Computational molecular engineering: Scalable simulation and reliable modelling

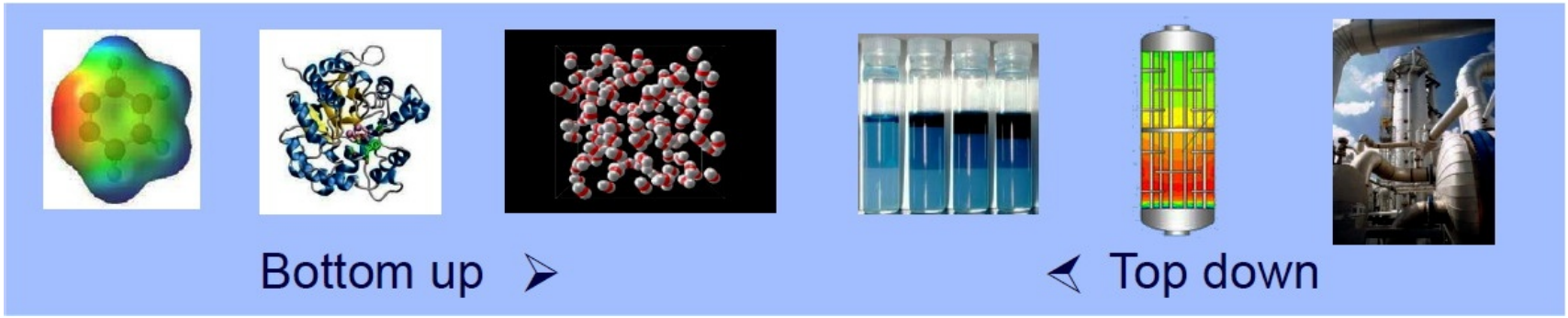
Martin Horsch

Laboratory of Engineering Thermodynamics, University of Kaiserslautern

Computational Molecular Engineering

Tianjin, 16th November 2015
Tianjin Center for Applied Mathematics

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

Force fields for molecular modelling

Geometry

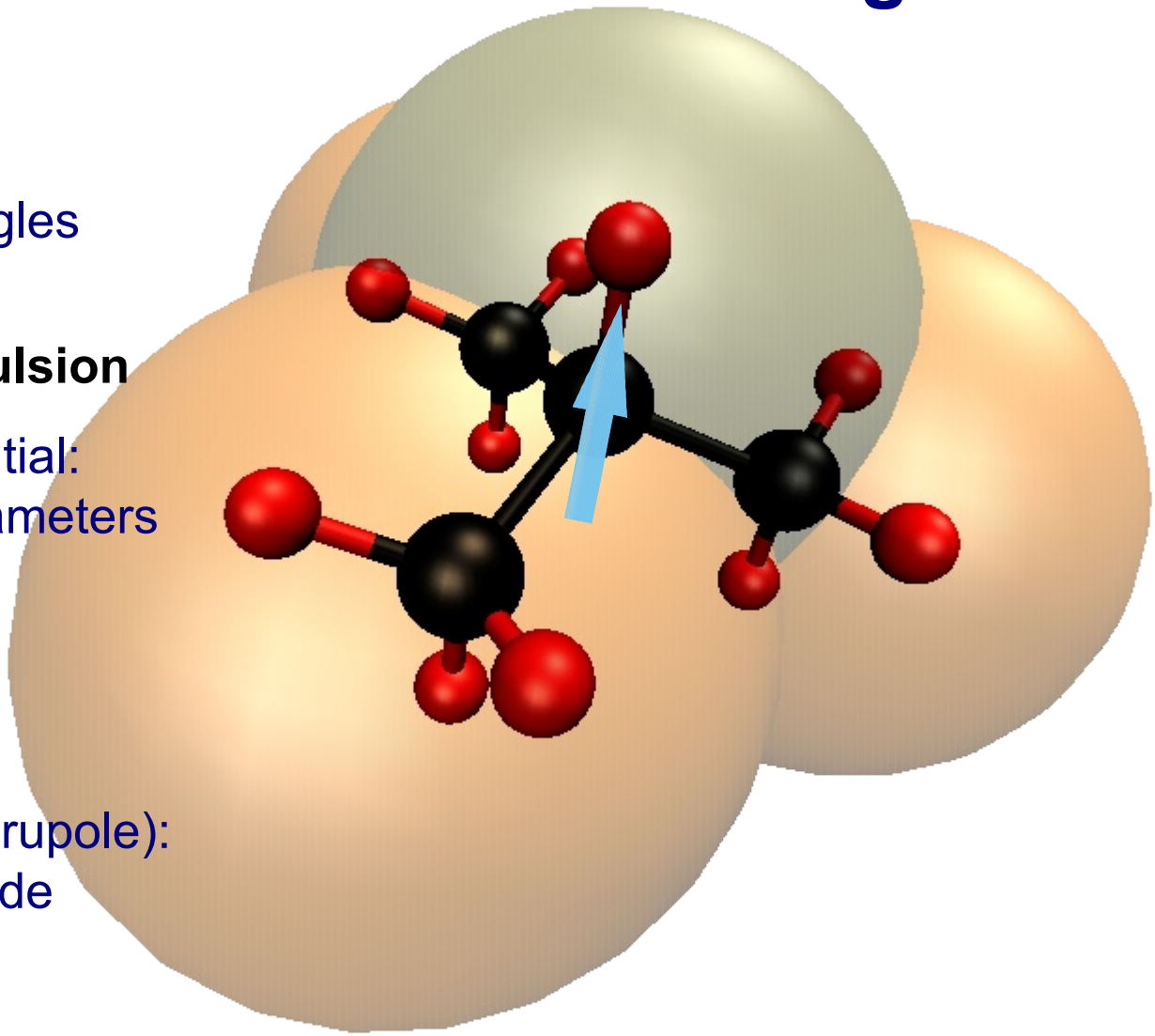
Bond lengths and angles

Dispersion and repulsion

Lennard-Jones potential:
Size and energy parameters

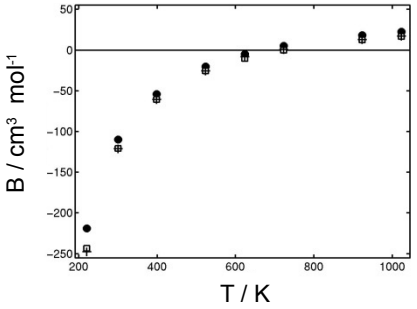
Electrostatics

Point polarities
(charge, dipole, quadrupole):
Position and magnitude

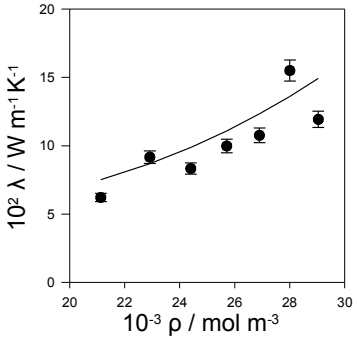


Simulation of bulk properties with ms2

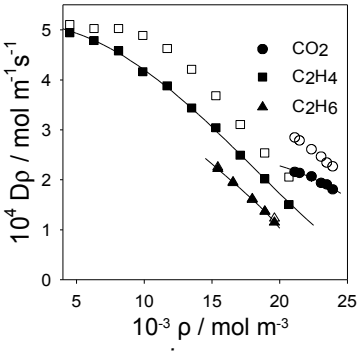
Second virial coefficient



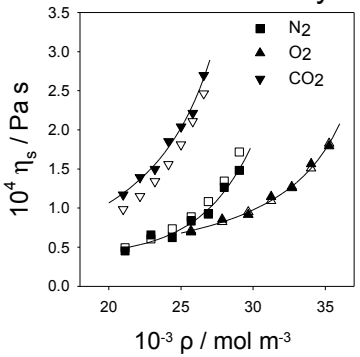
Thermal conductivity



Self-diffusion coefficient

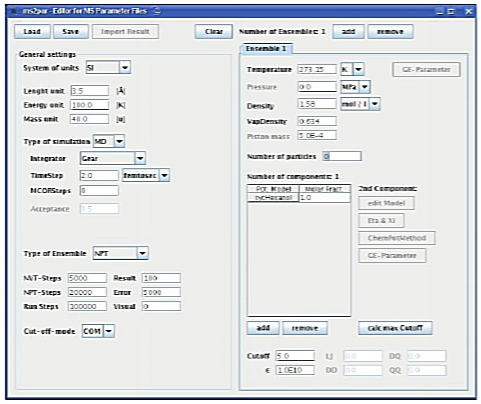


Shear viscosity



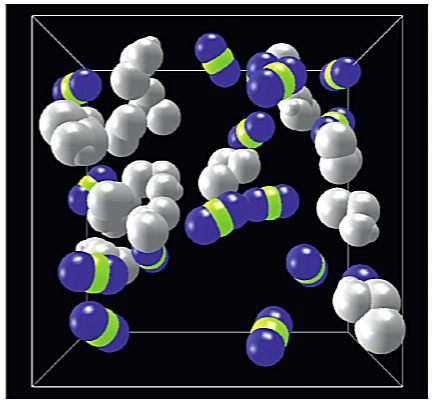
ms2 is freely available for academic use – register at http://www.ms-2.de/

ms2par



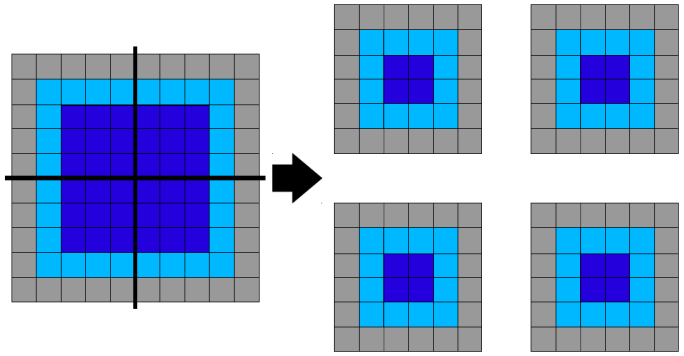
ms2chart

ms2molecules



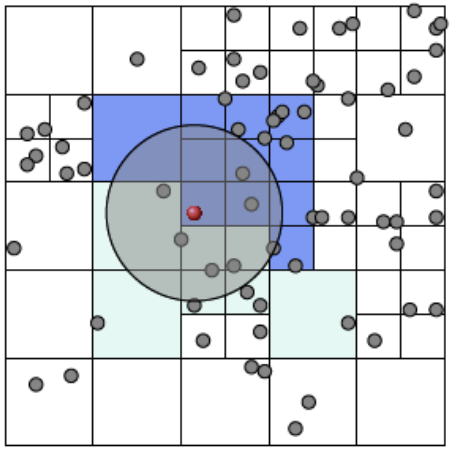
Efficient simulation of large systems

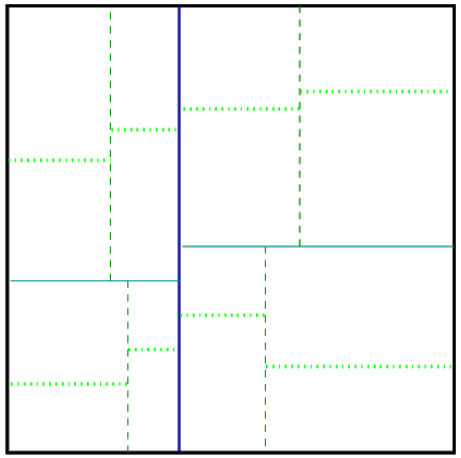
Linked-cell data structure suitable for spatial domain decomposition:



(non-blocking, over-lapping MPI send/ receive operations)

Methods for heterogeneous or fluctuating particle distributions:

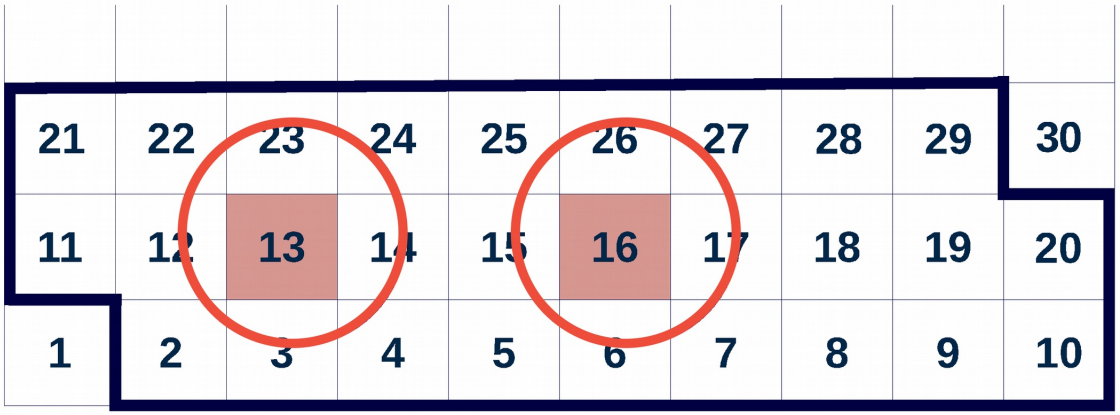




large systems “1”: molecular dynamics http://www.ls1-mardyn.de/

Efficient simulation of large systems

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



hyperthreaded sliding window

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

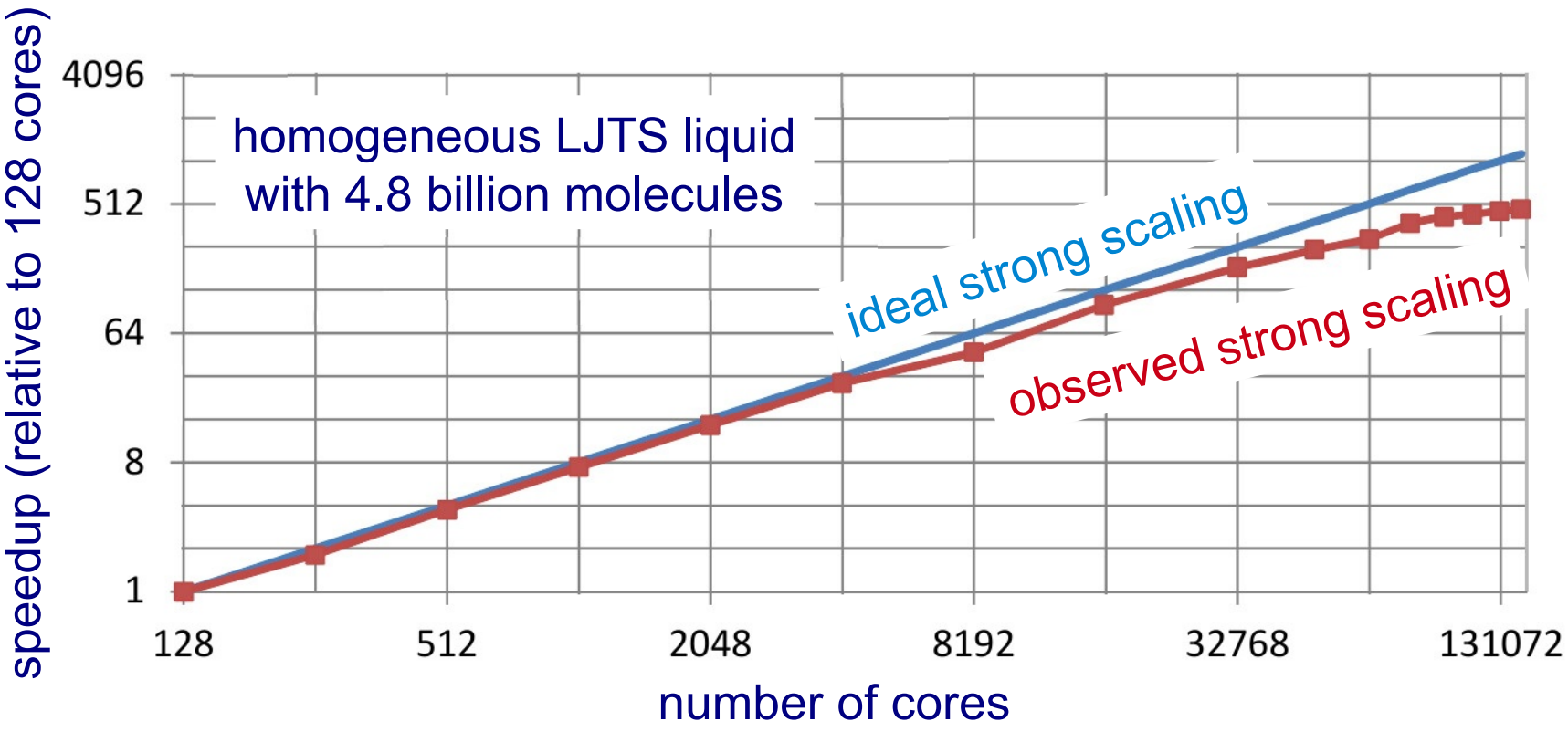
Efficient vectorization:

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

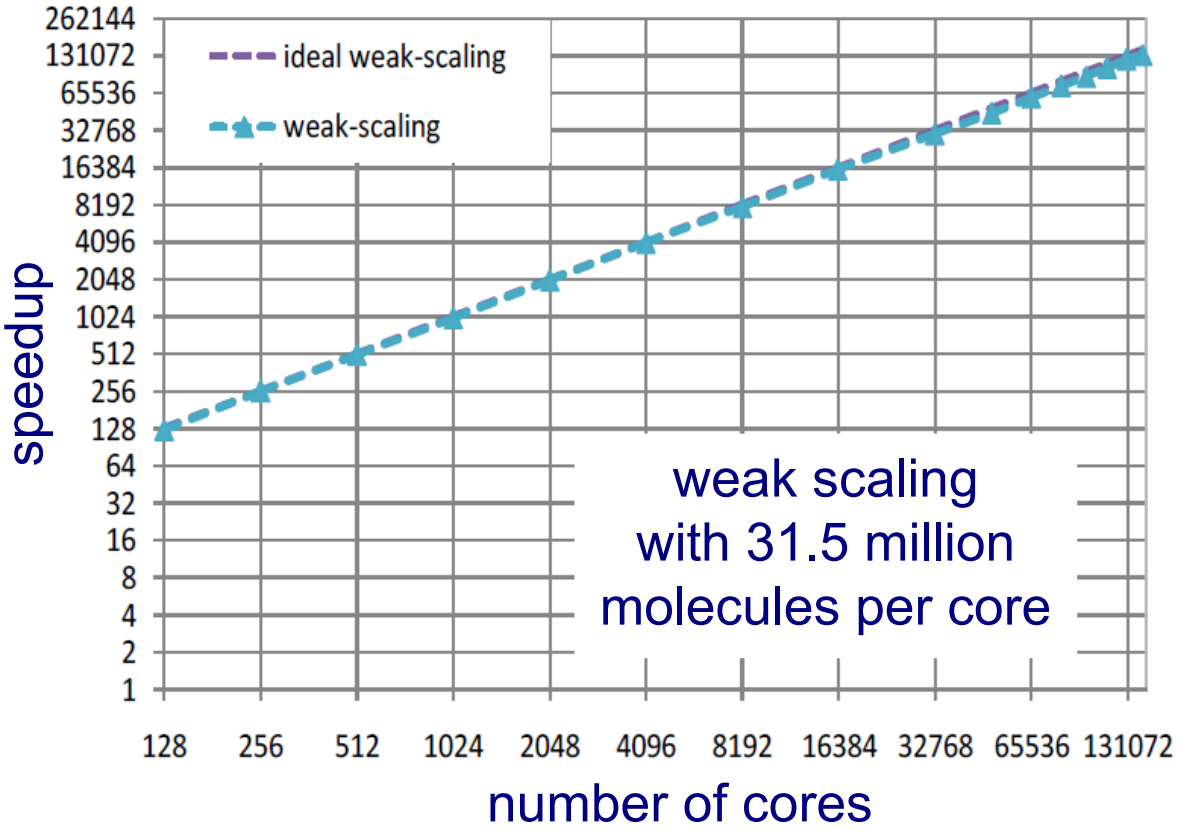
large systems “1”: molecular dynamics **http://www.ls1-mardyn.de/**

Large-scale MD simulation on SuperMUC

Scaling of *ls1 mardyn* examined on up to 146 016 cores, i.e. the whole SuperMUC at the Leibniz Supercomputing Centre, Garching, in 2013.



Large-scale MD simulation on SuperMUC



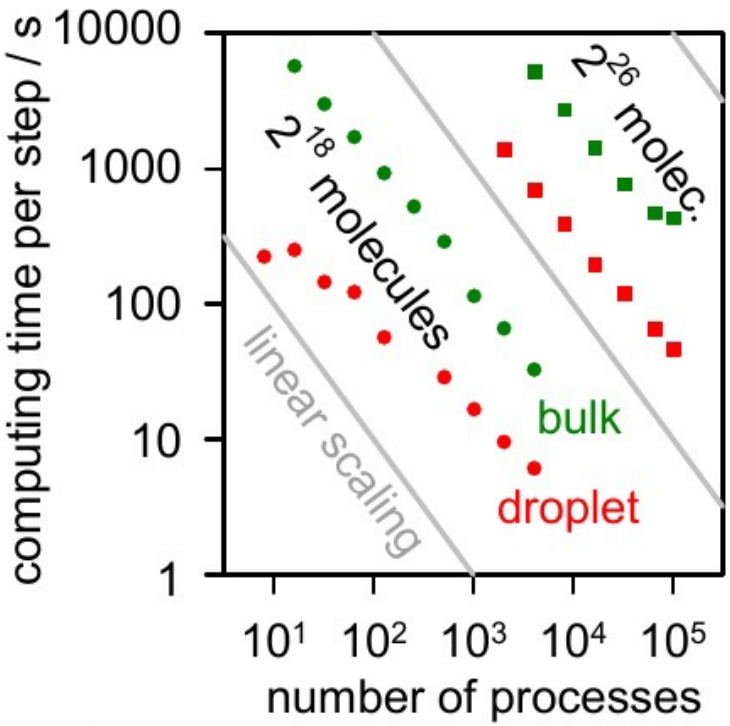
Up to $N = 4 \cdot 10^{12}$ on SuperMUC

large systems "1": molecular dynamics

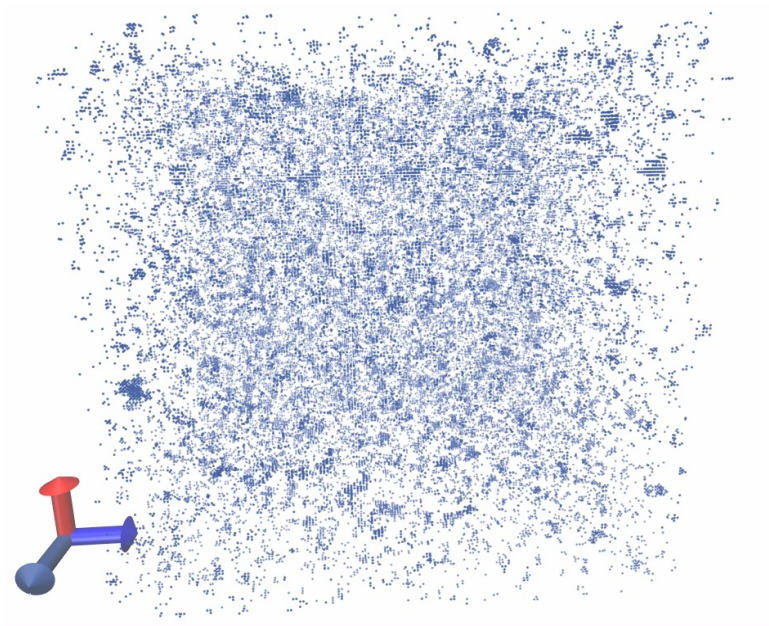
http://www.ls1-mardyn.de/

Large-scale MD simulation on hermit

strong scaling (Amdahl)



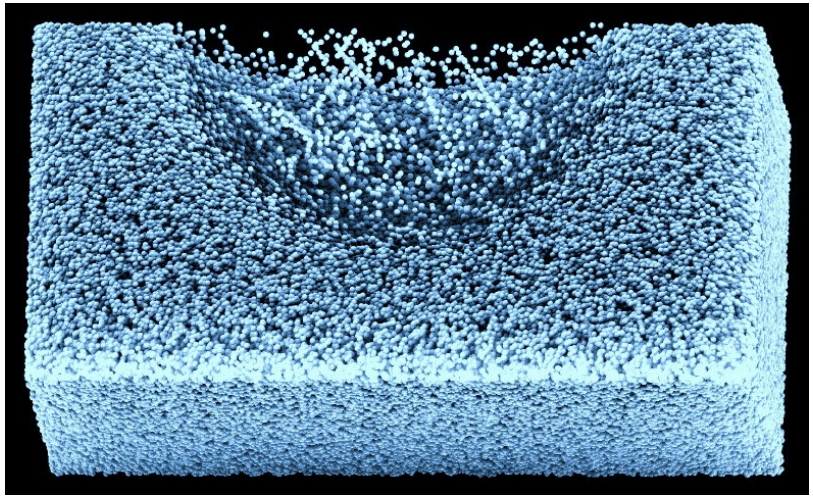
homogeneous cavitation

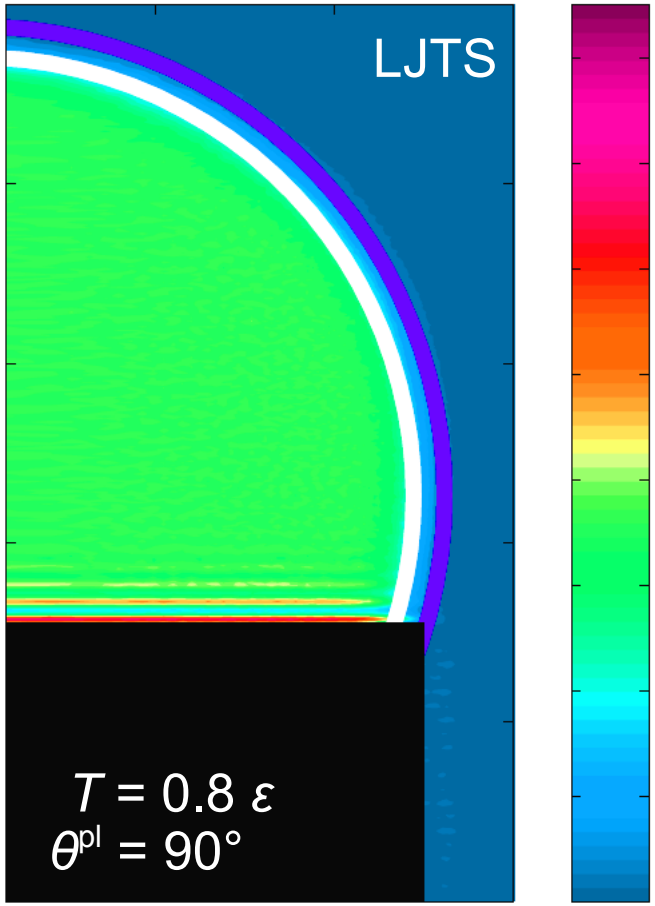


CO_2 (T = 280 K and ρ = 17.2 mol/l), 3CLJQ
100 million interaction sites, 110 592 cores

Molecular simulation of fluids at interfaces

- Vapour-liquid surface tension
- Nucleation and dispersed phases
- Adsorption (fluid-fluid and fluid-solid)
- Contact angle and contact line pinning

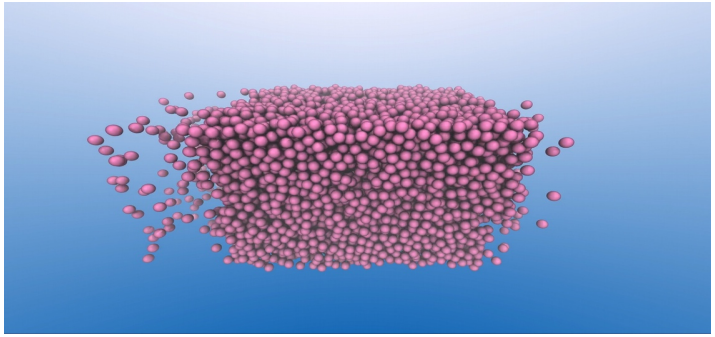


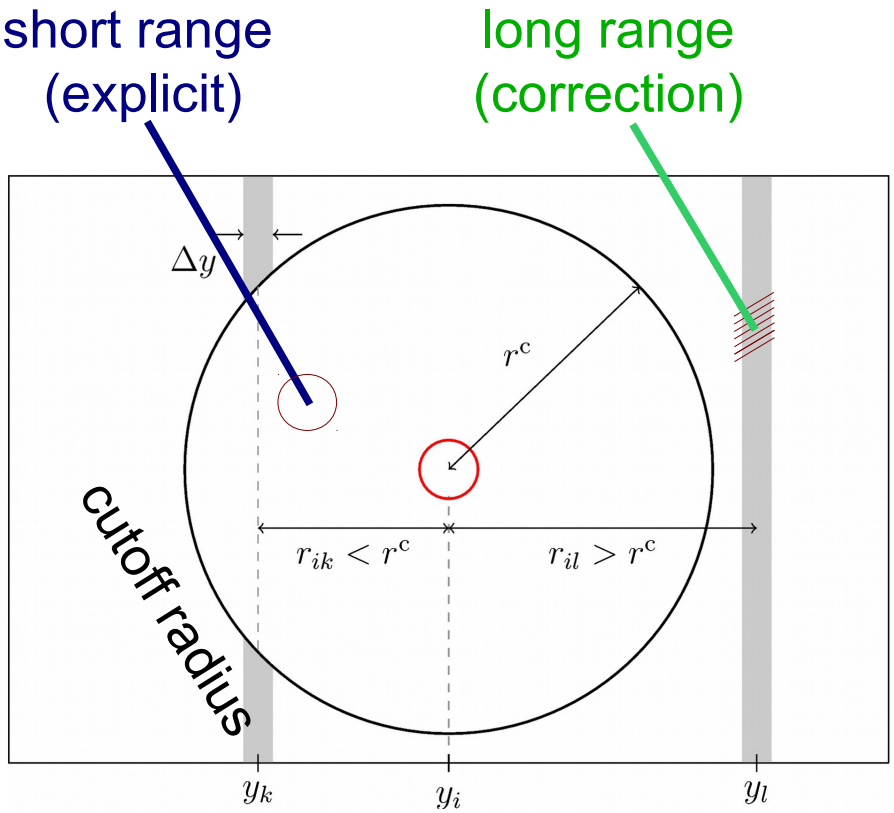


Long-range correction at planar interfaces

For planar interfaces:

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



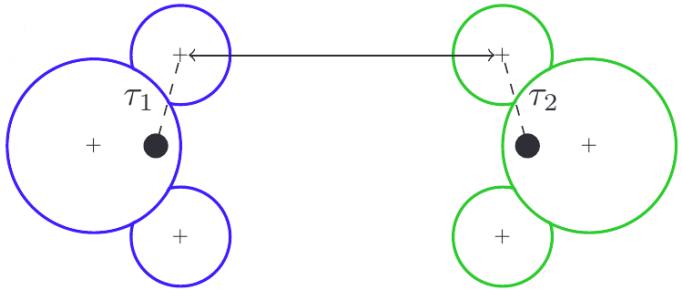


Full evaluation of all pairwise interactions is too expensive ...
... **short-range interactions** are evaluated only for **neighbours**.

Long-range correction at planar interfaces

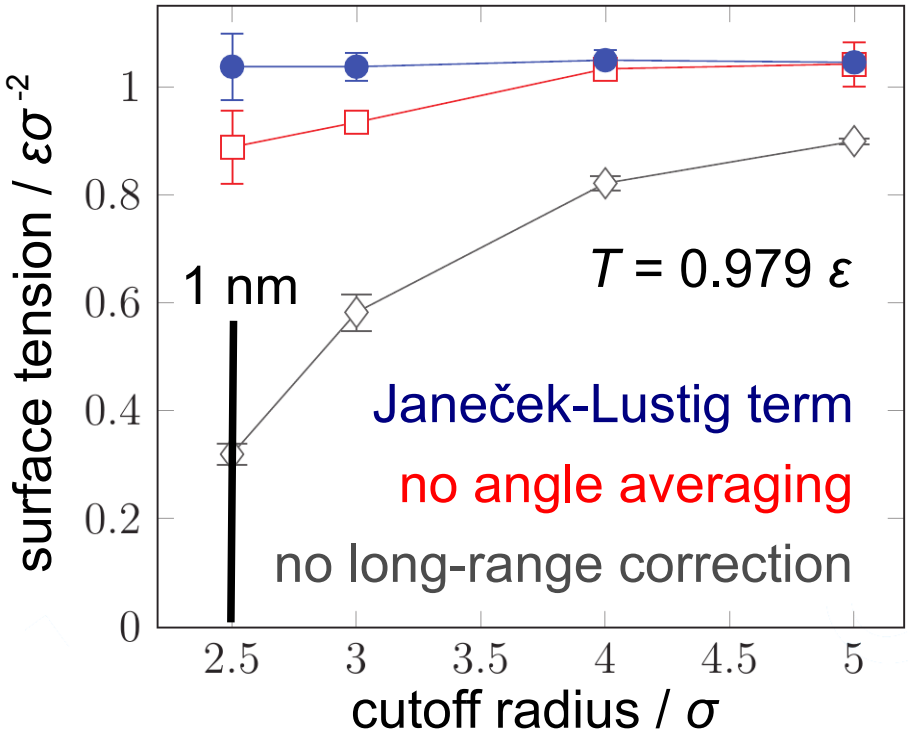
For planar interfaces:

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



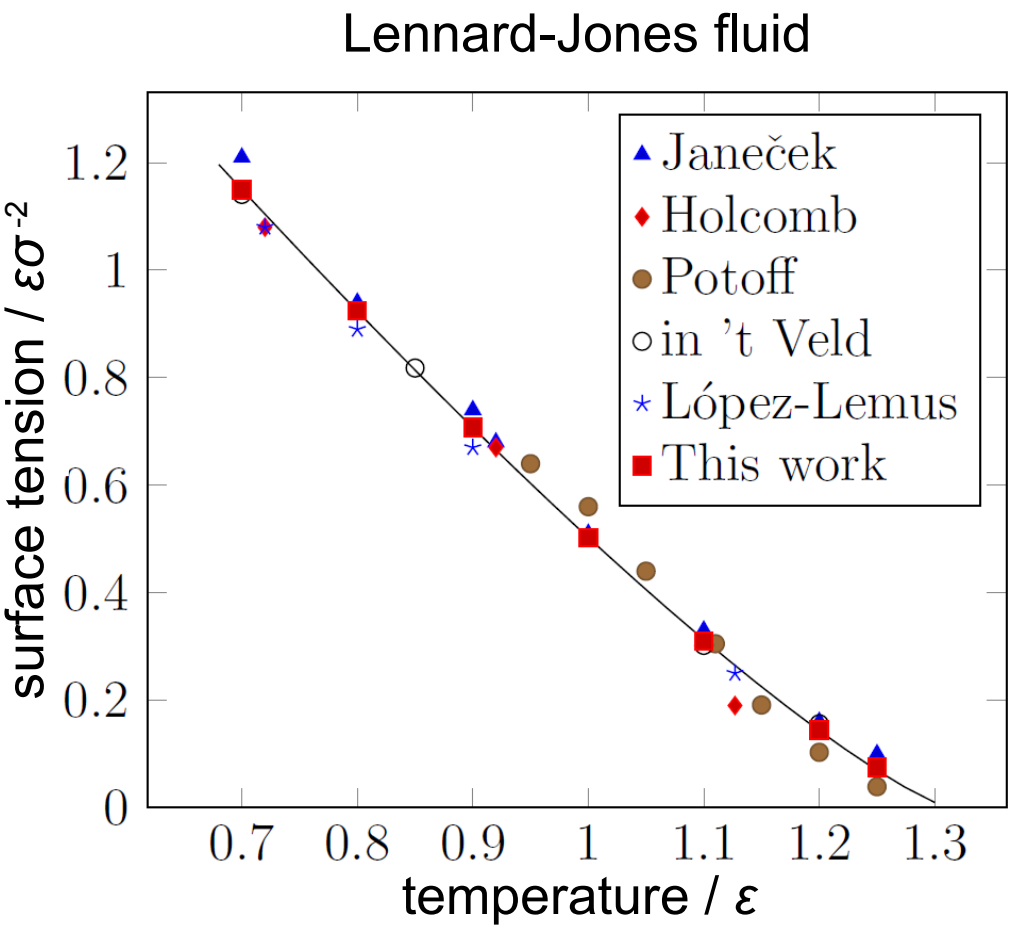
Angle-averaging expression for multi-site models, following **Lustig**.

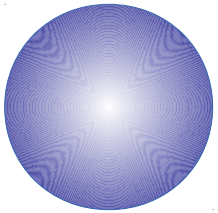
Two-centre LJ fluid (2CLJ)



Dipole and dispersion lead to analogous long-range correction expressions.
The long-range contribution of the quadrupole can be neglected.

Surface tension at high precision

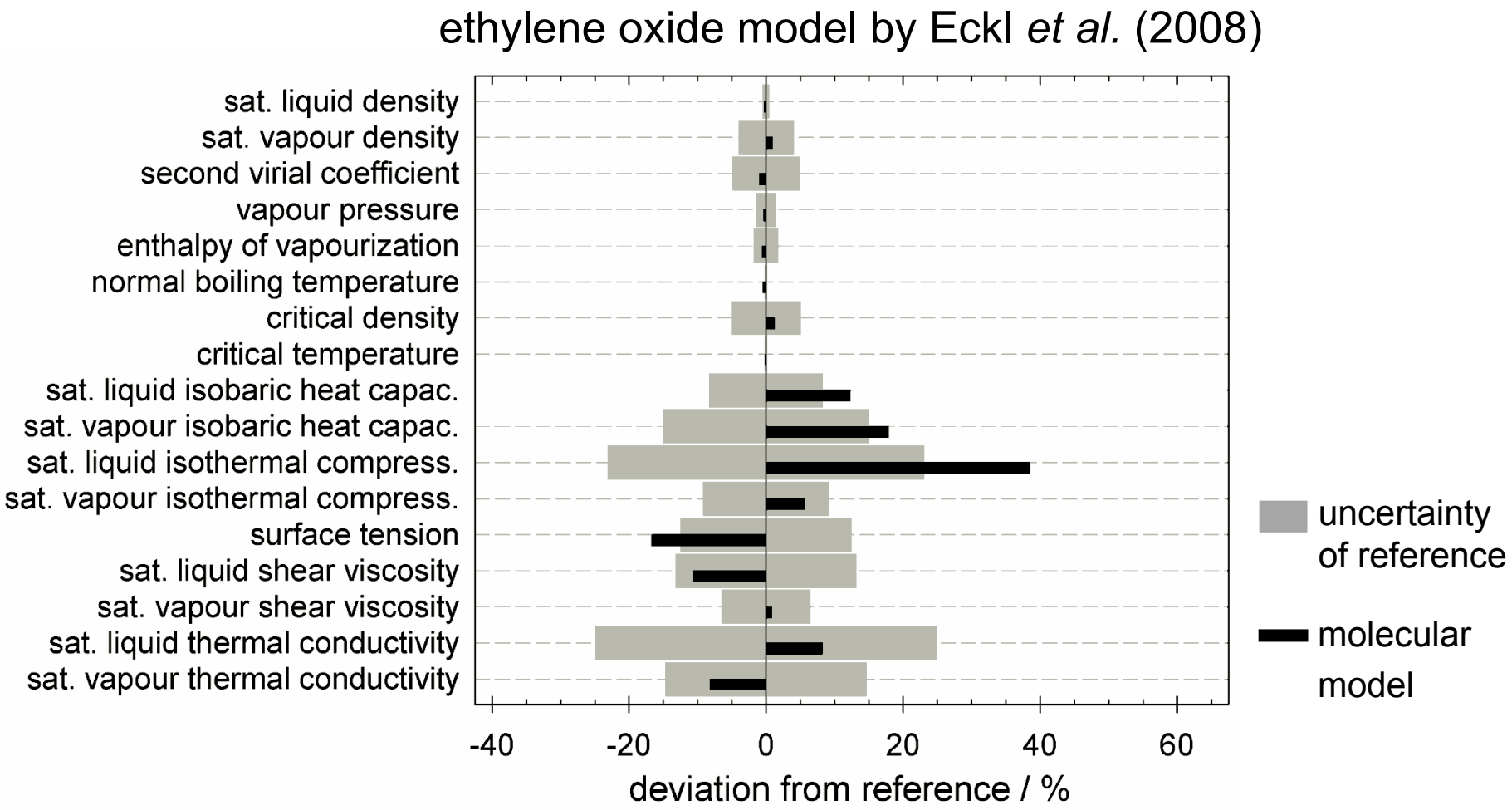




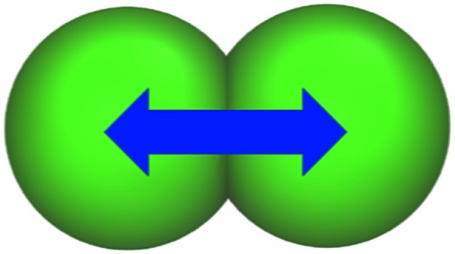
LJ

$$\gamma_0(T) = 2.94 \frac{\varepsilon}{\sigma^2}\left(1-\frac{T}{T_c}\right)^{1.23}$$

Objective: Accuracy for multiple properties



Force fields for molecular modelling



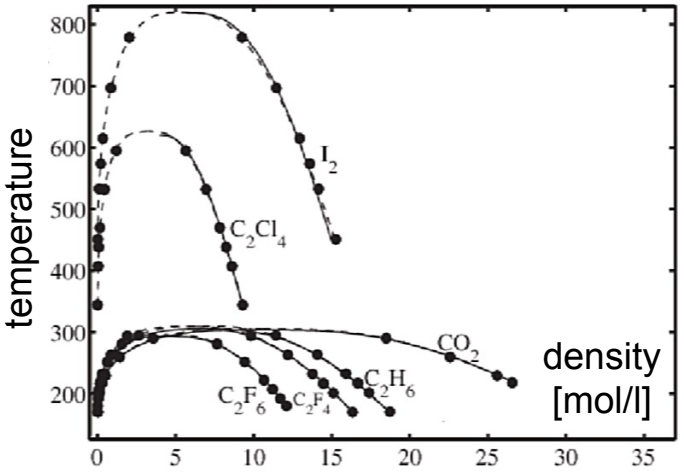
2CLJQ models:

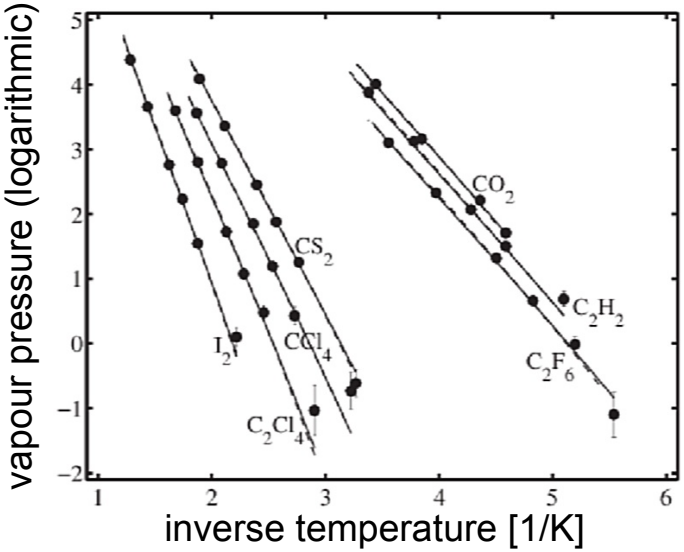
- 2 LJ centres
- Quadrupole

Fit of parameters σ, ε, L, Q to VLE data of 29 fluids by Stoll *et al.*

Deviation:

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



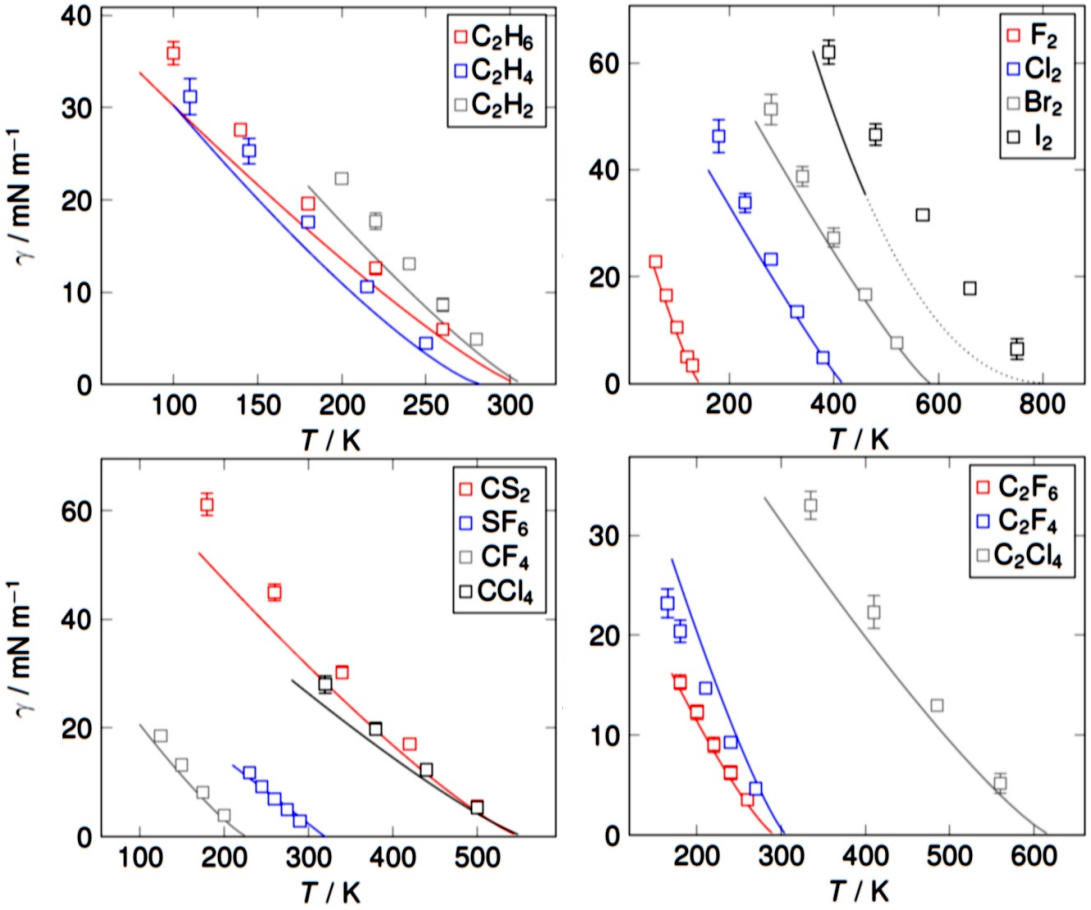


- simulation
— DIPPR correlation

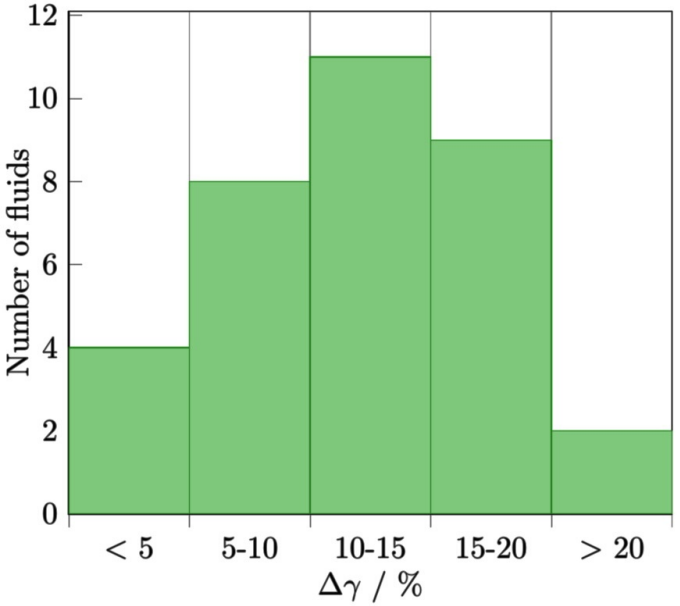
No interfacial properties were considered for the parameterization.

Predictive capacity of literature models

Two LJ + quadrupole (2CLJQ)



Two LJ + dipole (2CLJD)

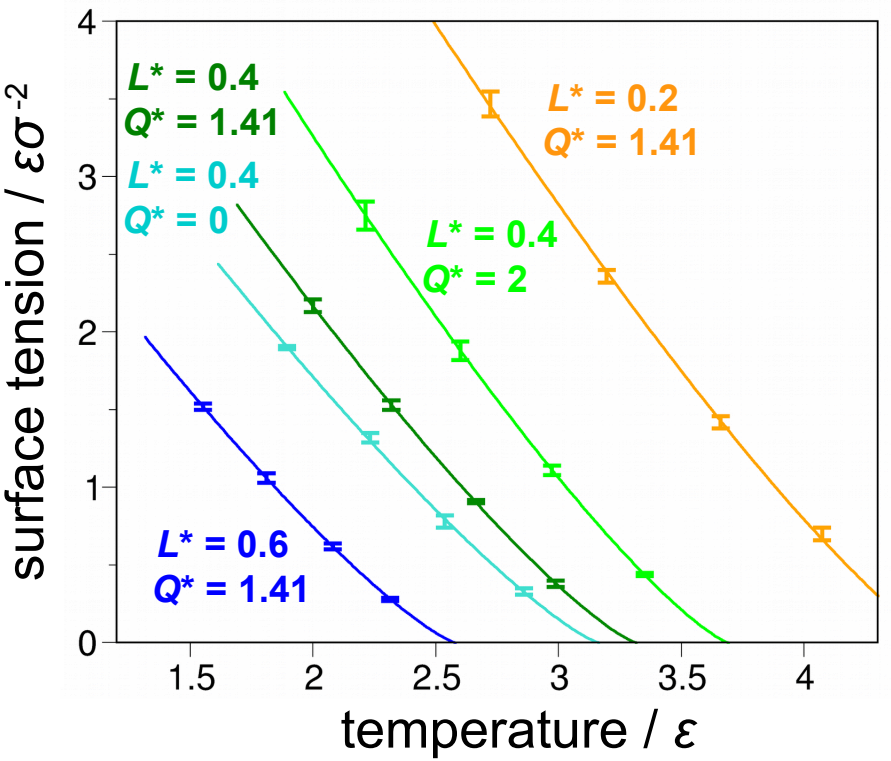


Fit to bulk properties

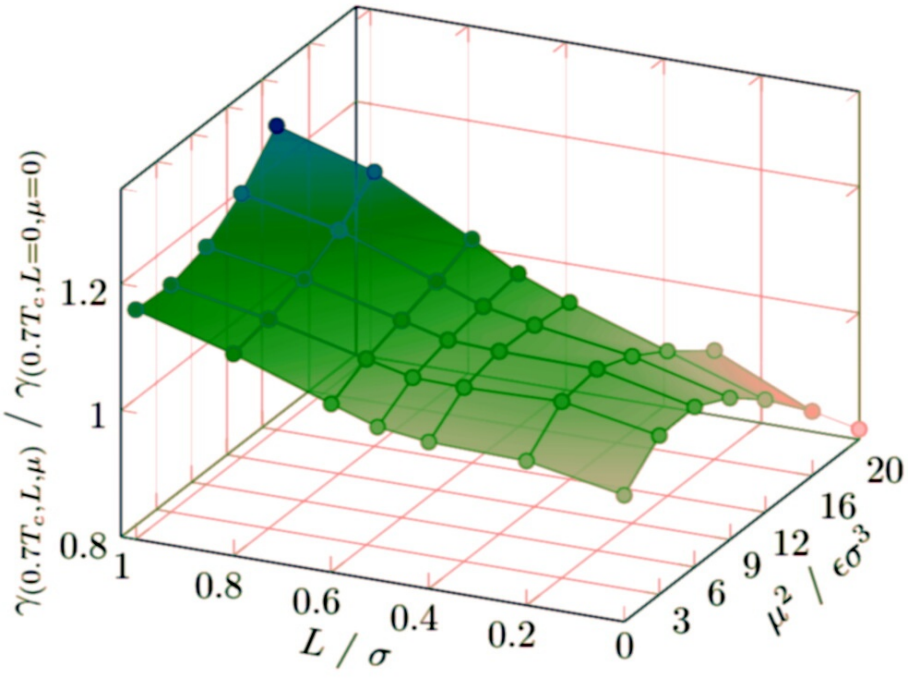
10 to 20 % overestimation of vapour-liquid surface tension

Massively parallel molecular modelling

Two LJ + quadrupole (2CLJQ)



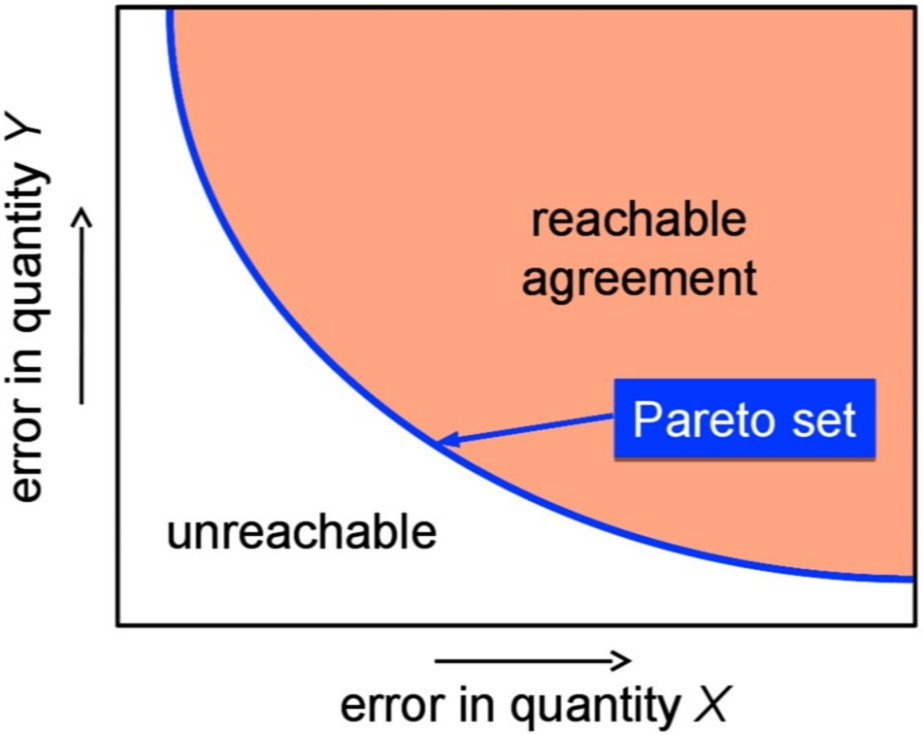
Two LJ + dipole (2CLJD)



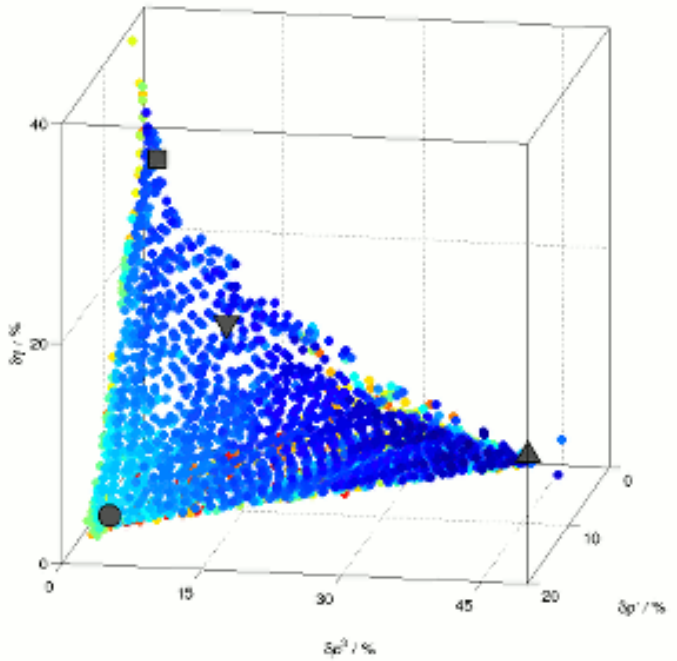
- Systematic exploration of the four-dimensional model parameter space
- Correlation of the surface tension by a critical scaling expression

Multicriteria model optimization

Pareto optimality criterion

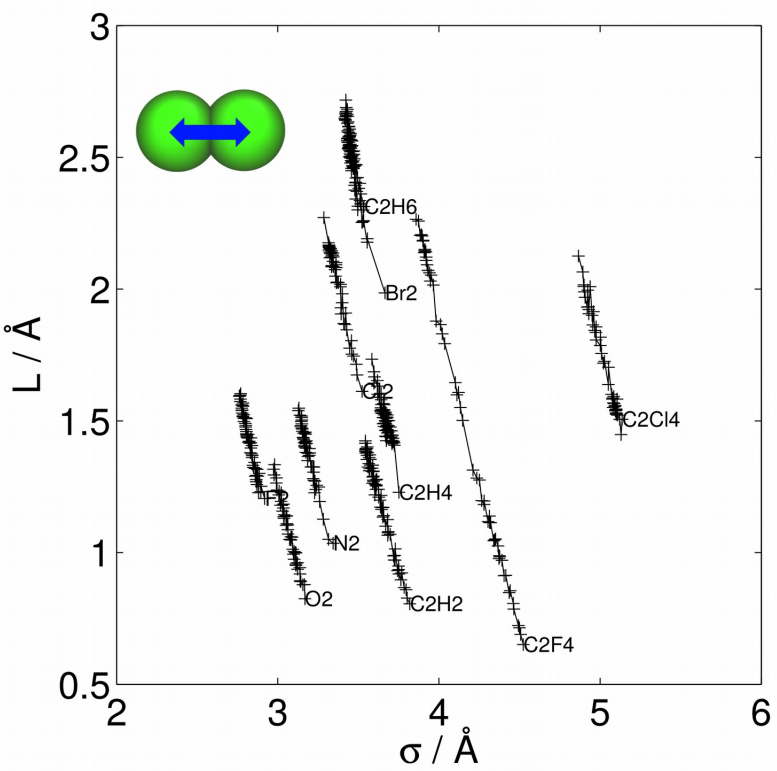


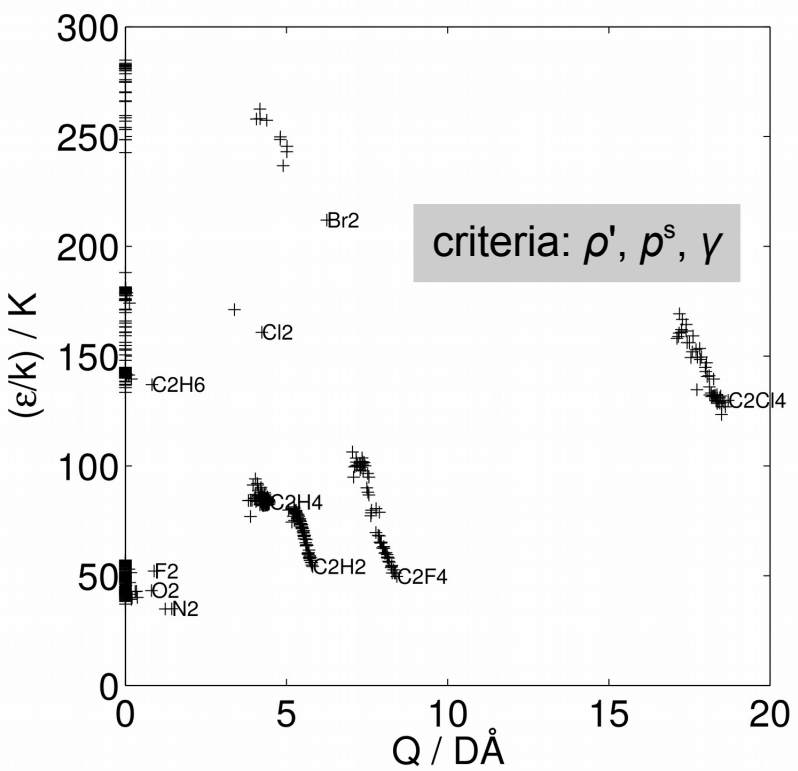
Pareto set for carbon dioxide



Multicriteria optimization requires massively-parallel molecular modelling.

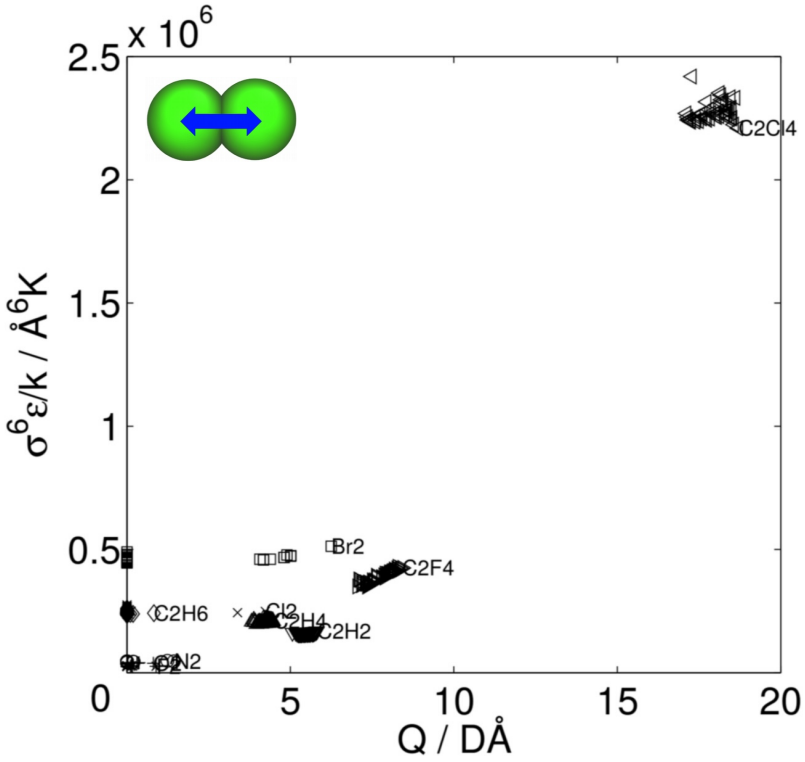
Pareto sets for 2CLJQ models of real fluids

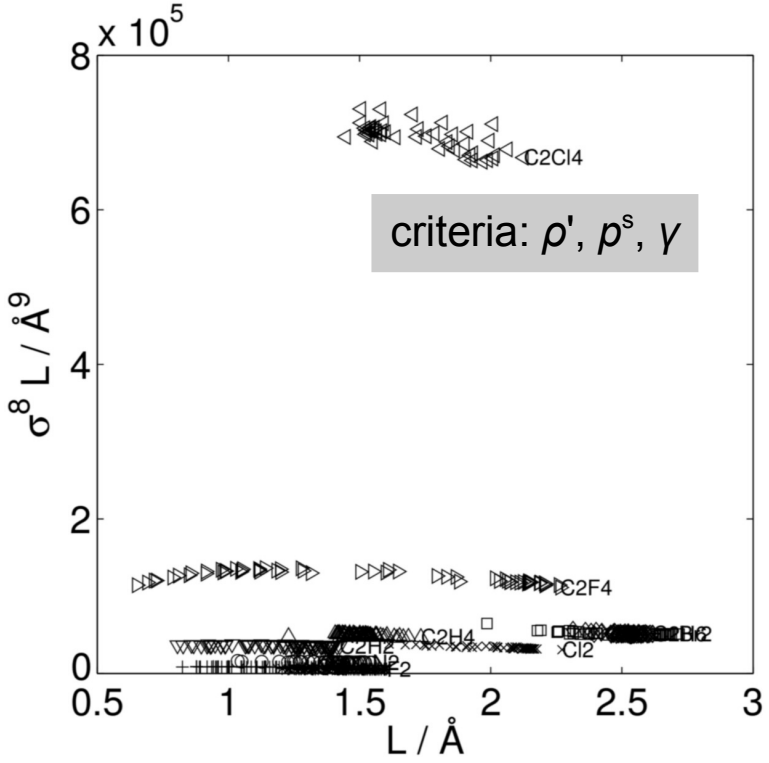




Projections of the Pareto set on the parameter space reveal intrinsic correlations between different model parameters, such as ε and Q.

Pareto sets for 2CLJQ models of real fluids

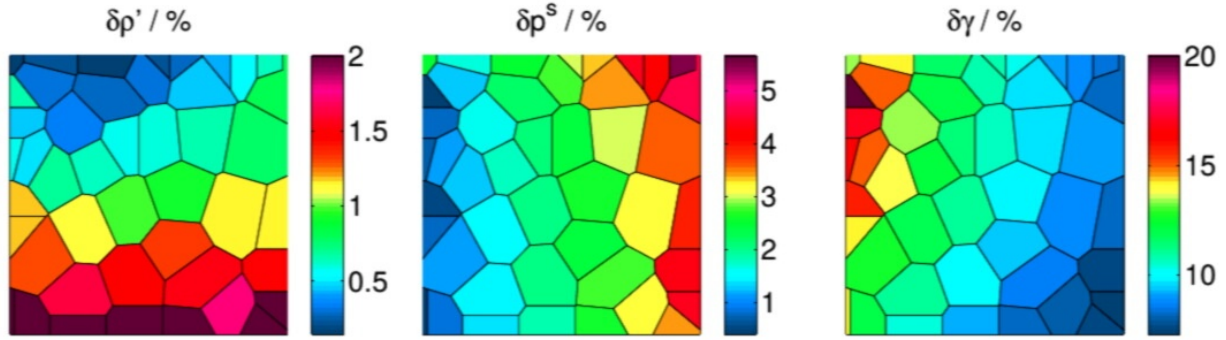




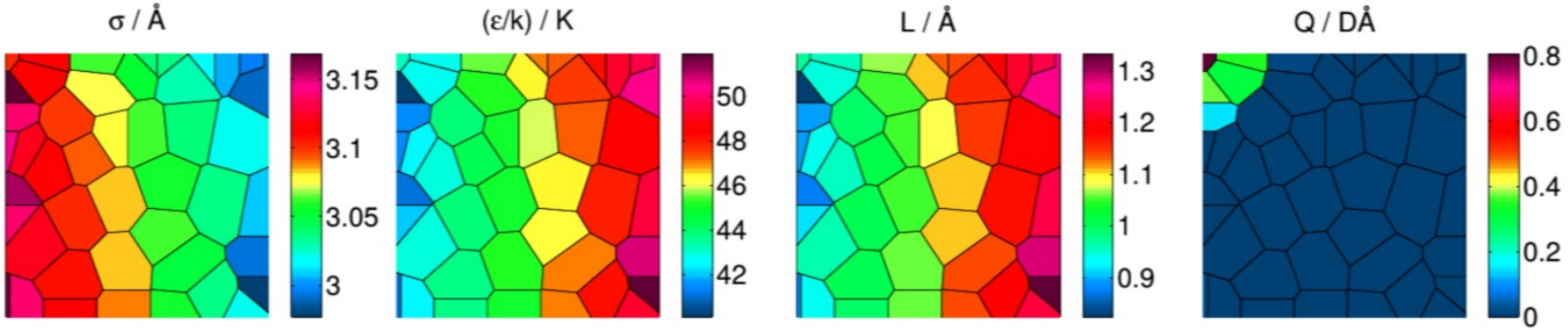
The dimension of the parameter space is effectively reduced, facilitating an efficient multicriteria optimization by navigating on the Pareto set.

Pareto sets for 2CLJQ models of real fluids

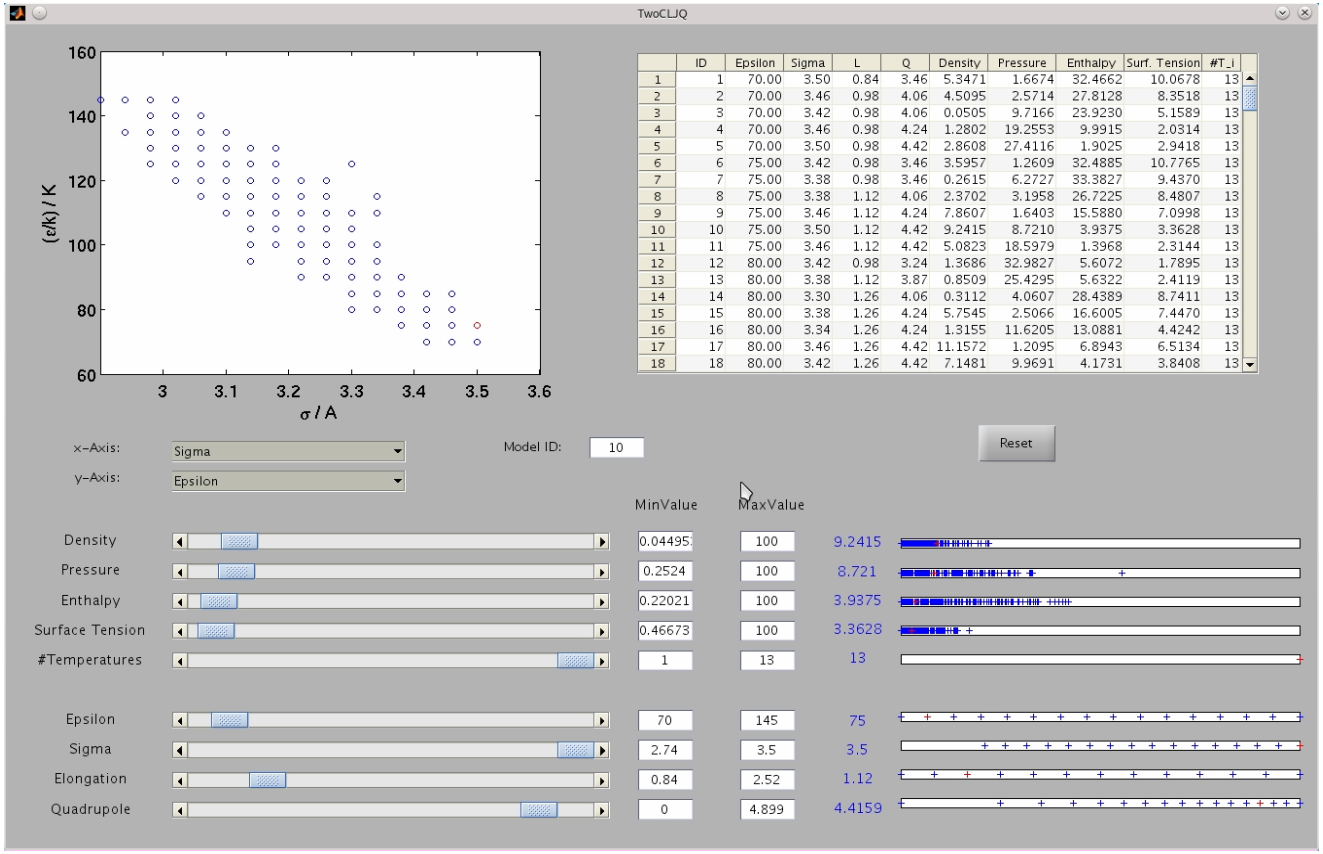
Representation of objective and parameter spaces by **patch plots**:



Pareto-optimal 2CLJQ models for molecular oxygen



Fast and simple model parameterization



	ID	Epsilon	Sigma	L	Q	Density	Pressure	Enthalpy	Surf. Tension	#T_i
1	1	70.00	3.50	0.84	3.46	5.3471	1.6674	32.4662	10.0678	13
2	2	70.00	3.46	0.98	4.06	4.5095	2.5714	27.8128	8.3518	13
3	3	70.00	3.42	0.98	4.06	0.0505	9.7166	23.9230	5.1589	13
4	4	70.00	3.46	0.98	4.24	1.2802	19.2553	9.9915	2.0314	13
5	5	70.00	3.50	0.98	4.42	2.8608	27.4116	1.9025	2.9418	13
6	6	75.00	3.42	0.98	3.46	3.5957	1.2609	32.4885	10.7765	13
7	7	75.00	3.38	0.98	3.46	0.2615	6.2727	33.3827	9.4370	13
8	8	75.00	3.38	1.12	4.06	2.3702	3.1958	26.7225	8.4807	13
9	9	75.00	3.46	1.12	4.24	7.8607	1.6403	15.5880	7.0998	13
10	10	75.00	3.50	1.12	4.42	9.2415	8.7210	3.9375	3.3628	13
11	11	75.00	3.46	1.12	4.42	5.0823	18.5979	1.3968	2.3144	13
12	12	80.00	3.42	0.98	3.24	1.3686	32.9827	5.6072	1.7895	13
13	13	80.00	3.38	1.12	3.87	0.8509	25.4295	5.6322	2.4119	13
14	14	80.00	3.30	1.26	4.06	0.3112	4.0607	28.4389	8.7411	13
15	15	80.00	3.38	1.26	4.24	5.7545	2.5066	16.6005	7.4470	13
16	16	80.00	3.34	1.26	4.24	1.3155	11.6205	13.0881	4.4242	13
17	17	80.00	3.46	1.26	4.42	11.1572	1.2095	6.8943	6.5134	13
18	18	80.00	3.42	1.26	4.42	7.1481	9.9691	4.1731	3.8408	13

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**.