3	N^1
CLaP[5]	0.21	0.21	0.13	-0.11	-0.13	-0.17	0.15
CAB	0.20	0.25	0.22	-0.09	-0.20	-0.53	0.265

Table 4.4.: Charges of [MMIM][Cl] derived by CAB are compared to the partial charges of the CLaP FF. A nomenclature of the atoms is given in Figure 4.3.

An introduction to the MDEC theory Although the reduced net-charge of the ions seems odd at first, a consistent mean field theory exists,[152,156] that can explain charge reduction for partial charges of classical force fields in terms of polarization.

In the Molecular Dynamics Electronic Continuum (MDEC) model of Leontyev and Stuchebrukhov[152], the charges of the force field are considered as effective $q^{\mathrm{eff}} = q/\sqrt{\epsilon_{\mathrm{el}}}$, screened by the electronic dielectric constant ϵ_{el} of the medium. The polarization $\boldsymbol{P}$ of the system is decomposed into a slow "inertial" part $\boldsymbol{P}_{\mathrm{in}}$ and a fast "noninterial" part $\boldsymbol{P}_{\mathrm{el}}$. While $\boldsymbol{P}_{\mathrm{in}}$ depends on the positions of the nuclei, $\boldsymbol{P}_{\mathrm{el}}$ is the contribution from the electrons, that react on a conformational change of the nuclei, nearly adiabatically. This approximation holds in the linear response regime, where the susceptibility χ is an additive quantity:

$$\chi = \chi_{\mathrm{nuc}} + \chi_{\mathrm{el}}, \tag{4.15}$$

which connects $\boldsymbol{P}$ to the electric field E:

$$\boldsymbol{P} = \chi \boldsymbol{E}, \tag{4.16}$$

and hence, $\boldsymbol{P}$ decomposes into a nuclear $\boldsymbol{P}_{\mathrm{in}}$ and an electronic part $\boldsymbol{P}_{\mathrm{el}}$:

$$\boldsymbol{P} = \chi_{\mathrm{nuc}}\boldsymbol{E} + \chi_{\mathrm{el}}\boldsymbol{E} \tag{4.17}$$

$$= \boldsymbol{P}_{\mathrm{in}} + \boldsymbol{P}_{\mathrm{el}}. \tag{4.18}$$

Thus the molecules are assumed to be immersed in a medium of dielectric constant ϵ_{el}, which in this case consists of the strongly delocalized electron cloud.

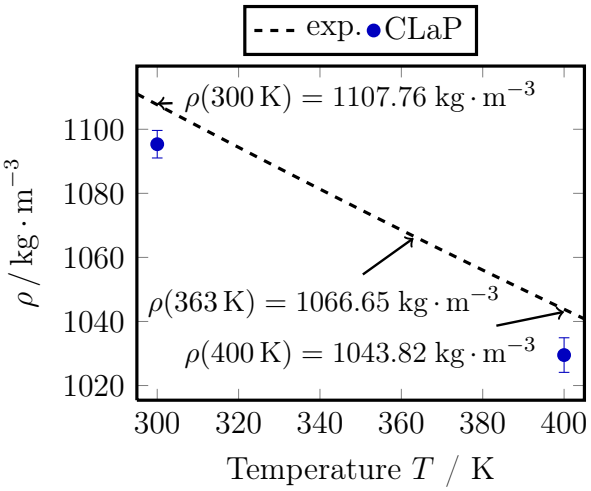
Leontyev and Stuchebrukhov[156] applied their approach to water and obtained very good agreement of experiment and theory. Moreover, it turns out that the parameterisation of the SPC/E

water model[137], partly includes the MDEC theory, which results in its good performance. While standard water models often fail to describe the correct dipole in the liquid phase, the proposed MDEC model is superior or equal to the standard models for all compared properties and has an electronic constant of $\epsilon_{\mathrm{el,water}} = 1.78$.

This implicit treatment of polarizability by an electronic dielectric constant ϵ_{el} results in an "effective" charge reduction for charged species. From the physical point of view, ϵ_{el} is a measure for the screening of the electrostatic interactions, arising from the almost inertialess reaction of the electron clouds to a force. In order to obtain unscreened charges, the results from the Blöchl analysis have to be scaled by a factor of $\sqrt{\epsilon_{\mathrm{el}}} = 1/\sqrt{q_{\mathrm{net}} \cdot q_{\mathrm{net}}}$.

4.3. Determining partial charges from *bulk-like* configurations

Generating a *bulk-like* configuration The first step for the calculation of partial charges for the liquid phase is the generation of corresponding quasi–independent configurations, which are then subjected to a Blöchl analysis. To this end, an *ab–initio* MD (AIMD) simulation is performed and 100 snapshots are extracted. In order to minimize the time for the equilibration of the computationally demanding AIMD, preequilibrated boxes have been prepared by classical MD simulations. Lastly, the global structure of the small system, given by the ion-ion radial distribution functions, has been compared to results of a larger system, in order to check if finite size effects are observable. The procedure is described for the IL [EMIM][DCA] below.



T / K	box length / nm	ρ^{exp} / kg · m^{-3}
300	1.997	1107.76
363	2.023	1066.65
400	2.037	1043.82

Figure 4.4.: On the left hand side the experimental density of [EMIM][DCA] measured by Fröba *et al.*[3] is plotted. Box samples containing 30 ion pairs were prepared for the AIMD simulations at temperatures indicated by the arrows, while the corresponding box sizes are summarized in the table on the right hand side.

Since all MD simulations were performed with GROMACS, appropriate input files are required. An easy starting point to obtain these files for complex molecules, such as molecular ionic liquids, is the PRODRUG server[157], which allows the generation of a PDB file containing

the structural information of a cation and an anion. These PDB files have been processed by the GROMACS tool `pdb2gmx`, which generates corresponding topology files, including the parameterisation of the system. Hence, all molecule specific information which is required for an MD simulation is present. In order to obtain a starting configuration for the AIMD, the cation and anion structures given by the PRODRUG server were energy minimized using the steepest descent method of GROMACS. The resulting structures were then placed in a cubic box of size $(2.5\,\mathrm{nm})^3$, randomly, such that a system of 30 ion pairs is obtained. This system has been subjected to a further energy minimization under periodic boundary conditions. To prepare the system for the final equilibration step, the system is heated up to the desired temperature. In the first 300 ps, the temperature T was increased from 100 K to 300 K, linearly. From then on, the slope of T was reduced to 50 K per 100 ps until a temperature of 400 K had been achieved. The error of the electrostatic forces, which were calculated using the SPME algorithm, had been estimated *a–priori* with the parameters tuned to achieve an accuracy of $\approx 10^{-4}\,\mathrm{kJ\cdot mol^{-1}\cdot nm^{-1}}$. For a box size $V = (2.5\,\mathrm{nm})^3$, this resulted in an interpolation order $n = 6$, $K = 128$ grid points in every dimension, and a splitting parameter β of $5.75\,\mathrm{nm^{-1}}$. The cutoffs of SPME in real space and of the LJ forces as well as the neighbor search cutoffs were set to 0.7 nm, while updating the neighbor list every five integration steps of 1 fs. For the annealing, the velocity-rescale thermostat by Bussi *et al.*[70] has been applied with a time constant $\tau = 0.1$ ps. Finally, the annealed system was equilibrated at $T = 300$, 363, and 400 K at a pressure $p = 1$ bar for 500 ps. The pressure is coupled with the barostat of Berendsen *et al.*[69]. A system state close to the experimental density is obtained. This allows us to scale the box vectors in order to create systems for the AIMD simulations, which are performed in an NVT ensemble at experimental density (see Figure 4.4). The resulting boxes have been energy minimized such that the maximal force was $10\,\mathrm{kJ\cdot mol^{-1}\cdot nm^{-1}}$ and afterwards equilibrated for 10 ns at the corresponding temperatures. This was how equilibrated starting structures for the AIMD simulations have been established.

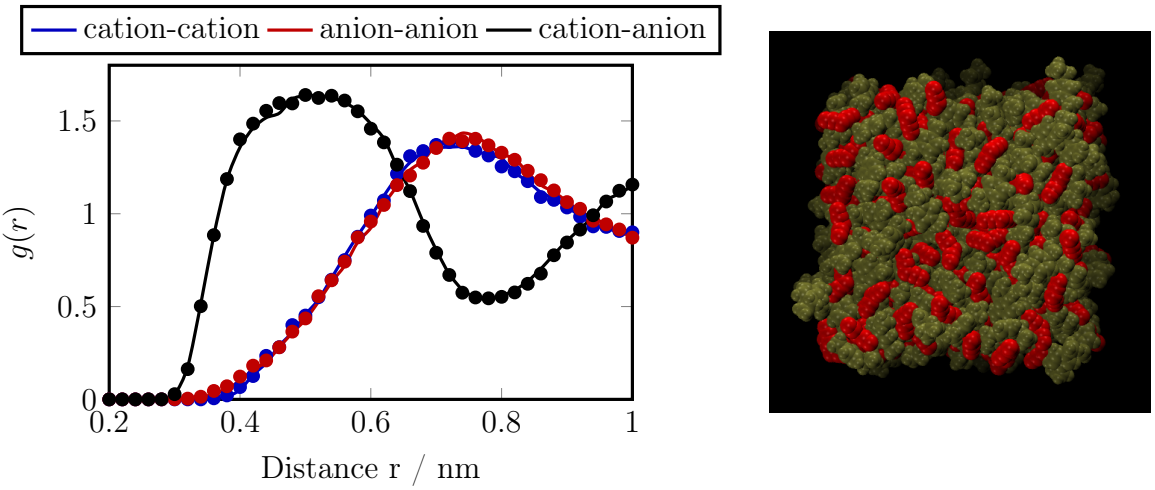


Figure 4.5.: While a comparison between the radial distribution functions of 30 (circles) and 240 (solid lines) ion pairs at $T = 400$ K is shown on the left hand side, a unit cell with 240 [EMIM][DCA] ion pairs is depicted on the right. The cations are drawn in ocre and the anions in red.

In order to benchmark, if structural finite size effects are observable, a box of 240 ion pairs was tested. To this end, a cubic unit cell with $V = (4.0\,\mathrm{nm})^3$ has been filled with 240 randomly placed ion pairs, chosen from the annealed 30 ion pair system. Due to the bigger box size the cutoffs for the SPME algorithm were increased to 1.2 nm in real space, but the reciprocal space cutoff was reduced, to a grid size of $K = 96$ in every spatial dimension. To distribute the error equally between real and reciprocal space, an Ewald splitting parameter $\beta = 3.1\,\mathrm{nm}^{-1}$ was applied, which to achieve an error of approximately $10^{-4}\,\mathrm{kJ \cdot mol^{-1} \cdot nm^{-1}}$, as before. At first, a steepest descent energy minimization was applied until the forces converged to a value below $100\,\mathrm{kJ \cdot mol^{-1} \cdot nm^{-1}}$. In order to equilibrate the system, NpT simulations with a Berendsen barostat[69] at $T = 300$ and 400 K were performed, until a stable value of the box size had been achieved, which required 500 ps. Afterwards, a further 1500 ps were simulated for data accumulation. The application of the Berendsen barostat for the data production is justified here, because only static observables are compared quantitatively. At both temperatures an underestimation of the experimental mass density by the simulations with 240 IP systems is observed (see Figure 4.4), but the deviation is only 1%. Hence, the CLaP force field provides a reliable description of the mass density, as already observed for other anions[9,10].

To detect possible finite size effects in the global liquid structure, the results of the 30 and 240 IP systems at $T = 400$ K are compared in terms of the radial distribution functions (RDF) between the anions, cations, and anion-cation (see Figure 4.5). Obviously, the structures compare very well, hence the global arrangement of the molecules is not influenced by the box size.

For this reason, the equilibrated 30 ion pair structures are expected to provide a solid basis for the AIMD simulations.

Calculating partial charges from *bulk-like* configurations To determine a set of partial charges for a classical force field, which contains the characteristics of a liquid, such as a polarized electron distribution, several aspects have to be considered. Firstly, the structural information of the IL given by the coordinates of the atoms must not suffer from finite size effects. Furthermore, a reliable method to assign the partial charges from a periodic charge density is required. The standard approaches RESP and CHELP are difficult to apply in dense systems, because a suitable choice of a surface for a fit of the electrostatic potential is hard. In addition, a certain amount of volume around the atoms has to be excluded, because in the vincinity of the atoms, a point charge is not able to model a continuous charge density. The introduced Blöchl method (CAB) (see section 4.2) has been especially developed for the periodic case and therefore is very suitable. Due to the sensitivity of the electron density to the configuration of the atoms, a sufficiently large set of configurations is required, to derive a set of partial charges that appropriately resembles the average charge distribution of the molecules in the polarized state. To this end, quasi-independent snapshots from an AIMD simulation have been chosen and CAB has been applied to every configuration. Finally, the partial charge of the atoms is derived by averaging over all snapshots and molecules. In order to introduce the procedure for the application of the Blöchl scheme, a set of 100 snapshots has been chosen from an AIMD simulation[7] of 30 [MMIM][Cl] ion pairs. The simulation had been performed at $T = 425$ K and integrated up to 30 ps simulation time with the plane-wave based approach of CPMD[51], using Troullier-Martins pseudopotentials and the PBE exchange-correlation functional. The snapshots were analyzed with the `CP2K` program package[52]. In

contrast to the AIMD simulations, the electron density for CAB is calculated with the hybrid Gaussian plane wave (GPW) method using a TZV2P basis set in combination with the PBE exchange correlation functional and Goedecker-Teter-Hutter (GTH) pseudopotentials[158]. In the Blöchl analysis, three Gaussians have been applied on every atomic site with a progression factor of 1.5 for the decay length $r_{c,i}$ of the Gaussians and $r_{c,0} = 0.265$ Å.

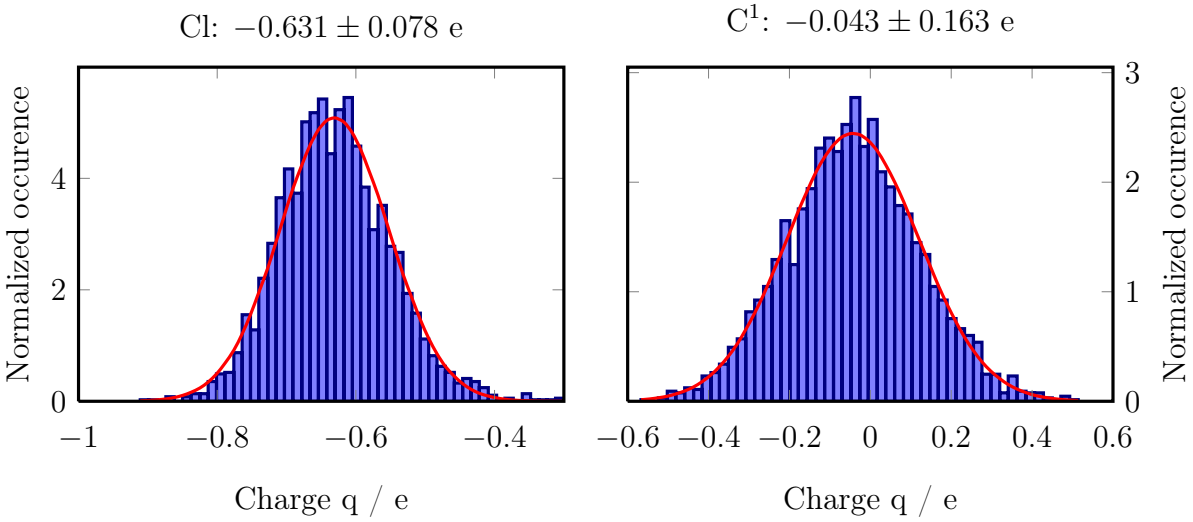


Figure 4.6.: Distribution of the partial charges determined by the Blöchl method from 100 snapshots of a 30 ps AIMD simulation with 30 [MMIM][Cl] ion pairs. The red line is a Gaussian distribution with the corresponding average charge and standard deviation.

In Figure 4.6, the normalized distribution of the derived partial charges for the anion and the C^1 atom of the cation are plotted (for nomenclature see Figure 4.3). The average chloride charge $q_{[Cl]^-} = -0.631$ e is very similar to the charges of the minimum energy geometries (see Table 4.1) and the distribution for the charges of the individual snapshots shows Gaussian behaviour. All distributions of cationic partial charges are dominated by a large spread of the order of the partial charge itself as shown in Figure 4.7 for the atoms of the methyl groups, except from C^1, with about four times larger fluctuations. The Gaussian character of the distributions is highlighted, by looking at the hydrogens H^3, which provide the largest number of samples. Hence, the derived set of partial charges is equivalent to a set of independent variables, because the law of large numbers applies and the individual charge distributions follow a Gaussian distribution. For this reason, the spread is not a relict of insufficient sampling, but the proof for a convergence of the probability distribution. The large spread emphasizes the high sensitivity of the partial charges to the environment resulting from polarization of the electron cloud.

Comparing the different sets of partial charges obtained in the isolated and bulk-like state allows to illustrate the rearrangement of the charges. The partial charges derived by averaging over the snapshots are considered as "effective", because they are expected to contain the characteristics the polarized electron cloud in the liquid phase. Thus, the reduced net-charge is a result from the strong polarizability of the charge density. The reduced net charge $q_{net} = \pm 0.631$ e, [MMIM][Cl] has an electronic dielectric constant $\epsilon_{el} = 2.51$.

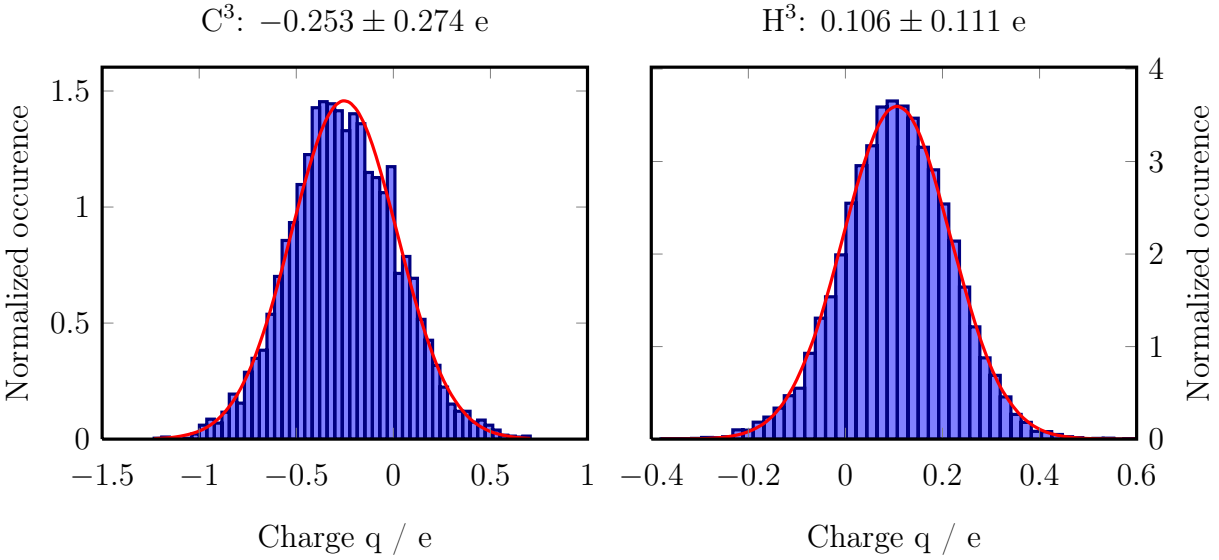


Figure 4.7.: Partial charges and their distribution derived from 100 AIMD snapshots for the methyl groups of [MMIM][Cl].

Comparing the charge distributions derived from a single ion given by CLaP, to the unscreened and "effective" *bulk* calculations clearly shows that negative charge is redistributed from the whole molecule to the C^2 atoms of the imidazolium ring. An important effect of this rearrangement is the expected increase in the molecular dipole moment poiting in the C^1–H^1 direction.[1] Moreover, if the "effective" charges are considered the changing interaction becomes clear. Apart from an increase of the dipole moment, the interaction H^1 with the anion becomes stronger, which has often been argued to play an important role for the dynamics of imidazolium based ionic liquids.[15,112,126,140,159–162] Although there is no significant difference in the effective charging of H^1, if the conditions are changed from a single ion to the bulk phase, the complete environment around is charged less positively, increasing the probability of hydrogen bonding at this specific atom. For this reason, the partial charges derived in bulk are favouring hydrogen bonding at H^1 and at the hydrogens of the methyl groups H^3, while decreasing the possibility at H^2, which has previously been observed.[112,160]

	partial charges q / e							
	H^1	H^2	H^3	C^1	C^2	C^3	N^1	Cl
CLaP[5]	0.21	0.21	0.13	-0.11	-0.13	-0.17	0.15	-1.0
"effective" CAB (30IP)	0.138	0.234	0.106	-0.043	-0.224	-0.253	0.193	-0.631
"effective" CAB / q^{net} (30IP)	0.219	0.371	0.168	-0.068	-0.355	-0.401	0.306	-1.0

Table 4.5.: Comparsion of the charges for [MMIM][Cl] derived by CAB and the CLaP FF.[5] A nomenclature of the atoms is given in Figure 4.3.

In summary, the charge distribution given by the Blöchl analysis shows that the ions are highly polarized. The redistribution of negative charge to the C^2 ring carbons and C^3 methyl carbons emphasizes the extended spatial distribution over the molecule, which is a characteristic of

the liquid state.

The influence of the exchange correlation functional approximation Since the accuracy of an electronic DFT calculation depends on the employed exchange–correlation (XC) functional, which is not exact, comparing with more accurate methods such as second-order Møller-Plesset (MP2) calculations is required to make a reasonable choice. In order to study the dependence of the partial charges on the exchange-correlation (XC) functional, the electron density of 100 snapshots of an "ab–initio" simulation[15] of [EMIM][DCA] (the scheme is shown in Figure 4.9) has been derived with a PBE and a Grimme-corrected BLYP (BLYP–G) XC functional. It has been shown, that the energetic order of the minimum energy structures, which were derived from MP2 calculations, are only reproducible with BLYP–G[15]. Hence, BLYP–G has been applied for the AIMD simulations to generate the configurations for a Blöchl analysis. In order to study, if the strong dependency of the energies is transferred to the results of the partial charges, two different XC functionals are compared.

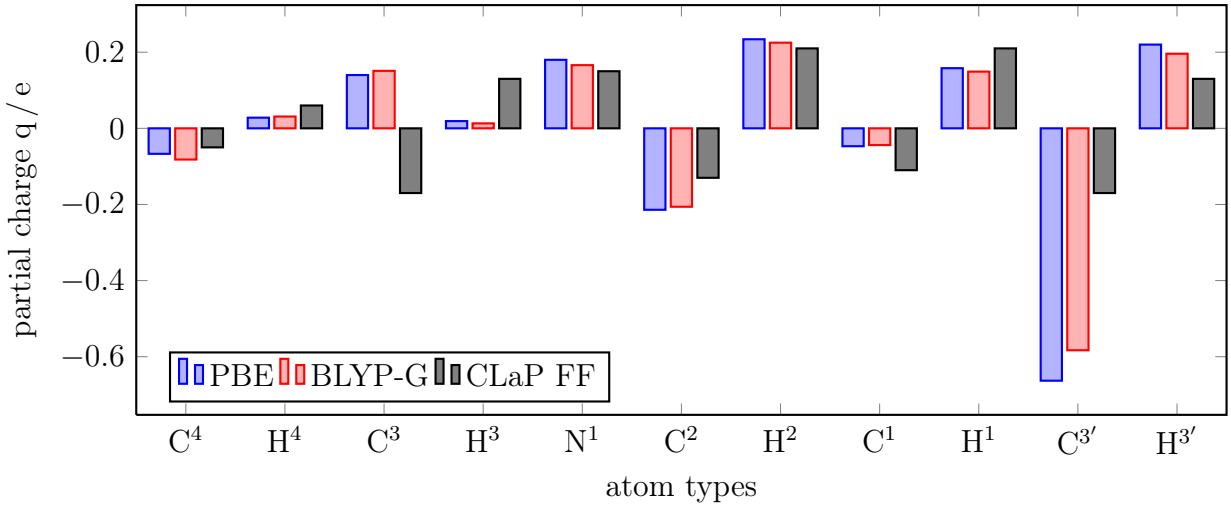


Figure 4.8.: Partial charges obtained from electron densities, based on different XC functionals, are compared to CLaP FF charges derived *in–vacuo*. A nomenclature of the atoms is given in Figure 4.9.

In Figure 4.8, the partial charges for [EMIM][DCA] derived by calculations, which differ in choice of the exchange-correlation functional only, are compared to the CLaP parameterisation of Canongia Lopes *et al.*[5] calculated from an isolated ion with the CHELP approach. As in the case of [MMIM][Cl], the charge distribution of $[EMIM]^+$ derived by the Blöchl method deviates significantly from the partial charges given by CLaP, which is discussed in detail in section 4.4. In contrast, the differences between the results with the PBE and BLYP–G XC functional are rather small.

Thus the calculation of the partial charge is only slighlty influenced by the exchange correlation functional. Most of the charges obtained with the different XC functionals are separated less than their standard deviation. However, as the BLYP–G functional has been shown to reproduce results obtained from higher-order QM methods more accurately, it is expected to allow a more accurate calculation of the partial charges as well.

Influence of small variations of the charge model on dynamical and structural properties
In contrast to [MMIM][Cl], which has a symmetric charge distribution, the partial charge assignment strategy for an asymmetric cation is not obvious. Moreover, since the final aim is the construction of a generic force field, it is expected, that small modifications of the partial charges are required. In order to study the influence on small changes in the charge distribution and the impact of the reduced net-charge different electrostatic models are compared for the IL [EMIM][DCA] (the scheme of the ions is shown in Figure 4.9). While bonded and short-range interactions were transferred from the CLaP FF, the average charge distribution derived by the Blöchl method (aCAB), and a version of aCAB with a symmetric charge distribution on the ring (CAB) were applied. For the FFs with a reduced net-charge, CAB and aCAB, the dihedral parameters were adapted to their partial charges as described in section 5.1.

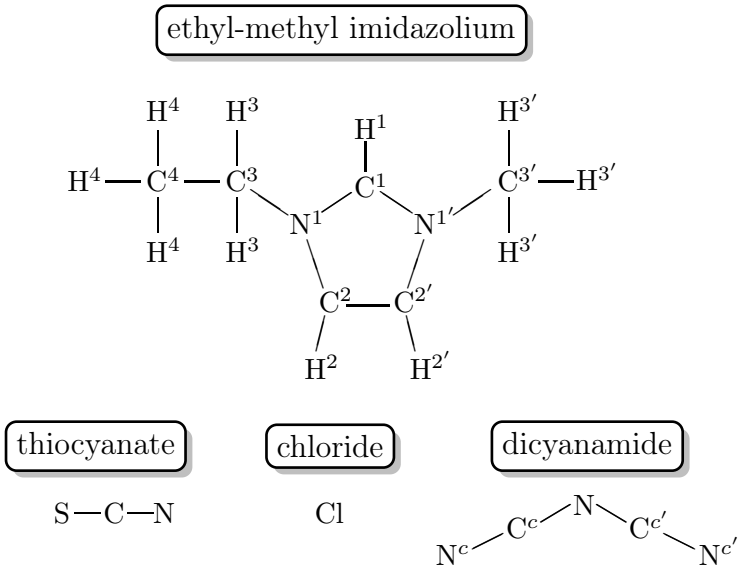


Figure 4.9.: Nomenclature scheme for the $[EMIM]^+$ cation and the $[SCN]^-$, $[Cl]^-$, and $[DCA]^-$ anions.

	partial charge q / e					
	C^1	$C^{2/2'}$	C^3	$C^{3'}$	C^4	$N^{1/1'}$
CLaP	-0.110	-0.130	-0.170	-0.170	-0.050	0.150
aCAB	-0.043	-0.179/-0.233	0.151	-0.583	-0.082	0.017 / 0.316
CAB	-0.043	-0.206	0.151	-0.583	-0.082	0.166
	H^1	$H^{2/2'}$	H^3	$H^{3'}$	H^4	
CLaP	0.210	0.210	0.130	0.130	0.060	
aCAB	0.149	0.247/0.203 (0.225)	0.013	0.196	0.031	
CAB	0.149	0.225	0.013	0.196	0.031	

Table 4.6.: CAB force field charges for [EMIM][DCA].

The partial charges are summarized in Table 4.6. The models aCAB and CAB differ only in the partial charge assignment of the symmetric ring carbons $C^{2,2'}$ and nitrogens $N^{1,1'}$, because

the difference between the charging of the symmetric ring hydrogens $H^{2,2'}$ are smaller than 0.04 e and their average has been used in the simulations.

In Figure 4.10, the distributions of the cationic dipole moments in respect to the center of mass are shown which were derived from the AIMD snapshots. Obviously, the averaged charge distribution aCAB does not reproduce higher electrostatic moments, but an acceptable modelling of the dipole distribution is achieved, if the imidazolium ring is made symmetric by averaging the symmetric carbons, nitrogens, and hydrogens, correspondingly. This is discussed in detail in section 4.4.

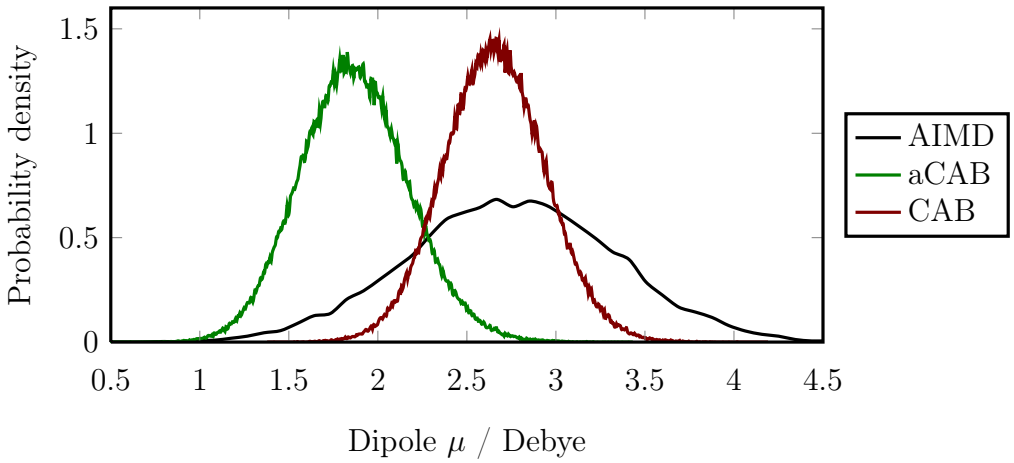


Figure 4.10.: Dipole moment distributions in respect to the center of mass for different charge assignments are compared to the Wannier analysis of the AIMD snapshots for [EMIM][DCA].

To consider a reasonable set of benchmark properties, the density and radial distribution functions are chosen to investigate structural properties, while the dynamics of the single ions is captured by their diffusion coefficient. Finally, collectivity is studied in terms of the static conductivity derived by the Einstein-Helfand method and the heat of vaporization $\Delta H^{\text{vap.}}$. To this end 240 [EMIM][DCA] ion pairs have been simulated with an extended-Lagrangian approach to sample an NpT ensemble at $T = 400$ K and $p = 1$ bar for 40 ns after an appropriate equilibration for 10 ns with a Berendsen barostat to achieve a stable pressure and further 10 ns equilibration time in the canonical ensemble. The Leap-Frog algorithm is applied with a timestep of 1 fs and the electrostatic error of SPME is kept below 10^{-3} kJ $\cdot$ mol^{-1} $\cdot$ nm^{-1}. The LJ interactions were calculated up to 0.9 nm and smoothly switched off within 1.2 nm by a corresponding scaling of the forces. A dispersion correction has been applied to energy and pressure.

While for many ILs, experimental measurements are rare, a wide spectrum of experimental and simulation data is present for [EMIM][DCA].[3,4,121,138,149,163] Since the mass density is easily obtained from an MD simulation, it very often acts as a target and validation property in the FF parameterisation process. Recently, Fröba *et al.*[3] have investigated [EMIM][DCA] and fitted the measured mass densities for different temperatures to a quadratic function:

$$\rho(T) \ / \ \text{kg} \cdot \text{m}^{-3} = 1339.847 - 0.87546 \cdot T + 0.33846 \cdot 10^{-3} \cdot T^2, \tag{4.19}$$

that predicts the experimental results within an error of ±0.02 % and agrees also with already published literature.[163] Furthermore, the studies by Fröba *et al.*[3] uncover a trend of an increasing density of the ionic liquid with increasing moleculare weight of the anion. For $T = 400$ K equation (4.19) predicts $\rho^{\text{exp.}} = 1043.8\ \text{kg} \cdot \text{m}^{-3}$ which is about 5 % above the simulation results. This results from a decreased electrostatic interaction for the reduced net–charge models, aCAB and CAB, and unsuitable short–range parameters. The difference between aCAB and CAB is negligible.

	ρ / kg · m^{-3}	D / 10^{-5} cm^2 · s^{-1}		D^+ / D^-	σ / S · m^{-1}	$\Delta H^{\text{vap.}}$ / kJ · mol^{-1}
		D^+	D^-			
aCAB	993.0	0.74	0.85	0.87	6.41	118.2
CAB	993.4	0.69	0.88	0.78	6.83	118.1

Table 4.7.: Some properties of [EMIM][DCA] obtained with different sets of partial charges at $T = 400$ K and $p = 1$ bar.

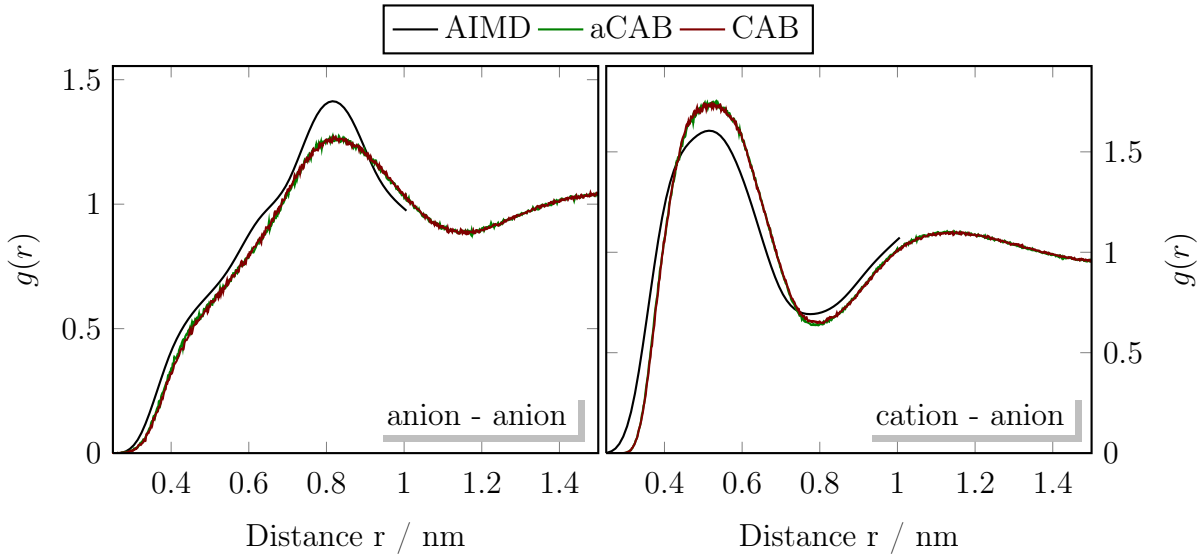


Figure 4.11.: Different radial distribution functions of [EMIM][DCA] for different charge assignments are compared to AIMD results.

In order to dissect the effect of the charge distribution on the local packing in more detail, radial distribution functions are considered. All parameterisations demonstrate the layering of oppositely charged ion species nicely, as already observed for other ionic liquids.[9,10,138] Corresponding to the results of the density, the cation-anion RDFs of aCAB and CAB, shown in Figure 4.11, coincide almost perfectly, but a deviation from reference AIMD simulations by Wendler *et al.*[15], are observed. The too low mass densities, reproduced by aCAB and CAB, are an artefact of a shift of their first maximum in the cation–anion RDF to larger distances compared to the reference AIMD RDF. In contrast, the anion-anion RDFs of aCAB and CAB agree very well with the AIMD results. This coincides with the results obtained for [MMIM][Cl][12], because the Cl–Cl RDF matches the reference AIMD results very well if

charges derived by the Blöchl method are inserted into the CLaP force field, just as done for the parameterisations CAB and aCAB. This shows that the balance between short and long range interactions depends strongly on the size of the ion. For small ions electrostatic interactions seem to dominate the global structural behaviour, while with an increasing size of the molecule the interplay between short–range and electrostatic interactions becomes more and more important for a correct modelling of the liquid structure.

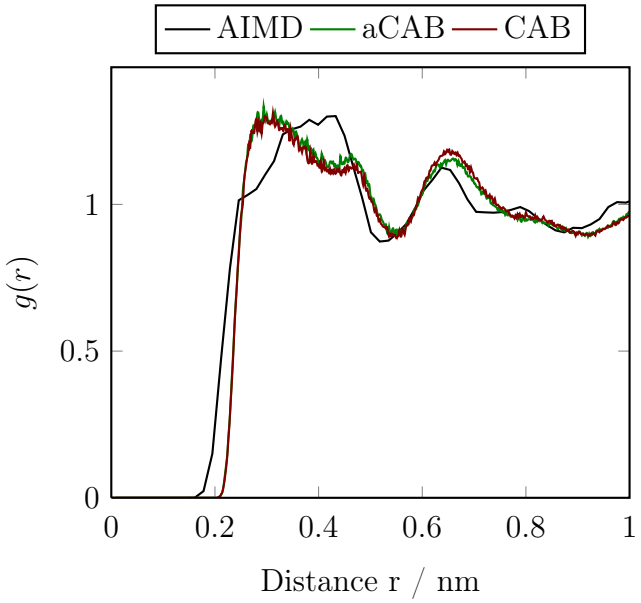


Figure 4.12.: Partial RDF between H^1 and the center nitrogen N of the anion.

A more detailled study of the local structure by a comparsion of the partial RDFs between H^1 and the center nitrogen N of $[DCA]^-$ is shown in Figure 4.12. No significant difference is observed between the results for aCAB and CAB, though the partial charges of the atoms around H^1 are completely assymetric in case of aCAB. Thus either the global nor the local structure is significantly influenced by slight changes of the partial charges. The results from the classical MD simulations with aCAB and CAB deviate from the AIMD reference RDF, showing that an optimization of the short–range parameters is inevitable.

While structural properties up until now will be applied for the calibration of the force field parameters, the following properties shed light on the dynamics and collectivity of the system. An indicator for the single particle dynamics of a system is the diffusivity. As shown in Figure 4.13 and summarized in Table 4.7, the parameterisations aCAB and CAB predict almost the same diffusion constant, but they are twice as large as the results from the polarizable model of Borodin[149]. Apart from a fast dynamics, a value for the ratio of the diffusion constants D^+ / D^- of smaller than 1 is observed, if reduced charges are applied, such as in the case of [MMIM][Cl].[12,129] However, the the ratio is not that small as for [MMIM][Cl], but the result for aCAB and CAB are smaller compared to the polarizable FF. This arises due to a subtle interplay between electrostatic and LJ interactions determining the molecular volume V_m of the ions[138], which seems to be of greater importance for small molecules like $[MMIM]^+$ and $[Cl]^-$. This is not unexpected, because V_m depends on the charging of the ion, as well as on its short-range parameterisation. Hence, the impact of charging decreases if the number of

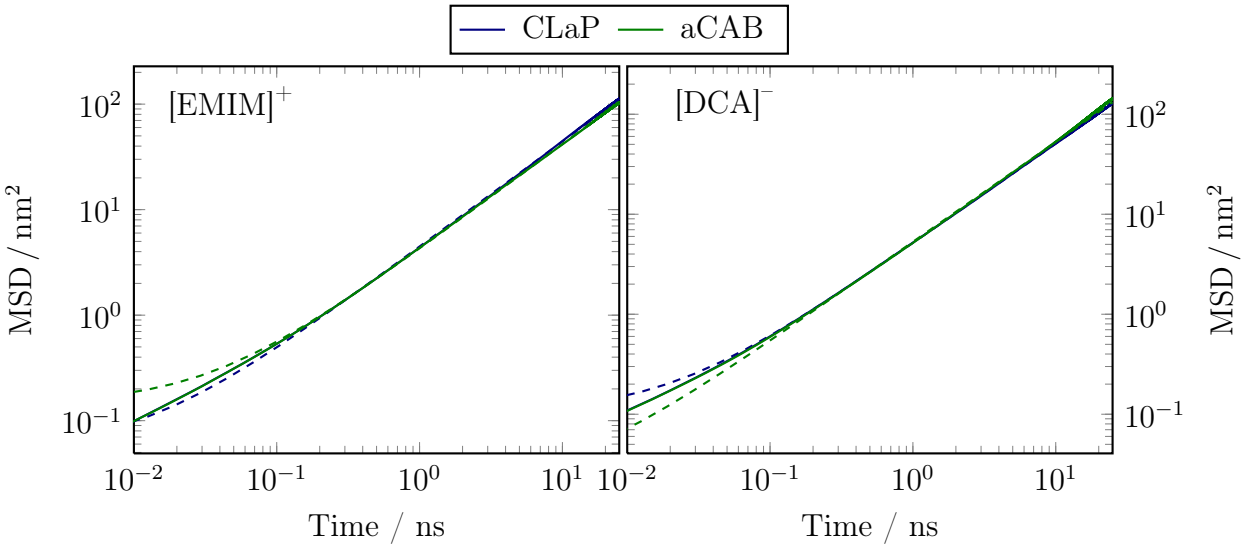


Figure 4.13.: Mean square displacement of $[EMIM]^+$ and $[DCA]^-$ for different sets of partial charges.

atoms constituting the molecule is increasing. With that, the ratio of the molecular volumes does not change drastically for [EMIM][DCA] if an electronic continuum model, such as aCAB or CAB, with a reduced net–charge is applied. This agrees with the simulations of Schröder and Steinhauser[138] performed at $T = 300$ K with different integer net–charge models, which give values of 0.88 and 0.92 for the ratio of the diffusion constants. Thus also slight variations in the charge model do not heavily disturb the single particle dynamics.

To study the correlation between the moving ions, the static conductivity σ is derived by the Einstein-Helfand method. As the simulation temperature is above the experimentally accessible range, corresponding values have to be derived from corresponding extrapolations. Since ILs are known to behave like supercooled liquids, the temperature dependence of σ is assumed to be reflected by the empirical Vogel-Fulcher-Tamman equation[164]:

$$\sigma = \sigma_0 \cdot \exp(\frac{B}{T - T_{\mathrm{VTF}}}). \tag{4.20}$$

While the parameters σ_0 and B are temperature-independent parameters, T_{VTF} depends on the regime, where equation (4.20) is applied. If the temperature range near the glass-transition temperature T_g is considered, T_{VTF} is related to T_g via the fragility m[165,166], that depends on the behaviour of the orientational relaxation near T_g. Therefore, T_{VTF} can be used either as a parameter to predict material properties, or, for decreasing the degrees of freedom of equation (4.20).

For [EMIM][DCA], Yoshida *et al.*[4] measured the static conductivity in the range from 300 K to 340 K. Yet, as the simulations were performed at $T = 400$ K, an extrapolation of the experimental results to higher temperatures is required. To this end, equation (4.20) is applied to the experimental data. The resulting fit is shown in Figure 4.14 on the left hand side and gives $\sigma(400\ \mathrm{K}) = 11.42\ \mathrm{S \cdot m^{-1}}$.

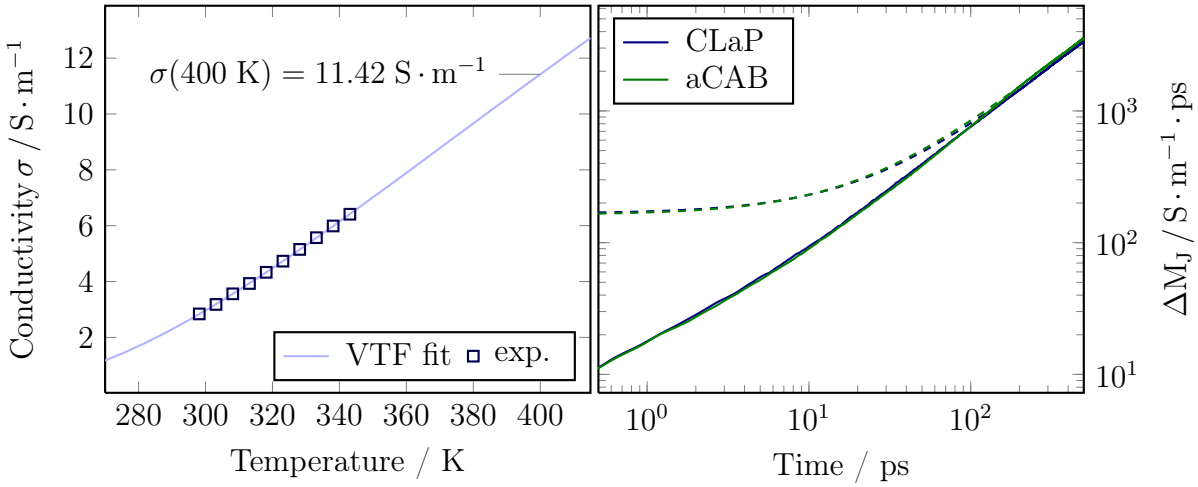


Figure 4.14.: Static conductivity for [EMIM][DCA] obtained from experiment[4] and different parameterisations.

On the right hand side of Figure 4.14, the mean square displacement of M_J is shown, which is used to determine σ from its slope. The results, summarized in Table 4.7, prove that the difference between the models with a reduced net-charge, aCAB and CAB, is rather small, but the experimental reference data is underestimated. Also the polarizable model of Borodin[149] predicts a too small value of 10.5 S · m^{-1} compared to experiment, but a larger value compared to the results of aCAB and CAB, although the diffusivity of aCAB and CAB is much faster than given by the polarizable model. But the collectivity of the dynamics is not sensitive to small changes in the partial charges.

Finally, the molar heats of vaporization $\Delta H^{\text{vap.}}$ are compared in Table 4.7, in order to analyze the influence of the partial charges on collective energetic properties. These are derived from the difference of the liquid $H^{\text{liq.}}$ and gas phase H^{gas} enthalpies:

$$\Delta H^{\text{vap.}} = H^{\text{gas}} - H^{\text{liq.}}. \tag{4.21}$$

The enthalpy of the liquid phase is easily calculated from the total energy $E^{\text{tot.}}$ and the term $P\langle V\rangle$ where P is the reference pressure of the NpT simulation and $\langle V\rangle$ the average volume. For the gas phase simulations, no pressure is accessible, because they are preformed in vacuum, but assuming an ideal behaviour, the equation of state can be utilized to obtain the term $P\langle V\rangle = nRT$ with the amount n of ion pairs in mol, the gas constant R, and the temperature T. Thus following relation is obtained:

$$\Delta H^{\text{vap.}} = E^{\text{tot.,gas}} + RT - H^{\text{liq.}}, \tag{4.22}$$

for the molar heats of vaporization.

No impact of the small differences in the charge distribution of aCAB and CAB on ΔH^{vap} is observable. The results of aCAB and CAB are slightly lower than those derived with a polarizable model at 298 K. Hence, the smaller value might also be attributed to the higher simulation temperature, but only an experimental measurement is able to conclusively validate

the results.

In summary, the comparison of different charge models for an [EMIM][DCA] force field highlighted that slight modifications of the charge distribution affect neither structural nor dynamical properties significantly. It has been shown that the local packing reproduced by aCAB and CAB does not resemble the results from the AIMD simulations. Neither diffusivity nor conductivity is influenced by small changes in the charge distribution, but fast dynamics compared to a polarizable model[149] for aCAB and CAB are observed, due to the reduced electrostatic interaction. If the collectivity of the system is studied in terms of the heat of vaporization, there is also no difference between aCAB and CAB is observable. Yet, neither structural, dynamical, nor energetic properties are strongly influenced by moderate changes of the partial charges, relieving a construction of a large transferable force field, because the number of different atom types may be reduced by "*educated*" averaging, such as performed in order to obtain a symmetric ring charge.

4.4. Mapping of partial charges

In order to study the mechanisms of the charge distribution for different cations and anions, snapshots of ethyl-methyl-imidazolium and buthyl-methyl-imidazolium cations in combination with the anions chloride $[Cl]^-$, dicyanamide $[DCA]^-$, and thiocyanate $[SCN]^-$ (see Figure 4.9) have been analyzed. The required configurations had been present from already existing studies of Wendler *et al.*[167] for $[DCA]^-$ and $[Cl]^-$, and from Thar *et al.*[168] for $[SCN]^-$ in combination with $[EMIM]^+$. Morever, the configurations for $[BMIM]^+$ based ILs had been exclusively provided by Wendler *et al.*[167]. While the snapshots from the simulations of Wendler *et al.*[15,167] are derived with plane wave based CPMD, the configurations from Thar *et al.*[168] are obtained by Born-Oppenheimer MD combined with the GPW method. All simulations were performed at a temperature $T = 400$ K. This variety of partial charges allows one to study the influence of the ion combination on the charge distribution. Since the Blöchl method is expected to provide a reliable description of the ionic charge state in the liquid, the obtained data will be a first step towards the construction of a transferable force field, which is desirable to allow a systematic investigation of ionic liquid properties with classical MD simulations. However, a proper understanding of the charge redistribution in the liquid is required in order to construct such a force field.

It is shown that the high charge density in an IL provides strong electrostatic screening, which causes a strong locality of the electrostatic properties. In order to derive a reliable set of partial charges, the influence of the combination of cation and anion is discussed, as well as the effect of small *a-posteriori* changes in the partial charges on the dipole moment. Finally a method is proposed that allows the construction of a set of partial charges transferable between different cations and anions.

Application of the MDEC theory to Blöchl charges The polarizability of an atoms determines the size of an induced dipole in respect to a certain electric field. If a dipole is induced a certain amount of charge is dislocated from the center of an atom. The charge on the center of the atom is reduced, assuming the amount of dislocated charge is constant, charge conservation in the whole system, and that various induced dipole moments arise from a delocalization

of the charge in different directions. Hence, if the charge for a polarized system is projected onto the center of the atoms and averaged over different configurations, a certain amount of charge is expected to be dislocated, constantly. To this end, the effective net-charges of the ions are always reduced in the polarized state and the electronic part of the polarization is described by the effective screening of the interactions. Charges $q_i^{\text{eff.}}$ derived with a Blöchl analysis describe the effective electrostatic forces. With that, the polarization state of the ion is not given directly, but included implicitly. Since the charge reduction given by $q_i^{\text{eff.}}$ allows to derive $\epsilon_{\text{el}} = \frac{1}{q_{\text{net}}^2}$, the polarized state of the charges $q_i^{\text{pol.}}$ is also accessible:

$$E^{\text{Coulomb}} = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}} \tag{4.23}$$

$$= \frac{1}{\epsilon_{\text{el}}}\frac{1}{4\pi\epsilon_0}\frac{q_i^{\text{pol.}} q_j^{\text{pol.}}}{r_{ij}}. \tag{4.24}$$

Yet, a measure for polarizability is the average dislocated charge per atom given by the root mean charge fluctuation δq:

$$\delta q = \sqrt{\frac{1}{N_{\text{atoms}}}\left(\langle q_{\text{net}}^2\rangle - \langle q_{\text{net}}\rangle^2\right)}, \tag{4.25}$$

where N_{atoms} is the number of atoms per ion. In Table 4.8, δq has been summarized for the studied imidazolium-based ionic liquids.

δq / e	[SCN]	[Cl]	[DCA]
[MMIM]		0.03, 0.08	
[EMIM]	0.05, 0.08	0.04, 0.10	0.04, 0.06
[BMIM]	0.05, 0.09	0.04, 0.09	0.04, 0.05

Table 4.8.: Root mean charge fluctuation of the cation and anion net charges of the studied ionic liquid systems, normalized with respect to the number of atoms per ion. The first number in a cell belongs to the cation in the row, while the second number belongs to the anion given by the column.

Obviously, δq of the cation is quite independent of the ion combination as expected, apart from small deviations for the $[SCN]^-$ systems and [MMIM][Cl]. If δq of the anions is considered, also small differences within a series of anions is observed, but without any trend, such that they are expected to be an artefact of poor statistics or numerical errors. However, δq of $[DCA]^-$ is quite close to δq of $[BMIM]^+$ and $[EMIM]^+$ for all ion combinations. This results from the elements which $[DCA]^-$ consists of, N and C, which are also present in imidazolium. The larger value might be a result of the missing hydrogens in $[DCA]^-$, which might decrease δq for the imidazolium-based cations, if the charge fluctuation at the hydrogens is smaller than on the heavier atoms. Since all cations consist of the same kinds of atoms, these results show, that the approximations for the linear response regime hold.

To compare our results to experiment, the relation between ϵ_{el} and the refractive index n_{D} of a medium is considered:

$$\epsilon_{\text{el}} = n_{\text{D}}^2. \tag{4.26}$$

For [EMIM][DCA], an experimental value of $\epsilon_{\text{el,exp.}} = 2.28$ has been obtained by Fröba *et al.*[3], which is in good agreement with $\epsilon_{\text{el,calc.}} = 2.23$ derived by the Blöchl method.

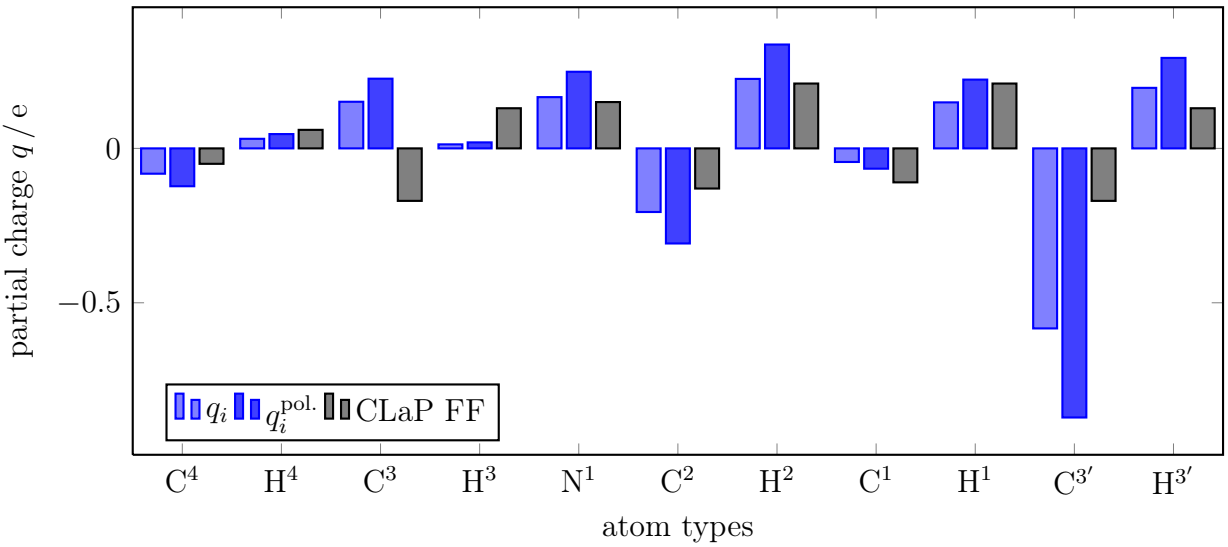


Figure 4.15.: Partial charges q_i for [EMIM][DCA] derived by CAB are compared to partial charges of the CLaP FF[5], which both describe the effective electrostatic forces. However to obtain the polarization state from the CAB charges $q_i^{\text{pol.}}$, a rescaling by the ionic net charge q_{net} is required.

The Blöchl method predicts the effective charges corresponding to the MDEC theory. For this reason, it is expected to be an appropriate choice for the calculation of the partial charges from systems under bulk conditions. In Figure 4.15 partial charges for [EMIM][DCA] derived by CAB and single ion calculations of Canongia Lopes *et al.*[5], which provide the parameterisation of the CLaP FF, are compared. The polarization state of the ion is included in the charge distribution of CLaP, which is also indicated by the integer net–charge. But the parameterisation has been performed in vacuum on single ion configurations, where no polarization effects arise. If the effective CAB partial charges q_i are compared to the CLaP FF, a delocalization of the charge over the whole molecule is observed for CAB, which is characteristic for the liquid phase. This becomes even more clear, if the charge distribution $q_i^{\text{pol.}}$ is considered, that includes P_{el} expliticly. Obviously, charge is distributed to over the ring and to the exterior regions of the cation, such that an expected increase in the dipole moment is observed. [1,14,15]

Hence a consistent combination of the MDEC theory and the results of a Blöchl analysis is a promising method to obtain a transferable set of partial charges for a large ionic liquid force field. By considering the polarizability of the medium in terms of the electronic dielectric constant ϵ_{el}, detailed insight into the redistribution of charge is obtained for a change from an unpolarized single ion state to a polarized one in the liquid phase. Moreover the separation of polarizability and charge distribution allows one to consider both quantities for the force field construction, which alleviates this time-consuming process.

Locality of charges and dipole moments Apart from the screening, which is reflected in the reduced ionic net-charge, also a strong locality of the dipole moments is observed in

[MMIM][Cl],[1,7] [EMIM]$^+$-based, and protic ILs.[14,15] This means that only the dipoles within the nearest neighborhood are correlated with each other, which arises from the high charge density in an ionic liquid and its polarizability. Hence it is also expected that the charge distribution is only influenced by its immediate neighborhood. For this reason, partial charges are derived for different ionic liquids from configurations of small, but periodic, and bulk like systems. From already existing studies, bulk configurations were available for [EMIM][DCA],[15] [EMIM][SCN][168] as well as [MMA][NO_3][169] and [MMIM][Cl].[7] Moreover snapshots of small corresponding counterparts, consisting of eight ion pairs, for all large systems were provided,[14] in order to perform a Blöchl analysis.

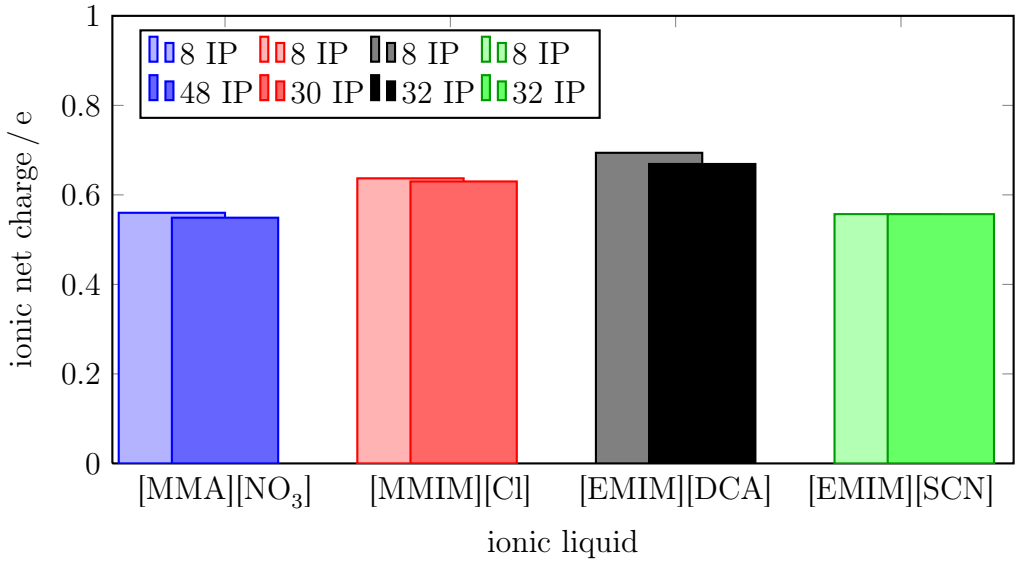


Figure 4.16.: Net charges of cations derived from samples of different size.

The ionic net charges are given in Figure 4.16 and Table 4.9. Apart for [EMIM][DCA] with a difference in charge of only 0.026 e, the effect of the system size on the total charge is negligible. If the fluctuations of the ionic partial charges are considered, a convergence to a value characteristic for the is observed, which is also independent of system size. This shows that already on a very local scale, the electrostatic properties converge to their equilibrium state and no finite–size effects are observable.

		q / e						
IL	IP	q_{net}	δq^+	δq^-	IP	q_{net}	δq^+	δq^-
[MMA][NO_3]	8	0.56	0.18	0.21	48	0.55	0.20	0.22
[MMIM][Cl]	8	0.64	0.13	0.08	30	0.63	0.12	0.08
[EMIM][Cl]	8	0.61	0.16	0.10			–	
[EMIM][DCA]	8	0.70	0.17	0.12	30	0.67	0.19	0.14
[EMIM][SCN]	8	0.56	0.22	0.14	32	0.56	0.22	0.14

Table 4.9.: Net charge q^{net} and charge fluctuation of the cations and anions $\delta q^{\pm}$ obtained from small and large systems.

Moreover, the difference of the partial charge distributions is also very small. For the anions $[DCA]^-$ and $[SCN]^-$, the results derived from a small and a large system are compared in Table 4.10 to force field parameters that were derived by single ion calculations. Since the reduced net–charge is a result of polarization, the vacuum calculations on single ions, which were performed for the force field of Chaumont and Wipff[8] (CWFF) for $[SCN]^-$ and Canongia Lopes *et al.*[5] (CLaP) for $[DCA]^-$, result in an integer net–charge for the ions. In order to compare the state of polarization of the ions $q^{\text{pol.}}$, the effective partial charges $q_i^{\text{eff.}}$ derived by CAB are scaled correspondingly. For both, $[DCA]^-$ and $[SCN]^-$, a shift of the electron density to the central atom of the anions is observed. Since the single partial charges for $[DCA]^-$ and $[SCN]^-$, as well as the total net–charges agree quite well for the large and small system, the redistribution of charge in a bulk–like environment is a characteristic property of an IL, that is already present on a local scale. This demonstrates that partial charges can be derived from small bulk–like systems.

q / e				
	$[DCA]^-$			
	N	C^c	N^c	q_{net}
8 IP	-0.600	0.506	-0.554	-0.696
32 IP	-0.584	0.490	-0.533	-0.670
32 IP / $q_{\text{net}}^{[DCA]^-}$	-0.875	0.631	-0.796	-1.0
CLaP	-0.760	0.640	-0.760	-1.0
	$[SCN]^-$			
	S	C	N	q_{net}
8 IP	-0.304	0.058	-0.312	-0.558
32 IP	-0.302	0.042	-0.297	-0.557
32 IP / $q_{\text{net}}^{[SCN]^-}$	-0.542	0.075	-0.533	-1.0
CW FF[8]	-0.757	0.480	-0.723	-1.0

Table 4.10.: Partial charges of the dicyanamide anion $[DCA]^-$ and thiocyanate anion $[SCN]^-$ derived by CAB from AIMD snapshots of 8 and 32 ion pairs (IP) are compared to the force field charges given by CLaP and Chaumont and Wipff[8], respectively.

Hence, this locality allows one to decrease the required computational effort for the calculation of an IL specific set of partial charges. In Figure 4.17, partial charges $q^{\text{pol.}} = q^{\text{eff.}}/q_{\text{net}}$ for [EMIM][Cl] are derived from 100 snapshots of 8 ion pairs by the Blöchl method and compared to the results of [EMIM][DCA] for both a small and large system. As one can see, the partial charges differ only very slightly between the different combinations of system size and ion combination. Apart from the terminal group of the long ethyl side chain, the charge is more delocalized over the ring and the attached groups compared to the unpolarized vacuum state given by the CLaP FF. This is consistent with the strong localization of the electron density of the anions. A decisive aspect is the small difference in the charge distribution of $[EMIM]^+$, though different anions are involved. Hence, an important requirement for a transferable force field is met, because the charge distribution is quite insensitive to the specific ionic liquid, in contrast to the overall charge reduction, which depends on the ion combination present in the system.

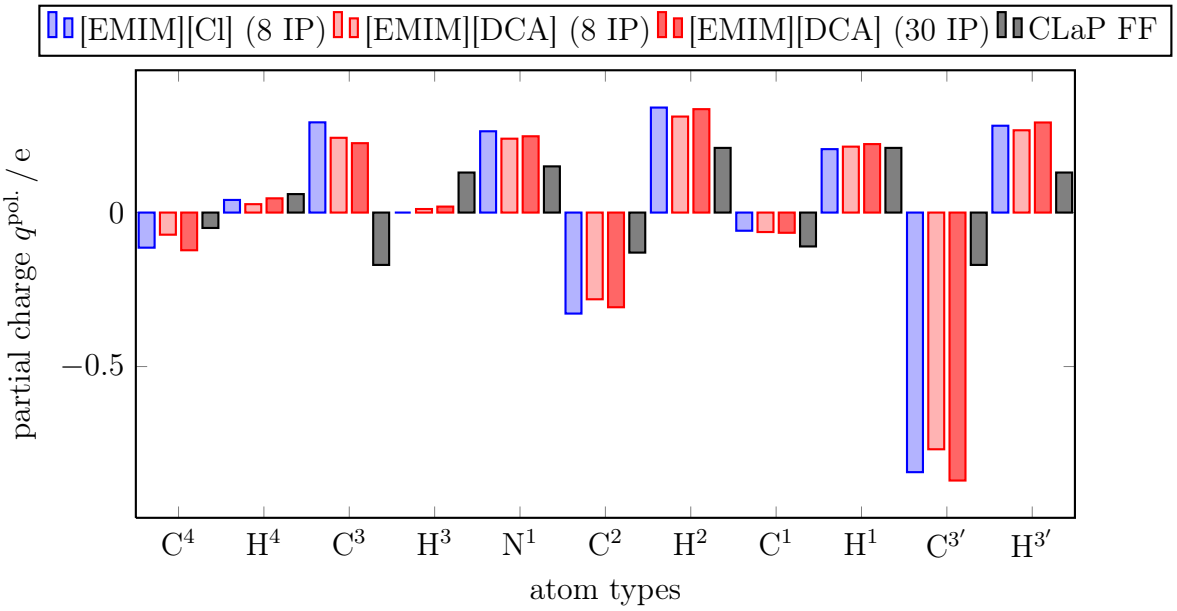


Figure 4.17.: Partial charges of $[EMIM]^+$ derived for different anions from small and large systems and a comparison to the integer charge model of CLaP. A nomenclature of the atoms is given in Figure 4.9.

But in addition to a correct description of the electric monopoles, an accurate modelling of the dipole moments is desirable, to describe the dielectric properties of the IL properly. Due to the symmetry of [MMIM] a mapping of the charges to different atom types is unique, but many reasonable possibilities exist to group the atoms of an asymmetric IL. For this reason the aspect of transferability is the guide for the definition of the different atom types, but the force field has to describe the properties of the IL with an acceptable level of accuracy. To this end, the dipole moment distributions given by a Wannier analysis[170–172] (WA) of the AIMD configurations and CAB charges, are shown in Figure 4.18, to study the reliability of the proposed partial charges. For this reason, the dipole moment distribution of [EMIM][DCA] given by WA is compared to dipole moment distributions derived by CAB charges, but based on different calculation methods. For the calculation of the direct distributions, the charges have been chosen corresponding to the configuration they were derived from. Though the direct distributions are much broader compared to the results of the WA, the averages do not differ drastically. This agreement is no coincidence, but a result of the Blöchl method that also fits higher electric moments. The difference in width is a result of the applied methods. For the calculation of the molecular dipoles with WA, the electron density is mapped to electron pairs, residing on the molecular orbitals (MO). In contrast, CAB projects the charges to the atoms dependent on their the actual configuration and is not restricted by MOs. For this reason, the charge on the ions is not given by an integer net–charge, but fluctuates strongly giving rise to the large width of the dipole moment distribution. Unfortunately, for $[DCA]^-$ the agreement is not as good as for $[EMIM]^+$, which is an artefact of the molecular geometry of the anion. Because the charge density around the central N atom is increased, it is expected, that a wide region around the molecule center is involved. Due to the overlap of the Gaussians constituting the model charge density, additional charge might be also assigned to side chain

carbons. Moreover, $[DCA]^-$ is quite linear and already small geometrical changes affect the dipole moment of the molecule strongly.

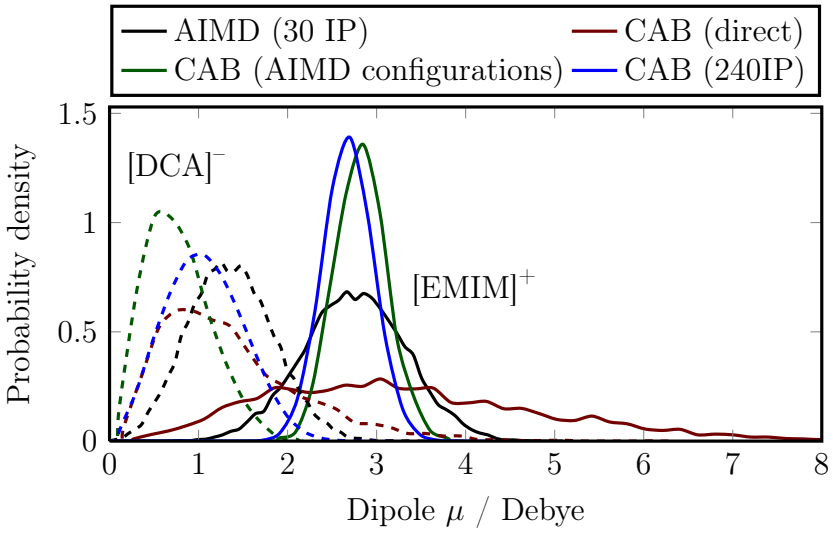


Figure 4.18.: The distribution of the dipole moments obtained from WA are compared to results given by CAB charges. The solid and dashed lines describe the dipole of $[EMIM]^+$ and $[DCA]^-$, respectively. The direct CAB distribution was obtained using the dipole moments given by the partial charge distribution of each snapshot. For the other cases, the partial charges were averaged over all snapshots first.

While P_{el} is considered in the dipole moment distribution, calculated directly from the snapshots, a distribution derived from the averaged effective force field charges describes polarizability on the mean–field level only. A Wannier analysis describes the extension of the electron cloud by a projection of the electron pairs on the correspondingly spatially extended MOs describing P_{el} explicitly. Yet, the dipole moments obtained from the calculation with the effective charges have to be scaled properly:

$$\mu = \sqrt{\epsilon_{el}} \cdot \mu^{eff}, \tag{4.27}$$

in order to obtain a comparable value for μ. The partial charges for the CAB parameterisation are obtained by averaging over all snapshots and molecules. In a further step, also the charges of the symmetric atoms of the imidazolium ring and the anion are averaged. If this parameterisation is applied to derive the dipole moment distribution from the AIMD snapshots, the averages also agree with the WA result for $[EMIM]^+$, but the width of the distribution given by CAB is smaller. Furthermore, the narrow distribution of the dipole moments of $[DCA]^-$is shifted to a lower average value compared to the AIMD results. This again shows the sensitivity of the dipole moment distribution of the almost linear $[DCA]^-$. However, if the WA results are compared to the dipole moment distributions obtained from an MD simulation of 240 IP, an improved description of the average dipole moment of $[DCA]^-$ is obtained. Although the width of the distributions agrees well, the average is below its correct value, which is attributed to the consideration mentioned above. Too little negative charge is assigned to the central

nitrogen, such that the dipole moment is weaker than it should. However, for $[EMIM]^+$ the difference between the results of the static partial charges applied to the AIMD snapshots and the 240 IP MD snapshots is quite small. Only a slight shift of the dipole moments to smaller values is observed. This is a result of the rather complex molecular structure, that allows us to fit the electron density accurately.

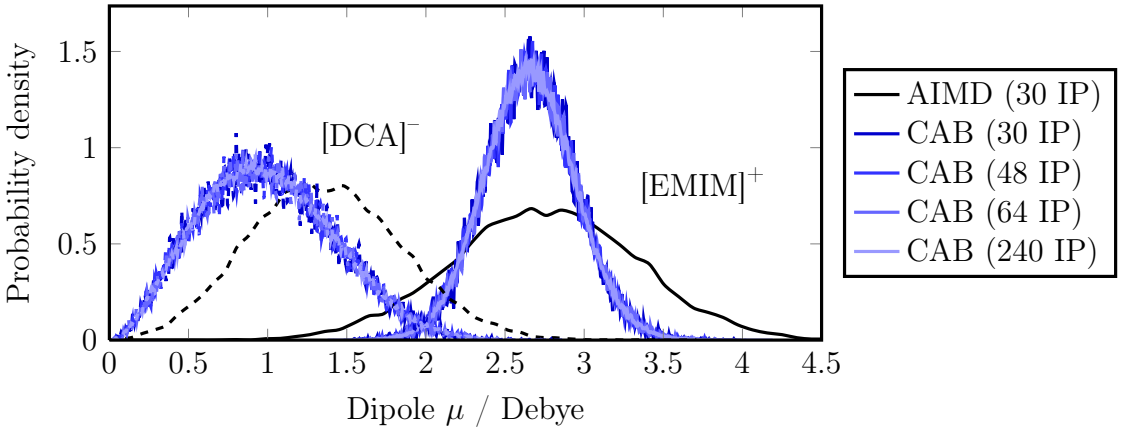


Figure 4.19.: Dipole moment distributions for different systems sizes are derived by CAB charges and compared to the AIMD results.

To verify by classical MD simulations that a system size of 30 IP does not contain any artificial structuring introduced by the periodic boundary conditions, the dipole moment distribution is calculated for systems consisting of up to 240 IP. The results, given in Figure 4.19, show that the average and shape of the distribution is already reproduced by the smallest system and with increasing size, only the statistics improve. Therefore, it becomes clear, that the width is not an artefact of too few statistics, but a result of the conformational changes, which are insufficient in order to describe the actual polarization state.

In summary, these results show that the application of the Blöchl method to IL systems allows a reliable description of the electronic properties by partial charges. The fact that the derived values already converge for small systems, proves that the electronic properties are very local, which is shown for the partial charge and dipole moment distributions. Finally, though the average is underestimated for $[DCA]^-$, as well as the distribution width for $[EMIM]^+$, the overall agreement with the WA is quite acceptable, such that the static partial charges are expected to provide a reliable description.

Partial charges from different combinations of cations and anions In order to construct a transferable force field, the number of atom types should be kept as low as possible. For this reason, CAB charges have been calculated from AIMD configurations taken from existing studies of the IL [MMIM][Cl][7] and all combinations of the cations $[EMIM]^+$[14,15,168] and $[BMIM]^+$[167] with $[Cl]^-$, $[DCA]^-$, and $[SCN]^-$, schematically depicted in Figure 4.9 and Figure 4.20. Hence, a data set is available, that allows the study of the differences in the charge distributions, in order to show that the polarized state of the molecule is ion combination independent, as predicted by the extended MDEC theory.

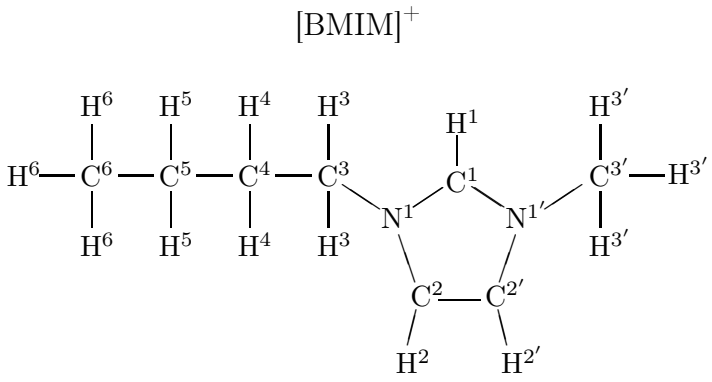


Figure 4.20.: Nomenclature scheme for the $[BMIM]^+$ cation.

In order to obtain an initial overview, the net–charges q_{net} and electronic dielectric constants $\epsilon_{el} = q_{net}^2$ are summarized in Table 4.11. Depending on the anion, the total charge varies from ± 0.56 e to ± 0.71 e, but within a series of systems, that differ only in the side chain length of the imidazolium cation, the differences are at most 0.04 e. This suggests, that ϵ_{el} is rather a quantity, which is specific for the cation-anion combination, and an indicator that the charge groups separated from the ring more than two sites are only slightly polarizable. Moreover, as the variation between the different anions is also small, a generic electronic dielectric constant might allow an appropriate description, which would significantly reduce the amount of required force field parameters.

	q_{net} / e, ϵ_{el}		
	[SCN]	[Cl]	[DCA]
[MMIM]		0.63, 2.52	
[EMIM]	0.56, 3.19	0.61, 2.69	0.67, 2.23
[BMIM]	0.60, 2.78	0.62, 2.60	0.71, 1.98

Table 4.11.: Net charges q_{net} and electronic dielectric constants ϵ_{el} of some imidazolium based ionic liquids.

A further criterion for a reduction of atom types is the coincidence of the charge distributions $q_i^{pol.} = \sqrt{\epsilon_{el}} q_i$ describing the polarized state of the ions. The charge distribution $q_i^{pol.}$ is chosen, because it considers polarization effects, explicitly, providing unbiased information about the charge distribution.

In Figure 4.21, the partial charges $q_i^{pol.}$ are compared for $[EMIM]^+$–based systems with the anions $[DCA]^-$, $[SCN]^-$, and $[Cl]^-$. Apart from the carbons C^3 and $C^{3'}$, the difference in charge is quite small. However, the distribution differs for most of the atoms significantly, if the electron density is derived from the liquid state or from calculations on a single molecule, which has been done for the parameterisation of the CLaP FF. Interestingly, C^3 even exhibits a change from negative to positive charge for all studied ion combinations.

However the positive charge of C^3 is not an $[EMIM]^+$ related property, but also observable for the C^4 and C^5 atoms of $[BMIM]^+$ for the same series of anions, as shown in Figure 4.22. In

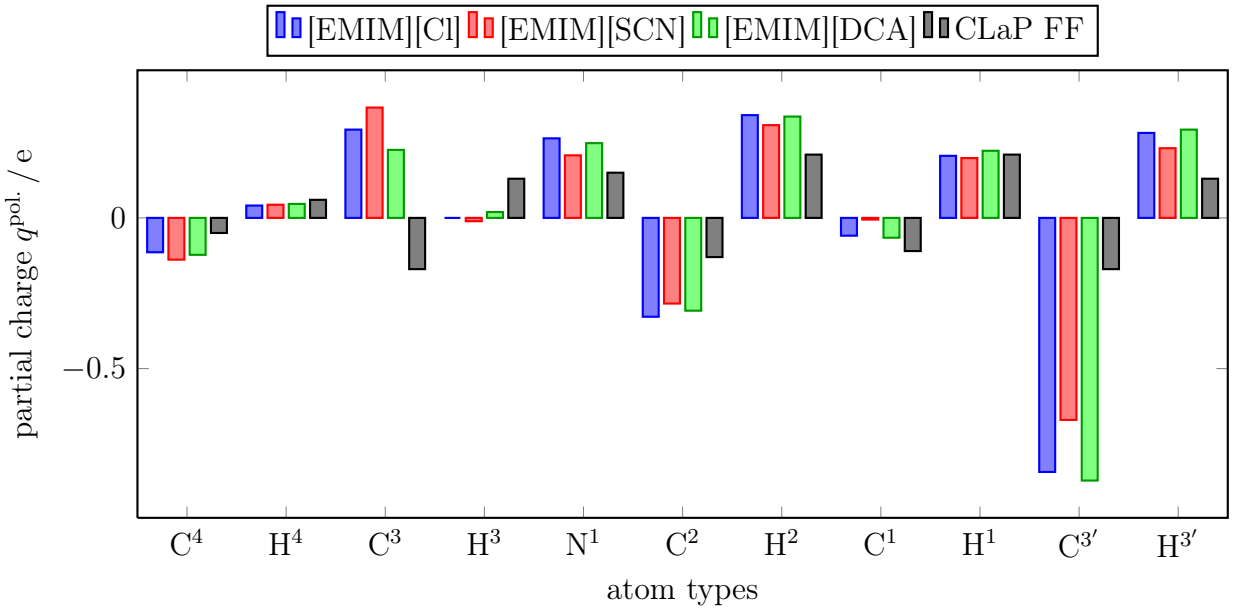


Figure 4.21.: Partial charges $q_i^{\text{pol.}} = q_i/q_{\text{net}}$ in the polarized state for $[\text{EMIM}]^+$ derived by CAB, are compared for different cation–anion combinations and to the results of single ion calculations from the CLaP FF.

contrast, the charge of C^3 for $[\text{BMIM}]^+$–based ionic liquids differs from −0.18 e over −0.10 e to 0.05 e in combination with the anions $[\text{DCA}]^-$, $[\text{Cl}]^-$, and $[\text{SCN}]^-$, respectively. A similar, but smaller fluctuation of the charge for different anions is observed for the upper part of the imidazolium ring, consisting of the nitrogens and C^1, and $C^{3'}$. This highlights the strong delocalization of the ring electrons and the sensitivity of the charges to the actual configuration. Hence, the observed fluctuation is a result of the strong polarizability of the imidazolium ring. For the remainder of the atom types, the partial charges vary only slightly, if different anions are combined with $[\text{BMIM}]^+$. Yet, compared to CLaP, a significant amount of negative charge redistributes from the alkyl chain's carbons C^4 and C^5 to the attached hydrogens H^4 and H^5, respectively. With that the electron cloud extends, which is expected for the liquid phase.

In Figure 4.23, the partial charges $q_i^{\text{pol.}}$ are compared for $[\text{C}_n\text{MIM}][\text{Cl}]$ with $n = 1, 2, 4$. Obviously, the polarization of the ring electrons coincides for the different systems, because the partial charges of the ring atoms and the attached hydrogens agree quite well. However due to increasing size of the cation from $[\text{MMIM}]^+$ to $[\text{BMIM}]^+$, the charge of the atoms attached to the ring differs, but characteristics remain. While all terminal carbon atoms are strongly negatively charged, the other carbons of the alkyl chain carry a positive or relatively small negative charge. Moreover, the charge of the alkyl chain hydrogens is decreased and the electron cloud is delocalized over the length of the side chain. In contrast, the charge transfer mechanism at the $CH^{3'}$ methyl group results in a localization of negative charge at the corresponding carbon, but no redistribution of the charge over the whole group occurs for neither $[\text{BMIM}]^+$, $[\text{EMIM}]^+$, nor $[\text{MMIM}]^+$.

A more global picture of the cationic charge distribution is provided, if the atoms are condensed in reasonable groups, which relieves the identification of the intra-molecular charge-transfer,

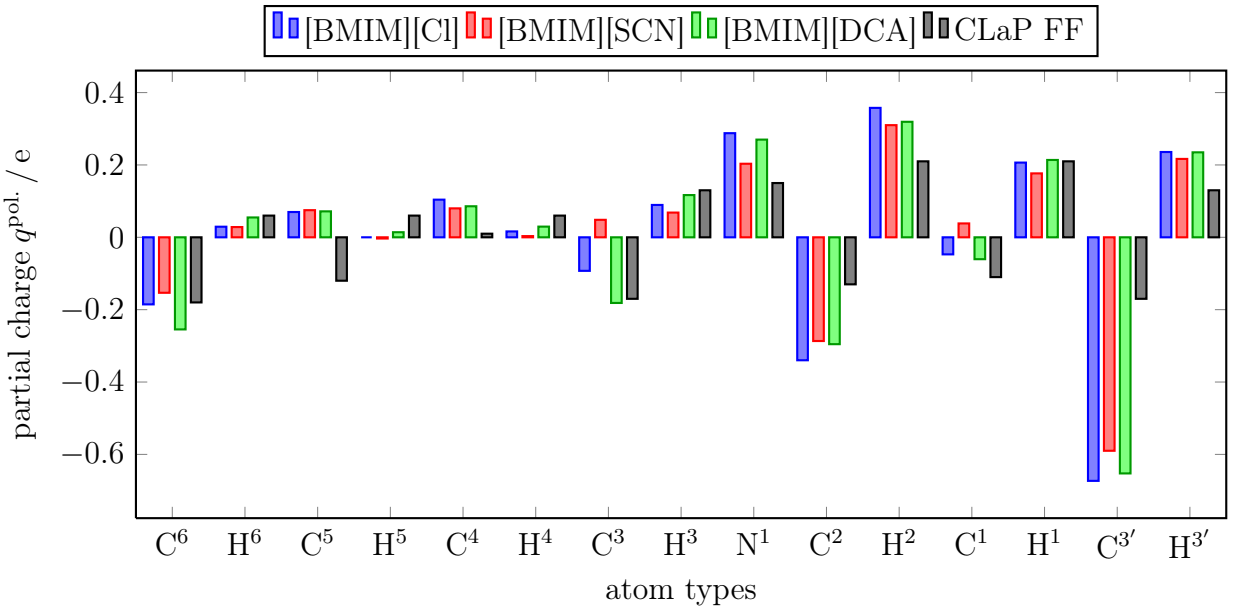


Figure 4.22.: Partial charges $q_i^{\mathrm{pol.}} = q_i/q_{\mathrm{net}}$ in the polarized state for $[\mathrm{BMIM}]^+$ derived by CAB, are compared for different cation–anion combinations to the results of single ion calculations from the CLaP FF.

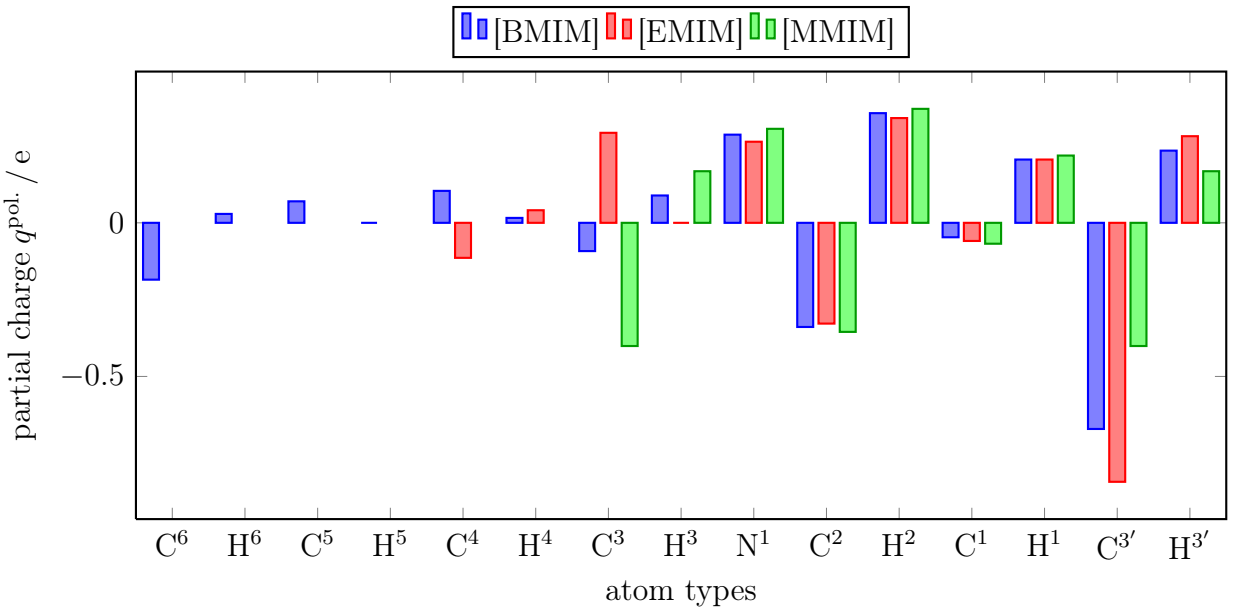


Figure 4.23.: Charge distribution $q_i^{\mathrm{pol.}}$ for a series of imidazolium-chloride based ionic liquids of different cationic alkyl chain length.

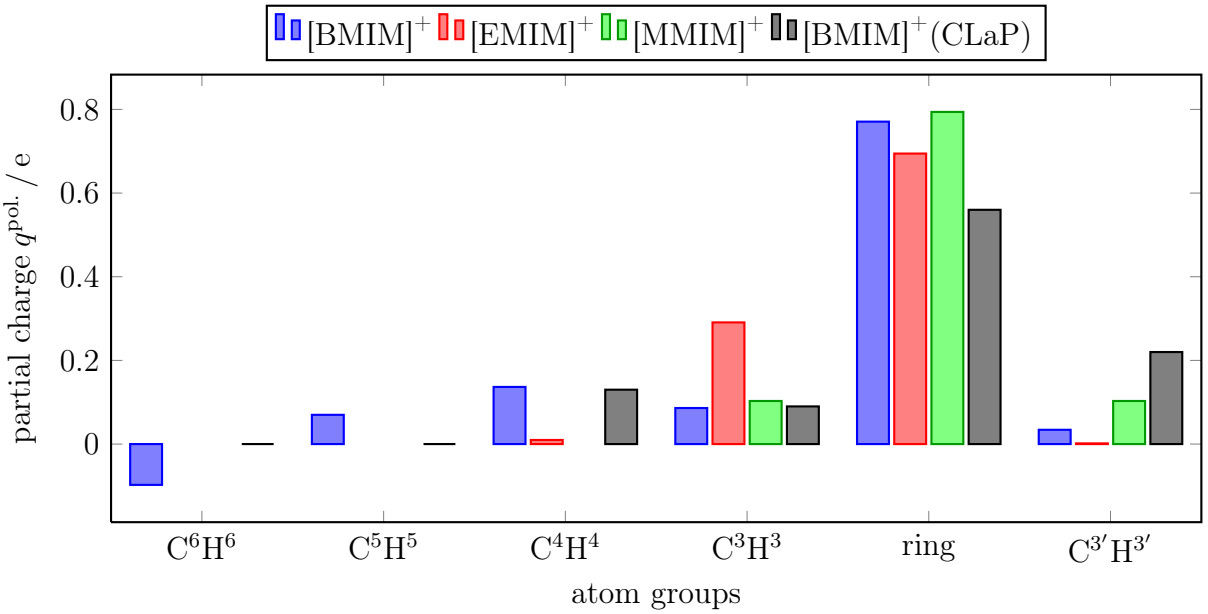


Figure 4.24.: Charge of atom groups for $[C_nMIM][Cl]$ for n=1, 2, and 4.

which occurs, if the charge distribution alters from a gas to a liquid phase state. In Figure 4.24, the Blöchl bulk charges of the alkyl groups and the imidazolium ring are compared to the *in–vacuo* CLaP parameterisation for different side chain length of $[C_nMIM][Cl]$ with n=1, 2, and 4. As expected, all results derived from bulk show the characteristic for a liquid phase, because the charge is more delocalized over whole molecule compared to CLaP, which even assigns a zero net–charge to the atom groups, which are separated by more than two carbons from the ring. Most of the charge is localized around the ring and this amount is quite independent of the alkyl chain length of the cation. However the other atom groups react to the size of the cation sensitively, because the space for a redistribution of charge is increasing. While the charge of $[MMIM]^+$ is symmetric due to the molecular cation structure, the increasing anisotropy of the cations $[EMIM]^+$ and $[BMIM]^+$ introduces an asymmetric charge to the cation. However, since $[EMIM]^+$ is still a rather small molecule, the global charge distribution shows a symmetry with respect to a zero charge of the terminal groups CH^4 and $CH^{3'}$. Hence, $[EMIM]^+$ is separated in two small neutral groups at the termini and two unlike charged and unlike sized domains in between.

This partition is also applicable to $[BMIM]^+$, if the grouping is adapted by considering CH^6 and CH^5 as one terminal domain and the CH^4 and CH^3 groups as the inner chain domain. The resulting charge distribution, provided in Figure 4.25, shows good agreement between the $[BMIM]^+$ and $[EMIM]^+$ domains. Hence, the charge transfer from the imidazolium ring to the long alkyl chain does not increase linearly with chain length, but converges already for $[EMIM]^+$, also observable in the small differences between the corresponding ϵ_{el}. For a side chain length of $n > 2$, also only the inner domains attached to the ring are charged positively, while the terminal domain and the $CH^{3'}$ group do not carry a net–charge. For this reason, it is expected that imidazolium based ILs with $n > 4$ behave similarly, such that only the ring domain and the first two atom groups of the side chain carry the whole net–charge of the ion.

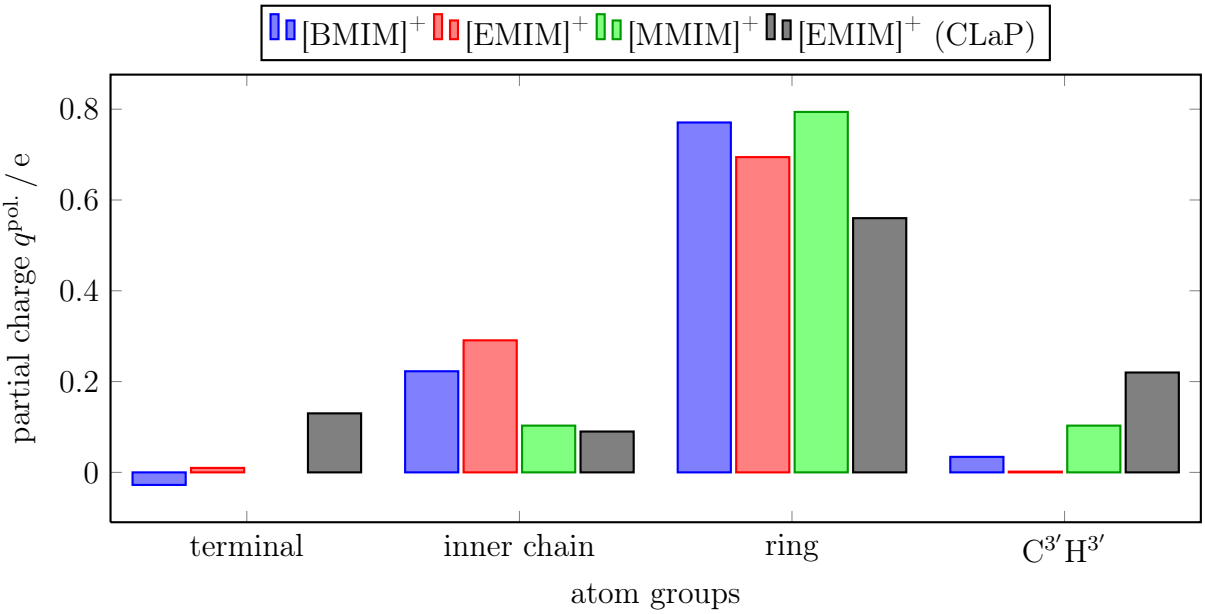


Figure 4.25.: Charge of molecular domains of $[C_nMIM][Cl]$ for n=1, 2, and 4.

To prove that this behaviour is generic for imidazolium–based ILs, the charge distribution of the atomgroups for $[BMIM]^+$–based systems with the anion $[Cl]^-$, $[SCN]^-$, and $[DCA]^-$ is provided in Figure 4.26. Though the charge of [BMIM][SCN] is more strongly localized around the imidazolium ring, the charge of all other atom groups agrees between the different anions. Hence the charging of the domains also remains constant.

Summarizing, the analysis of the cationic charge distribution shows that the polarized state of the cation is only weakly affected by the anion and results from the limited range of accessible configurations. Yet, it is not an ion specific effect. Though differences arise for a varying size of the side chain length, a scheme is observable for all studied ILs. If the alkyl side chain is elongated, only the ring and domains of the long side chain, which are attached to the molecule directly, carry a positive charge. The terminal domain of the chain and the $CH^{3'}$ atom group are neutral or only slightly charged. Hence, it is also expected that this trend continues for imidazolium–based ILs with $n > 4$, such that the additional atom groups in a longer side chain are also uncharged.

To finally prove, that the charge distribution $q^{pol.}$ is not specific for an ion combination, it is necessary that the anionic charge distributions also show good agreement.

In Figure 4.27 the effective charge distributions and the polarized states of $[Cl]^-$, $[SCN]^-$, and $[DCA]^-$ derived by a Blöchl analysis are compared. As already observed for the polarized state of the cationic charges, the charge distribution of the anions is insensitive to the ion combination. If the effective and polarized states of [BMIM][DCA] and [BMIM][SCN] are compared a decrease of the difference between the partial charges for the majority of atoms is observable as shown in Figure 4.28. Hence, the fluctuations in the charge distributions are negligible, which shows that the anionic and cationic charge distribution is not coupled to the ion combination, as expected.

Finally, considering all results which were discussed for the different cation–anion combina-

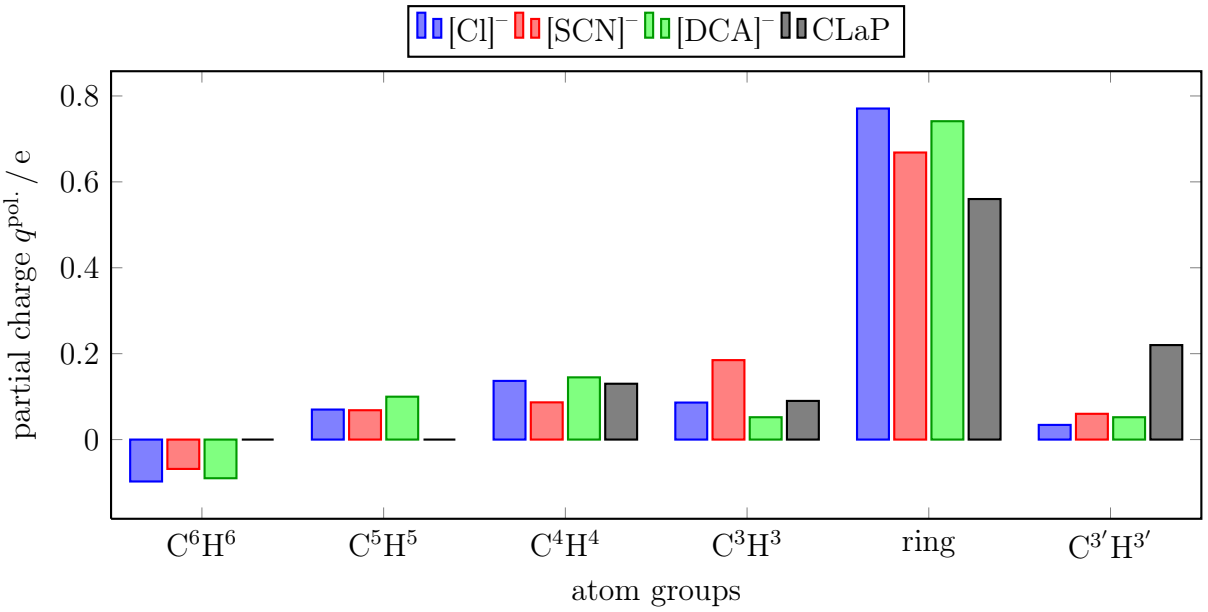


Figure 4.26.: Charge of atom groups for $[BMIM]^+$ combined with different anions.

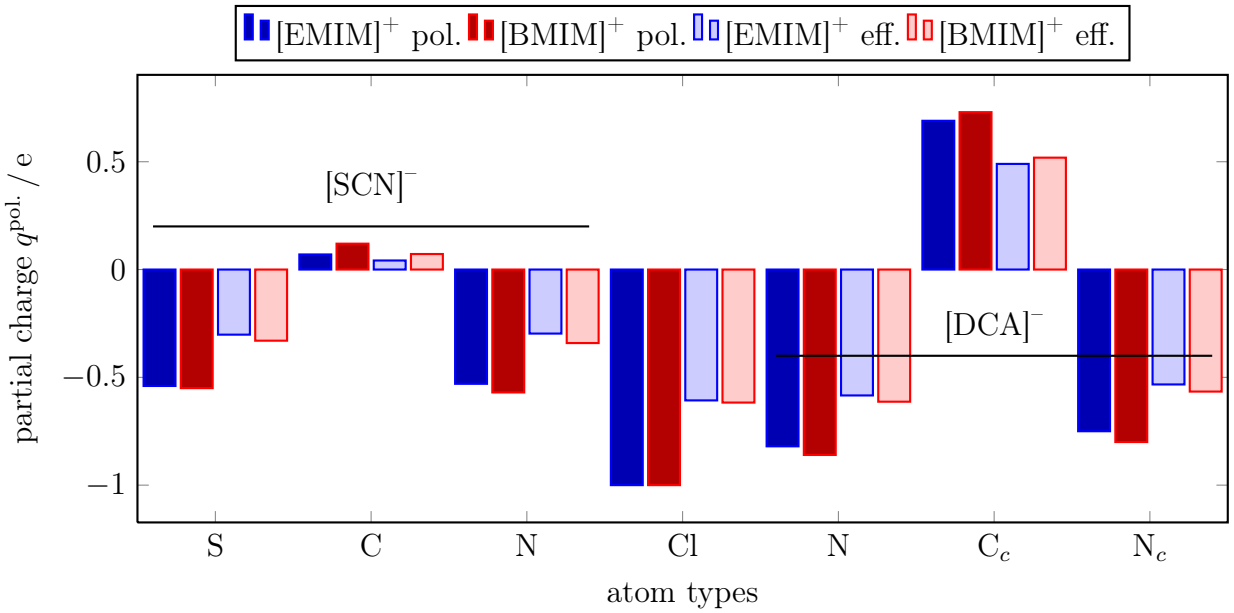


Figure 4.27.: Partial charges of $[Cl]^-$, $[SCN]^-$, and $[DCA]^-$for $[EMIM]^+$ and $[BMIM]^+$ based ionic liquids derived by the Blöchl method under bulk conditions are given for the polarized state (pol.). In addition the effective distribution is illustrated, that includes the polarization effects implicitly (eff.).

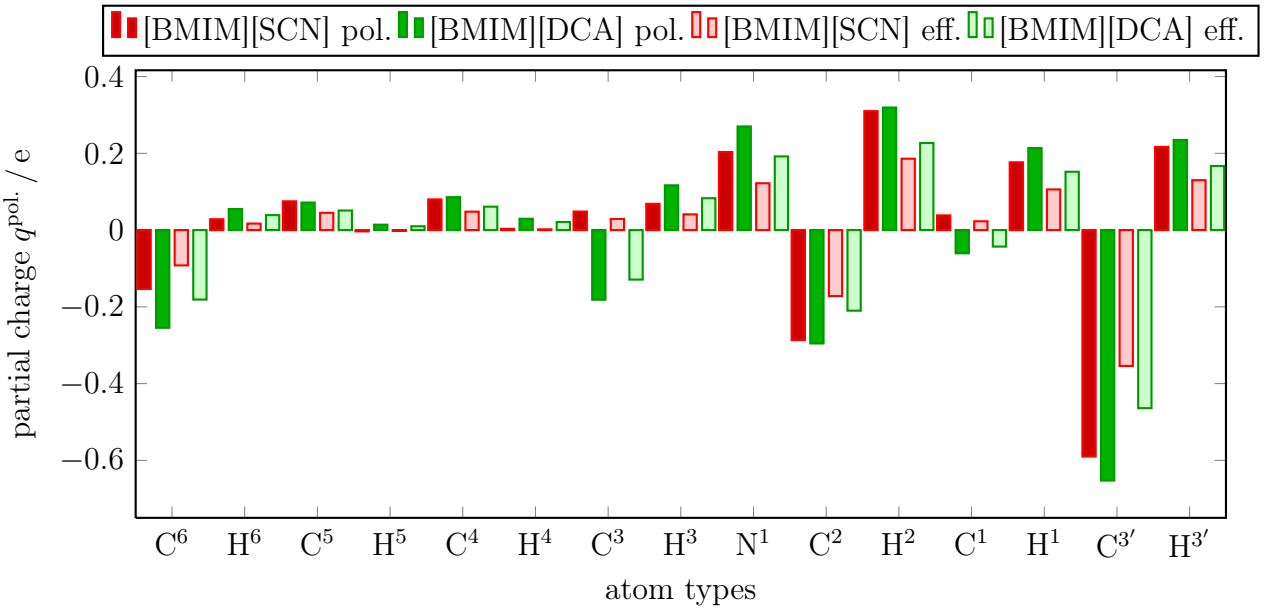


Figure 4.28.: Polarized and effective partial charges q_i, which are derived from bulk configurations, are compared for [BMIM][DCA] and [BMIM][SCN].

tions, it turns out that in the polarized state, the global partial charge distributions are unaffected by the choice of the anion as well as the side chain length of the imidazolium cation. Hence a gateway to a large transferable force field opens, because a consistent construction of generic atom types may allow a reasonable description of a large set of ionic liquids. Moreover it has been shown in section 4.3 that small changes in the partial charge distribution only slightly affect the performance of the FF. Yet, even a framework for a model, which relies on a generic ϵ_{el} could potentially be constructed, a topic which is discussed in the next paragraph.

A transferable electrostatic model for ionic liquids As shown in the preceding paragraphs, the variation of the cationic charge distribution with respect to different anions is small, if the MDEC model is applied in order to implicitly consider polarizability by an electronic dielectric constant ϵ_{el}. The variation of the charge distribution is small, compared to the fluctuations of the partial charges arising from the sensitivity of the actual configuration and the limited accuracy of the XC functional. To counterbalance this deficiencies the cationic partial charges are averaged over a series of different ILs. With that a set of generic partial charges is constructed, such that only the charge reduction is a system-specific parameter, but as observed, not too sensitive within a series of anions. In order to establish a FF for arbitrary chain length, it is necessary to determine the charging of atom groups that are separated more than four sites away from the imidazolium ring. Though these groups are expected to be of hydrophobic character, it has been shown, that the terminal group of the long $[\text{BMIM}]^+$ side chain (SC) still carries a small dipole in the terminal domain (see Figure 4.25). Yet, it is currently not known how much the charge distribution alters with increasing SC length. Though only small or negligible differences are expected, further calculations are needed to clarify this aspect. However, as soon as a convergence of the SC charge distribution is observed, a solid basis for a force field, transferable to arbitrary SC length is created. Inhere, a sufficient

large data set is available for $[EMIM]^+$, with which a procedure to establish a set of generic partial charges is introduced. Finally also a generic set of partial charges for imidazolium based cations is constructed assuming that the alkyl groups in the side chain separated more than four carbons from the ring are completely hydrophobic.

	charge q / e				
FF	CH^4	CH^3	ring	$CH^{3'}$	ϵ_{el}
[EMIM][Cl]	0.01	0.29	0.70	0.00	2.69
[EMIM][DCA]	0.02	0.26	0.71	0.01	2.23
[EMIM][SCN]	-0.01	0.34	0.65	0.02	3.29

Table 4.12.: Charging of the atom groups defined in section 4.4 for different imidazolium based ILs, based on charges not biased by the corresponding electronic dielectric constant ϵ_{el}.

One of the main guidelines for the force field construction is the minimization of atom types, because it significantly reduces the amount of work required for the final short-range parameterisation. For this reason, atom groups with equal amount of charge are identified first. For $[EMIM]^+$ the charging of the atom groups, defined in the previous paragraph, indicates that the charge is mainly located at the imidazolium ring and the first atom group of the ethyl chain, while the terminal groups are apolar domains of the molecule, shown in Table 4.12.

	partial charge q / e				
atom	$[Cl]^-$	$[SCN]^-$	$[DCA]^-$	ILEC	CLaP
C^4	−0.11	−0.14	−0.12	−0.12	−0.05
H^4	0.04	0.04	0.05	0.04	0.06
C^3	0.29	0.36	0.23	0.29	−0.17
H^3	0.00	−0.01	0.02	0.00	0.13
N^1	0.26	0.20	0.25	0.24	0.15
C^2	−0.33	−0.28	−0.31	−0.30	−0.13
H^2	0.34	0.30	0.34	0.33	0.21
C^1	−0.06	−0.01	−0.07	−0.04	−0.11
H^1	0.21	0.19	0.22	0.21	0.21
$C^{3'}$	−0.84	−0.66	−0.87	−0.81	−0.17
$H^{3'}$	0.28	0.23	0.29	0.27	0.13

Table 4.13.: Parameters of CAB charges for $[EMIM]^+$-based ILs and the constructed generic ILEC FF are compared to the CLaP parameterisation.

The charge of the single carbons and hydrogens belonging to different groups varies, disallowing a reduction of atom types. However, as a symmetric ring has been shown to describe the dipole moment distributions, sufficiently accurately, it provides the basis for a further construction of the force field. Therefore by averaging, a symmetric ring is established in the $[EMIM]^+$charge distributions, derived in combination with the anions $[SCN]^-$, $[Cl]^-$, and $[DCA]^-$. In the following steps, the polarized states $q^{pol.} = \sqrt{\epsilon_{el}} q^{eff.}$ of the different ILs have been averaged

and finally a charge of 0.02 e is subtracted from $C^{3'}$. This was required in order to correct for numerical errors and achieve an integer net-charge. The $C^{3'}$ carbon was chosen, because it carries the largest amount of charge, which minimizes the distortion due to the correction. Furthermore, the resulting charge provides an apolar $CH^{3'}$ terminal group, which directly arises for CH^4. This parameterisation is denoted by ILEC, as it it based on an Ionic Liquids Electronic Continuum model and is given in Table 4.13.

In Figure 4.29, the dipole moment distribution is shown for [EMIM][DCA] calculated with WA, CAB charges derived from the corresponding configurations, and the ILEC model. Though the ILEC dipole moment distribution overestimates the average from AIMD, it deviates only slightly from CAB. Yet, apart from the monopoles, the next higher electric moment is modelled, appropriately too.

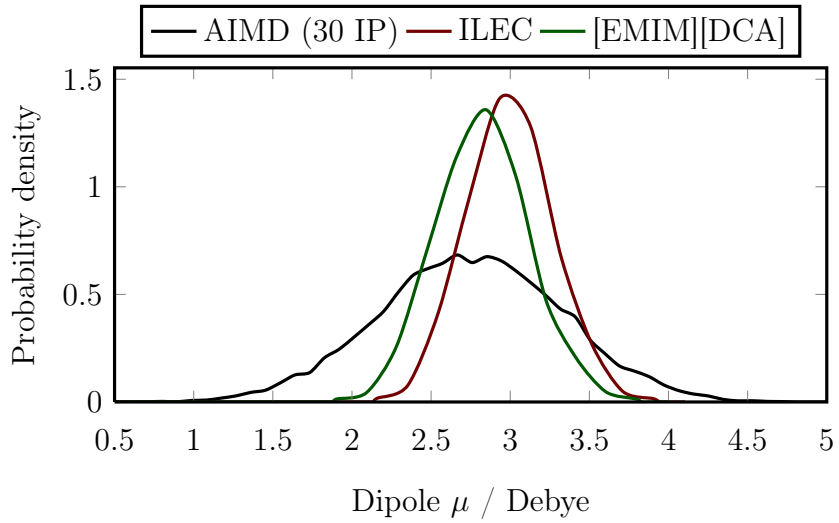


Figure 4.29.: The distribution of the cationic dipole moments in [EMIM][DCA] obtained from WA are compared to result with CAB charges, derived directly for this IL, and the generic ILEC charges.

This good agreement of the dipoles, is a first proof, that ILEC provides the basis for a generic $[EMIM]^+$ force field, transferable within different anions. Moreover, it represents a first validation of the introduced method to construct a generic force field. Hence, it also represents the theoretical background for a treatment of imidazolium-based ILs with longer side chain.

With the present data, a set of partial charges can also be constructed for imidazolium cations with a side chain length $n > 2$. The same procedure is applied as already suggested for $[EMIM]^+$, but to make the model applicable to arbitrary n, the partial charges have to be modified slightly and it has to be assumed, that the alkyl groups separated more than four carbons from the ring have to be neutral. This is achieved by a shift of the positive charge from the C^5 to the C^6 atom and a transfer of 0.01 e from C^4 to C^6. The resulting partial charges are summarized in Table 4.14 and denoted $[C_nMIM]$.

A further step towards an even more generic force field is the application of a common reduced net–charge q_{net}. It is obtained by averaging over all results from the $[EMIM]^+$ and $[BMIM]^+$ based ILs and results in $q_{net} = 0.63$ e which also corresponds to the value for [MMIM][Cl].

atom	[BMIM]/[EMIM]	[C_nMIM]	eff. [C_nMIM]	CLaP
C^6	−0.20	-0.12	-0.09	−0.18
H^6	0.04	0.04	0.03	0.06
C^5	0.07	0.00	0.00	−0.12
H^5	0.00	0.00	0.00	0.06
C^4	0.09/-0.12	0.08	0.05	0.01
H^4	0.02/0.04	0.02	0.01	0.06
C^3	-0.08/0.29	−0.08	−0.05	−0.17
H^3	0.09/0.00	0.09	0.06	0.13
N^1	0.25	0.25	0.16	0.15
C^2	−0.31	−0.31	−0.20	−0.13
H^2	0.33	0.33	0.21	0.21
C^1	−0.02	−0.02	−0.01	−0.11
H^1	0.20	0.20	0.13	0.21
$C^{3'}$	-0.63/-0.82	−0.63	-0.39	−0.17
$H^{3'}$	0.23/0.27	0.23	0.14	0.13

Table 4.14.: A generic set of partial charges for imidazolium–based cations.

Multiplying the partial charges by this generic factor finally gives a generic set of partial charges for arbitrary side chain length and counterions. The result is given in Table 4.14 and denoted effective [C_nMIM]. Thus, a model for the partial charges is present.

Finally, a technique to optimize the remaining force field parameters remains elusive, preventing a consistent and transferable model. For this reason, a suitable computational and theoretical framework is established in chapter 5.

5. Adaption and construction techniques for a force field

As shown for [EMIM][DCA], a drastic change of the partial charges requires an appropriate adaption of the short range interactions, which is a time-consuming and not straightforward task. For this reason, a strategy is desired to minimize this effort, besides the already described parametrization of the partial charges. The major steps for the force field optimization are summarized in Figure 5.1, but the last important stages have not been performed yet. In this chapter, the computational and theoretical methods for this task are developed, such that an application to arbitrary systems is possible.

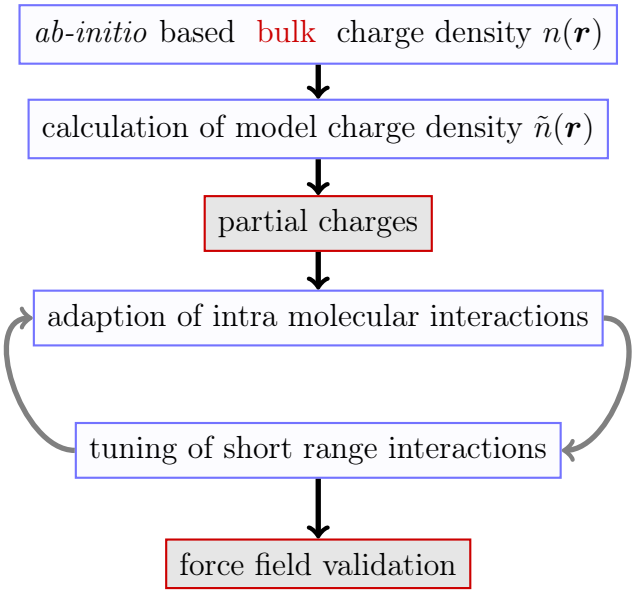


Figure 5.1.: Major steps of the force field parametrization

One of the main goals is to reduce the number of iterations in the final optimization loop, shown by the grey arrows, to a minimum, as this is one of the steps consuming the main amount of computational resources. Therefore the number of atom types has to be kept as small as possible, because the amount of CPU time scales linearly or even quadratically with the number of atom types, depending on the applied optimization method. The bonded and angle potential do not need to be adapted, as they are not affected by a change of the partial charges or LJ parameters and the adaption of the dihedral interactions is a straightforward task. However, more elaborated methods are required to optimize the LJ interactions, which is in this case performed by fitting to reference data. Finally, a reliable validation of the FF is required, which considers different kinds of system properties.

5.1. Adapting dihedral parameters

The force field parameters, describing the interactions for the covalent bonds and the angles between them, are fitted in respect to their ground state modes and hence are independent from the charge distribution, because ground state dominance is assumed. For this reason, these intramolecular force field parameters are independent of the short range parametrization. However, Canongia Lopes and Padua[121] already outlined, that the dihedral potentials describe the interactions of atoms that are separated by exactly three bonds, which is the same separation criterion as for the 1–4 interactions. These are given by the non–bonded interactions between the first and last atom of a dihedral, scaled by a factor ϵ^{1-4} that is determined by the force field. Hence, if a force field defines 1–4 interactions, the overall dihedral profile $E = E^{\text{dih.}} + E^{1-4}$ is affected in case the non-bonded parameters change. Different ϵ^{1-4} have been suggested, as for example $\frac{1}{2}$ and $\frac{1}{1.2}$ for the OPLS[53] and AMBER[82] force field, respectively, but they are based on empirical findings. Finally, the dihedral parameters have to be consistent with the 1–4 interaction and require adaption to maintain the appropriate intramolecular energy landscape in case the non–bonded parameters change.

Although, most of the force fields use a sum of cosine terms with different phase shifts and amplitudes, for the sake of computational efficiency and flexibility Ryckaert-Bellemans (RB) potentials[88] are applied for parametrization of the dihedral interaction:

$$E^{RB}_{dihedral} = \sum_{i=0}^{5} V_i \cos^i(\Psi). \tag{5.1}$$

The dihedral angle Ψ is given in the polymer convention, with $\Psi = 0$ corresponding to a *trans* configuration and the force field potentials V_i for the dihedral interaction have to be adapted in respect to a dihedral profile given by a QM calculation. As the CLaP parameters constitute the framework for the intramolecular parameters of an optimized force field, the dihedral parameters were transferred into a set of RB parameters by applying the theorems for the trigonometric functions. Moreover, the sum of 1–4 and dihedral interactions provides the dihedral profile of the QM calculation, and therewith, the reference energies. If an initial guess for a new force field (BLFF) is set up by inserting Blöchl charges into CLaP, the 1–4 interactions change and the dihedral profiles requires an adaption to the reference energies:

$$E^{\text{BLFF}}_{1-4} + E^{\text{BLFF}}_{dih.} = E^{\text{CLaP}}_{1-4} + E^{\text{CLaP}}_{dih.}, \tag{5.2}$$

$$E^{\text{BLFF}}_{dih.} = E^{\text{CLaP}}_{1-4} + E^{\text{CLaP}}_{dih.} - E^{\text{BLFF}}_{1-4}. \tag{5.3}$$

To this end, a simple least–square minimization for every type of dihedral has been applied to match V_i to the corresponding set of short range interactions. This has been achieved by sampling a certain number of dihedral angles and fitting the derived energies with equation (5.3) to the reference values.[17]

As example, the dihedral profile of $N^1{-}C^1{-}N^{1'}{-}C^{3'}$ given by equation (5.2) is shown on the left hand side of Figure 5.2. While the unfitted potential predicts a more favorable energy for a *cis* configuration of the molecule, the adapted parameters for BLFF match the reference potential given by CLaP, perfectly. Hence the equilibrium geometry of BLFF corresponds to the prediction of QM calculations. On the right hand side of Figure 5.2, the corresponding

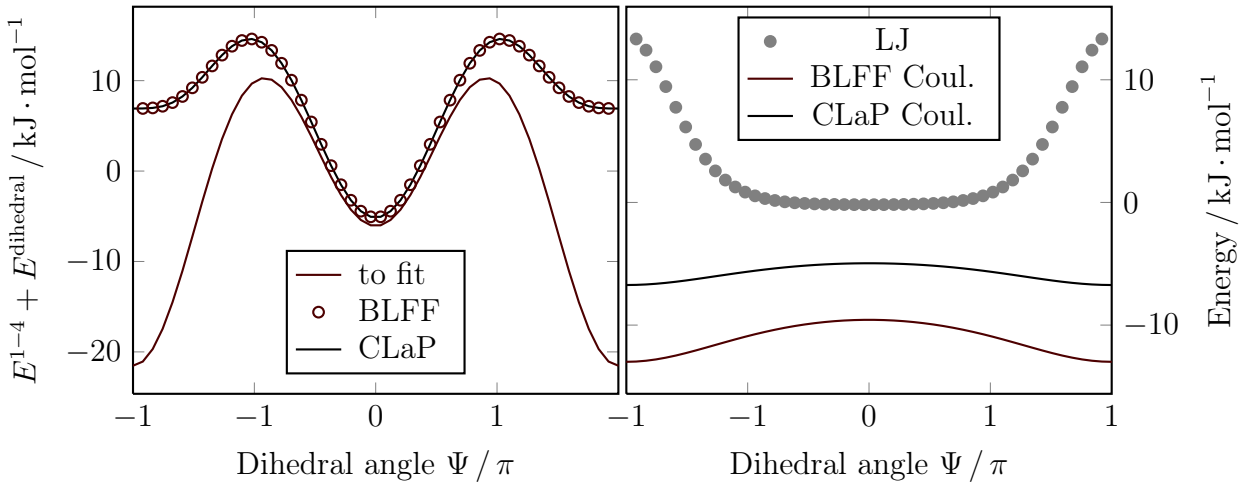


Figure 5.2.: On the left hand side the energy profile E^{dihdral} of the dihedral interactions for $N^1-C^1-N^{1'}-C^{3'}$ including the corresponding 1-4 interactions E^{1-4}. The solid lines depict the reference energies (CLaP, black) and the energies derived from our set of charges and LJ parameters (to fit, red). The fit is plotted in red circles. In the plot on the right hand side, E^{1-4} is decomposed into the electrostatic and LJ contributions, where the latter is equal for BLFF and CLaP.

contributions to the 1–4 interactions are drawn. Although the difference between CLaP and BLFF differs only within a few kJ · mol^{-1}, the overall dihedral profile is affected strongly. Since the 1–4 Coulomb interaction of BLFF is shifted to lower energies, the contribution to the dihedral profile is only positive for $\Psi \sim \pi$, which is the reason for the large difference. Hence, an adaption of the dihedral interactions is necessary to obtain energies able to model the dihedral behaviour predicted by QM calculations.

5.2. Optimization of short range interactions

Due its melting point at 398 K, dimethylimidazolium chloride [MMIM][Cl] it is not a Room Temperature Ionic Liquid, but the chemical compounds are very similar to common RTILs like ethylmethylimidazolium chloride or other imidazolium based ILs with varying side chains and anions. However due to its size, it is perfectly suited to develop and establish a multiscale framework of computional techniques, ranging from the theory of quantum mechanics to classical physics, to study RTILs by classical molecular dynamics simulations. Although there is only one experimental publication available by Fannin *et al.*[6] studying thermodynamic properties of [MMIM][Cl], enough information is contained to tune and discuss the validation of a classical MD force field. The non–bonded short range interactions between the particle types i and j are modeled by a LJ potential:

$$E_{LJ}^{ij} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]. \tag{5.4}$$

Different kinds of physical properties have to be consistently mapped to a 2-parameter pairwise–additive potential, but actually, equation (5.4) depends on 4 parameters, ϵ_i, ϵ_j, σ_i, and σ_j, present for every atom involved in the pair interaction. The parameters for equation (5.4) are derived from a combination rule, that is in this case of Lorentz-Bertelot type, as suggested in the CLaP force field:[5,121,122]

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_i + \sigma_j\right), \tag{5.5}$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}. \tag{5.6}$$

In this section methods for an adaption of this parameters are discussed and depicted on the example [MMIM][Cl].[17]

5.2.1. The conjugate gradient based approach

Though many parameters have to be determined to setup a force field for classical molecular dynamics simulations, the final adaption of the short range interactions modeled by Lennard–Jones (LJ) or Buckingham (BH) potentials is one of most challenging parts. For this reason, a method is desired, which allows an easy determination of any kind of force field parameters. This problem is tackled with a conjugate gradient (CG) approach that is iteratively minimizing an error function ε, derived from certain reference properties. Moreover, the Python module PyPaTEGro is developed, described in Appendix A, that automatically performs the required simulations for the error analysis and provides a new set of parameters. Furthermore, the workload is distributed over many CPUs in order to reduce the amount of real waiting time to a minimum.

Setup of an error function ϵ The use of a CG method requires a suitable choice of an error function $\epsilon(\boldsymbol{p})$, which depends on the fit parameters $\boldsymbol{p} \in \mathbb{R}^n$. It is defined by the mean square error Δ of the simulated data $y(\boldsymbol{p})$ from given reference data y_0:

$$\Delta = \sum_k \Delta_k, \text{ with,} \tag{5.7}$$

$$\Delta_k = \left(y_k(\boldsymbol{p}) - y_{k0}\right)^2. \tag{5.8}$$

A Taylor expansion of the elements $\Delta_k(\boldsymbol{p})$ is suitable to define a set of linear equations that can be solved by a CG method efficiently.

The target set of parameters $\boldsymbol{p}_{target}$ has to define the minimum of ϵ:

$$\boldsymbol{\nabla}_i \Delta_k = 2 \cdot \left(y_k(\boldsymbol{p}) - y_{k0}\right) \left(\frac{\partial y_k}{\partial p_i}\right) = 0 \tag{5.9}$$

Normally only a few points of ϵ are known. For this reason, a Taylor expansion up to first

order is applied to $y_k(\boldsymbol{p})$:

$$y_k(\boldsymbol{p}) = y_k(\boldsymbol{p}_0) + \sum_i \left(\frac{\partial y_k}{\partial p_i}\right)\Big|_{p=p_0} \cdot (p_i - p_0). \tag{5.10}$$

Combining equations (5.9) and (5.10), while assuming $\partial_{p_i} y_k$ is constant within the space of the allowed parameters yields:

$$\boldsymbol{\nabla}_i \Delta_k = 2 \left[\sum_j \left(\frac{\partial y_k}{\partial p_j}\right)\Big|_{p_0} \cdot (p_j - p_0) + (y_k(\boldsymbol{p}_0) - y_{k0})\right] \cdot \left(\frac{\partial y_k}{\partial p_i}\right), \tag{5.11}$$

so that from equation (5.9) we get:

$$\boldsymbol{\nabla}_i \Delta = 2 \left[\sum_j \sum_k \left(\frac{\partial y_k}{\partial p_i}\right)\Big|_{p_0} \left(\frac{\partial y_k}{\partial p_j}\right)\Big|_{p_0} \cdot (p_j - p_0) + (y_k(\boldsymbol{p}_0) - y_{k0}) \left(\frac{\partial y_k}{\partial p_i}\right)\Big|_{p_0}\right] = 0. \tag{5.12}$$

The set of linear equations 5.12 for $i = \{1 \dots n\}$ can be written as $A\boldsymbol{x} + \boldsymbol{b} = 0$ with:

$$a_{ij} = \sum_k \left(\frac{\partial y_k}{\partial p_i}\right)\Big|_{p_0} \left(\frac{\partial y_k}{\partial p_j}\right)\Big|_{p_0} \tag{5.13}$$

$$b_i = \sum_k (y_k(\boldsymbol{p}_0) - y_{k0}) \left(\frac{\partial y_k}{\partial p_i}\right)\Big|_{p_0} \tag{5.14}$$

An alternative error function ϵ_{alt} is defined as:

$$\epsilon_{alt} := \frac{1}{2}\boldsymbol{x}^T A \boldsymbol{x} + \boldsymbol{b}\boldsymbol{x}, \tag{5.15}$$

where $\boldsymbol{x} = \boldsymbol{p} - \boldsymbol{p}_0$. Equation (5.15) has the same derivatives and hence minima as well as maxima as ϵ. This definition of ϵ_{alt} has some nice properties, which are necessary to apply the CG method, as proven in reference[173] (p. 603-607):

- $A_{alt} \in \mathbb{R}^{n \times n}$ is symmetric, and with that,
- $A_{alt} = A$ is semi-positive definite, because it is the Hessian of the quadratic function ϵ_{alt}.

Search of the parameter space $\mathbb{P}^n$ In order to locate the minimum $\boldsymbol{p}_{target}$ of ϵ within $\mathbb{P}^n$ efficiently, some considerations have to be taken into account:

1. Force field parameters, e.g the LJ interactions, only make sense if the parameters are within a certain range, e.g. $p_i \geq 0$ or $\|p_i - p_{i0}\| \leq \delta_i$ for $\boldsymbol{p} = (p_1, \dots, p_n) \in \mathbb{P}^n$.
2. Since the elements of the error function are Taylor expanded, they are valid within a certain range.
3. It is often desirable to weight the different elements of the error function with a certain weighting factor ω_k.

The last point can be straightforwardly incorporated into the approach, because if weighting the different Δ_i with a factor ω_k, equations (5.13) and (5.14) turn into:

$$a_{ij} = \sum_k \left(\frac{\partial y_k}{\partial p_i}\right)\Big|_{p_0} \left(\frac{\partial y_k}{\partial p_j}\right)\Big|_{p_0} \omega_k, \text{ and,} \tag{5.16}$$

$$b_i = \sum_k \left(y_k(\boldsymbol{p}_0) - y_{k0}\right) \left(\frac{\partial y_k}{\partial p_i}\right)\Big|_{p_0} \omega_k. \tag{5.17}$$

The points 1 and 2 can be handled easily, too. In case a parameter p_i changes too drastically its value is fixed corresponding to the boundary conditions. Afterwards a further CG step with the fixed parameters allows to check if the left parameters are able to further reduce the amount of the error function.

Improvement of the sampling Obviously, the accuracy of the introduced method is increased, if the error function Δ is expanded up to second order:

$$\Delta = \Delta|_{x=0} + \boldsymbol{x}^T \, \boldsymbol{\nabla}\Delta|_{x=0} + \frac{1}{2}\boldsymbol{x}_T \, \boldsymbol{\nabla} \left(\boldsymbol{\nabla}\Delta\right)|_{x=0} \boldsymbol{x}, \tag{5.18}$$

$$\boldsymbol{\nabla}\Delta = H\boldsymbol{x} + \boldsymbol{d}, \text{ with } H_{ij} = \frac{\partial^2 \Delta}{\partial x_i \partial x_j}\Big|_{x=0} \text{ and } d_i = \frac{\partial \Delta}{\partial x_i}\Big|_{x=0}. \tag{5.19}$$

The matrix H is symmetric and semi positive definite, too. Hence, the CG method is applicable to solve the equation $\boldsymbol{\nabla}\Delta = 0$. The elements of H and $\boldsymbol{d}$ are obtained from the difference quotients derived from the simulation results, but now besides the first also the second derivatives of Δ are necessary to construct the respective Hessian H:

$$\frac{\partial^2 f(x,y)}{\partial x \partial y} = \frac{1}{h_x h_y}\left[f(x+h, y+h) - f(x+h, y) - f(x, y+h) + f(x,y)\right] \tag{5.20}$$

Obviously, simulations of all combinations of modified parameters N are necessary to construct H with an computational effort $\sim \frac{1}{2}N \cdot (N+3) + 1$:

$$d_i = \frac{\Delta(\dots, h_i, \dots) - \Delta(\dots, -h_i, \dots)}{2h_i}, \tag{5.21}$$

$$H = \begin{cases} \dfrac{\Delta(\dots, h_i, \dots) - 2\Delta(0) + \Delta(\dots, -h_i, \dots)}{h_i^2} & \text{for } (i = j), \\ \dfrac{1}{h_x h_y}\left[\Delta\left(0, \dots, h_i, h_j, \dots, 0\right) - \Delta\left(0, \dots, h_i, \dots, 0\right) - \right. & \text{for } (i \neq j). \\ \qquad \left. -\Delta\left(0, \dots, h_j, \dots, 0\right) + \Delta\left(0\right)\right] & \end{cases} \tag{5.22}$$

Although high computational effort is required compared to the first order method, correlations between the different derivatives are included explicitly.

A method in between is to increase the accuracy of the first derivatives by simulations, which symmetrically changed parameters around the reference topology. With that, the number of

required simulations scales as $2N + 1$ for N parameters to be optimized.

Minimizing an error function for [MMIM][Cl] Assigning point charges to the atoms is quite straightforward, compared to the adaption of the short range interactions. The optimization of LJ parameters, by using the error function Δ, given by equation (5.15), is the most tedious part, because a numerous simulations have to be set up and performed in order to sample the parameter space $\mathbb{P}^n$. The number of simulations depends on the number of temperatures n^T involved in the error function and the number of parameters n that should be tuned. Finally also the accuracy of derivatives plays a role. A second-order finite differences schemes is applied[173], where the derivative of a function $f(x)$ is obtained by:

$$\frac{\partial f}{\partial x} = \frac{f(x+h) - f(x-h)}{2h}. \tag{5.23}$$

Thus for every iteration step $(2 \cdot n+1)n^T$ simulations are required to obtain the derivatives and the error function at the expansion point. Since the FF shall be applicable to a certain range of temperatures, corresponding information has to be involved, but the required computation time should be kept at a minimum. To this end, only target properties are chosen, which require a minimal sampling of the phase space, such that most of the computation time is spent for equilibration.

In order to match the temperature dependence of the parameters, the experimental mass densities[6] at $T = 425$, 440, and 465 K are used as target properties. Hence, the respective elements of the error function, defined by equation (5.7), are:

$$y_{00} = \rho(425\text{K}) = 1123.4 \text{ kg} \cdot \text{m}^{-3}, \tag{5.24}$$
$$y_{10} = \rho(440\text{K}) = 1115.0 \text{ kg} \cdot \text{m}^{-3}, \tag{5.25}$$
$$\text{and, } y_{20} = \rho(465\text{K}) = 1101.1 \text{ kg} \cdot \text{m}^{-3}. \tag{5.26}$$

Furthermore, the weighting parameters $\omega_i = 1/y_{i0}^2$ are set to the square of the inverse reference density y_{i0}, for the establishment of a relative and unit–less error.

Information about the local structure is provided by the radial distribution functions (RDFs) between the hydrogens H^1, H^2, and H^3 and the chloride anion, given by AIMD simulations[7]. For this case, the mean square error of the relative deviation from the reference RDF $g^{\text{ref}}(r)$ is considered as corresponding term Δ_i in the error function equation (5.15):

$$\Delta_i = \frac{1}{N} \sum_{i=1}^{N} \delta g(r_i), \text{ with} \tag{5.27}$$

$$\delta g(r_i) = \begin{cases} \left(\frac{g(r_i)}{g^{\text{ref}}(r_i)} - 1.0 \right)^2 & \text{if } g^{\text{ref}}(r_i) > 0, \\ \left(\frac{g(r_i)}{g^{\text{ref}}_{\text{min}}} - 1.0 \right)^2 & \text{if } g^{\text{ref}}(r_i) = 0. \end{cases} \tag{5.28}$$

To allow a normalization of the single errors in case the reference value is zero, the smallest non-zero element $g^{\text{ref}}_{\text{min}}$ of the RDF is applied. Depending on the location of the maxima, equation (5.27) is applied to different ranges of the RDFs. Although all calculations of the MSE start at 2 Å, the end points differ within 4 Å, 6 Å, and 8 Å for H^1-Cl, H^3-Cl, and

H^2-Cl, respectively. All errors of the RDFs are weighted equally by corresponding weight factors $\omega_i = 1$.

Hence, the error function Δ carries no unit and all target properties are weighted equally, because they all have a relative measure. As the minimization procedure is guided by the matrix A and the vector $\boldsymbol{b}$, the partial derivatives of the error function terms are necessary. To this end, a second–order finite differences scheme is applied,[173] where the errors for different sets of parameters were derived by corresponding simulations. So for every minimization step, one simulation is performed with the set of start parameters $\boldsymbol{p}_0 \in \mathbb{P}^n$, where n is the number of parameters to be tuned. The number n varies at the different stages of tuning and is given in the next paragraph for every tuning stage. Additionally, $2\,n$ sets of simulations with slightly decreased and increased parameters provide the necessary data for the calculation of the partial derivatives. During the minimization procedure the step width δ_i for the finite–differences is adapted in respect to the change of the parameters p_i. If the minimization predicts a change of p_i larger than $2\delta_i$, δ_i is decreased by a factor of 0.5. For a change of p_i smaller than δ_i, δ_i is increased by a factor of 2 for the next set of simulations. This criteria are applied, to construct a balance between the accuracy of the derivatives and to reduce the risk to get trapped in a local minima. To assure, that the applied Taylor expansion is valid, the changes of the parameters are bound to a value of $4\delta_i$. Since the different simulations are independent, the iteration procedure runs massively parallel and the tuning time is dominated by the number of accessible processors. This parallelization allows to parametrize the short range interactions within days or weeks which is an acceptable time for this kind of process.

For the initial guess, the short range parameters and bonded parameters from CLaP have been chosen. Partial charges have obtained from a Blöchl analysis[11], and finally, a refit of the dihedral parameters, as described in section 5.1, yields the starting parametrization BLFF. The optimization of short–range interactions is separated in four major steps. At first, only the parameters of the hydrogens have been tuned until the error converged, so $n = 6$. In the second step, the tuning was restricted to the parameters of the chloride with $n = 2$, but after one iteration step, no decrease of the error has been observed, and reference properties were reduced to the mass densities and the H^1-Cl RDF. As, the resulting deviation of the mass densities were still in the order of several percent, the target properties were restricted to the mass densities and only the parameters of chloride were optimized. After two iteration steps a maximal error of 1 % is achieved for the mass densities. In the last part of the optimization procedure, the parameters of H^1 and Cl are tuned, thus $n = 4$, and the full set of target properties is considered. With this setup, the error nearly converged after three steps and the result is chosen as the tuned force field BTFF with Blöchl charges. In every iteration step the dihedral parameters were adapted to the corresponding 1–4 interactions.

In order to allow an efficient optimization, the Python[133] module PyPaTEGRO has been implemented, which is described in Appendix A detailedly. The package acts as an interface between the user and GROMACS, which allows to setup and perform the simulations for the error minimization. Moreover, the package includes several scripts to analyze important thermodynamic properties, such that it is very useful for non-optimization studies, too. Its strength is especially, the distribution of the workload and that the whole simulation process from the equilibration to the analysis is performed by PyPaTEGRO with a single submission command to the queue of a supercomputer.

Target properties: mass density and radial distribution functions In order to optimize the performance of the force field, the LJ parameters were adapted to minimize the error of the mass density ρ and different radial distribution functions (RDFs). The results show, that the optimization routine works and is robust, also outside the regime of tuning. However, a perfect adaption to all target properties is quite time consuming, especially for the three RDFs between the hydrogens and the choride, and is not performed, here. This task will be necessary if a generic charge model has been established, and the LJ parameters have to be adapted to this generic set of charges.

A comparison of the mass density ρ for the temperatures T, which were used for the parameter tuning, is shown in 5.3. The finally tuned force field BTFF shows very good agreement with the experimental data and is also able to recover the trend over a temperature range of 50 K quite well. This exhibits the reliability of the proposed approach, because the mass densities simulated with the initial parametriation BLFF are far off the experimental values. Hence, obviously a good agreement with ρ is achieved easily over the range of temperatures, that were used for the parameter optimization. However, compared to the results of CLaP, which predicts the course of ρ with T quite well, a stronger decrease in ρ is observed for increasing T.

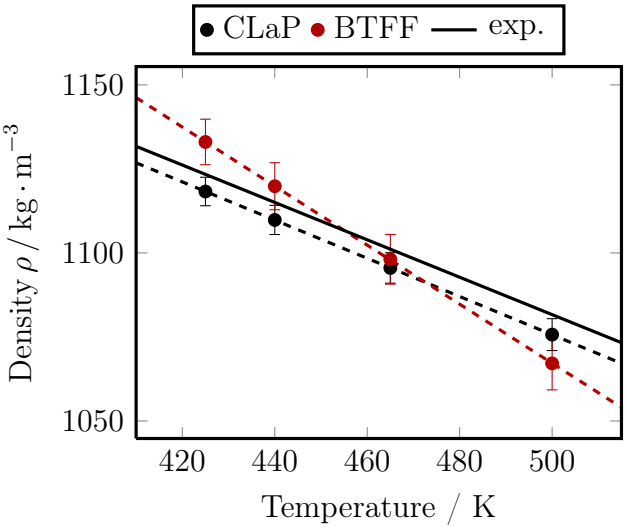


Figure 5.3.: Density of mass obtained with different FF parametrizations and by experiment[6].

In 5.4, the RDF for different stages of the parameter adaption are shown. While the results of the simulations with CLaP deviate from the AIMD results, significantly, the parameter optimization improves the agreement between the RDFs from the classical MD and the AIMD. Hence the shape and, especially, the location of the maxima of the RDFs given by BTFF agree with the AIMD ones. However, compared to the error of the mass density, the deviation of the RDFs is still large.

Yet, the optimization procedure allowed to derive LJ parameters, that model the mass density within an error of maximal 1 % in only a few iteration steps. The RDFs are not matched optimally compared to ρ, but a considerable improvement is achieved. However, the location of the first maximum of the H^1-Cl RDF is expected to play an important role for the dynamics

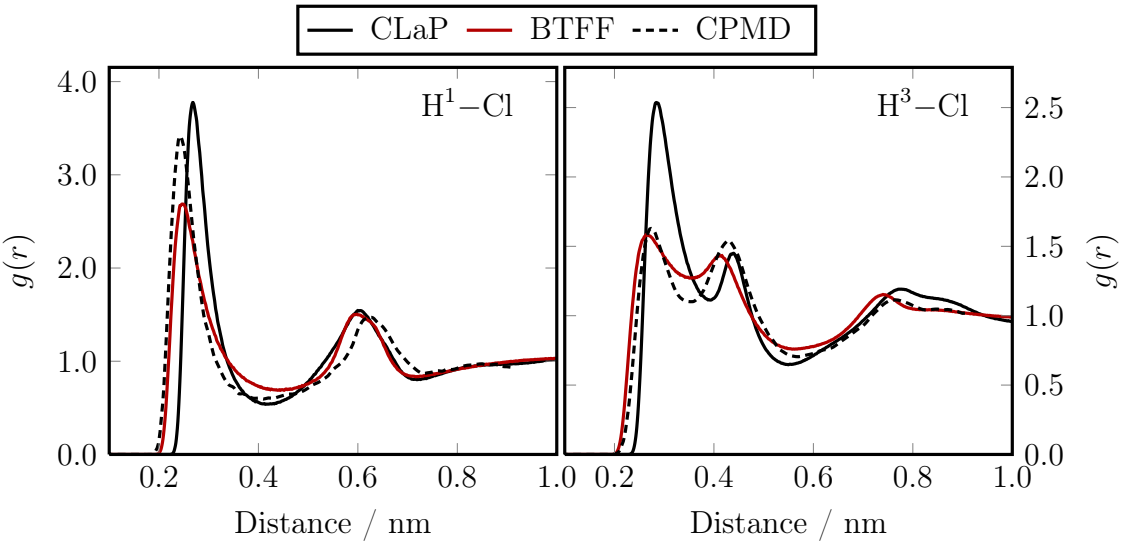


Figure 5.4.: The radial distribution functions (RDFs) derived from simulations with the CLaP parametrization and the tuned FF BTFF are compared to the reference data from AIMD simulations[7].

of the system[126,127,168,174]. For this reason, it is of certain interest, how the dynamic properties given by a certain FF compare to the experimental measurement, in respect to the deviation of the H^1-Cl RDF of the FF from the reference RDF. To this end, a benchmark of CLaP and BTFF is described in the following section.

5.3. Validation of the derived parameters

So far, the procedure for the tuning of a force field has been described. In the following paragraphs, the reliability of the parametrization is discussed by the study of properties that have not been used as targets for the parametrization. Energetic as well as different dynamic quantities are derived at temperatures $T = 425$, 440, 465, and 500 K to dissect the performance of the tuned force field BTFF, also outside the tuning regime, and detect insufficiencies.[17]

Dynamic properties at different stages of tuning Since no dynamical properties were involved in the parameter tuning procedure, a validation of the tuned force field in terms of dynamical properties is inevitable. Unfortunately, experimental data for [MMIM][Cl] is rare and the static conductivity is the only available dynamical property, that has been measured[6]. However, it allows to study the collective system dynamics. Though many possibilities exist to calculate the conductivity σ from an MD simulation, a careful choice is inevitable to allow comparison to experiment. For instance, the Nernst-Einstein method allows to calculate σ in terms of the diffusion constants of the cations and anions, but it completely neglects mutual correlations, which are included in experimental measurements. However, as the Nernst-Einstein method relies on single particle properties, the computational effort is quite small compared to correlation including methods. Therefore, it is applied to obtain a first rough

approximation. Moreover, classical MD simulation studies have been published, that discuss the diffusion of [MMIM][Cl][129,153], which allows at least comparison of our results to different parametrization methods. To consider the mutual correlation of the ions, the Einstein-Helfand (EH) method is applied to derive the static conductivity from the mean square displacement of the translational dipole moment $\boldsymbol{M}_\mathrm{J}$:

$$\boldsymbol{M}_\mathrm{J} = \sum_{i=1}^{N} q_i^\mathrm{net} \cdot \boldsymbol{x}_i^\mathrm{com}, \tag{5.29}$$

where q_i^net and $\boldsymbol{x}_i^\mathrm{com}$ are the net charge and the center of mass of the ion, respectively, and N is the number of ions. At this point it has to be clarified, if the unscreened or effective charges have to be applied for the calculation of the conductivity. The conductivity is defined via the current response $\boldsymbol{j}$ to an electric field $\boldsymbol{E}$:

$$\boldsymbol{j} = \sigma \boldsymbol{E}. \tag{5.30}$$

As $\boldsymbol{j} = \boldsymbol{j}_\mathrm{n} + \boldsymbol{j}_\mathrm{e}$ decomposes into a nuclear $\boldsymbol{j}_\mathrm{n}$ and an electronic contribution $\boldsymbol{j}_\mathrm{e}$, the static limit obeys a special case. While the atomic nuclear contribution arises from the movement of the nucleus carrying effective charges $\boldsymbol{j}_\mathrm{n} = q^\mathrm{eff}\boldsymbol{v}_\mathrm{n}$, $\boldsymbol{j}_\mathrm{e} = q^\mathrm{el}\boldsymbol{v}_\mathrm{e}$ is the contribution of the constantly polarized electron cloud. For a static $\boldsymbol{E}$ field the Born-Oppenheimer approximation predicts an instantaneous reaction of the electron cloud to the configuration of the nuclei. Yet, the electronic part of the current for the whole system vanishes, because the steadily polarized charge q^el does not move compared to the nuclei. Hence for a calculation of the static conductivity, effective charges have to be applied, which constitute a model with a reduced net-charge. However, if a frequency dependent spectrum is of interest, static charges are not applicable, because they are not able to include the characteristics of the spectrum arising due to the time-dependency of $\boldsymbol{j}_\mathrm{e}$.[175]

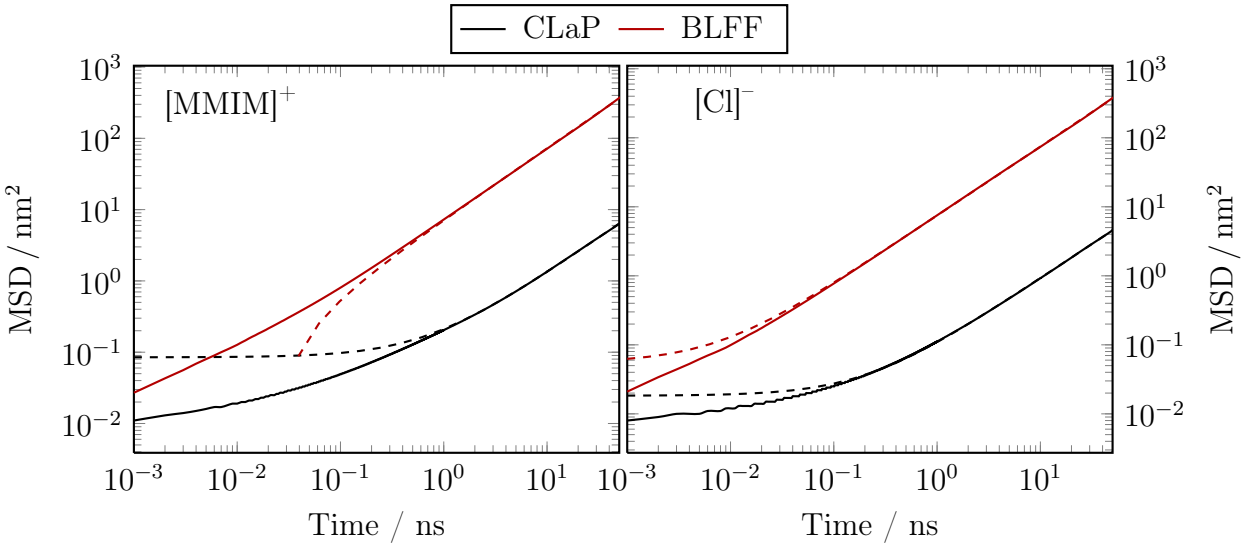


Figure 5.5.: Mean square displacement of the ion coordinates of the 239 ion pairs system at $T = 425\,\mathrm{K}$.

In order to apply the NE method, at first the Einstein relation has been used to calculate the diffusion constants $D^{\pm}$ from a linear fit to the mean square displacement (MSD) of $\boldsymbol{x}_i^{\mathrm{com}}$:

$$\lim_{t\to\infty} \frac{\left\langle [\boldsymbol{x}(t) - \boldsymbol{x}(0)]^2 \right\rangle}{t} = 6 \cdot D, \tag{5.31}$$

as shown in Figure 5.5. In Table 5.1, the results for the simulations at $T = 425\,\mathrm{K}$ are summarized to show the sensitivity of the dynamics to the force field parameters. BTFF, considered to be the most accurate force field, predicts diffusion constants which are two orders of magnitude largen than CLaP. However, if the diffusion is too fast or not is hard to decide, because published force fields give different results and no experimental data is available. While the UA10 united-atom model of Liu *et al.*[153] with a reduced net charge of 0.8 e predicts roughly six times smaller values, the model of Youngs and Hardacre[129] with a reduced net charge of 0.6 e, gives comparable but still smaller values as BTFF. As the H^1 hydrogen is known to play an important role for the dynamics of ILs[126,127,168,174], a reduction of the diffusivity in BTFF might occur if the corresponding RDFs is matched more accurately. This increases the probability for hydrogen bonding and, therewith, the interaction between the ions. The obviously fast diffusion might be an artefact of the smaller height of the first H^1-Cl RDF maximum compared to the AIMD reference.

A significant difference between CLaP and the FFs including Blöchl charges is the ratio γ between the diffusion constants of the cations and anions, $\gamma = D^+/D^-$. While CLaP predicts $\gamma > 1$, BTFF shows $\gamma \sim 1$. This has also been observed for [MMIM][Cl] by Youngs and Hardacre[129] while studying models differing in the scaling of the ion net charge.[129] There, all models with a net charge smaller than ± 0.7 e gave a ratio of the diffusion constants of $\gamma < 1$. In terms of Stoke's law, the faster diffusion of the anions can be understood from their smaller hydrodynamic radii which account for the interaction range of particles within their environment. Unfortunately, no experimental data for diffusion is available for [MMIM][Cl]. However, experimental measurements of other ILs, with an approximately spherical anion like [EMIM][BF_4][141] or [BMIM][PF_6][176], give ratios of $\gamma = 1.24$ or $\gamma = 1.34$, and therefore suggest a larger hydrodynamic radius for the anion compared to the cation. Though the anions' hydrodynamic radii were calculated to be smaller than the cations' ones[177]. Hence, Stoke's law is not applicable. Due to the stronger electrostatic interaction compared to reduced charge models, CLaP tends to model a too rigid hydrogen bond network, which causes the slow dynamics of CLaP, that is also observed for other ILs, like [EMIM][BF_4][10]. In case of decreased intermolecular interaction given by scaled net charge parametrizations as BTFF, or in the studies by Youngs and Hardacre[129], Stoke's law is valid as it shows slower diffusion for larger ions. So, though [MMIM][Cl] consists of a molecular, but symmetric, cation and a symmetric atomic anion, it seems that it can be considered as a simple liquid in terms of Stoke's law.

To derive the conductivity σ_{NE} from the diffusion coefficients, the Nernst-Einstein equation has been applied:

$$\sigma_{\mathrm{NE}} = \frac{Nq^2}{Vk_{\mathrm{B}}T}\left(D^+ + D^-\right), \tag{5.32}$$

with the volume V, the temperature T, the number of ion pairs N, the effective net charge of the ions q, and the Boltzmann constant k_{B}. The conductivities, derived by equation (5.32) with the diffusion constants obtained at $T = 425\,\mathrm{K}$, are summarized in Table 5.2. While CLaP

FF	D / 10^{-5} cm$^2 \cdot$ s^{-1} D^+	D^-	γ
CLaP	0.021	0.015	1.40
BTFF	1.208	1.240	0.98

Table 5.1.: Diffusion constants and their ratio $\gamma = D^+ / D^-$ for the different parametrizations derived from simulations at temperature $T = 425\,\mathrm{K}$.

underestimates the experimental conductivity[6] of $10.65\,\mathrm{S \cdot m^{-1}}$ by an order of magnitude, BTFF overestimates the value from experiment strongly. However, these numbers do not include the collective contribution of the ions' mutual correlations to the conductivity.

FF	σ / S$\cdot$m^{-1} σ_{NE}	σ_{EH}	κ
CLaP	0.80	0.67	0.84
BTFF	21.94	14.68	0.67

Table 5.2.: Conductivities at $T = 425\,\mathrm{K}$ calculated by the correlation neglecting Nernst Einstein approach and the correlation including Einstein Helfand technique are summarized, as well as the ratio $\kappa = \sigma^{\mathrm{EH}}/\sigma^{\mathrm{NE}}$.

Hence, an appropriate method such as the Green-Kubo approach being based on integration of the current autocorrelation function is necessary. Unfortunately, very long simulations are necessary to gain reliable results, as shown in section 3.5. But an equivalent technique reducing the required amount of simulation time significantly is available. Corresponding to the Einstein method for the diffusion constant, the Einstein-Helfand[147] (EH) technique allows to determine σ^{EH} from the slope of the mean square displacement of the translational dipole moment $\boldsymbol{M}_{\mathrm{J}} = \sum_{i=1}^{N} q_i \cdot \boldsymbol{x}_i^{\mathrm{com}}$ involving the sum over all center of mass coordinates $\boldsymbol{x}_i^{\mathrm{com}}$ of the N ions with an effective net-charge q_i:

$$\lim_{t\to\infty} \langle [\boldsymbol{M}_{\mathrm{J}}(t) - \boldsymbol{M}_{\mathrm{J}}(0)]^2 \rangle = \lim_{t\to\infty} \langle \Delta \boldsymbol{M}_{\mathrm{J}} \rangle = 6 V k_{\mathrm{B}} T \sigma t. \tag{5.33}$$

In other words, equation (5.33) is the integrated version of the Green-Kubo relation for the conductivity, such as the Einstein method for the diffusion constant. The EH method allows to consider the collective current of the system and to derive an accurate value for the conductivity with much less statistics than required if the current autocorrelation function is integrated explicitly.

Unfolding of the ion coordinates is necessary for the calculation of $\Delta \boldsymbol{M}_{\mathrm{J}}$ to avoid discontinuities caused by PBC. In Figure 5.6, $\Delta \boldsymbol{M}_{\mathrm{J}}$ is given for different parametrizations at $T = 425\,\mathrm{K}$. While the linear regime starts around 50 ps for BTFF and BLFF, approximately 100 ps are necessary to reach the linear regime for CLaP. A straight line is fitted to the $\Delta \boldsymbol{M}_{\mathrm{J}}$ from the beginning of the linear regime up to 500 ps, and the resulting conductivities are summarized in Table 5.2. Though the same trends arise compared to the Nernst-Einstein method, EH provides consistently smaller conductivities, but BTFF predicts still a slightly too large value. Considering not only the difference, but also the ratio between the conductivities

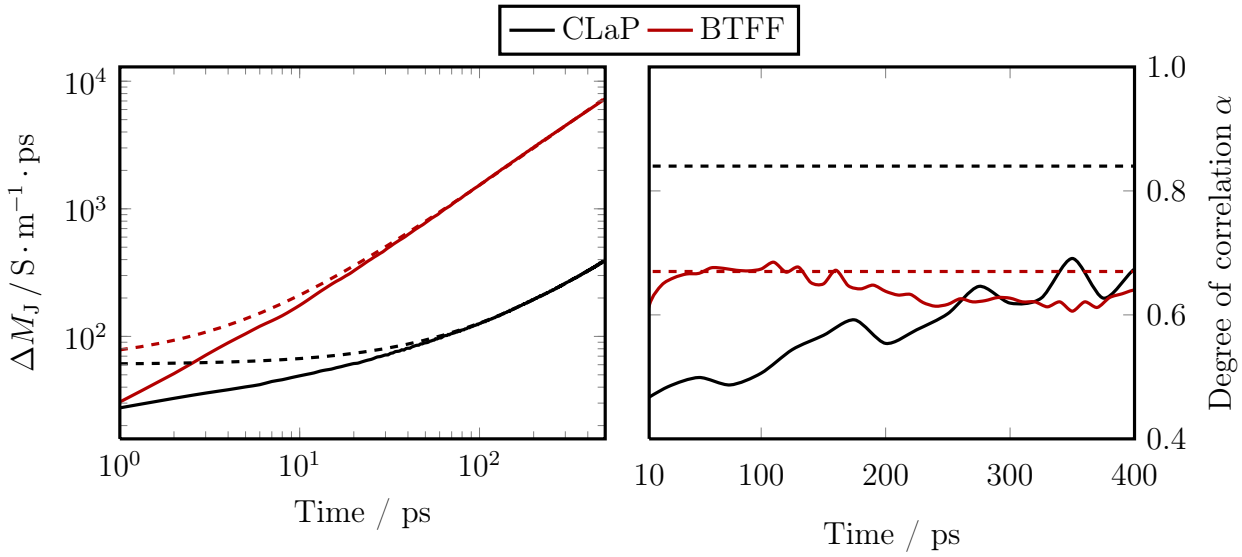


Figure 5.6.: Mean square displacement of the translational dipole moment M_J as a function of time t on a double-logarithmic scale for different FF parametrizations at T = 425 K and time-dependent correlation factor α as given by eq. 5.34.

$\kappa = \sigma_{EH}/\sigma_{NE}$, two levels of correlation κ are identified. While the force field with Blöchl charges, BTFF, predict $\kappa \approx 0.7$, a larger value of $\kappa \approx 0.8$ is observed for CLaP. Borodin[149] suggested a similar, but time dependent correlation $\alpha(t)$:

$$\alpha(t) = \frac{\sigma^{\mathrm{EH}}(t)}{\sigma^{\mathrm{NE}}(t)}, \tag{5.34}$$

where $\sigma^{\mathrm{EH}}(t)$ and $\sigma^{\mathrm{NE}}(t)$ are derived from linear fits with a time span of 10 ps, and hence, $\lim_{t\to\infty} \alpha = \kappa$. In their polarizable model for ILs[149,150], α converges already within the sub-diffusive regime of $\Delta \boldsymbol{M}_J$. This finally allows to obtain an accurate prediction of the correlation including conductivity with less statistics than required if EH is applied only to the linear regime of $\Delta \boldsymbol{M}_J$.

On the right hand side of Figure 5.6, α is plotted for CLaP and BTFF at T = 425 K to judge if similar procedure can be applied here. As expected, the fast dynamics of BTFF is reflected in a corresponding fast convergence behaviour. For CLaP, the convergence is not observable with the present statistics of $\Delta \boldsymbol{M}_J$. However, for BTFF, α converges before 200 ps and agrees with the corresponding long time limit κ. Hence, as polarizable models,[149,150] BTFF also predicts this fast convergence of α. This is an important fact, because it is related for example to the "caging" in ILs.[123] Finally, though BTFF overestimates σ also with EH, the value of κ and the quite fast convergence of α is comparable to polarizable models.[149,150]

Yet, although the optimized force field BTFF predicts the density in the temperature range from 425 to 500 K accurately, the conductivity is overestimated at T = 425 K. As a proper modeling of hydrogen bonding strongly influences the dynamics, and the important role of the acidic H^1 hydrogen is well known,[126,127,168,174] a more accurate matching of the RDF between the hydrogens and the anion seems to be required for an appropriate modelling of the ionic

liquid. However, compared to a polarizable model the qualitative behaviour of the dynamics is properly predicted by the classical non-polarizable force field.

Temperature dependence of dynamical properties On the one hand side, it has been shown, that the mass density is predicted very accurate for within temperature range, which has been used during the parameter optimization. However, on the other hand side, it has been observed, that the decrease of density with increasing temperature is overestimated. Hence, it is of importance to validate the temperature dependence of the dynamics. To this end, the diffusion and conductivity for CLaP and BTFF are compared at temperatures $T = 425$, 440, 465, and 500 K. Since ionic liquids are known to behave like supercooled liquids, the temperature dependence of the dynamical properties $P(T)$ is assumed to be reflected by the empirical Vogel-Fulcher-Tamman equation[164]:

$$\log P(T) = \log A + \left(\frac{B}{T - T_{\mathrm{VTF}}} \right). \tag{5.35}$$

While the parameters A and B are always temperature-independent, T_{VTF} can be handled in two different ways. If the temperature range near the glass-transition temperature T_g is considered, T_{VTF} is related to T_g via the fragility m[165,166], that depends on the behaviour of the orientational relaxation near T_g. This allows to reduce the degrees of freedom for the fit. If a wider temperature range is taken into account, T_{VTF} is used only a fit parameter, but not related to the glass transition temperature. Here, the VTF equation is fitted to a wide temperature range in order to study the temperature trend of the dynamic properties.

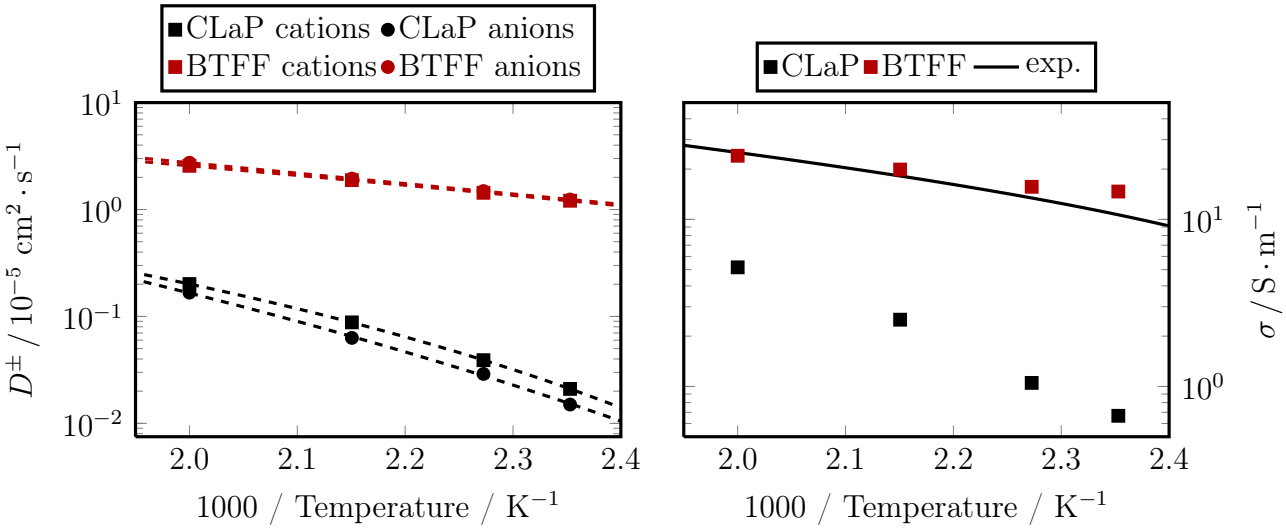


Figure 5.7.: Diffusion constants calculated from the MSD and conductivities are summarized for different temperatures. As guide for the eye corresponding VTF fits (loosely dashed lines) are provided for the MSD.

In Figure 5.7, the diffusion constants obtained by CLaP and BTFF for different temperatures and the corresponding VTF fits, given by the dashed lines, are shown. Besides the expected prediction of a faster diffusion by BTFF compared to the results of CLaP, a completely different

behaviour in respect to the temperature is observed. While BTFF gives almost straight lines, a significant curvature arises for CLaP. As the curvature is related to T_{VTF}, this implies small values of T_{VTF} or even, that instead of VTF behaviour, Arrhenius behaviour is given, which corresponds to $T_{VTF} = 0$. However, one has to keep in mind, that the conductivity at $T = 425$ K is overestimated by BTFF, which implies also a too fast diffusion at $T = 425$ K. For this reason, a more accurately tuned version of BTFF is expected to show a curvature in the temperature dependence of the diffusion coefficients, too, because the diffusion seems to be overestimated at low temperatures.

	D / 10^{-5} cm$^2 \cdot$ s^{-1}					
	CLaP			BTFF		
T / K	D^+	D^-	γ	D^+	D^-	γ
425	0.021	0.015	1.40	1.208	1.240	0.97
440	0.039	0.029	1.34	1.433	1.488	0.96
465	0.088	0.063	1.40	1.883	1.947	0.97
500	0.201	0.167	1.20	2.563	2.738	0.94

Table 5.3.: Diffusion constants derived from simulations with the force fields CLaP and BTFF.

The diffusion coefficients and their corresponding ratio $\gamma = D^+ / D^-$ are summarized in Table 5.3. At this point, it should be mentioned, that the statistical error of all values is at least an order of magnitude smaller than the number of significant digits, which is of importance for the discussion of the results of γ. Though CLaP and BTFF show a qualitatively different behaviour, because for all studied temperatures CLaP and BTFF predict $\gamma > 1$ and $\gamma < 1$, respectively, a common behaviour of the course of γ with T is observable. While γ is decreasing, if the temperature is increased from $T = 425$ K to $T = 440$ K, a further increase of the temperature to $T = 465$ K gives the same ratio for γ as for $T = 425$ K. Finally, if the highest studied temperature $T = 500$ K is considered, γ reaches its significantly smallest value. This fluctuations of γ indicate, that the energy barriers, which have to be passed for a increasing diffusion, differ for the cation and the anion at low temperatures, but vanish with increasing temperature. Since chloride is has no rotational degrees of freedom, the increase in diffusion is dominated by the breaking of the hydrogen bond network. In contrast, a breaking of the hydrogen bond formation influences the rotation of $[MMIM]^+$, and hence, the diffusion of the cation. This subtle interplay between the short and long range interactions, which governs the rotational and translation motion, respectively, is a well known characteristic behaviour of glass formers like [MMIM][Cl].

In order to validate the tuned force field BTFF also outside the temperature regime, which has been applied during the parameter optimization, the conductivity is studied, because it can be compared to experimental results[6].

In Figure 5.7 the experimental and simulation data is shown and summarized in Table 5.4. The solid black line describes the experimentally derived course of the conductivity, which is given by a least-square fit to a parabolic function[6]. A comparison of the simulation results shows, that BTFF predicts the conductivity quite well over a wide temperature range, while CLaP strongly underestimates it. Again two level of correlations are observed for the different force fields, where BTFF is correlated more strongly, but the correlation is temperature

	σ / S · m^{-1}				
	CLaP		BTFF		
T / K	σ	κ	σ	κ	$\sigma^{\text{exp.}}$
425	0.67	0.84	14.68	0.67	10.65
440	1.05	0.72	15.67	0.63	13.42
465	2.51	0.84	19.87	0.65	18.19
500	5.16	0.77	24.03	0.63	25.13

Table 5.4.: Conductivities σ at different temperatures obtained by the force fields are compared to experimental data[6]. The correlation κ is also summarized for CLaP and BTFF.

independent. The too fast dynamics at low temperatures, where high densities arise and short range interactions strongly contribute to the overall interaction, highlights, that the dispersion energy given by the short range interactions is too small. In high temperature regimes, where the electrostatic long range interaction dominates, this contribution becomes negligible and the force field describes the dynamics of the system very accurately.

In summary, the validation process has shown, that though the tuned force field BTFF overestimates the conductivity at low temperatures around $T = 440$ K, but a convergence to the correct behaviour is achieved in the high temperature regime starting around $T = 470$ K. This indicates, that a too small dispersion interaction is modelled, which is furthermore underlined by the quite fast diffusion compared to other force fields. Hence, for the construction of a transferable and accurate FF for ILs, a more proper modeling of the very local structure, describe by the first maxima in the RDFs, is inevitable.

Heats of Vaporization As sensitive indicator for the single particle dynamics of a system as well as its collectivity, expressed by the shear viscosity or ionic conductivity, the heat of vaporization ΔH_{vap} is often used as a target property for the force field parameterization. Furthermore, no large amount of simulation time is necessary to sample the phase space properly, which is also of great importance for the parameter optimization. Though a much less amount would be required, the statistics of the liquid phase are obtained from 90 ns simulation time, while the single ion pair has been simulated for 9 ns in the gas phase, which would also be a sufficient time to obtain enough statistics in the liquid phase.

The heat of vaporization ΔH_{vap} is the difference in enthalpy $H = U + PV$ between the liquid and the gas phase. The enthalpy of the liquid state $H_{\text{liq.}} = \langle E^{\text{tot.}} \rangle + P \cdot \langle V \rangle$ is obtained for the average total energy E^{tot} and the product of the reference pressure P with the average volume $\langle V \rangle$[89]. However, since no volume is defined for the single ion pair vacuum calculations, the term PV is undefined, too. For this reason the ion pairs in the gas phase are considered to behave ideal, such that the missing term is given by the corresponding equation of state:

$$PV = nRT, \tag{5.36}$$

with the gas constant R and the amount of substance n in mole. Therefore, the heat of vaporization per mole is given by:

$$\Delta H_{\text{vap}} = E^{\text{tot.}}_{\text{gas}} + RT - H_{\text{liq.}}. \tag{5.37}$$

Moreover, during the phase transition, the polarization of a molecule alters, which is not described by classical non-polarizable FFs. Thus, the energy contribution E^{pol} from the change of the polarization state has to considered explicitly. While this is mostly not taken into account, it has been shown that an explicit consideration improves the parametrization of SPC/E water[156]. Standard force fields for ILs usually model the polarizability in the LJ interactions. Hence, E^{pol} is only partly captured by the decreasing number of short range interactions of an ion moving from the liquid to the gas state. Moreover, the partial charges mostly rely on the charge distribution of an isolated single ion. Thus, the explicit calculation of E^{pol}, actually given by the change of the dipole moment, is impossible. In contrast, our approach allows to directly access E^{pol} because ϵ_{el} has been derived for the gas and liquid phase. From the MP2 gas phase calculations[11], $\epsilon_{\mathrm{el}} = 1.67$ is calculated by averaging of the net ion charges of the four ground state geometries, corresponding to an average ionic net charge of ± 0.77 e. However, the results of the MP2 calculations are only a rough estimate, because they have been performed at 0 K and only 4 geometries are sampled. To improve the estimate, 100 snapshots are chosen from an MD simulation of ion pair and simulated with AIMD at a finite temperature $T = 425$ K until the average charge of the chloride has been converged. In this case an average net charge of 0.65 e is obtained, corresponding to an electronic dielectric constant of $\epsilon_{\mathrm{el}} = 2.37$, which differs only slightly from the bulk value $\epsilon_{\mathrm{el}} = 2.51$.

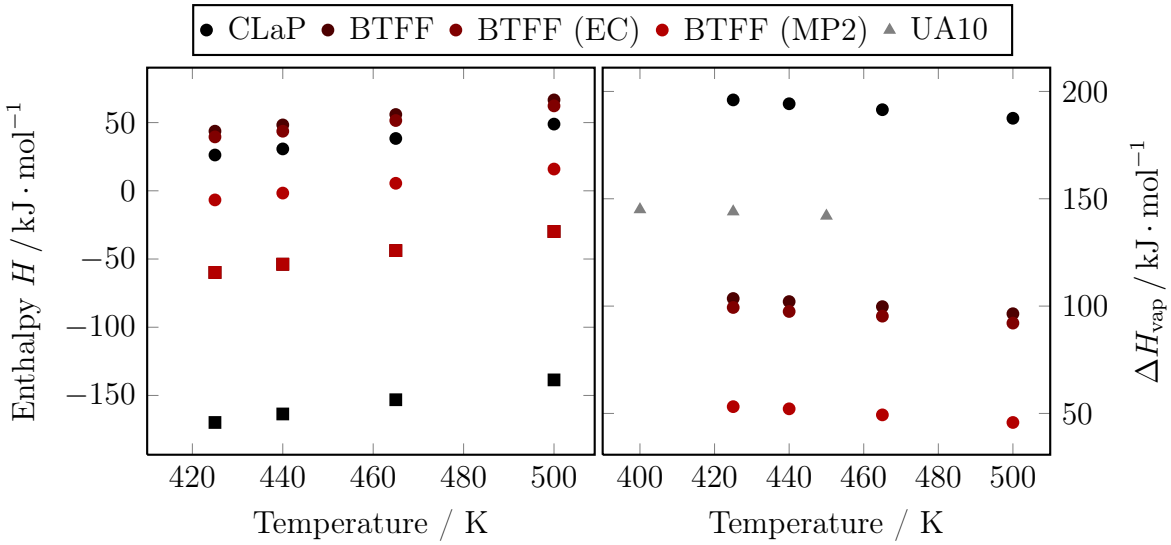


Figure 5.8.: The enthalpies H of the gas and liquid phase, as well as heats of vaporization ΔH^{vap} are derived with CLaP and BTFF. For BTFF also the result for the gas phase calculations in the electronic continuum model (EC) are shown, to point out the contribution of E^{pol} to ΔH^{vap}. The circles in the left plot depict H in the liquid phase, while the circles show the values for the gas phase state. The heat of vaporization is also compared to a united atom model for [MMIM][Cl] (UA10)[153].

The obtained enthalpies H for BTFF and CLaP in the gas and liquid phase and the vaporization enthalpies are presented in Figure 5.8. Comparing the enthalpies of the liquid phase, depicted by the squares, reveals a dramatic increase in energy if the net charge decreases. Considering the gas phase enthalpies, shown by the circles, the large impact of $E^{\mathrm{pol.}}$ becomes obvious. If the electronic continuum (EC) model, which is only valid for the liquid phase, is

still taken into account for the gas state calculations, the gas phase enthalpy is larger than for CLaP due to the reduced electrostatic interaction. However, the gas phase enthalpy decreases by $E^{\mathrm{pol}} \approx 50\ \mathrm{kJ \cdot mol^{-1}}$, if the partial charges of the liquid phase are scaled up to the average net charge of 0.77 e as obtained by the MP2 calculations at 0 K. But the application of a set of partial charges rescaled to a net charge of 0.65 e, as obtained from the AIMD simulations of the single ion pairs at $T = 425$ K, results in a gas phase enthalpy, which coincides almost with the results, where the EC model is used and gives $E^{\mathrm{pol.}} \approx 4\ \mathrm{kJ \cdot mol^{-1}}$. Hence, the common method to use the liquid state charges for the calculation of the vacuum energies introduces only a small error, and the partial charges are expected to describe the gas phase of the ionic liquid, too. Moreover, this small contributation might be easily incorporated into the LJ interaction. Yet, as already shown for SPC/E[137] and other water models[156], the contribution of $E^{\mathrm{pol.}}$ to the enthalpy is observable, but much larger for water, than for an ionic liquid pair.

Unfortunately, experimental data for ΔH^{vap} is lacking for [MMIM][Cl]. Therefore in the following, the results are compared to already published simulation data for [MMIM][NTf_2] by Köddermann *et al.*[126] (KPL) and Chaban[130] (QSC), and for [MMIM][Cl] by Liu *et al.*[153] (UA10). While KPL and QSC are reparametrized versions of CLaP, UA10 provides an united-atom model. Moreover, the parametrization methods differ[16]. KPL has been derived by an optimization of short range interactions with respect to several thermodynamic properties such that bulk polarization effects are modelled exclusively by the LJ interactions. In contrast, QSC is a result from a global charge scaling optimized with respect to a certain experimental property, such as the conductivity. Finally, UA10 also constitutes a model with a reduced ionic net-charge, but motivated by *ab-initio* ion pair calculations. Furthermore, the LJ parameters of UA10 are adapted by force-matching to all-atom simulations of the liquid phase,[124] while QSC relies on the short-range parametrization of CLaP. Therefore, KPL and UA10 model polarizability in a similar manner, though UA10 gives a reduced ion net charge. The partial charges of UA10 are based on gas phase calculation, which are not able to describe the bulk polarization. Yet, a large part of the polarization effects are included in the short-range parameters. The parametrization QSC includes the major part of the polarization within the charge-scaling, such that the short-range interactions are mainly responsible for the dispersion interaction and repulsion, but the charge distribution relies on an isolated state. However, neither of the studies, explicitly considers the contribution of E^{pol} for the calculation of ΔH^{vap}.

	ΔH^{vap} / kJ · mol^{-1}			
T / K	CLaP	BTFF(MP2)	BTFF(EC)	BTFF
425	196.05	53.16	99.36	103.54
440	194.26	52.15	97.50	102.11
465	191.49	49.33	95.32	99.75
500	187.50	45.79	92.08	96.45

Table 5.5.: Heats of vaporization $\Delta H^{\mathrm{vap.}}$ of [MMIM][Cl] obtained with CLaP, BLFF and BTFF. Obviously the energy is underestimated, if the change of the polarization state is not considered, when the molecule is moved from the liquid (EC) to the vacuum.

In Figure 5.8 and Table 5.5, the enthalpies of vaporization calculated with different parameterisations are summarized. While for [MMIM][NTf_2], $\Delta H_{\mathrm{vap}} \approx 130\ \mathrm{kJ \cdot mol^{-1}}$ has been

obtained around $T = 300$ K by Köddermann *et al.*[178] with KPL and with QSC by Chaban[130], UA10 predicts enthalpies around 144 $\mathrm{kJ \cdot mol^{-1}}$ [153] for [MMIM][Cl] at $T = 400$ K. In contrast, CLaP predicts a heat of vaporization, that is approximately 50 $\mathrm{kJ \cdot mol^{-1}}$ larger, which is also reflected in the very slow dynamics. In contrast, the enthalpies of vaporization derived by BTFF are up to 150 $\mathrm{kJ \cdot mol^{-1}}$ smaller, depending on the values, which are used for the gas phase enthalpies. However, if the appropriate method is chosen, which uses the set of partial charges rescaled to a net charge of 0.65 e, BTFF predicts $\Delta H^{\mathrm{vap.}}$ around 100 $\mathrm{kJ \cdot mol^{-1}}$, which is roughly 40 $\mathrm{kJ \cdot mol^{-1}}$ smaller than the results from UA10. This difference in $\Delta H^{\mathrm{vap.}}$ is also reflected in the behaviour of the diffusion constants, because BTFF gives a three times faster diffusion than UA10.

Though CLaP predicts a thermodynamic quantity as the mass density within an error of less than 0.5 %, the enthalpy of vaporization is much larger compared to the results of classical MD simulations with other force fields or experimental data for similar ILs. This is in agreement with the too slow dynamics modelled by CLaP, because the ions in the system are too strongly bound, which is a relict of an insufficient parametrization of the LJ interactions, that are transferred from OPLS and AMBER, but not adapted to the polarizabilities of the atom types. In contrast, KPL[126] shows that careful inclusion of polarizability into LJ interactions is possible and accurately describes energetic, static, and dynamic properties of $[\mathrm{C_nMIM}][\mathrm{NTf_2}]$ ionic liquids. Though with that, polarization effects can appropriately be modelled by solely LJ interactions also for ionic liquids, the FF has been derived in respect to dynamical variables. Therefore, the required amount of computational resources is significantly larger than for the procedure porposed in this work, which in contrast, does not rely on a large experimental data set. This classical MD study of ionic liquids also considers the contributions of $E^{\mathrm{pol.}}$ for the calculation of ΔH^{vap} explicitly, and reveals a contribution of roughly 4 $\mathrm{kJ \cdot mol^{-1}}$. However, BTFF predicts unusual small $H^{\mathrm{vap.}}$, which is consistent with the overestimation of the dynamics at low temperatures. Hence, only a final adaption of the dispersion interactions, modelled by the LJ parameters, seems to be required, in order to obtain an accurate force field over the whole temperature range. Therefore even a single experimental data point for $H^{\mathrm{vap.}}$ might be sufficient, but is not accessible, unfortunately. The mismatch of the maxima of the RDFs is also an indicator for the bonding of the atoms. Yet, an appropriate adaption is expected to allow a very accurate modelling of the ionic liquid.

Hence, a first iterative cycle to refine force fields for classical molecular dynamics simulations is closed. The proposed method for the parameter tuning is quite promising as it is able to adapt the set of parameters in a reasonable time due to the massive parallelization of the method. Furthermore, the proposed approach to derive the partial charges from liquid state configurations allows to include only static properties like the density of mass and radial distribution functions during the parameterisation process reducing the computational effort as shorter simulation times are required. Though, the validation process shows that the dynamics in terms of diffusion and conductivity is overestimated at low temperatures, the deviations from experiment vanish at high temperatures and the correct trend is achieved. This is a quite promising result, because dynamic quantities have never been involved in the parametrization process. Though the study of the heat of vaporization reveals, that the tuned force field BTFF predicts unusual small value, it is shown that a correct inclusion of the polarization energy $E^{\mathrm{pol.}}$ contributes roughly 5 % to $\Delta H^{\mathrm{vap.}}$ and is therefore not negligible. Yet, an important level in the multiscale approach is reached, because a force field parametrization machinery is established that requires minimal experimental data, while the remaining information is

carefully extracted from calculations and simulations. Furthermore, the computational framework, given by the Python module PyPaTEGro is now present, which allows to include all properties into the parameterisation process, that were introduced in this study. This is an important step towards an efficient way for identifying, classification, and understanding of this class of solvents, because it supplies the basis for an efficient design of a consistent and transferable force field for RTILs.

6. Summary and outlook

In this work, a multiscale method has been introduced and elaborated in order to optimize or construct force fields (FFs) for ionic liquids (ILs). A FF consists of a variety of interaction terms. Some of them, such as the bond and angle potentials, can be straightforwardly derived, but various methods for the calculation of the point charges are accessible and also the determination of the short-range interactions is a non-trivial task that usually requires an iterative adaption to experimental reference data. A FF is a highly correlated parameter set, such that small changes of single parameters may drastically affect the results. For this reason, a large parameter set providing descriptions of various atom types and molecules, is required which is also transferable between different states of the phase diagram.

Ionic liquids are of major interest in this thesis, therefore two famous FFs for [EMIM][BF_4], namely CLaP[5] and LHW[124], were benchmarked.[10] These FFs have been parametrized *in-vacuo*, so it is not guaranteed that they also provide an accurate model of the liquid phase. If only static properties, like the mass density, the static dielectric constant, or the global liquid structure, are considered both FFs perform well, but subtle differences arise in the local structure if the radial and spatial distribution and the orientation is studied in more detail. In terms of dynamics, both FF fail in a correct modeling of the system. This is observable in underestimation of the diffusivity and conductivity. The conductivity is a collective property, thus a reliable and time-saving method for its calculation had to be identified. It has been shown that the Einstein-Helfand technique fulfills this requirements. For this reason, it has been implemented in the program `g_current`, which became part of the molecular dynamics simulation package GROMACS.

The benchmark showed that methods to optimize IL FFs are necessary, which is the major topic of this thesis and discussed in chapter 4 and chapter 5. Apart from an optimization of the FF, also efficient simulations are desired, because a large number of them has to be performed during the tuning and validation of the FF parameters. Most of the computing time is spent for the calculation of the electrostatic forces. For large systems under periodic boundary conditions (PBC) the Smooth-Particle-Mesh-Ewald (SPME) algorithm allows an efficient treatment of the electrostatics, but an *a-priori* error estimate had been lacking. An error estimate for SPME recently became available,[13] but is only applicable to even interpolation orders. Based on the connection of the various branches of mesh-based Ewald techniques and error estimates[104], an error estimate for an arbitrary set of SPME parameters was derived in section 3.3, which also includes an *a-priori* error estimate of self-forces. This provides a round figure of the error estimates. In order to numerically validate the estimate, the algorithm has been implemented in the GROMACS tool `g_pme_error`, which also became part of the official release. The tool allows one to estimate the error and tune the parameters in order to a achieve a certain error limit.

For a minimization of the tuning time, a Message Passing Interface parallelization is also provided. Yet GROMACS has been extended, such that highly optimized simulations in terms of accuracy and speed are easily achieved and efficient analysis tools have been made available. These both aspects considerably affect the FF parameterisation and validation process, described in chapter 5, because a large number of simulations is required in order to adapt the parameters. In addition, the trajectories for the final validation have to be integrated over a large number of steps for the sampling of a sufficiently large part of the phase space. Yet already small gains in performance considerably contribute to the whole parameterisation process.

Equipped with the computational framework for highly efficient simulations, the first step in the FF optimization process was the refinement of partial charges, as described in chapter 4. It was shown that partial charges, which have been derived from isolated states, are not able to correctly model the bulk effects found in the liquid phase. Characteristics of the charge distribution of a liquid state are an increased dipole moment and reduced ionic net-charges, compared to the isolated state, which has been shown by QM calculations[1,7,15,167] and NMR experiments[154]. For this reason, an alternative method by Blöchl[55] (CAB) was introduced in section 4.2.[11] It is applicable to bulk systems which describe the liquid state. A liquid state is only accessible by periodic boundary conditions, which actually enables the use of CAB, because it relies on the charge density in reciprocal space $n(\boldsymbol{k})$. Partial charges are obtained from an adaption of a model charge density to the multipole expansion and width of the calculated charge density $n(\boldsymbol{k})$. A comparison of the CAB charges to partial charges derived with standard approaches like RESP or CHELP, highlights the high sensitivity of all results to the configuration of the atoms, especially if polarization plays a role such as in the liquid phase. A comparison of CAB charges, derived from the bulk phase, to partial charges of the CLaP FF, derived from isolated states, showed that CAB provides an improved description of liquid characteristics given by charge transfer and polarization. The partial charges capture these aspects by a reduced net-charge and a strong fluctuation of their average.

These effects can be quantified by the MDEC theory of Leontyev and Stuchebrukhov[152], showing that the reduced ionic net–charge is an effective and implicit description of polarization. The MDEC approximation is only applicable, if the medium is strongly screening, such that the electrostatic interactions converge within the local environment. For ILs, this has been observed, because the correlations of the dipole moments already vanish beyond the nearest neighborhood.[15,167] In addition, the influence of the system size on the partial charges was investigated and also a strong screening has been observed. This shows that the MDEC theory is applicable to ILs and allows the reduction of the computational effort to determine the partial charges, because even small systems contain enough information for an appropriate fit. It was also shown that CAB charges appropriately describe the average molecular dipole moment in the liquid compared to the results of a WA of the AIMD charge density.

In order to establish a generic FF moderate changes to the CAB charges are required. Thus classical MD simulation were performed with slightly different sets of CAB based partial charges. The effect on neither structural nor dynamical properties is strong. As expected,[12] a decreased mass density and faster dynamics are observed, if reduced net charges are involved, but the CAB-based models do not differ significantly. Yet small modifications of the CAB charges affect the MD simulations only slightly and a comparison of the partial charges for different cation–anion combinations showed that the distributions are quite independent of the ion combination. With that, a generic set of partial charges is constructed that allows a considerable description of the cation dipole moment and is transferable to different imidazolium based cations and anions.

The missing pieces for a multi-scale optimization technique were finally described in chapter 5. A method for an adaption of the short-range parameter was developed,[17] which provides the theoretical basis for the construction of a transferable FF. Besides a theoretical framework, the Python[133] module PyPATEGRO, described in Appendix A, was implemented to allow one a simple and user-friendly application. Two kinds of interactions, which are not completely

independent from each other, are involved in this tuning step. The dihedral interactions are connected to the 1–4 interactions and an appropriate adaption of the corresponding potentials is required, if the partial charges or LJ parameters are changed. This is a comparably straightforward task, while the optimization of the LJ parameters is more time-consuming and is based on the minimization of an error function. Polarization strongly contributes to the dynamics and is already parameterised in the partial charges, so the error function is constructed of solely static properties, such as different RDFs and the mass density at different temperatures. For this reason, only short simulation times are required allowing a very fast tuning.

As example system, the IL [MMIM][Cl] was used and considerable improvement of the accuracy in terms of the reference data has been obtained. During the validation of the optimized FF, an overestimation of dynamical properties is observed at low temperatures, but the behaviour is predicted qualitatively correct compared to a polarizable FF[149,150]. At higher temperatures the predicted conductivity converges to the experimental reference. This is expected to arise from a too small dispersion interaction, which is also indicated by a too small first maximum of the H^1-Cl RDF. For this reason, very accurate modelling of the local structure is expected to improve the overall description significantly. This is underlined by results for the heats of vaporization which are quite small. Thus the applied technique allows a considerable improvement of the FF, but insufficiencies are still present, which are expected to vanish with a decreasing deviation of the RDFs.

In summary, theories and computational tools were introduced, developed, and validated in this thesis, which provide the foundation for the construction of a transferable RTIL FF. A large data set has also been made available and the next steps are rather clear. Enough computer time has to be allocated in order to optimize the short-range parameters for the suggested set of generic force field charges. Also many possibilities for an improvement of the parameter space sampling for the short-range interactions exists, but the basic framework, provided here, is accessible and easily extensible.

Acknowledgement The author thanks Katharina Wendler, Yuan-Yuan Zhao, Baofu Qiao, Christian Krekeler, Jochen Schmidt, Luigi Delle Site, and Robert Berger for fruitful discussion and for providing the geometries of the ion pairs and of the AIMD bulk simulations. Han Wang is acknowledged for his nice cooperation for the SPME error estimate. The Deutsche Forschungsgemeinschaft is acknowledged for funding within the priority program SPP 1191 Ionic Liquids as well as the SimTech cluster of excellence. Computing time was obtained by the Höchstleistungsrechenzentrum Stuttgart (HLRS). At last but not least, the author thanks his supervisor Christian Holm and the whole ICP for the inspiring and nice atmosphere, especially Owen Hickey for a careful reading of the manuscript, and all my family and friends for their persistent support.

Supporting Information All graphs in this thesis have been generated with the LaTex packages PGF, Ti*k*Z, and PGFPLOTS, if not stated otherwise. In order to provide an adequate representation of the information in the period of ebooks, all graphs are also available as clickable PDFs, which allow to extract detailed data from the graph by a simple click on the corresponding point. Including all this data into the thesis would have generated a too large file, but if the informations is required it can be downloaded from the publication server of the University Stuttgart.

Bibliography

[1] C. Krekeler, J. Schmidt, Y. Zhao, B. Qiao, R. Berger, C. Holm and L. Delle Site, *J. Chem. Phys.*, 2008, **129**, 174503.

[2] Y. Zhao, C. Krekeler, B. Qiao, J. Schmidt, C. Holm, L. D. Site and R. Berger, *submitted*, 2012.

[3] P. A. Fröba, H. Kremer and A. Leipertz, *J. Phys. Chem. B*, 2008, **112**, 12420–12430.

[4] Y. Yoshida, O. Baba and G. Saito, *J. Phys. Chem. B*, 2007, **111**, 4742–4749.

[5] J. N. Canongia Lopes, J. Deschamps and A. A. H. Padua, *J. Phys. Chem. B*, 2004, **108**, 2038–2047.

[6] A. A. Fannin, L. A. King, J. A. Levisky and J. S. Wilkes, *J. Phys. Chem.*, 1984, **88**, 2609–2614.

[7] C. Krekeler, F. Dommert, J. Schmidt, Y. Y. Zhao, C. Holm, R. Berger and L. Delle Site, *Phys. Chem. Chem. Phys.*, 2010, **12**, 1817–1821.

[8] A. Chaumont and G. Wipff, *Inorg. Chem.*, 2009, **48**, 4277–4289.

[9] B. Qiao, C. Krekeler, R. Berger, L. Delle Site and C. Holm, *J. Phys. Chem. B*, 2008, **112**, 1743–1751.

[10] F. Dommert, J. Schmidt, B. Qiao, Y. Zhao, C. Krekeler, L. Delle Site, R. Berger and C. Holm, *J. Chem. Phys.*, 2008, **129**, 224501.

[11] J. Schmidt, C. Krekeler, F. Dommert, Y. Zhao, R. Berger, L. Delle Site and C. Holm, *J. Phys. Chem. B*, 2010, **114**, 6150–6155.

[12] F. Dommert, J. Schmidt, C. Krekeler, Y. Y. Zhao, R. Berger, L. Delle Site and C. Holm, *J. Mol. Liq.*, 2010, **152**, 2–8.

[13] H. Wang, F. Dommert and C. Holm, *J. Chem. Phys.*, 2010, **133**, 034117.

[14] K. Wendler, S. Zahn, F. Dommert, R. Berger, C. Holm, B. Kirchner and L. Delle Site, *J. Chem. Theory Comput.*, 2011, **7**, 3040–3044.

[15] K. Wendler, F. Dommert, Y. Y. Zhao, R. Berger, C. Holm and L. Delle Site, *Faraday Discuss.*, 2012, **154**, 111–132.

[16] F. Dommert, K. Wendler, R. Berger, L. Delle Site and C. Holm, *ChemPhysChem*, 2012, **13**, 1625–1637.

[17] F. Dommert and C. Holm, *Phys. Chem. Chem. Phys.*, 2013, **15**, 2037–2049.

[18] I. Krossing, J. M. Slattery, C. Daguenet, P. J. Dyson, A. Oleinikova and H. Weingärtner, *J. Am. Chem. Soc.*, 2006, **128**, 13427–13434.

[19] U. Domanska, M. Krolikowska and M. Krolikowski, *Fluid Phase Equilib.*, 2010, **294**, 72–83.

[20] C. P. Fredlake, J. M. Crosthwaite, D. G. Hert, S. N. V. K. Aki and J. F. Brennecke, *Journal of Chemical & Engineering Data*, 2004, **49**, 954–964.

[21] A. Lewandowski and A. Swiderska-Mocek, *Z. Phys. Chem.*, 2009, **223**, 1427–1435.

[22] P. Walden, *Bull. Acad. Imper. Sci. (St. Petersburg)*, 1914, **8**, 405–422.

[23] P. Bonhôte, A.-P. Dias, N. Papageorgiou, K. Kalyanasundaram and M. Grätzel, *Inorg. Chem.*, 1996, **35**, 1168–1178.

[24] J. D. Holbrey and K. R. Seddon, *Clean Prod. Processe*, 1999, **1**, 223–236.

[25] T. Welton, *Chem. Rev. (Washington, DC, U. S.)*, 1999, **99**, 2071–2083.

[26] R. Hagiwara and Y. Ito, *J. Fluorine Chem.*, 2000, **105**, 221–227.

[27] J. G. Huddleston, *Green Chemistry*, 2001, **3**, 156–164.

[28] C. Margulis, H. Stern and B. Berne, *J. Phys. Chem. B*, 2002, **106**, 12017–12021.

[29] H. Olivier-Bourbigou and L. Magna, *J. Mol. Catal. A: Chem.*, 2002, **182-183**, 419–437.

[30] *Ionic Liquids in Synthesis*, ed. P. Wasserscheid and T. Welton, Wiley-VCH, 2002.

[31] J. S. Wilkes, *J. Mol. Catal. A: Chem.*, 2004, **214**, 11–17.

[32] J. Ding and D. Armstrong, *Chirality*, 2005, **17**, 281–292.

[33] H. Tokuda, S. Tsuzuki, M. A. B. H. Susan, K. Hayamizu and M. Watanabe, *J. Phys. Chem. B*, 2006, **110**, 19593–19600.

[34] H. Weingärtner, *Angew. Chem. Int. Ed.*, 2007, **47**, 654–670.

[35] R. P. Swatloski, S. K. Spear, J. D. Holbrey and R. D. Rogers, *J. Am. Chem. Soc.*, 2002, **124**, 4974–4975.

[36] S. Tan and D. MacFarlane, in *Topics in Current Chemistry*, 2010, vol. 290, ch. Ionic Liquids in Biomass Processing, pp. 311–339.

[37] C. Wang, G. Cui, X. Luo, Y. Xu, H. Li and S. Dai, *J. Am. Chem. Soc.*, 2011, **133**, 11916–11919.

[38] P. Wasserscheid and W. Keim, *Angew. Chem., Int. Ed.*, 2000, **39**, 3772–3789.

[39] A. Yokozeki and M. B. Shiflett, *Applied Energy*, 2007, **84**, 1258–1273.

[40] S. Seki, T. Kobayashi, Y. Kobayashi, K. Takei, H. Miyashiro, K. Hayamizu, S. Tsuzuki, T. Mitsugi and Y. Umebayashi, *J. Mol. Liq.*, 2010, **152**, 9–13.

[41] D. R. Hartree, *Math. Proc. Cambridge*, 1928, **24**, 89–110.

[42] V. Fock, *Z. Phys. A Hadron Nucl.*, 1930, **61**, 126–148.

[43] C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–.

[44] P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, 864–871.

[45] W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–.

[46] A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098.

[47] C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.

[48] B. Miehlich, A. Savin, H. Stoll and H. Preuss, *Chemical Physics Letters*, 1989, **157**, 200–206.

[49] J. Perdew, K. Burke and M.Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.

[50] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103–128.

[51] *CPMD, http://www.cpmd.org/, Copyright IBM Corp. 1990-2008 Copyright MPI für Festkörperforschung Stuttgart 1997-2001*.

[52] J. Hutter, *Computer code* `CP2K`, 2000-2008, `http://cp2k.berlios.de`.

[53] W. Jorgensen, D. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

[54] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. R. Ross, I. T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, *Comp. Phys. Comm.*, 1995, **91**, 1–41.

[55] P. E. Blöchl, *J. Chem. Phys*, 1995, **103**, 7422–7428.

[56] M. Born and R. Oppenheimer, *Ann. Phys. Berlin*, 1927, **389**, 457–484.

[57] D. C. Rapaport, *The Art of Molecular Dynamics Simulation*, Cambridge University Press, 2nd edn., 2004.

[58] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577.

[59] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, IOP, London, 1988.

[60] L. Landau and E. Lifschitz, *Lehrbuch der Theoretischen Physik, Band I: Mechanik*.

[61] M. E. Tuckerman and G. J. Martyna, *J. Phys. Chem. B*, 2000, **104**, 159–178.

[62] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, Academic Press, San Diego, 2nd edn., 2002.

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

[64] H. F. Trotter, *Proc. Amer. Math. Soc.*, 1959, **10**, 545–551.

[65] M. Creutz and A. Gocksch, *Phys. Rev. Lett.*, 1989, **63**, 9.

[66] R. M. Wilcox, *J. Math. Phys.*, 1967, **8**, 962–982.

[67] L. Verlet, *Phys. Rev.*, 1967, **159**, 98.

[68] W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, *J. Chem. Phys.*, 1982, **76**, 637–649.

[69] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

[70] G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

[71] S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.

[72] W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

[73] B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

[74] E. Lindahl, B. Hess and D. van der Spoel, *J. Mol. Mod.*, 2001, **7**, 306–317.

[75] D. V. D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.

[76] H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.

[77] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen and H. J. C. Berendsen, *Gromacs User Manual version 4.5*, 2005.

[78] G. J. Martyna, M. L. Klein and M. Tuckerman, *J. Chem. Phys.*, 1992, **97**, 2635–2643.

[79] B. L. Holian, A. F. Voter and R. Ravelo, *Phys. Rev. E*, 1995, **52**, 2338–.

[80] M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

[81] G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177–4189.

[82] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz J.R, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

[83] C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.

[84] C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.*, 1990, **11**, 361–373.

[85] M. P. Tosi, *Solid State Phys.: Adv. Res. App.*, 1964, 1.

[86] F. G. Fumi and M. P. Tosi, *J. Phys. Chem. Solids*, 1964, **25**, 31–43.

[87] M. Hülsmann, T. Köddermann, J. Vrabec and D. Reith, *Comput. Phys. Commun.*, 2010, **181**, 499–513.

[88] J.-P. Ryckaert and A. Bellemans, *Faraday Discuss. Chem. Soc.*, 1978, **66**, 95–106.

[89] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1st edn., 1987.

[90] S. Liem, D. Brown and J. Clarke, *Computer Physics Communications*, 1991, **67**, 261–267.

[91] P. P. Ewald, *Ann. Phys.*, 1921, **369**, 253–287.

[92] L. Greengard and V. Rhoklin, *J. Comput. Phys.*, 1987, **73**, 325.

[93] L. Greengard and V. Rokhlin, *Acta Numerica*, 1997, **6**, 229–269.

[94] R. Sperb, *Molecular Simulation*, 1998, **20**, 179–200.

[95] A. Arnold and C. Holm, *Comput. Phys. Commun.*, 2002, **148**, 327–348.

[96] A. Arnold and C. Holm, *J. Chem. Phys.*, 2005, **123**, 144103.

[97] D. Wolf, P. Keblinski, S. R. Phillpot and J. Eggebrecht, *J. Chem. Phys.*, 1999, **110**, 29.

[98] M. Neumann and O. Steinhauser, *Chem. Phys. Lett.*, 1983, **102**, 508.

[99] J. Kolafa and J. W. Perram, *Mol. Simul.*, 1992, **9**, 351–368.

[100] G. Hummer, *Chem. Phys. Lett.*, 1995, **235**, 297–302.

[101] T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

[102] M. Deserno and C. Holm, *J. Chem. Phys.*, 1998, **109**, 7678.

[103] M. Deserno and C. Holm, *J. Chem. Phys.*, 1998, **109**, 7694.

[104] V. Ballenegger, J. J. Cerdà and C. Holm, *J. Chem. Theory Comput.*, 2012, **8**, 936–947.

[105] V. Ballenegger, J. J. Cerdà and C. Holm, *Comput. Phys. Commun.*, 2011, **182**, 1919–1923.

[106] F. Weik, *M.Sc. thesis*, University Stuttgart, 2011.

[107] D. S. Cerutti, R. E. Duke, T. A. Darden and T. P. Lybrand, *J. Chem. Theory Comput.*, 2009, **5**, 2322–2338.

[108] A. Neelov and C. Holm, *J. Chem. Phys.*, 2010, **132**, 234103.

[109] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999–2012.

[110] A. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

[111] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.

[112] C. Hanke, S. Price and R. Lynden-Bell, *Mol. Phys.*, 2001, **99**, 801.

[113] L.-Y. Hsu and D. E. Williams, *Acta Crystallogr. A*, 1980, **36**, 277–281.

[114] D. E. Williams and S. R. Cox, *Acta Crystallogr. B*, 1984, **40**, 404–417.

[115] D. E. Williams and D. J. Houpt, *Acta Crystallogr. B*, 1986, **42**, 286–295.

[116] T. A. Halgren, *J. Am. Chem. Soc.*, 1992, **114**, 7827–7843.

[117] J. de Andrade, E. Böes and H. Stassen, *J. Phys. Chem. B*, 2002, **106**, 13344–13351.

[118] J. de Andrade, E. Boes and H. Stassen, *J. Phys. Chem. B*, 2002, **106**, 3546–3548.

[119] T. I. Morrow and E. J. Maginn, *J. Phys. Chem. B*, 2002, **106**, 12807–12813.

[120] J. N. Canongia Lopes, A. A. H. Padua and J. Deschamps, *J. Phys. Chem. B*, 2004, **108**, 11250–11250.

[121] J. N. Canongia Lopes and A. A. H. Padua, *J. Phys. Chem. B*, 2004, **108**, 16893–16898.

[122] J. N. Canongia Lopes and A. A. H. Padua, *J. Phys. Chem. B*, 2006, **110**, 19586–19592.

[123] T. Yan, C. Burnham, M. DelPopolo and G. Voth, *J. Phys. Chem. B*, 2004, **108**, 11877–11881.

[124] Z. Liu, S. Huang and W. Wang, *J. Phys. Chem B*, 2004, **108**, 12978–12989.

[125] T. Youngs, M. DelPopolo and J. Kohanoff, *J. Phys. Chem. B*, 2006, **110**, 5697–5707.

[126] T. Köddermann, D. Paschek and R. Ludwig, *ChemPhysChem*, 2007, **8**, 2464–2470.

[127] B. L. Bhargava and S. Balasubramanian, *J. Chem. Phys.*, 2007, **127**, 114510.

[128] W. Zhao, H. Eslami, W. L. Cavalcanti and F. Müller-Plathe, *Z. Phys. Chem.*, 2007, **221**, 1647–1662.

[129] T. G. A. Youngs and C. Hardacre, *ChemPhysChem*, 2008, **9**, 1548–1558.

[130] V. Chaban, *Phys. Chem. Chem. Phys.*, 2011, **13**, 16055–16062.

[131] V. V. Chaban, I. V. Voroshylova and O. N. Kalugin, *Phys. Chem. Chem. Phys.*, 2011, **13**, 7910–7920.

[132] V. V. Chaban and O. V. Prezhdo, *Phys. Chem. Chem. Phys.*, 2011, **13**, 19345–19354.

[133] *Python*, www.python.org, www.python.org.

[134] W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

[135] VMD, *Visual Molecular Dynamics – Homepage*, 2003.

[136] B. Hess, *J. Chem. Theory Comput.*, 2008, **4**, 116–122.

[137] H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.

[138] C. Schröder and O. Steinhauser, *J. Chem. Phys.*, 2008, **128**, 224503.

[139] L. Laaksonen, *J. Mol. Graphics Modell.*, 1992, **10**, 33–34.

[140] B. L. Bhargava and S. Balasubramanian, *Chem. Phys. Lett.*, 2006, **417**, 486–491.

[141] A. Noda, K. Hayamizu and M. Watanabe, *J. Phys. Chem B*, 2001, **105**, 4603–4610.

[142] C. Wakai, A. Oleinikova, M. Ott and H. Weingärtner, *J. Phys. Chem. B*, 2005, **109**, 17028–17030.

[143] V. A. Bright and G. A. Baker, *Comment on "How Polar Are Ionic Liquids? Determination of the Static Dielectric Constant of an Imidazolium-based Ionic Liquid by Microwave Dielectric Spectroscopy". J. Phys. Chem. B, 110(11):5822 - 5823*, 2006.

[144] C. Wakai, A. Oleinikova and H. Weingärtner, *Reply to "Comment On 'How Polar Are Ionic Liquids? Determination of the Static Dielectric Constant of an Imidazolium-based Ionic Liquid by Microwave Spectroscopy'"*, *J. Phys. Chem. B*, 110(11): 5824 - 5824, 2006.

[145] C. Daguenet, P. J. Dyson, I. Krossing, A. Oleinikova, J. Slattery, C. Wakai and H. Weingärtner, *J. Phys. Chem. B*, 2006, **110**, 12682–12688.

[146] H. Weingärtner, P. Sasisanker, C. Daguenet, P. Dyson, I. Krossing, J. Slattery and T. Schubert, *J. Phys. Chem. B*, 2007, **111**, 4775–4780.

[147] S. W. de Leeuw and J. W. Perram, *Physica A*, 1981, **107**, 179–189.

[148] C. Schröder, M. Haberler and O. Steinhauser, *J. Chem. Phys.*, 2008, **128**, 134501.

[149] O. Borodin, *J. Phys. Chem. B*, 2009, **113**, 11463–11478.

[150] O. Borodin, W. Gorecki, G. D. Smith and M. Armand, *J. Phys. Chem. B*, 2010, **114**, 6786–6798.

[151] R. M. Lynden-Bell, *Phys. Chem. Chem. Phys.*, 2010, **12**, 1733–1740.

[152] I. V. Leontyev and A. A. Stuchebrukhov, *J. Chem. Phys.*, 2009, **130**, 085102.

[153] Z. Liu, T. Chen, A. T. Bell and B. Smit, *J. Phys. Chem. B*, 2010, **114**, 10692–10692.

[154] T. Cremer, C. Kolbeck, K. R. J. Lovelock, N. Paape, R. Wölfel, P. S. Schulz, P. Wasserscheid, H. Weber, J. Thar, B. Kirchner, F. Maier and H.-P. Steinrück, *Chem. Eur. J.*, 2010, **16**, 9018–9033.

[155] G. Lippert, J. Hutter and M. Parrinello, *Mol. Phys.*, 1997, **92**, 477–487.

[156] I. V. Leontyev and A. A. Stuchebrukhov, *J. Chem. Theory Comput.*, 2010, **6**, 3153–3161.

[157] A. Schüttelkopf and D. van Aalten, *Acta Crystallogr D Biol Crystallogr.*, 2004, **60(Pt 8)**, 1355–63.

[158] S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B*, 1996, **54**, 1703.

[159] C. Hardacre, S. E. J. McMath, M. T. Nieuwenhuyzen, D. Bowron and A. K. Soper, *J. Phys.: Condens. Matter*, 2003, **15**, S159–166.

[160] C. Hardacre, J. D. Holbrey, S. J. McMath, D. T. Bowron and A. K. Soper, *J. Chem. Phys.*, 2003, **118**, 273–278.

[161] B. Bhargava and S. Balasubramanian, *J. Chem. Phys.*, 2005, **123**, 144505.

[162] S. Zahn, F. Uhlig, J. Thar, C. Spickermann and B. Kirchner, *Angew. Chem. Int. Ed.*, 2008, **41**, 3639–3641.

[163] S. V. Dzyuba and R. A. Bartsch, *Chem. Phys. Chem.*, 2002, **3**, 161–166.

[164] P. G. Debenedetti and F. H. Stillinger, *Nature*, 2001, **410**, 259–267.

[165] R. Böhmer, K. L. Ngai, C. A. Angell and D. J. Plazek, *J. Chem. Phys.*, 1993, **99**, 4201–4209.

[166] R. Richert and C. A. Angell, *J. Chem. Phys.*, 1998, **108**, 9016–9026.

[167] K. Wendler, M. Brehm, F. Malberg, B. Kirchner and L. Delle Site, *J. Chem. Theory Comp.*, 2012, **8**, 1570–1579.

[168] J. Thar, M. Brehm, A. P. Seitsonen and B. Kirchner, *J. Phys. Chem. B*, 2009, **113**, 15129–15132.

[169] S. Zahn, J. Thar and B. Kirchner, *J. Chem. Phys.*, 2010, **132**, 124506.

[170] G. H. Wannier, *Phys. Rev.*, 1937, **52**, 191–197.

[171] N. Marzari, I. Souza and D. Vanderbilt, *Psi-K Scient. Highlight of the Month*, 2003, **57**, 129–168.

[172] C. Sagui, P. Pomorski, T. A. Darden and C. Roland, *J. Chem. Phys.*, 2004, **120**, 4530–4544.

[173] H. R. Schwarz, *Numerische Mathematik*, B. G. Teubener Stuttgart, 1997.

[174] P. A. Hunt, B. Kirchner and T. Welton, *Chem.–Eur. J.*, 2006, **12**, 6762–6775.

[175] C. Schröder, *Phys. Chem. Chem. Phys.*, 2012, **14**, 3089–3102.

[176] H. Tokuda, K. Hayamizu, I. Kunikazu, A. Susan and M. Watanabe, *J. Phys. Chem. B*, 2004, **108**, 16593 – 16600.

[177] H. Tokuda, K. Hayamizu, K. Ishii, M. Susan and M. Watanabe, *J. Phys. Chem. B*, 2005, **109**, 6103–6110.

[178] T. Köddermann, D. Paschek and R. Ludwig, *Chem. Eur. J. of Chem. Phys.*, 2008, **9**, 549–555.

A. PyPaTEGRO: A Python Parameter Tuning Engine for GROMACS

A.1. Introduction and Installation

PyPATEGRO is a set of Python scripts and modules which has been developed to alleviate force field parameterisation with the MD package GROMACS. It allows one to easily generate input files that are required for the simulation, analysis, and submission to a queuing system. Only a few requirements have to be met in order to use this software:

1. a Python interpreter (version 2.5 till 3.0),
2. the NumPy and SciPy packages,
3. a GROMACS distribution (version 4.4 or later), and
4. an OpenTorque queuing system.

The module is installed by extraction of the compressed source code into the `$PYTHONPATH`. Some of the module files are also executable and contained in the directory `pypategro/bin`. For this reason, the `$PATH` variable has to be set accordingly. After this steps, the tool is ready to use.

A.2. Working mechanisms

To optimize short-range parameters, an error function is minimized by a conjugate-gradient approach as described in subsection 5.2.1 detailledly. A short introduction is given here, as well as some remarks on the implementation in the code.

Construction of the total error function ϵ_{tot} For reference data, that only consists of a single value, such as the mass density or diffusion constants, the norm of the error is easily given by the relative deviation:

$$\epsilon = \frac{y - y_{ref}}{y_{ref}} = \frac{y}{y_{ref}} - 1.0. \tag{A.1}$$

Though this is a reliable error norm in most of the cases, problems arise, if the reference properties has a value of zero. Moreover for properties, which are characterized by a large set

of points, such as the radial distribution functions, an error measure that consists just of a single number is desired. To this end, the root mean square error between every point of the reference and simulation data is considered:

$$\epsilon_{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\epsilon_i^2}. \tag{A.2}$$

Furthermore, to achieve consistency with the errors of the single point properties, for all points of the data set the relative error ϵ_i is chosen, such that equation (A.2) reduces to equation (A.1), if the data set consists just of a single point.

Finally only a reasonable normalization is required. Several possibilities are implemented and determined by the function `rse_func`. One possibility, which is also applied this thesis, is to use the minimum of the absolute values of $y_{i,ref}$ which is larger than zero. This assures, that the error is compatible to cases with $y_{i,ref} \gtrsim 0$, to maintain the relative character of the errors and it is also applicable to functions, that may have negative values, which is not the case for the RDF. For the RDF also the possiblity exists to leave out the normalization, because it converges to 1, which can used as normalization factor. So far, one has to comment and uncomment the function `rse_func` in order to use the different normalization methods for equation (A.2).

Finally, the total error ϵ_{tot}, which has to be minimized, is defined by the sum of the squares of all single errors for the reference properties α:

$$\epsilon_{tot} = \sum_{\alpha}\epsilon_{\alpha}^2. \tag{A.3}$$

For the single value properties ϵ_α is given by equation (A.1) and equation (A.2) applies to properties that are defined by a whole set of data points.

Adaptive step width control Since the error minimization routine is based on the derivatives, which are obtained by the finite differences method, the estimate of the error depends on the accuracy, heavily. While a too large step width δ decrease the accuracy of the derivative at the applied point $\boldsymbol{p}_0$ in the parameters space $\mathfrak{P}^k$, a decrease of δ shrinks the range of the validity of the error expansion around $\boldsymbol{p}_0$. For this reason, an adaptive step width control is applied, which automatically decreases δ for the finite differences method and keeps the change of the parameters within given bounds, to assure that the error estimate is applicable. Hence, if the derivatives are accurate enough, the error should decrease monotonically. However, δ may also not become to small, because the risk of being trapped in a local minimum increases with decreasing δ. To improve the sampling of the error function in the high-dimensional parameter space, δ is adapted for the subsequent simulations, corresponding to the changes of the parameters.

Order of the finite differences scheme In PyPATEGRO, a first– and a second–order finite differences schemes can be applied to obtain the derivatives for the Taylor expansion. The choice is controlled by the option `-n` of the main script `ppgrun.py`.

A.3. Input files

A.3.1. The ppg input file

The `ppg` input file controls almost all options for PyPATEGRO. It is divided into different sections that are described in the following. One always has to bear in mind, that if an option is defined twice, the latter one will be chosen without warning.

`[ defaults ]` This section controls the main setup for the simulation environment.

- `ppn_apbs`: processor per node for analysis
- `wallt_pbs`: walltime for analysis
- `rdf_ndx`: GROMACS indexfile for RDF analysis
- `ppdir`: directory of PAPAteGRO and templates
- `wdir`: base directory for the runs and location of index and mdp files
- `source`: optional argument, if given a corresponding line will be inserted into the queue scripts
- `dt`: restart time for MSD calculation and dt for g_rdf
- `dtconv` (0.0): time interval for frames of the converted tractory, containing only whole molecules, 0.0 corresponds to every frame
- `max_delta` (1.0): maximum change of the parameters in units of the stepwidth; if set to -1, the parameters are adapted without restrictions
- `qprefix` (PBS): Different queuing system usually allow to set the job options via command lines starting with a special prefix, such as `#PBS` for the OpenTorque system. `qprefix` allows to choose the prefix behind the hash symbol.
- `mdp`: a GROMACS mdp file that determines the main simulations parameters such as cut-offs, PME options, saving frequency of the trajectories, ...; the file has to be fully functional and only some options will be changed by PyPATEGRO
- `nmol_ene`: number of molecules/ion pairs contained in the system; important for the calculation of energies in mol

`[ templates ]` Template files for the simulations and analysis for the OpenTorque queueing system that **allows to submit jobs from the compute nodes**.

- `simulation`: simulation template, usually `gmx.pbs.templ`
- `analysis`: analysis template, usually `analysis.apbs.templ`

`[ exp_fit ]` Polynomial coefficients for a temperature dependency of the properties density and MSD can be provided and the reference data is chosen from this function if no value is

given in section `[ reference ]`. The polynomial will be of the form:

$$f(T) = \sum_{i=0}^{n} a_i T^i, \tag{A.4}$$

and the order depends on the number n of given parameters. The temperature T is assumed to be in K. For the MSD also the molecule name has to be set as given in the GROMACS topology file.

- `density` a_0 a_1 ...
- `msd molecule` a_0 a_1 ...

`[ fit_dihedrals ]` If a fit of the dihedral interactions as described in section 5.1 should be performed, the molecule names `mol` and the reference topologies `molref.top` have to be specified in this section for every molecule separately:

```
mol molref.top
```

If the section is not given, the fit will not be performed.

`[ atomtypes ]` This section controls the tuning of the Lennard-Jones parameters ϵ and σ. Corresponding to the naming in the GROMACS topology file, the atomtypes, kind of LJ parameter (`eps` or `sig` for ϵ and σ, respectively), and stepwidth for the finite differences method in units of parameters.

```
Cl      sig     0.05
Cl      eps     0.1
```

If no parameters are provided in the section input files for simulations with the provided topology files are generated, which is very useful for the preparation of validation runs.

`[ mdrun ]` Options for `mdrun` and `mpiexec`

- `-np`: number of processors
- `-npme`: number of dedicated PME nodes
- `-maxh`: maximal runtime for the queue

`[ mdp ]` Common options for the simulations.

- `ref-t`: temperature for the different simulations
- `ref-p`: pressure for the simulations in corresponding order to the temperatures. If not given, a pressure of 1 bar will be used.
- `nsteps_em`: number of steps for the energy minimization

- `emstep`: maximal step size for the steepest descent energy minimization step
- `emtol`: if the maximal force in the system is smaller than this value (force is given in kJ/(mol nm) the energy minimization will terminate
- `nsteps_eq`: number of steps for the equilibration stage
- `nsteps_md`: number of steps for the production stage
- `nsteps_hvap`: number of steps for the vacuum calculation if the heats of vaporization should be calculated
- `equi_time`: fraction of the production run, which should be omitted for the analysis

`[ simulation ]`

- `top_include`: files which are included in the topology files. If the files are not found in the given path, which is assumed to be relative to the `ppg` input file, also the `topol` directory relative to the `ppg` input file will be scanned.
- `structure`: structure files with a start configuration

`[ reference ]` So far, PyPATEGRO is able to include the mass density and RDFs into the error function for an optimization of the SR interactions. The input in this section is always organized as follows. In the first column, a keyword followed by a number without space determines which property should be calculated from the simulation. If the same property for different temperatures are involved the numbers have to differ, otherwise only the last input line will be considered. In the second column the weight for the error of this property in the error function is set. The third column always determines the temperature for which the property should be analyzed. In the following columns property specific options are set, apart from the last column, which provides the reference data for the error function. The input for the mass density and RDF is organized as follows:

- Density
  - reference values are expected in SI-units, kgm^{-3}
  - sample input line

    ```
    density1  <weight> <temperature>  <reference value>
    ```

- for the calculation of the RDF between two groups a sample input line looks like:

  ```
  rdf1 <weight> <temperature> <group1> <group2> <reference file>
  ```

- `group1` and `group2` correspond to the group numbers in the index file, which is set in the defaults section, starting with 0 and should have a certain content structure as described in subsection A.3.2.
- the range where the error will be calculated is determined by the range provided in the `reference file`

Apart from parameter optimization, PyPATEGRO also is able to create a set of simple MD simulations at different temperatures and pressures. The following options for the refer-

ence section prepare the data for an easy calculation of different properties. In the future, PyPATEGRO will be extended such that these properties can also be included in the error function. Therefore the option for the the weight and reference data is already included, but not of use.

- MSD
  - the reference values has to be in units of $10^{-5}\,\mathrm{cm}^2/\mathrm{s}$
  - sample input line:

    ```
    msd1 <weight> <temperature> <molecule> <nmol> <reference value>
    ```

  - the option `molecule` determines which molecule species should be considered, an the naming has to be consistent with the name given in the topology
  - `nmol` sets the number of molecules which have to be analyzed
  - the ouput of the analysis is in the directory `g_msd`, which is a subdirectory of the simulation directory for a certain temperature
  - the tool `calc_msd.py` allows to fit the data and obtain the diffusion constants
- Conductivity
  - derived from fit of straight line to MSD of translational current → Einstein-Helfand method
  - the SI-unit for the conductivity, Sm^{-1}, is expected as reference value
  - sample input line:

    ```
    cond1 <weight> <temperature> <nshift> <reference value>
    ```

  - the parameter `nshift` determines after how many frames a new start point for the averaging should be set
  - the tool `calc_eh.py` allows to fit a straight line to the results, given in the file `dsp.xvg` which can be found in the subdirectory `g_c_xtc` of the simulation directory
- Heat of vaporization
  - derived from the enthalpy difference of the liquid and gas phase
  - ΔH_{vap} is derived in kJmol^{-1}
  - sample input line:

    ```
    hvap1 <weight> <temperature> <conf> <q_liq> <q_gas> <reference value>
    ```

  - for the calculation of the enthalpy in the gas phase, a corresponding configuration `conf` is required
  - if the charging of the liquid and gas phase is very different, the respective net–charges can be set with `q_liq` and `q_gas`, respectively
  - the tool `hvapor.py` postprocesses the files `liq_enthalpy.xvg` and `vac_energy.xvg`, which are contained in the subdirectory `g_energy` of the simulation direcory, and

derives the heats of vaporization

A.3.2. Indexfiles

For some kinds of analysis, index files are required for the corresponding GROMACS tools. These have to contain a certain structure, which is described in the following

RDF For the calculation of the RDF, the index file has to start with following groups:

1. system,
2. cation,
3. anion,

but afterwards the groups can be in arbitrary order.

MSD For the calculation of the MSD, index files are required for the different molecule species. They have to be named like the molecule as given in the topology file with the suffix `.ndx`. Each index group has to contain a single molecule. Thus for a system of n molecules, n groups should be contained in the index file.

A.4. Example for a working cycle

The main script of PyPATEGRO is `ppgrun.py`. Depending on the provided options and input, it is able to generate a set of simulations, analyze the error from reference data and minimize the error function. For the generation of the simulation set, a PyPATEGRO input file `ppg.in` is needed, which controls various options as described above, as well as a GROMACS topology file `topol.top`. For this example, it is assumed that we are in the working directory, determined by the option `wdir`.With a command like:

```
ppgrun.py -i ppg.in -t topol.top -n 2 -m generation -o run1
```

the directories `run1a` and `run1b` are created containing several subdirectories depending on the parameters which should be tuned. The option `-n 2` determines that a second-order finite differences scheme should be applied for the error minimization. If the standard templates for the simulation and analysis scripts, which are submitted to the queueing system. For the submission of the simulations, the script `submit_run.sh` has been prepared and the generated runs are started like this:

```
submit_run.sh run1a
submit_run.sh run1b
```

The simulation is divided into three parts. At first an energy minimization is performed, followed by an equilibration run. Finally the production run is automatically submitted.

When a simulation has finished the analysis script is submitted which finally processes the trajectory and derives the desired data. If everything has finished the parameters can be optimized.

```
ppgrun.py -i ppg.in -n 2 -m analysis -d run1 -o run2
```

Now `ppgrun.py` is analyzing the data and writes a logfile `run1.log` and a topology file `run1.top` with the optimized parameters in the directory `run1a`. Finally a new set of simulations based on the optimized parameters in `run1.top` is generated in the directories `run2a` and `run2b`. Thus everything is ready for the next iteration step. If the option `-o` is omitted no new output directories will be created. Examples of input files can be found in the directory `pypategro/examples`.

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Stuttgart, 20.12.2012 Florian Dommert