Information Technology

Computational Molecular Engineering als aufstrebende Technologie in der Verfahrenstechnik

Computational Molecular Engineering as an emerging technology in process engineering

Martin Horsch: Laboratory of Engineering Thermodynamics (LTD), TU Kaiserslautern, Erwin-Schrödinger-Str. 44, D-67663 Kaiserslautern, Germany
Tel: +49 631 205 4028, Fax: +49 631 205 3835, E-Mail: martin.horsch@mv.uni-kl.de
*1982 (Stuttgart); 2006 Dipl.-Inf. (U. Stuttgart); 2010 Dr.-Ing. (U. Paderborn); 2011 Junior Professor for Computational Molecular Engineering (TU Kaiserslautern)

Christoph Niethammer: High Performance Computing Centre Stuttgart (HLRS), Nobelstr. 19, D-70569 Stuttgart, Germany
Tel: +49 711 685 87203, Fax: +49 711 685 65832, E-Mail: niethammer@hlrs.de
*1983 (Stuttgart); 2009 Dipl.-Phys. (U. Stuttgart)

Jadran Vrabec: Thermodynamics and Energy Technology (ThEt), University of Paderborn, Warburger Str. 100, D-33098 Paderborn, Germany
Tel: +49 5251 602 421, Fax: +49 5251 603 522, E-Mail: jadran.vrabec@upb.de
*1967 (Basel); 1992 Dipl.-Ing. (U. Bochum); 1996 Dr.-Ing. (U. Bochum); 1997–1999 OMC GmbH (Düsseldorf); 2007 Habilitation (U. Stuttgart); 2009 Professor for Thermodynamics and Energy Technology (U. Paderborn)

Hans Hasse: Laboratory of Engineering Thermodynamics (LTD), TU Kaiserslautern, Erwin-Schrödinger-Str. 44, D-67663 Kaiserslautern, Germany
Tel: +49 631 205 3497, Fax: +49 631 205 3835, E-Mail: hans.hasse@mv.uni-kl.de
*1960 (Landau); 1984 Dipl.-Ing. (U. Karlsruhe, TH); 1990 Dr.-Ing. (TU Kaiserslautern); 1995 Habilitation (TU Kaiserslautern); 1995–1998 BASF AG (Ludwigshafen); 1998 Professor for Thermodynamics and Thermal Process Engineering (U. Stuttgart); 2008 Professor for Thermodynamics (TU Kaiserslautern)

Abstract

The present level of development of molecular force field methods is assessed from the point of view of simulation-based engineering, outlining the immediate perspective for further development and highlighting the newly emerging discipline of *Computational Molecular Engineering (CME)* which makes basic research in soft matter physics fruitful for industrial applications. Within the coming decade, major breakthroughs can be reached if a research focus is placed on *processes at interfaces*, combining aspects where an increase in the accessible length and time scales due to massively parallel high-performance computing will lead to particularly significant improvements.

Zusammenfassung

Der aktuelle Entwicklungsstand molekularer Kraftfeldmethoden wird vom Standpunkt des simulationsgestützten Ingenieurwesens beurteilt, indem Perspektiven für die unmittelbare Zukunft herausgearbeitet werden. Dabei ist insbesondere die neu entstehende Disziplin des *Computational Molecular Engineering (CME)* zu beachten, die Ergebnisse aus der physikalischen Grundlagenforschung für die industrielle Anwendung nutzbar macht. Im kommenden Jahrzehnt sind hier größere Durchbrüche zu erwarten, wenn in der Forschung ein Schwerpunkt auf *Prozesse an Grenzflächen* gesetzt wird. Die Kombination aus den größeren Längen- und Zeitskalen, die durch massiv-paralleles Höchstleistungsrechnen erschlossen werden, wird auf diesem Gebiet zu besonders maßgeblichen Fortschritten führen.
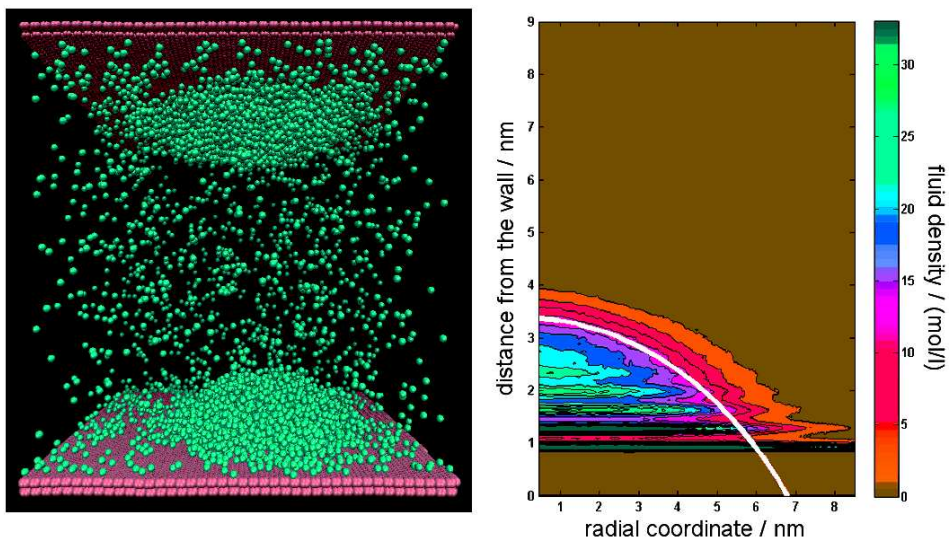
Figure 1: MD simulation snapshot (left) and average fluid density contour plot (right) for a sessile argon droplet on a solid substrate. The simulation was conducted with *ls1 mardyn*.

1 Introduction

Computational Molecular Engineering (CME) is a novel discipline within the broad spectrum of simulation-based engineering and high-performance computing, aiming at adapting molecular force field methods, which were developed within the soft matter physics and thermodynamics communities [1], to the needs of industrial users in chemical and process engineering.

Only in recent years have sufficiently accurate molecular models become available for a wide variety of fluids, and it is only now that, due to massive parallelization, molecular simulations of complex nanoscopic systems (approaching the microscale) are feasible with a reasonable computational effort [7, 14]. The present article briefly comments on the present level of development in CME (Section 2) as well as the massively parallel molecular simulation program *ls1 mardyn* (Section 3). In Section 4, the conclusion is drawn that an increase in the length scale accessible by molecular simulation also requires an extension of the simulated time scale to facilitate a substantial progress.

2 Computational Molecular Engineering today

We witness today the progress from molecular simulation as a theoretical and rather academic method to CME as an arsenal of robust tools intended for practical use, which e.g. supplements or replaces experiments that are hazardous or hard to conduct [15]. This follows the general pattern by which engineering increasingly drives scientific development in areas originating from pure chemistry and physics, building on substantial basic research efforts, as soon as they have become ripe for technical application.

The degree of sophistication of molcular force field methods and the complexity of the simulated systems varies considerably between the various fields of application. In particular, the interdependence of elementary thermodynamic properties such as pressure, density, temperature, enthalpy, and composition can be reliably obtained by simulating homogeneous systems that contain up to 1 000 molecules [1]. With relatively little additional effort, higher-order derivatives of the free energy (e.g. heat capacities or the speed of sound) are accessible as well [9]; the case is similar for mechanical properties of solid materials [13]. By Grand Equilibrium or Gibbs ensemble simulation, vapour-liquid equilibria between homogeneous bulk phases, i.e. without an interface between them, can be efficiently and accurately sampled [4, 15]. Systems where a phase boundary is explicitly present can also be treated, cf. Fig. 1. Such simulations require more molecules, so that finite-size effects can be isolated [3], and longer computations (i.e., with more simulation steps) need to be carried out, since fluid interfaces often relax more slowly than the homogeneous bulk fluid and exhibit more significant fluctuations, e.g. capillary waves, on a long time scale.

This facilitates a modelling approach that has turned out to be particularly fruitful in recent years: Thereby, the electrostatic features of a molecular model, i.e. the choice of parameters for point charges, dipoles or quadrupoles, are determined from quantum chemical calculations. United-atom sites interacting by the Lennard-Jones potential are employed for intermolecular repulsion and dispersive London forces [1], also known as van der Waals forces. The corresponding potential parameters are adjusted to optimize the overall agreement with experimental data [5]. These models, hence, though simple, enable accounting for the most important types of molecular interactions separately, including hydrogen bonding [7]. Furthermore, they account for the structuring effects of the interactions in the fluid (local concentrations, radial distribution functions etc.) in a self-consistent way. This distinguishes them from other approaches for describing fluid properties and explains the fact that such models yield reliable extrapolations with respect to two different aspects: First, to conditions far beyond those where the experimental data for the parameter fit were determined; second, to a wide variety of fluid properties which were not considered during parametrization at all [5].

Once pure component models are available, the extension to mixtures is straightforward. Mixing rules are available for predicting the unlike interaction parameters. If suitable experimental data are available, adjustable binary parameters can be employed to improve mixture models. This concept can also be applied to modelling fluid-wall interactions, cf. Fig. 1. Furthermore, transferable pair potentials are available which directly map functional groups to the model parameters of corresponding single-atom or united-atom interaction sites [10]. In this way, molecular simulation can deploy its predictive power, on the basis of a physically sound modelling approach, even where the available set of experimental data reaches its limits.
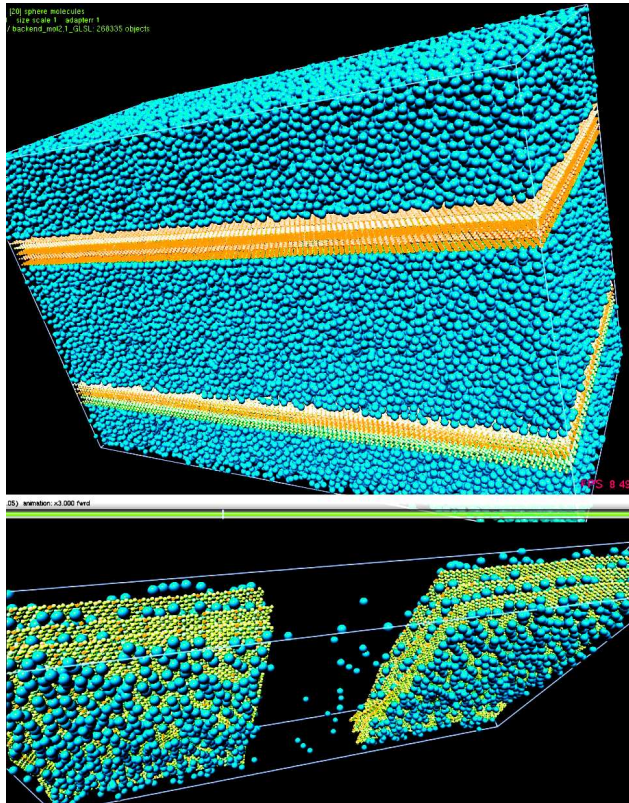
4

Figure 2: Top: MD simulation snapshot for Couette shear flow of methane in a graphite nanopore [8]. Bottom: Entrance effects, adsorption/desorption kinetics, and permeability of fluid methane in nanoporous carbon, employing non-equilibrium MD simulation [8]. The simulations were conducted with *ℓs1 mardyn.*

Both Monte Carlo (MC) and molecular dynamics (MD) simulation are suitable for determining most thermophysical properties: MC simulation evaluates an ensemble average by stochastically generating a representative set of configurations, i.e. position and momentum coordinates of the molecules. Thereby, MC simulation uses the Metropolis algorithm (which is randomized), whereas MD simulation computes a trajectory segment by integrating Newton's equations of motion (which are deterministic). If the same molecular force field is used, temporal and ensemble averaging lead to consistent results, since all thermodynamically relevant systems are at least quasi-ergodic [1]. MC simulation does not rely on time and does not require an explicit computation of momentum coordinates, which is advantageous for simulating adsorption and phase equilibria [15]; in these and similar cases, the most effective methods involve grand-canonical or quasi-grand-canonical ensembles with a varying number of molecules [3], where MD simulation has the disadvantage that momentum coordinates have to be determined for molecules that are inserted into the system. For more complex prop-

5

erties, however, e.g. regarding non-equilibrium states and the associated relaxation processes, time-dependent phenomena become essential, so that MD is the preferred simulation approach (cf. Fig. 2).

Scientifically and technically, all preconditions for the introduction of molecular simulation in an industrial environment are now fulfilled [7]. Organizational aspects relevant for this process include institutional support, the active interest and involvement of both corporate and academic partners, and channelling of the effort to a few simulation codes, at most, rather than reinventing the HPC wheel again and again. In this respect, the development in Great Britain can serve as a positive example, where a community centered around the *Computational Collaboration Project 5* develops and applies the *DL_POLY* program. An example for successful collaboration between academia and industry can be found in the United States, where the *Industrial Fluid Properties Simulation Challenge* also attracts international attention and participation [5]. However, the corresponding programming efforts are highly fragmented: Parallel developments are attempted based on the *Amber*, *CHARMM*, *LAMMPS*, *NAMD* and *MC-CCS Towhee* codes, among many others [10, 11, 12].

At present, however, the German CME community arguably constitutes the best environment for mastering the qualitative transition of molecular simulation from a scholarly academic occupation to a key technology in industrial-scale fluid process engineering. Its institutional structure guarantees an orientation towards industrial use and successfully integrates engineering with high performance computing. It is within this framework that a consistent toolkit encompassing two major components is developed: The *ms2* program (i.e. <u>m</u>olecular <u>s</u>imulation: <u>second</u> generation), intended for thermophysical properties of bulk fluids – the interested reader is referred to Deublein et al. [4] – and *ℓs1 mardyn*, which is briefly introduced below.

3   Large systems in molecular dynamics

The novel MD simulation code *ℓs1 mardyn* (i.e. <u>ℓ</u>arge <u>s</u>ystems <u>f</u>irst: <u>m</u>olecular <u>d</u>ynamics) aims at expanding the temporal and spatial range of scales accessible to molecular simulation, with a focus on inhomogeneous systems (e.g. at interfaces) and non-equilibrium thermodynamics. In many relevant cases, such as those involving vapour liquid coexistence, the molecule distribution may be very heterogeneous and change over time in an unpredictable way (e.g. during phase decomposition or multi-phase flow). The design of *ℓs1 mardyn* was oriented towards three major objectives: Modularity of the structure, interdisciplinar collaboration within the development process, and transferability of the code base to diverse and heterogeneous HPC architectures.

Molecular simulations with system dimensions that far exceed the cut-off radius, beyond which a mean-field approach is employed for the intermolecular interactions, are most efficiently parallelized by space decomposition schemes. Thereby, the simulation volume is subdivided into smaller subvolumes (one for each process) that ideally carry the same load [2]. Finding an optimal balance requires a method that estimates the load corresponding to the possible decompositions on the fly, since the particle density distribution can vary significantly over simulation time.

In *ℓs1 mardyn*, an interface class for the domain decomposition scheme permits the generic implementation of different load balancing strategies operating on spatial subdomains. The linked-cell data structure, which is used to find neighboring molecules, introduces a subdivision of the simulation volume into small cells (with dimensions on the order of the cut-off radius) that are also used as basic volume units for the decomposition.

On the basis of the computational cost for each of the cells, load balancing algorithms can group cells together such that N_p subvolumes with approximately equal load are created, where N_p is the number of processing units. A hierarchical tree-based approach (similar to k-dimensional trees) turned out to be the most adequate dynamic load balancing method; this concept was previously known to be suitable for parallelizing particle-based simulations with short-range interactions [2]. Thereby, the simulation volume is recursively bisected by planes with alternating orientation (normal to the x, y, z, x, ... axis). This process is repeated until each process is assigned one cuboid subdomain. With this technique, an excellent scalability was obtained even for heterogeneous scenarios on up to 10^5 processing units of the *hermit* supercomputer at the High Performance Computing Centre Stuttgart (HLRS), cf. Fig. 3.

4   Conclusion

From a computational point of view, large MC or MD simulations can better be tackled than MD simulations of processes over a relatively long time span. By far the largest part of the numerical effort is required for evaluating the force field, a task which can be efficiently distributed over multiple processes, as discussed above. In contrast, the temporal evolution along a trajectory through the phase space cannot be parallelized due to its inherently sequential nature. In the past, this has repeatedly led developers of highly performant simulation codes to boast of the number of (low-density gas) molecules that they succeeded in loading into memory as the single considered benchmark criterion [6].

However, large and *interesting* systems generally also require more simulation *time*. Industrial users will hardly care how many trillions of molecules can be simulated for a dilute homogeneous gas over a few picoseconds, or even less. From the point of view of thermodynamics and fluid process engineering, the criterion for the world record in molecular simulation should not be the number of molecules N, but rather an exponent ℓ such that e.g. within a single day, at least $N = 10^{3\ell}$ molecules in a condensed state were simulated over at least $10^{\ell+4}$ time steps. This would promote a proportional increase of the accessible length and time scales, which is what real-life applications require.

By pushing this frontier forward, a wide spectrum of novel scale-bridging simulation approaches will become feasible, paving the way to a rigorous investigation of many size-dependent effects, which on the microscale may be qualitatively different from the nanoscale. Following this route, major breakthroughs will be reached within the coming decade, assuming that a research focus is placed on processes at interfaces. By focussing on such applications, cf. Fig. 2, an increase in

the accessible length and time scale due to massively parallel high-performance computing will lead to particularly significant improvements, opening up both scientifically and technically highly interesting fields such as microfluidics (including turbulent flow), coupled heat and mass trasfer, and design of functional surfaces to an investigation on the molecular level.

References

[1] ALLEN, M. P., AND TILDESLEY, D. J. *Computer Simulation of Liquids.* Oxford University Press, 1989.

[2] BERNARD, P.-E., GAUTIER, T., AND TRYSTRAM, D. *Large Scale Simulation of Parallel Molecular Dynamics.* IEEE Computer Society, Washington, D.C., 1999, pp. 638–644.

[3] BINDER, K. Applications of Monte Carlo methods to statistical physics. *Rep. Prog. Phys. 60*, 5 (1997), 487–559.

[4] DEUBLEIN, S., ECKL, B., STOLL, J., LISHCHUK, S. V., GUEVARA CARRIÓN, G., GLASS, C. W., MERKER, T., BERNREUTHER, M. F., HASSE, H., AND VRABEC, J. ms2: A molecular simulation tool for thermodynamic properties. *Comput. Phys. Comm. 182*, 11 (2011), 2350–2367.

[5] ECKL, B., VRABEC, J., AND HASSE, H. *Fluid Phase Equilib. 274*, 1–2 (2008), 16–26.

[6] GERMANN, T. C., AND KADAU, K. Trillion-atom molecular dynamics becomes a reality. *Int. J. Mod. Phys. C 19*, 9 (2008), 1315–1319.

[7] GUEVARA CARRIÓN, G., HASSE, H., AND VRABEC, J. Thermodynamic properties for applications in chemical industry via classical force fields. In *Multiscale Molecular Methods in Applied Chemistry* (Heidelberg, 2012), no. 307 in Topics in Current Chemistry, Springer, pp. 201–249.

[8] HORSCH, M., VRABEC, J., BERNREUTHER, M. F., AND HASSE, H. Poiseuille flow of liquid methane in nanoscopic graphite channels by molecular dynamics simulation. In *Proc. VI. Int. Symp. THMT* (2009), K. Hanjalić et al., Eds., Begell House, pp. 89–92.

[9] LUSTIG, R. Direct molecular NVT simulation of the isobaric heat capacity, speed of sound, and Joule-Thomson coefficient. *Molec. Sim. 37*, 6 (2011), 457–465.

[10] MARTIN, M. G., AND SIEPMANN, J. I. Novel configurational-bias Monte Carlo method for branched molecules. transferable potentials for phase equilibria. 2. United-atom description of branched alkanes.

[11] PHILLIPS, J. C., BRAUN, R., WANG, W., GUMBART, J., TAJKHORSHID, E., VILLA, E., CHIPOT, C., SKEEL, R. D., KALE, L., AND SCHULTEN, K. Scalable molecular dynamics with NAMD. *J. Computat. Chem. 26*, 16 (2005), 1781–1802.

[12] PLIMPTON, S. Fast parallel algorithms for short-range molecular dynamics. *J. Computat. Phys. 117*, 1 (1995), 1–19.

[13] Rösch, F., and Trebin, H.-R. Crack front propagation by kink formation. *EPL 87* (2009), 66004.

[14] Stone, J. E., Hardyn, D. J., Ufimtsev, I. S., and Schulten, K. GPU-accelerated molecular modeling coming of age. *J. Molec. Graphics Modeling 29*, 2 (2010), 116–125.

[15] Ungerer, P., Nieto-Draghi, C., Rousseau, B., Ahunbay, G., and Lachet, V. Molecular simulation of the thermophysical properties of fluids: From understanding toward quantitative predictions. *J. Molec. Liq. 134* (2007), 71–89.
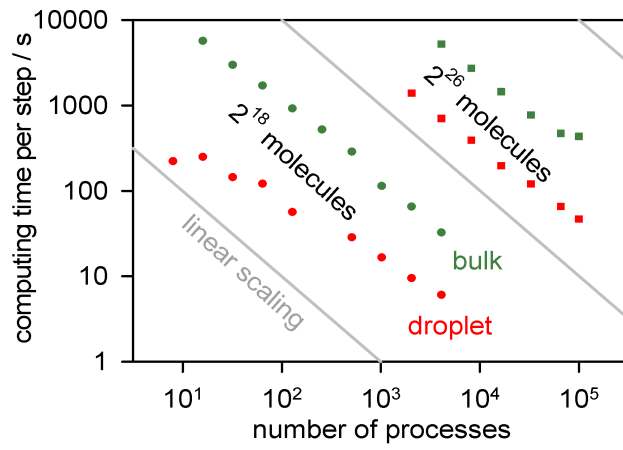
9

Figure 3: Strong scaling of the *ℓs1 mardyn* program on the *hermit* supercomputer for homogeneous (bulk) and heterogeneous (droplet) scenarios.