Martin Thomas Horsch
Scientific Computing Department
STFC Daresbury Laboratory
UK Research and Innovation

Multicriteria optimization of molecular force fields: Bulk properties and the surface tension of fluids

VIMMP
VIRTUAL MATERIALS MARKETPLACE

CCP5 Annual Meeting

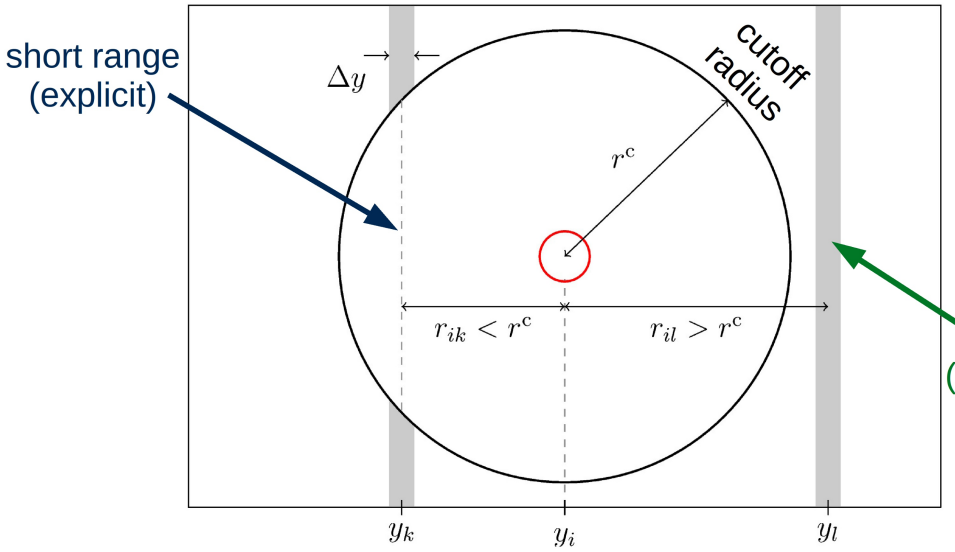
11th September 2018

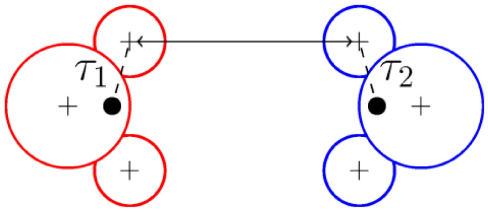
Manchester, UK

MD simulation of vapour-liquid interfaces

Molecular simulation of heterogeneous systems:

- Greater length scale and greater time scale;
- Long-range contribution from dispersion.

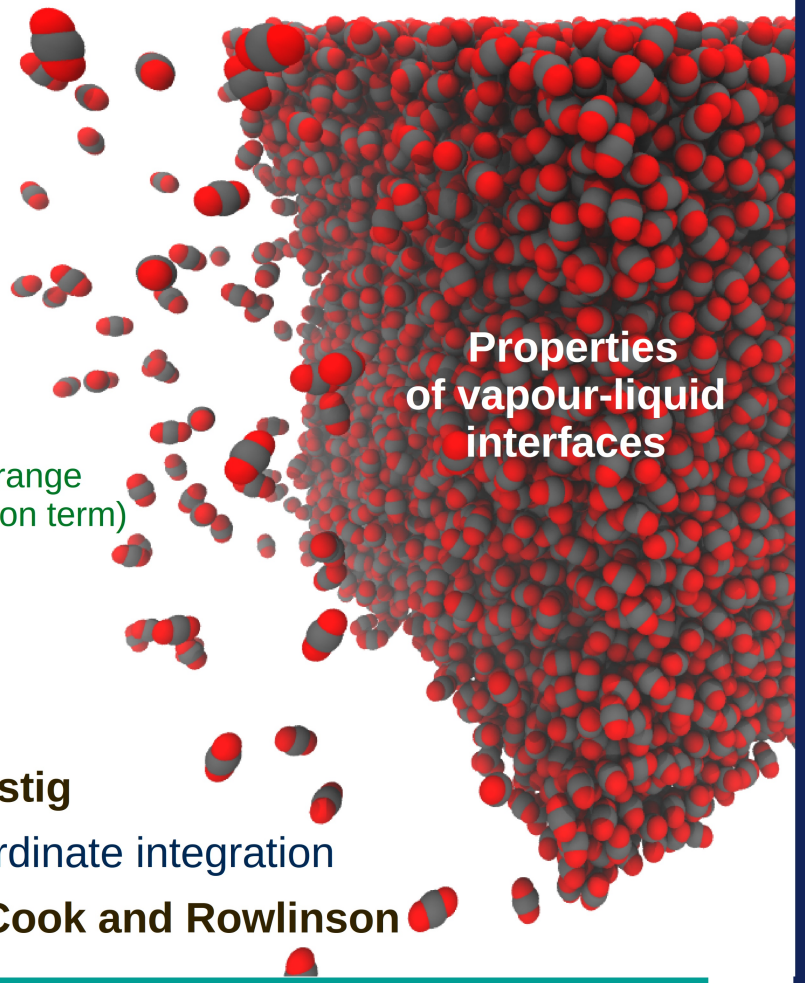




Angle averaging following **Lustig**

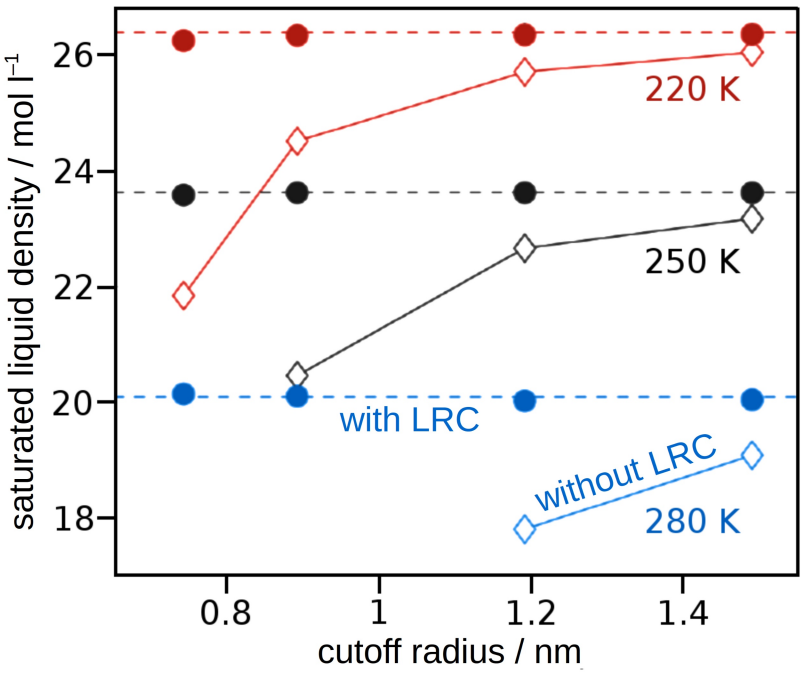
Janeček LRC by normal-coordinate integration

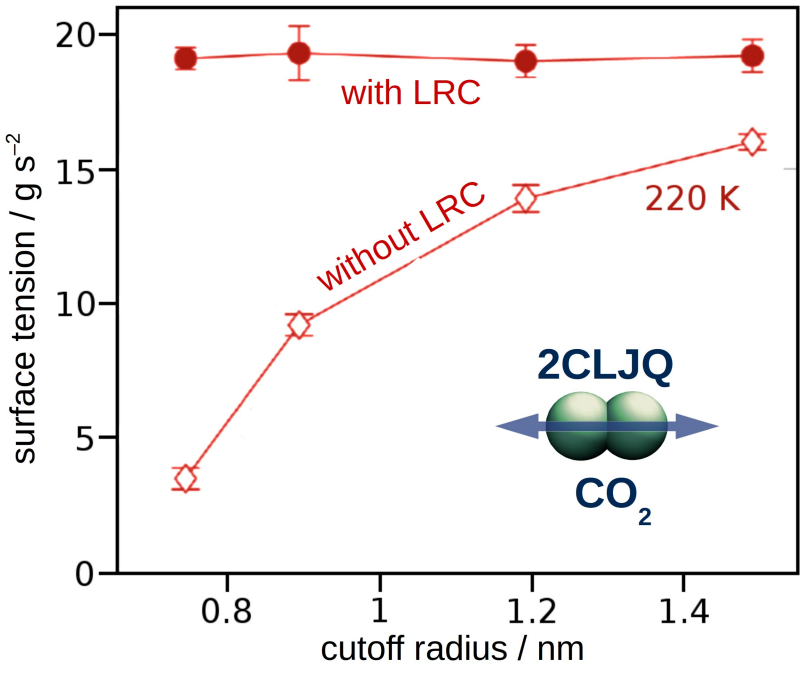
Dipolar interaction following **Cook and Rowlinson**



MD simulation of vapour-liquid interfaces

Validation by sensitivity check: Influence of the cutoff radius on simulation results.





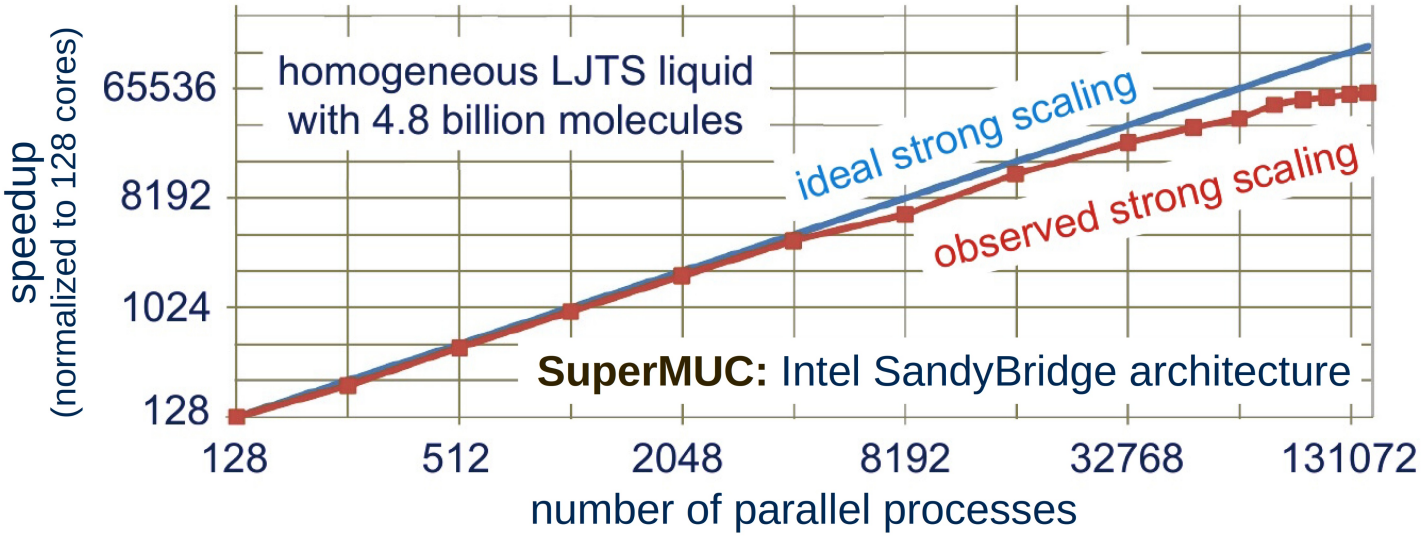
(S. Werth *et al.*, *Mol. Phys.* 112, 2227, **2014**)

Computing time requirements for the explicitly computed pairwise interactions scale cubically in r_c and are significantly reduced by a LRC following Janeček and Lustig.

Massively parallel molecular dynamics simulation

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

(large systems 1: molecular dynamics)[1, 2]



N = 4 125 000 000 000

2013 molecular dynamics world record[2]

[1]C. Niethammer *et al.*, *J. Chem. Theory Comput.* 10(10), 4455 – 4464, **2014**;
[2]W. Eckhardt *et al.*, *Proc. ISC 2013*, LNCS 7905, 1 – 12, **2013**.

TECHNISCHE UNIVERSITÄT MÜNCHEN

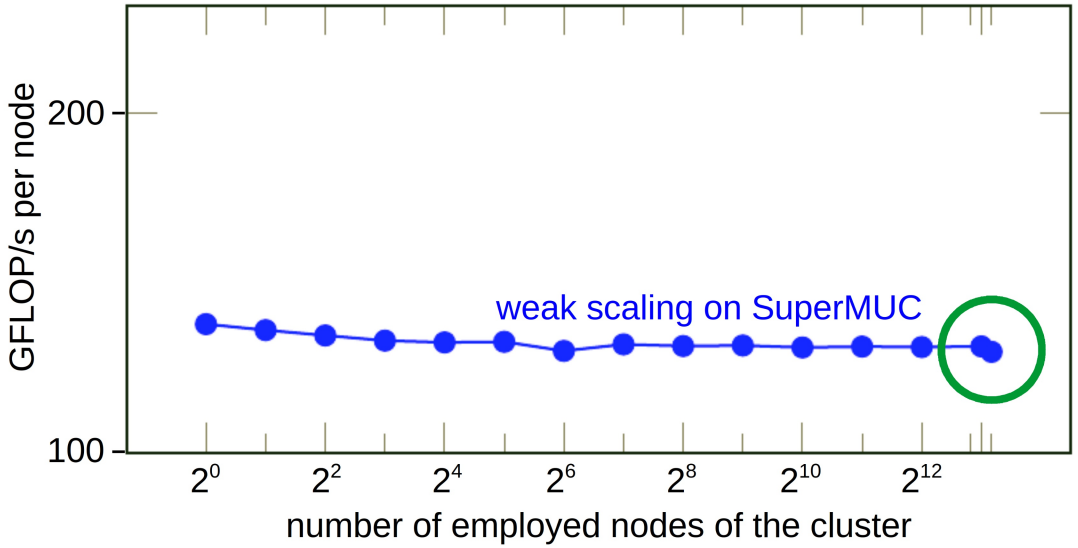
Computational Molecular Engineering

Massively parallel molecular dynamics simulation

SuperMUC (Garching):
SandyBridge architecture

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

(large systems 1: molecular dynamics)[1, 2]



N = 4 125 000 000 000
2013 molecular dynamics world record[2]

SuperMUC
weak scaling

TECHNISCHE UNIVERSITÄT MÜNCHEN

Computational Molecular Engineering

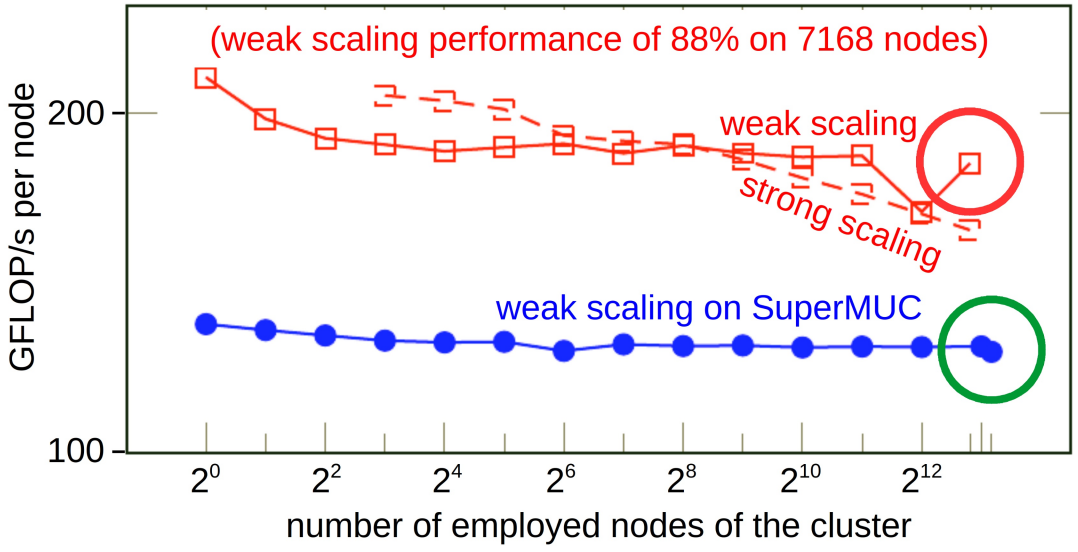
[1]C. Niethammer *et al.*, *J. Chem. Theory Comput.* 10(10), 4455 – 4464, **2014**;
[2]W. Eckhardt *et al.*, *Proc. ISC 2013*, LNCS 7905, 1 – 12, **2013**.

Massively parallel molecular dynamics simulation

Hazel Hen (Stuttgart):
Haswell architecture

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

(large systems 1: molecular dynamics)[1, 2]



(cores per node: 24 for Hazel Hen, 16 for SuperMUC)

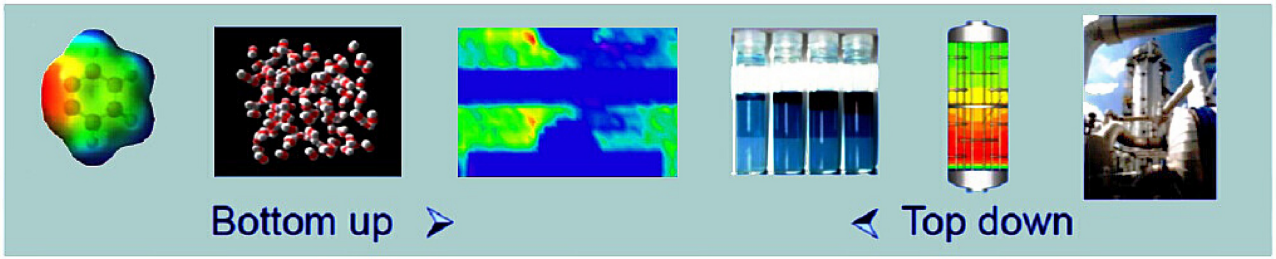
N = 21 000 000 000 000
2018 molecular dynamics world record[2]

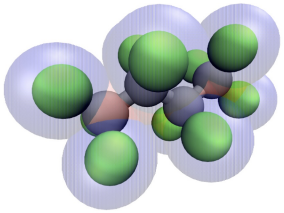
Hazel Hen
weak scaling

Computational
Molecular Engineering

[1]C. Niethammer *et al.*, *J. Chem. Theory Comput.* 10(10), 4455 – 4464, **2014**;
[2]P. Neumann, S. Seckler, N. Tchipev, *et al.*, private communication, **2018**.

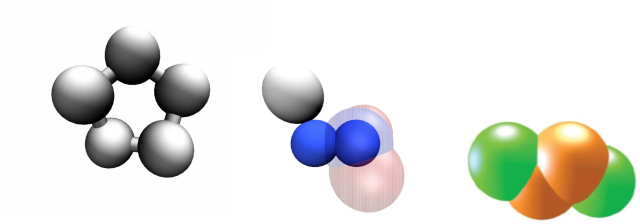
Computational molecular engineering





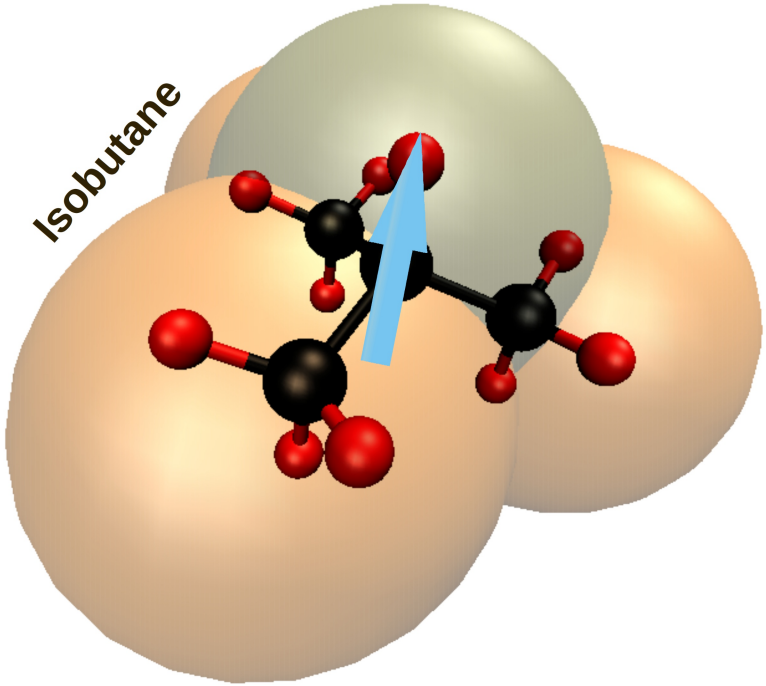
physics-driven aspects (qualitative validity) ⟷ **data-driven aspects (quantitative reliability)**

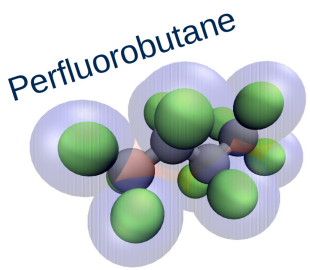
- Realistic representation of the underlying physical features

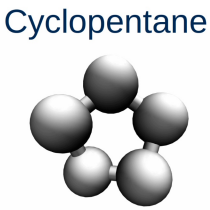


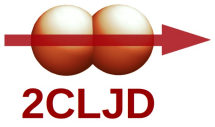
- Models with free parameters, which can be adjusted to experimental data
- Reliable interpolation, extrapolation and prediction of thermophysical properties

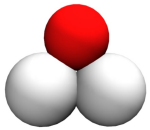
Model data and metadata repositories











Molecular Model Database

http://molmod.boltzmann-zuse.de/

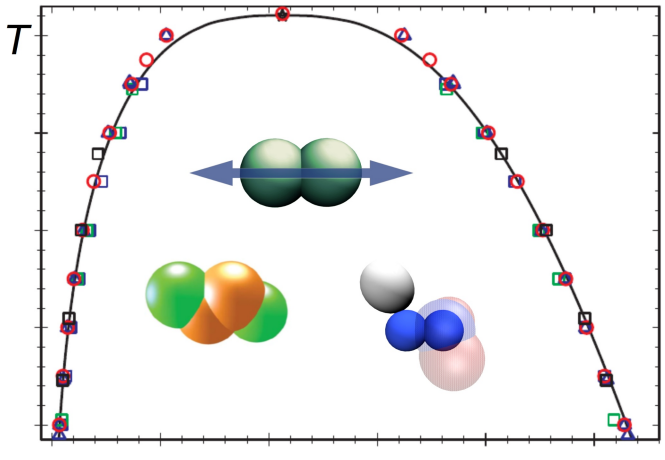
Repository maintained by Simon Stephan *et al.* (Boltzmann-Zuse Society)

Literature models by J. Stoll, H. Hasse, J. Vrabec *et al.*, **2001 – 2018**

Model validation and uncertainty quantification

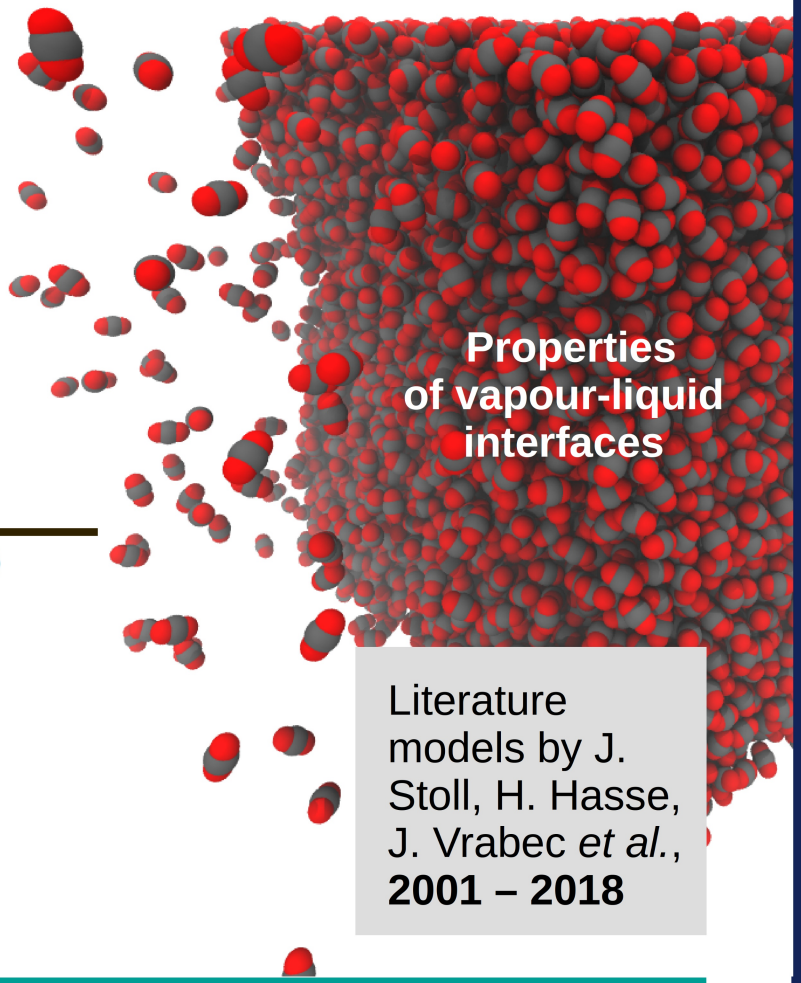
How accurate are these molecular models?

VLE bulk phase properties



prediction
(model property)

validation
(model accuracy)



Literature models by J. Stoll, H. Hasse, J. Vrabec *et al.*, **2001 – 2018**

Molecular model developers adjusted the force field parameters to bulk VLE properties.

No interfacial properties were considered for the parameterization.

Model validation and uncertainty quantification[1–3]

Non-polar: LJ

Neon (Ne), Argon (Ar)
Krypton (Kr), Xenon (Xe)
Methane (CH_4)

Dipolar: 2CLJD

Carbon monoxide (CO)
R11 ($CFCl_3$)
R12 (CF_2Cl_2)
R13 (CF_3Cl)
R13B1 ($CBrF_3$)
R22 (CHF_2Cl)
R23 (CHF_3)
R41 (CH_3F)
R123 ($CHCl_2$-CF_3)
R124 ($CHFCl$-CF_3)
R125 (CHF_2-CF_3)
R134a (CH_2F-CF_3)
R141b (CH_3-$CFCl_2$)
R142b (CH_3-CF_2Cl)
R143a (CH_3-CF_3)
R152a (CH_3-CHF_2)
R40 (CH_3Cl)
R40B1 (CH_3Br)
Methyl iodide (CH_3I)
R30B1 (CH_2BrCl)
R20 ($CHCl_3$)
R20B3 ($CHBr_3$)
R21 ($CHFCl_2$)
R32 (CH_2F_2)
R30 (CH_2Cl_2)

+12%

Dipolar: 2CLJD (contd.)

R30B2 (CH_2Br_2)
Methylene iodide (CH_2I_2)
R12B2 (CBr_2F_2)
R12B1 ($CBrClF_2$)
R10B1 ($CBrCl_3$)
R161 (CH_2F-CH_3)
R150a ($CHCl_2$-CH_3)
R140 ($CHCl_2$-CH_2Cl)
R140a (CCl_3-CH_3)
R130a (CH_2Cl-CCl_3)
R160B1 (CH_2Br-CH_3)
R150B2 ($CHBr_2$-CH_3)
R131b (CH_2F-CCl_3)
R123B1 ($CHClBr$-CF_3)
R112a (CCl_3-CF_2Cl)
R1141 (CHF=CH_2)
R1132a (CF_2=CH_2)
R1140 ($CHCl$=CH_2)
R1122 ($CHCl$=CF_2)
R1113 ($CFCl$=CF_2)
R1113B1 ($CFBr$=CF_2)

Quadrupolar: 2CLJQ

Fluorine (F_2)
Chlorine (Cl_2)
Bromine (Br_2)
Iodine (I_2)
Nitrogen (N_2)
Oxygen (O_2)
Carbon dioxide (CO_2)

+20%

Quadrupolar: 2CLJQ (contd.)

Carbon disulfide (CS_2)
Ethane (C_2H_6)
Ethylene (C_2H_4)
Acetylene (C_2H_2)
R116 (C_2F_6)
R1114 (C_2F_4)
R1110 (C_2Cl_4)
Propadiene (CH_2=C=CH_2)
Propyne (CH_3-C≡CH)
Propylene (CH_3-CH=CH_2)
R846 (SF_6)
R14 (CF_4)
R10 (CCl_4)
R113 ($CFCl_2$-CF_2Cl)
R114 (CF_2Cl-CF_2Cl)
R115 (CF_3-CF_2Cl)
R134 (CHF_2-CHF_2)
R150B2 (CH_2Br-CH_2Br)
R114B2 ($CBrF_2$-$CBrF_2$)
R1120 ($CHCl$=CCl_2)

Multi-site models

Isobutane (C_4H_{10})
Cyclohexane (C_6H_{12})
Methanol (CH_3OH)
Ethanol (C_2H_5OH)
Formaldehyde (CH_2=O)
Dimethylether (CH_3-O-CH_3)
Acetone (C_3H_6O)

+22%

Multi-site models (contd.)

Ammonia (NH_3)
Methylamine (NH_2-CH_3)
Dimethylamine (CH_3-NH-CH_3)
R227ea (CF_3-CHF-CF_3)
Sulfur dioxide (SO_2)
Ethylene oxide (C_2H_4O)
Dimethylsulfide (CH_3-S-CH_3)
Hydrogen cyanide (HCN)
Acetonitrile (NC_2H_3)
Thiophene (SC_4H_4)
Nitromethane (NO_2CH_3)
Phosgene ($COCl_2$)
Benzene (C_6H_6)
Toluene (C_7H_8)
Chlorobenzene (C_6H_5Cl)
Dichlorobenzene ($C_6H_4Cl_2$)
Cyclohexanol ($C_6H_{11}OH$)
Cyclohexanone ($C_6H_{10}O$)
Cyanogen (C_2N_2)
Cyanogen chloride (CClN)
Formic acide (CH_2O_2)
Ethylene glycol ($C_2H_6O_2$)
Water (H_2O)
Hydrazine (N_2H_4)
Methylhydrazine (CH_6N_2)
Dimethylhydrazine ($C_2H_8N_2$)
Perfluorobutane (C_4F_{10})
Ethyl acetate ($C_4H_8O_2$)
Hexamethyldisiloxane ($C_6H_{12}OSi_2$)
Octamethylcyclotetrasiloxane ($C_8H_{24}O_4Si_4$)

[1]Werth *et al.*, *Chem. Eng. Sci.* 121, 110, **2015**; [2]Werth *et al.*, *J. Chem. Phys.* 144, 054702, **2016**; [3]Werth *et al.*, *J. Mol. Liq.* 235, 126, **2017**.

Process of modelling and simulation standardization

Review of Materials Modelling VI (edited by de Baas), 2017

RoMM VI **MODA** **CWA 17284** **Ontologies**

Semi-formalized terminology or vocabulary

Graph language & formalized terminology

CEN European standard

Ontologies

(Ontology development is work in progress at present.)

EMMC

(1) User Case
(2) Model
(3) Solver
(4) Processing

MODA
(graph language)

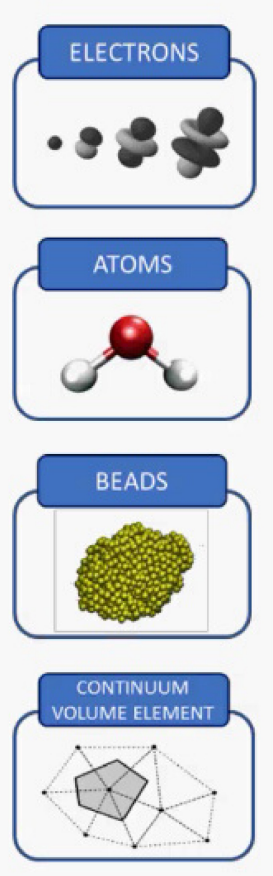
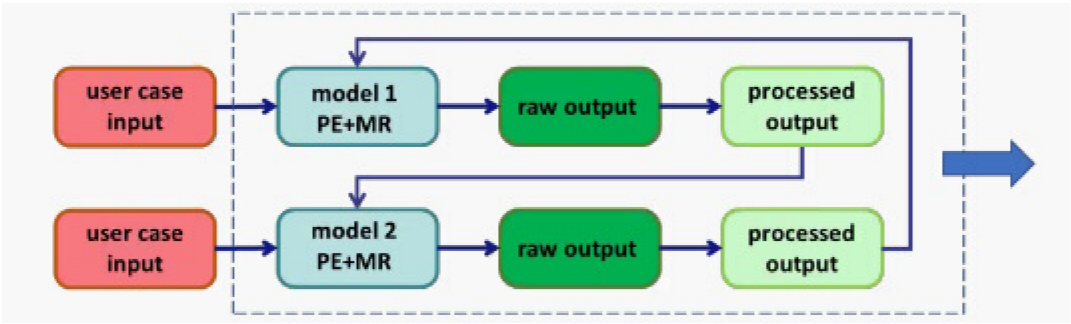
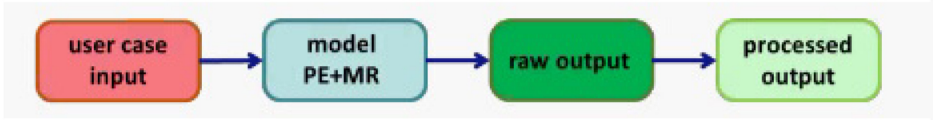
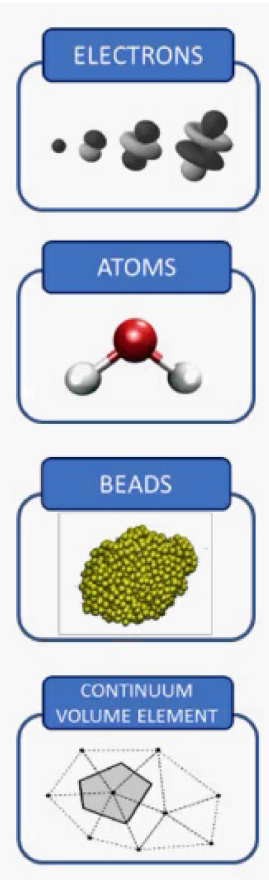


Diagram representation of simulation workflows

MODA: “Modelling Data” graph language & CEN Workshop Agreement 17824







electronic

atomistic

mesoscopic

continuum

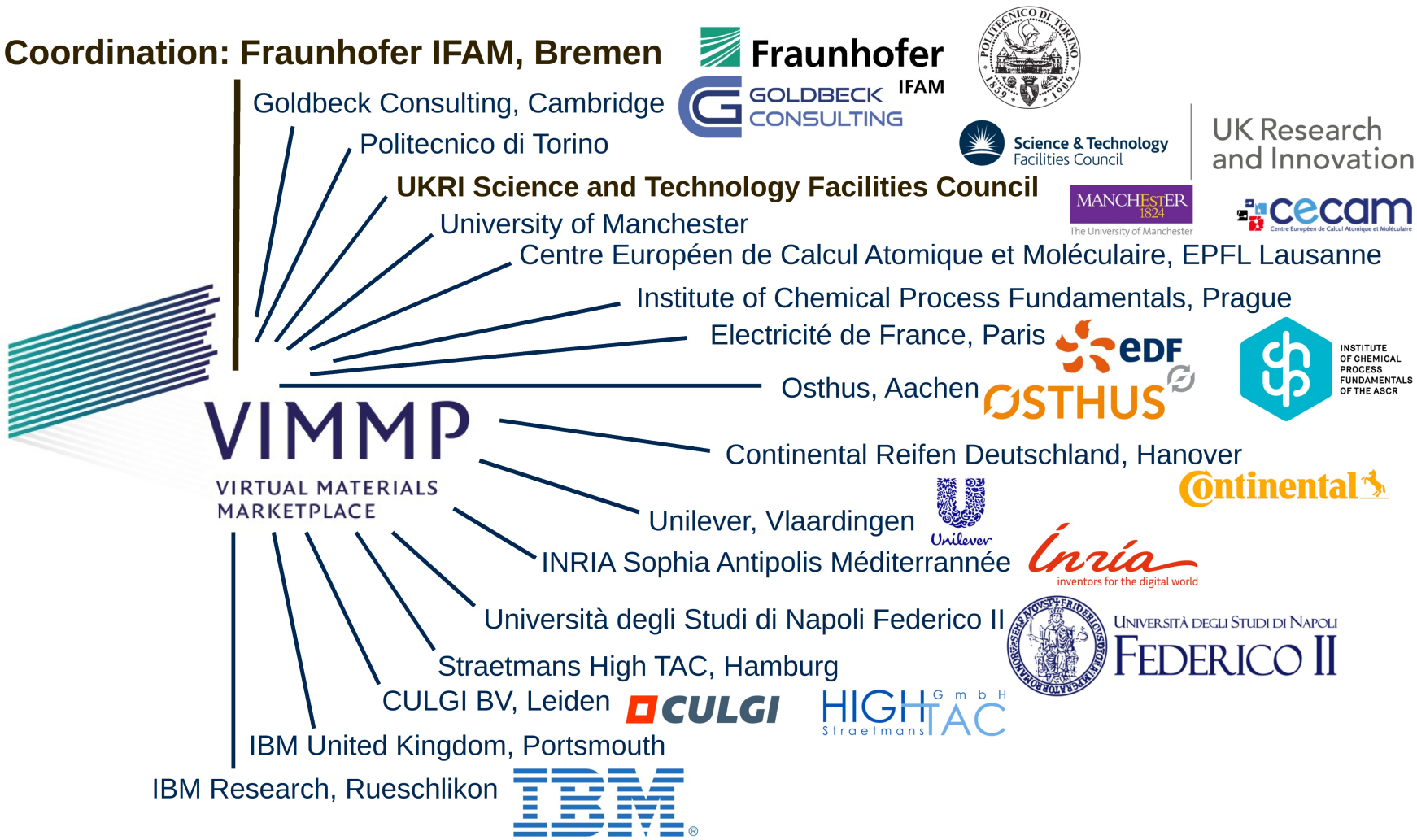
EMMC

(1) User Case
(2) Model
(3) Solver
(4) Processing

cen

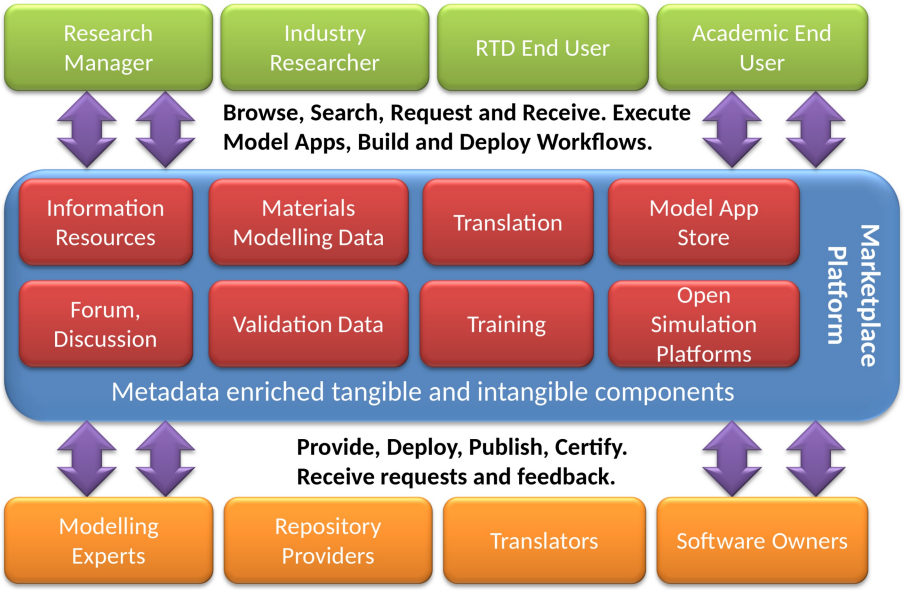
EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

Virtual Materials Marketplace: VIMMP (Horizon 2020)



Virtual Materials Marketplace: VIMMP (Horizon 2020)

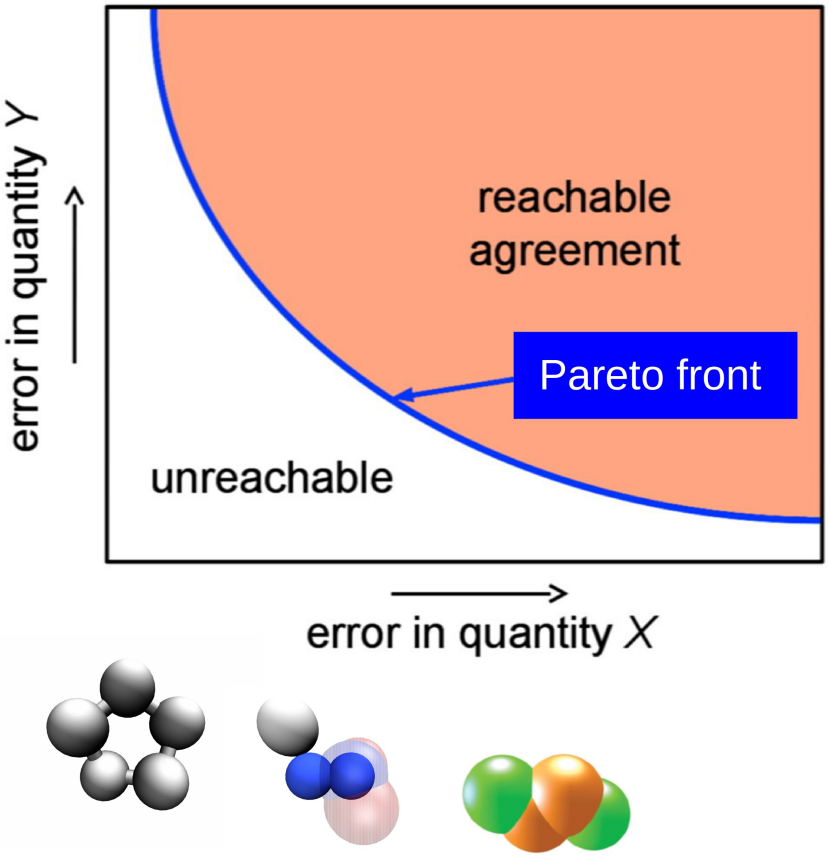
VIMMP Marketplace concept: To serve its participants and facilitate exchange, e.g., between materials **model providers**, industrial & academic **client** end users, and **translators**.



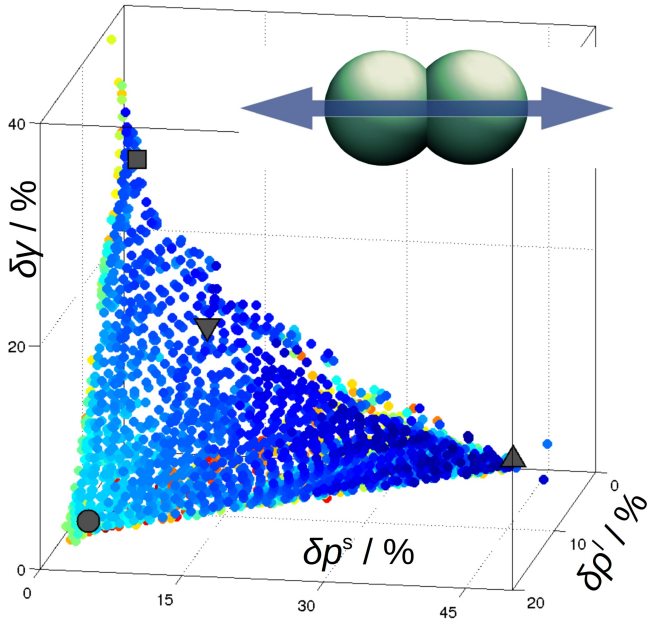
The **VIMMP Marketplace** will provide end-user interfaces to information resources, discussion forums, databases and repositories, translation and training services, validated models and modelling software, and the ability to utilise open simulation platforms to build and deploy workflows via cloud-based computing resources.

Multicriteria optimization of molecular models

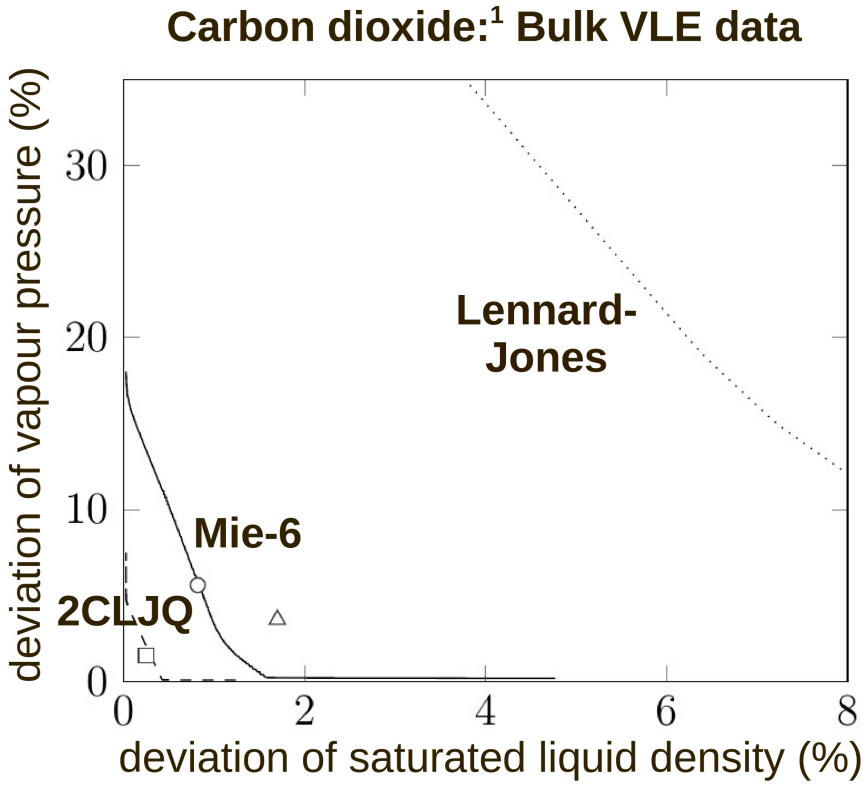
Multidimensional objective space



Example: 2CLJQ molecular models of carbon dioxide



Comparison between model classes



LJ potential (2 parameters)

Mie-6 potential (3 parameters)

2CLJQ potential (4 parameters)

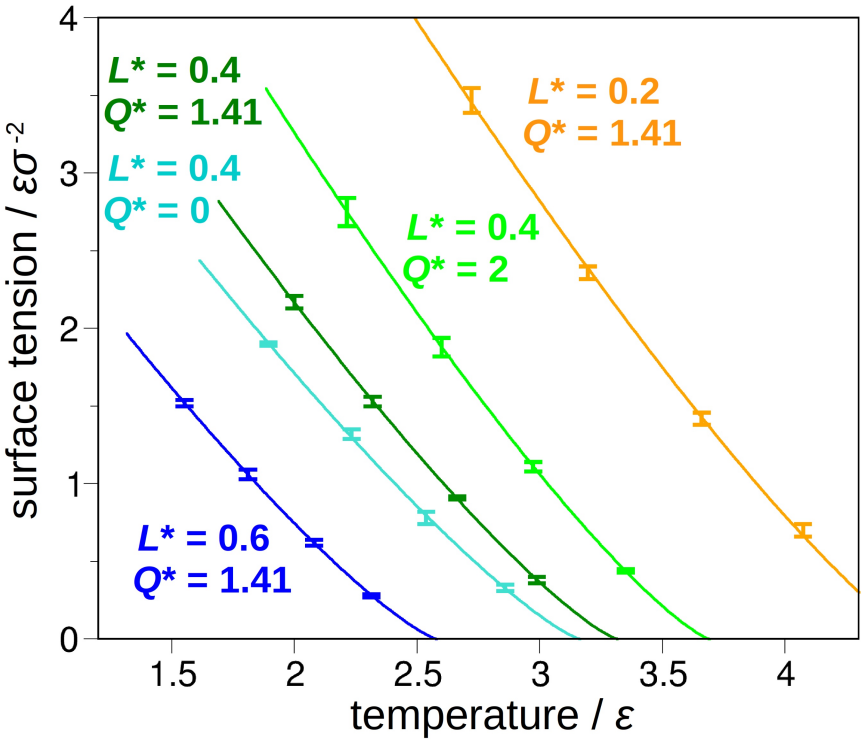
Vrabec *et al.* (2001) 2CLJQ model

Prerequisite: Thermodynamic characterization of the investigated model classes

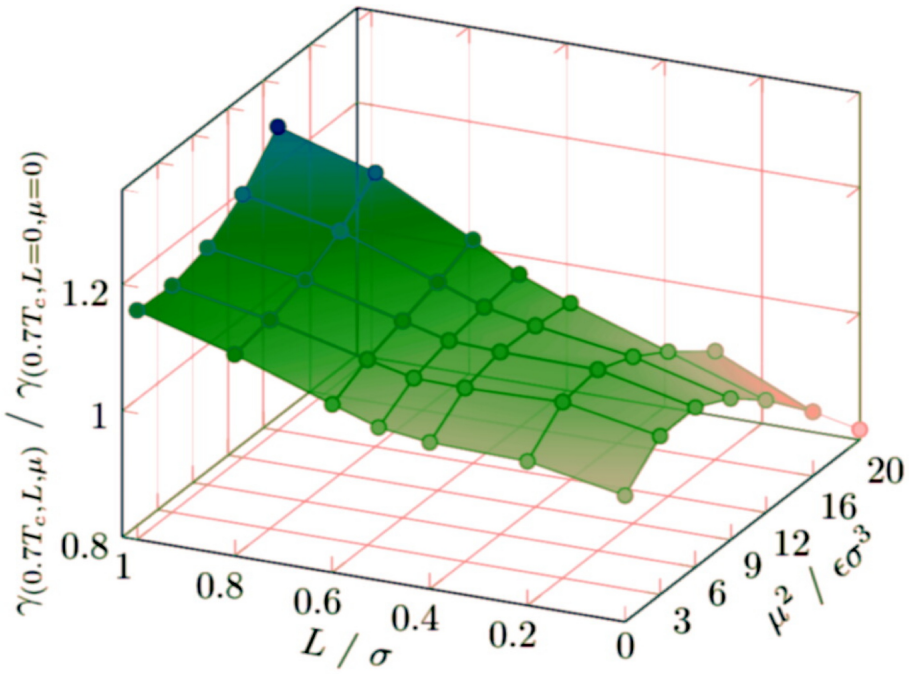
[1]S. Werth, K. Stöbener, M. Horsch, H. Hasse, *Mol. Phys.* 115(9–12), 1017–1030, **2017**.

Model class characterization

Two LJ + quadrupole (2CLJQ)



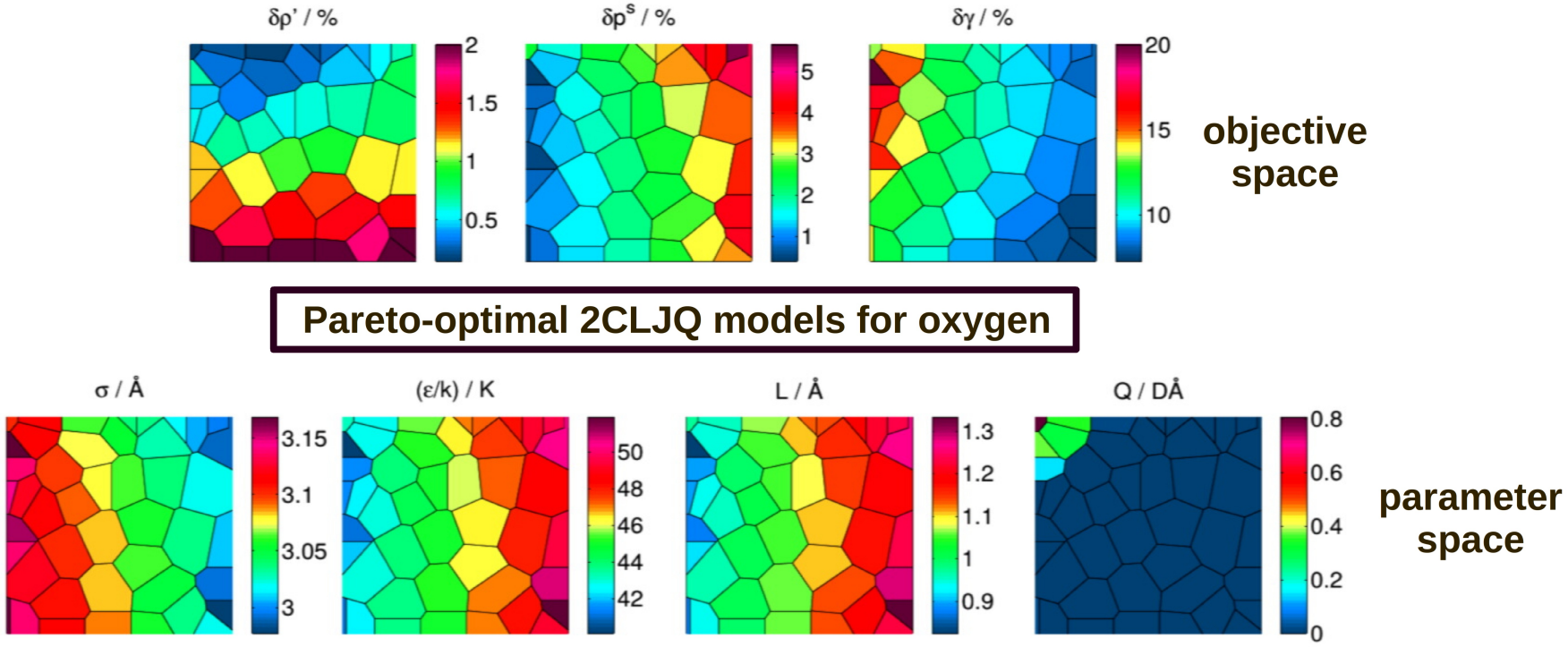
Two LJ + dipole (2CLJD)



- Systematic exploration of the physically relevant part of the model parameter space
- Correlation of *y* by critical scaling expressions (2CLJQ, 2CLJD, Mie-6)

Bespoke model selection by client end users

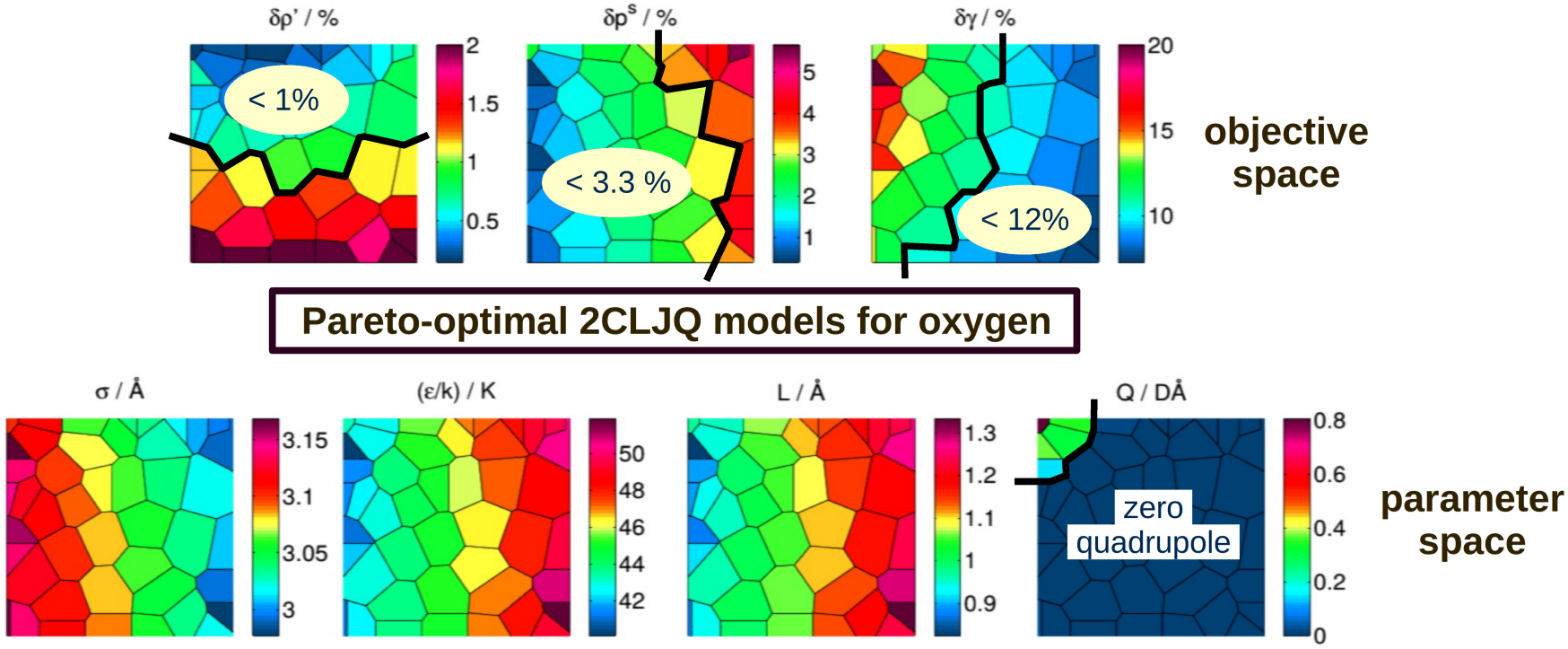
Self-organized patch plots[1] visualizing the Pareto front and the Pareto-optimal models:



[1]K. Stöbener *et al.*, *Fluid Phase Equilib.* 411, 33 – 42, **2016**.

Bespoke model selection by client end users

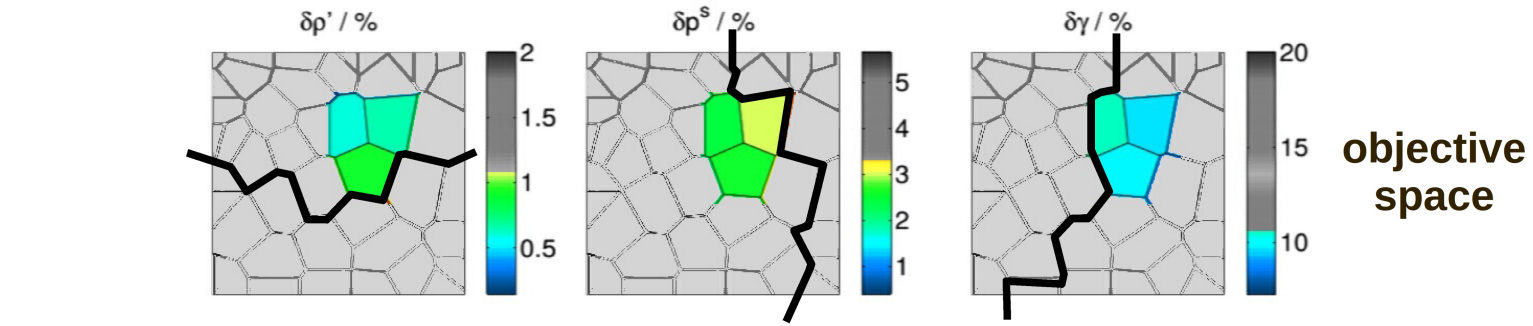
Self-organized patch plots[1] visualizing the Pareto front and the Pareto-optimal models:



[1]K. Stöbener *et al.*, *Fluid Phase Equilib.* 411, 33 – 42, **2016**.

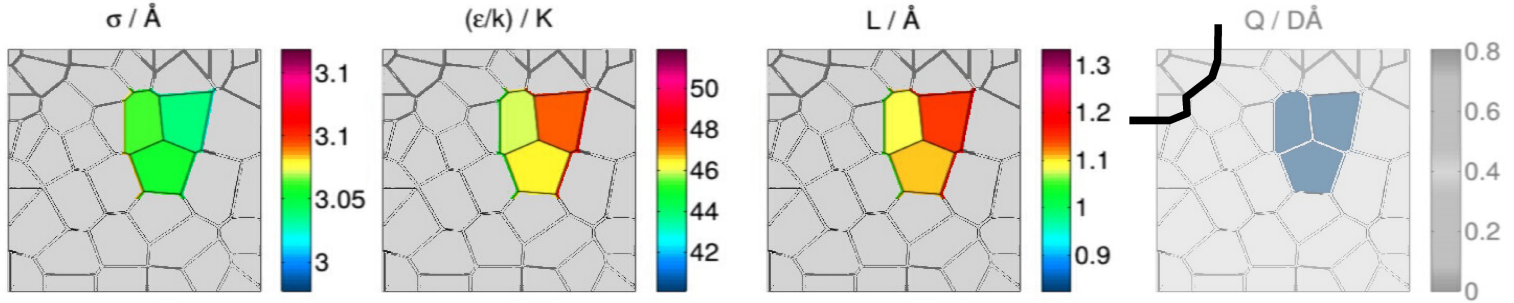
Bespoke model selection by client end users

Self-organized patch plots[1] visualizing the Pareto front and the Pareto-optimal models:



objective space

Pareto-optimal 2CLJ models satisfying all constraints



parameter space

[1]K. Stöbener *et al.*, *Fluid Phase Equilib.* 411, 33 – 42, **2016**.

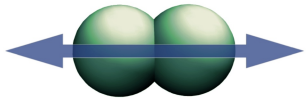
Paradigm shift in computational molecular engineering

Molecular modelling as an art practised by specialized academics

- A very limited community of people is qualified to design molecular models;
- for each fluid, a researcher develops a model, to be looked up from literature;
- it is usually not disclosed how the model was designed and optimized.

Users need a background (e.g., in molecular thermodynamics and statistical mechanics), to assess the model quality, and they **cannot actively contribute** to adjusting the model without repeating the expert work.

Molecular modelling as a technology accessible to all industrial engineers



- Models are flexibly adjustable, e.g., by multicriteria optimization, based on a preceding characterization of the model class;
- bespoke models for specific user cases can be obtained with little effort; or automatically, interoperating with other platforms (e.g., process simulation)
- reliable statements on the model uncertainty are available to the user.

Users who may not possess a background in molecular modelling and simulation are enabled to **actively guide** computational molecular engineering, supported by **translators**.

Significant collaboration and contributions acknowledged:

Berlin	–	Jadran Vrabec
Bremen	–	Welchy Leite Cavalcanti, Michael Hoffmann
Daresbury	–	Silvia Chiacchiera, Michael Seaton, Ilian Todorov
Garching	–	Hans-J. Bungartz, Steffen Seckler, Nikola Tchipev
Hamburg	–	Colin W. Glass, Philipp Neumann
Kaiserslautern	–	Hans Hasse, Peter Klein, Simon Stephan, Katrin Stöbener, Stephan Werth

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 760907.

Computational Molecular Engineering

Science & Technology Facilities Council | UK Research and Innovation

VIMMP VIRTUAL MATERIALS MARKETPLACE