Quantitatively reliable massively-parallel molecular modelling and simulation of vapour-liquid interfaces

M. Horsch, K. Stöbener, S. Werth, P. Klein, K.-H. Küfer, H. Hasse

1 Laboratory of Engineering Thermodynamics, University of Kaiserslautern
2 Fraunhofer Institute for Industrial Mathematics, Kaiserslautern

Kanpur, 30th December 2015
IITK Chemical Engineering Seminar
Computational molecular engineering

From Physics (qualitative accuracy)

- Physically realistic modelling of intermolecular interactions
- Separate contributions due to repulsive and dispersive as well as electrostatic interactions

To Engineering (quantitative reliability)

- No blind fitting, but parameters of effective pair potentials are adjusted to experimental data
- Physical realism facilitates reliable interpolation and extrapolation

Bottom up 🔄

Top down ←
Force-field models of low-molecular fluids

Geometry
Bond lengths and angles

Dispersion and repulsion
Lennard-Jones potential:
Size and energy parameters

Electrostatics
Point polarities
(charge, dipole, quadrupole):
Position and magnitude
Vapour-liquid equilibria

Bulk properties

Interfacial properties

vapour pressure, saturated densities, composition, enthalpy of vaporization, etc., by Grand Equilibrium simulation

heterogeneous systems with finite-size effects and long-range interactions
Vapour-liquid equilibria: Grand equilibrium

- temperature
- liquid composition
- partial molar volumes
- chemical potentials
- pressure
- vapour composition
Vapour-liquid equilibria: Grand equilibrium

**Given:** Temperature $T$, liquid composition $x$

**First step:** $NpT$ simulation of the liquid phase

An estimate, which may deviate from $p^{\text{sat}}(T)$, is used for $p$ in this simulation. The chemical potential and its first and second derivatives with respect to pressure are determined.

**Second step:** Pseudo-$\mu VT$ vapour simulation

Grand-canonical simulation where the value of $\mu$ is determined on the fly from the pressure.

**Obtained:** Pressure $p$, vapour composition $y$
Concurrency of multiple Markov chains

strong scaling of for MC simulation

runtime per MC loop / s

number of parallel processes

5488 molecules
2744 molecules
1372 molecules
**Thermodynamic properties of bulk fluids**

Second virial coefficient

**Thermal conductivity**

Self-diffusion coefficient

Shear viscosity

---

*ms2 is freely available for academic use: register at www.ms-2.de*

---

Vapour-liquid equilibria: Saturated densities and vapour pressures
Vapour-liquid interfaces

Bulk properties

\[ T \]

\[ \rho \]

Interfacial properties

vapour pressure, saturated densities, composition, enthalpy of vaporization, etc., by Grand Equilibrium simulation

heterogeneous systems with finite-size effects and long-range interactions
Molecular dynamics of fluids at interfaces

Interfacial properties

heterogeneous systems with finite-size effects and long-range interactions
Molecular dynamics of large systems

The development of *ls1 mardyn* is jointly driven by
- [High Performance Computing Center Stuttgart (HLRS)](http://www.hlrs.de), University of Stuttgart,
- [Laboratory for Engineering Thermodynamics (LTD)](http://www.ls1-mardyn.de/), University of Kaiserslautern,
- [Scientific Computing in Computer Science (SCCS)](http://www.tum.de), Technische Universität München,
- [Thermodynamics and Energy Technology (ThEt)](http://www.tu-dortmund.de), University of Paderborn,

under the auspices of the Boltzmann-Zuse Society for Computational Molecular Engineering (BZS).

Please cite the work of Niethammer et al. (2014), *J. Chem. Theory Comput.*, 10: 4455, in all publications containing the results of MD simulations with the *ls1 mardyn* program.

The development team can be contacted via the *ls1 mardyn* contact point at the University of Kaiserslautern.

[http://www.ls1-mardyn.de/](http://www.ls1-mardyn.de/)
Parallelization by volume decomposition

Linked-cell data structure suitable for spatial domain decomposition:

Methods for heterogeneous or fluctuating particle distributions:

Large systems “1”: molecular dynamics

http://www.ls1-mardyn.de/
Hyperthreading and vectorization

Memory-efficient implementation based on the linked-cell data structure:

Optionally, forces acting on molecules are only stored until their cell leaves the sliding window.

Efficient vectorization:
- Optimization by hand, using advanced vector extensions (AVX).
- Conversion from array of structures (AoS) to structure of arrays (SoA).

large systems “1”: molecular dynamics

http://www.ls1-mardyn.de/
World record with ls1 mardyn on SuperMUC

Weak scaling with 31.5 million molecules per core

Up to $N = 4 \cdot 10^{12}$ on SuperMUC

Large systems “1”: molecular dynamics

http://www.ls1-mardyn.de/

30th December 2015
Martin Horsch, Katrin Stöbener, Stephan Werth, Peter Klein, Karl-Heinz Küfer, and Hans Hasse
MD simulation of homogeneous cavitation

liquid CO$_2$ at 220 K and 22.6 mol/l

Scale-up to the entire *hermit* cluster for canonical simulation of cavitation in carbon dioxide.

Evaluation of local density at 180 x 180 x 180 grid points:

$\leq 5$ ?

(6 mol/l)

Liquid phase detected for more than 5 neighbours within a radius of 6.9 Å around the grid point.

13 x 10$^6$ molecules (52 x 10$^6$ sites)
MD simulation of homogeneous cavitation

liquid CO$_2$ at 220 K and 23.9 mol/l

Three consecutive regimes:

- relaxation (equilibration)
- homogeneous cavitation
- growth beyond critical size

These and other simulations of interfacial phenomena crucially depend on an accurate surface tension.

13 x 10$^6$ molecules (52 x 10$^6$ sites)
Long-range correction at planar interfaces

For planar interfaces:

**Long-range correction** from the density profile, following Janeček.

Full evaluation of all pairwise interactions is too expensive ...

... **short-range interactions** are evaluated only for **neighbours**.
Long-range correction at planar interfaces

For planar interfaces:

**Long-range correction** from the density profile, following Janeček.

**Angle-averaging** expression for multi-site models, following Lustig.

Dipole and dispersion lead to analogous long-range correction expressions. The long-range contribution of the quadrupole can be neglected.
The density at the centre of a liquid nanofilm deviates from that of the saturated bulk liquid at the same temperature (scaling $\sim 1 / d^3$).
The deviation of the surface tension of a nanofilm from the macroscopic value exhibits the same tendency (scaling with $1/d^3$).
Surface tension at high precision

Lennard-Jones fluid

\[ \gamma_0(T) = 2.94 \frac{\varepsilon}{\sigma^2} \left(1 - \frac{T}{T_c}\right)^{1.23} \]
Validation of molecular force field models

2CLJQ models:
- 2 LJ centres
- 1 quadrupole

Fit of parameters $\sigma$, $\epsilon$, $L$, $Q$ to VLE data of 29 fluids by Stoll et al.

Deviation:
- $\delta \rho' \approx 1\%$
- $\delta P_{\text{sat}} \approx 5\%$

No interfacial properties were considered for the parameterization.
Validation of molecular force field models

2CLJD models:
- 2 LJ centres
- 1 dipole

Fit of parameters $\sigma$, $\varepsilon$, $L$, $\mu$ to VLE data of 46 fluids by Stoll et al.

Deviation:
- $\delta \rho' \approx 1\%$
- $\delta P^{\text{sat}} \approx 5\%$

present simulations ○
models by Stoll et al. X
other literature models +
Massively parallel molecular modelling

Two LJ + quadrupole (2CLJQ)

Model parameters:
- LJ size parameter $\sigma$
- LJ energy parameter $\varepsilon$
- Elongation $L$
- Quadrupole moment $Q$

- Systematic exploration of the four-dimensional model parameter space
Massively parallel molecular modelling

- Systematic exploration of the four-dimensional model parameter space
- Correlation of the surface tension by a critical scaling expression
Validation of molecular force field models

Two LJ + quadrupole (2CLJQ)

Two LJ + dipole (2CLJD)

Fit to bulk properties

10 to 20 % overestimation of vapour-liquid surface tension
Model optimization with multiple objectives

ethylene oxide model by Eckl et al. (2008)

- sat. liquid density
- sat. vapour density
- second virial coefficient
- vapour pressure
- enthalpy of vapourization
- normal boiling temperature
- critical density
- critical temperature
- sat. liquid isobaric heat capac.
- sat. vapour isobaric heat capac.
- sat. liquid isothermal compress.
- sat. vapour isothermal compress.
- surface tension
- sat. liquid shear viscosity
- sat. vapour shear viscosity
- sat. liquid thermal conductivity
- sat. vapour thermal conductivity

Graph showing deviation from reference values for various properties with uncertainty of reference and molecular model.
Multicriteria optimization requires massively parallel molecular modelling.
Multicriteria optimization

Literature example: LDPE synthesis


- \( N_{\text{syn}} = 400 \)
- \( N_{\text{syn}} = 600 \)
- \( N_{\text{syn}} = 700 \) (Reference case)
- \( N_{\text{syn}} = 800 \)
- Operating point

(11 design parameters)

Monomer conversion fraction vs. amount of side products

(2CLJQ for carbon dioxide)

Multiple objectives

High-dimensional parameter spaces require stochastic exploration methods.
Parameter space and objective space

\( p \) model parameters

(here, \( p = 4 \))

- LJ size parameter \( \sigma \)
- LJ energy parameter \( \varepsilon \)
- Model elongation \( L \)
- Multipole moment \( \mu \) or \( Q \)

Dimension of Pareto set \( d \leq p \).

\( q \) optimization criteria

(here, \( q = 3 \))

- Saturated liquid density \( \rho' \)
- Saturated vapour pressure \( \rho^s \)
- Vapour-liquid surface tension \( \gamma \)

Dimension of the Pareto set cannot be greater than \( q - 1 \).

In general, \( d = \min(p, q - 1) \).

(here, \( d = 2 \))
Computation of the Pareto set

Multicriteria optimization problem

Simultaneously minimized objective functions $f_{\xi}$ with $\xi \in \{\rho', p^s, \gamma\}$ given by

$$f_{\xi} = \left\langle \delta \xi^2 \right\rangle_{0.55T_c^{\text{exp}} < T < 0.95T_c^{\text{exp}}} = \lim_{N \to \infty} \frac{1}{N+1} \sum_{i=0}^{N} \left(1 - \frac{\xi_{\text{sim}}(T)}{\xi_{\text{exp}}(T)}\right)^2 T/T_c = 0.55 + 0.4i/N$$

(here: $N = 9$).

Sandwiching

Alternating construction of inner (reachable) and outer (unreachable) approximations, assuming local convexity of the Pareto set.

Hyperboxing

In non-convex regions ("hyperboxes"), Pascoletti-Serafini scalarization is employed to obtain a suitable local single-criterion optimization problem.
Invariants of Pareto-optimal models

For obtaining a rough approximation of the Pareto set, the dimension of the parameter space can be reduced from four to three (or even two).
Model tailoring by the end user

Representation of objective and parameter spaces by patch plots:

Pareto-optimal 2CLJQ models of molecular oxygen
Model tailoring by the end user

For each specific application, accuracy requirements can be specified:

Restrictions imposed on 2CLJ models of molecular oxygen
Model tailoring by the end user

Intersection of the highlighted areas within all replicas of the patch plot:

2CLJ models of molecular oxygen fulfilling all requirements
Overall compromise models: Pareto knee

A single-criterion optimization algorithm would often miss the Pareto knee.
Comparision between model classes

Carbon dioxide: Mie–$n,6$ potential

$$u(r) = \frac{n}{n-6} \left( \frac{n}{6} \right)^{6} \left( \frac{\sigma}{r} - \left( \frac{\sigma}{r} \right)^{6} \right)$$

parameter space

objective space
Comparision between model classes

Carbon dioxide: Mie–\( n,6 \) potential ./ other model classes

![Graph showing deviation of saturated liquid density (\%) vs. deviation of vapour pressure (%)](image)

- **LJ potential**
  - (2 parameters)

- **Mie–\( n,6 \) potential**
  - (3 parameters)

- **2CLJQ potential**
  - (4 parameters)

- **Vrabec et al. (2001)**
  - 2CLJQ model

- **Avendaño et al. (2011)**
  - Mie–\( n,m \) model
Summary

The traditional art of molecular modelling

An expert modelling artist designs and publishes

- a single optimized model for a particular fluid,
- according to his choice of criteria (often unknown to the public),
- users are passive, they have to live with the artists' decision.
Summary

The traditional art of molecular modelling

An expert modelling artist designs and publishes
  • a single optimized model for a particular fluid,
  • according to his choice of criteria (often unknown to the public),
  • users are passive, they have to live with the artists' decision.

Scientific modelling by multicriteria optimization

For established model classes and multiple thermodynamic criteria,
  • the dependence of thermodynamic properties on the model parameters is determined and correlated,
  • the deviation between model properties and real fluid behaviour is characterized, and the Pareto set is published,
  • users can design their own tailored model with minimal effort.