Automatized determination of fundamental equations of state based on molecular simulations in the cloud

Andreas Köster [a], Tao Jiang [b], Gábor Rutkai [a], Colin W. Glass [b], Jadran Vrabec [a, *]

[a] Lehrstuhl für Thermodynamik und Energietechnik, Universität Paderborn, Warburger Str. 100, 33098 Paderborn, Germany
[b] Höchstleistungsrechenzentrum Stuttgart (HLRS), Universität Stuttgart, 70550 Stuttgart, Germany

ARTICLE INFO

ABSTRACT

An automatized procedure for the parameterization of fundamental equations of state (EOS) that are explicit in terms of the Helmholtz energy and are based on molecular simulation data is presented. The simulation runs are carried out via a cloud-based framework that combines multiple, distributed computing resources. A user-friendly graphical user interface ensures that minimal knowledge about the background operations is required. In order to exemplify the capabilities of this approach an EOS for ethylene oxide is created and compared to data from the literature.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Thermodynamic data are essential for the design and optimization of chemical engineering processes. Typically, the chemical industry is relying on experimental investigations to generate such data. Depending on the type of fluid, this undertaking is difficult to conduct or in the case of hazardous substances, e.g. being explosive, toxic or mutagenic, this might even be practically impossible. As an attractive alternative route, molecular modelling and simulation on the basis of classical force fields has been proven to be an adequate method due to the steady developments of the last decades. Along with its powerful predictive capabilities, large sets of thermophysical data can be rapidly produced with considerably less effort than laboratory measurements.

It has recently been shown that molecular simulation data are useful for the construction of empirical fundamental equation of state (EOS) correlations that are explicit in terms of a thermodynamic potential [1—4], which is most often the Helmholtz energy [5]. Thermodynamic potentials are essential building blocks of thermodynamics, because every other time-independent thermodynamic property can be obtained as a combination of their partial derivatives with respect to their independent variables. Unfortunately, the exact mathematical expression of a thermodynamic potential, with the exception of very simple systems (e.g. ideal gas or some hard-body fluids), is unknown. The construction of an EOS therefore necessarily means that the mathematical form of such a correlation has to be fitted to thermodynamic data. Once the parameterized correlation is available, it provides thermodynamic data in a consistent manner and can be used to inter- and extrapolate between or beyond the underlying data points. More importantly, EOS have the ability to yield all thermodynamic data, and not just the ones considered in their fitting process. Naturally, the more thermodynamically independent properties are considered during the construction of EOS, the better the representation of other properties. Therefore, in order to create such a correlation, a large number of independent thermodynamic data points has to be available. This scenario is the application field for the statistical mechanical formalism proposed by Lustig [6,7]. With this formalism, the derivatives of the Helmholtz energy up to arbitrary order can be obtained at a given state point by a single canonical (NVT) ensemble molecular simulation run [6,7]. This formalism also exists for the entropy derivatives in the microcanonical (NVE) ensemble [6,7]. Since any thermodynamic property is just a combination of these derivatives, this represents the full set of thermodynamically independent data. Due to this improvement on availability, reliable EOS can very efficiently be fitted on the basis of

* Corresponding author.
E-mail address: jadran.vrabec@upb.de (J. Vrabec).

molecular simulation data only.

However, setting up molecular simulation runs and fitting EOS requires considerable manual effort and expertise. Maginn [8] pointed out that molecular simulation has to overcome some barriers in order to become a mainstream tool in chemical engineering. One of these barriers is the time which is necessary to set up and analyze a molecular simulation run compared to its actual runtime. Furthermore, there are few software packages which are able to yield all properties that are important for the chemical industry (e.g. phase equilibria, gas solubility, heat capacity, density, etc.) [8]. Hence, the goal of this work is to automatize and optimize this workflow by combining the molecular simulation tool *ms*2 [9] with an EOS fitting algorithm and embedding this into a cloud-based simulation framework which requires minimal knowledge about the background operations, and has a simple graphical user interface (GUI) for creating the necessary input for simulations. In the background, a job manager efficiently allocates the molecular simulation runs to the available high performance computing (HPC) resources. Once the simulation runs are terminated and the results are ready, the empirical EOS correlation is fitted and its parameters are returned, along with the quality of representation of the simulation data. Thus, the development of EOS is essentially becoming an automatic process. In this work, the capabilities of this approach and the resulting EOS are compared for ethylene oxide with experimental data from the literature and with the recently published EOS by Thol et al. [4] for this fluid, that we consider as a reference here.

2. Fundamental equation of state correlation

The present EOS correlation is explicit in terms of the Helmholtz energy a. It is common to separate a into an ideal contribution a^o and a residual contribution a^r

$$\alpha(T,\rho) = \frac{a^o(T,\rho) + a^r(T,\rho)}{RT} = \alpha^o(T,\rho) + \alpha^r(T,\rho)$$
$$= \alpha^o(\tau,\delta) + \alpha^r(\tau,\delta), \tag{1}$$

with the reduced inverse temperature $\tau = T_c/T$, the reduced density $\delta = \rho/\rho_c$ and the molar gas constant $R = 8.3144621\,\text{J}\,\text{mol}^{-1}\,\text{K}^{-1}$ [10], where T_c is the critical temperature and ρ_c is the critical density. The ideal contribution a^o, which only depends on the properties of single molecules, can usually be obtained from spectroscopic data or ab initio calculations, whereas the much more demanding residual contribution is the target of molecular simulation because it depends on the interactions between the molecules. The empirical equation that represents the residual contribution to the reduced Helmholtz energy α^r consists of polynomial and exponential terms

$$\alpha^r(\tau,\delta) = \sum_{k=1}^{N_{\text{pol}}} n_k \tau^{t_k} \delta^{d_k} + \sum_{k=N_{\text{pol}+1}}^{N_{\text{pol}}+N_{\text{exp}}} n_k \tau^{t_k} \delta^{d_k} \exp\left(-\delta^{l_k}\right). \tag{2}$$

In general, the simultaneous optimization of coefficients n_k, exponents t_k, d_k, l_k as well as the number and type of polynomial and exponential terms requires expert use of complex non-linear fitting algorithms [5,11]. Assuming that a sufficient amount of thermodynamic data is available, such an effort usually takes months or years, requiring a high level of human intervention and experience. The recently published EOS of Thol et al. [4] for ethylene oxide is such a model, and was therefore chosen as a reference in the present work. In the fitting process of that EOS [4], both experimental and molecular simulation data were used. Nonetheless, such an approach is currently not feasible in an automatized EOS development process. There are, however,

recommendations in the literature for the functional form and the values of the exponents t_k, d_k, l_k of correlations that have proven to be a good choice for a large number of fluids. These are for example the 10- and 12-term EOS by Span and Wagner [5,12], the 14-term EOS by Sun and Ely [13] and the modified Benedict-Webb-Rubin (MBWR) EOS [14], which consists of 40 terms when transformed into the Helmholtz energy representation [5]. In this work, the MBWR EOS [14] was chosen to represent the molecular simulation data set. Such generalized EOS have the advantage that, since the exponents t_k, d_k, l_k in Eq. (2) are known, the task at hand reduces to the fitting of the coefficients n_k. This can be done in a fully automatized manner and has a response time of seconds, e.g., with the algorithm for weighted multiproperty fits of Hust and McCarty [15], in which the weight corresponds to the statistical uncertainty of the underlying molecular simulation data. The values of the coefficients t_k, d_k, l_k and n_k for the present EOS can be found in Table 1.

3. Software architecture

Due to the volume of calculations in the field of molecular simulation, researchers typically rely on large computing infrastructure. Here, a single molecular simulation run is relatively time intensive and the database required to parameterize an EOS should

Table 1
Parameters for the residual part of the reduced Helmholtz energy $\alpha^r(\tau,\delta)$ according to Eq. (2). Note that only the coefficients n_k were fitted in the automatized procedure, the exponents t_k, d_k, l_k were taken from the literature [14].

k	n_k	t_k	d_k	l_k
1	−0.444761941	0	1	0
2	4.295500322	0.5	1	0
3	−6.304073590	1	1	0
4	2.495663889	2	1	0
5	−1.413871625	3	1	0
6	0.122634301	0	2	0
7	−0.403993247	1	2	0
8	0.953802102	2	2	0
9	3.626258062	3	2	0
10	−0.005286229	0	3	0
11	0.167724139	1	3	0
12	−0.261302669	2	3	0
13	0.037473360	1	4	0
14	−0.148085338	2	5	0
15	−0.112161698	3	5	0
16	0.056427170	2	6	0
17	−0.005390223	2	7	0
18	0.010228306	3	7	0
19	−0.001370416	3	8	0
20	−15.487760440	3	0	0
21	−0.079331908	4	0	0
22	0.006185345	5	0	0
23	15.486057470	3	0	2
24	0.081647538	4	0	2
25	−0.007000321	5	0	2
26	10.769184560	3	2	2
27	1.108403574	4	2	2
28	−0.210854465	5	2	2
29	3.440671002	3	4	2
30	0.841327903	4	4	2
31	−0.256828769	5	4	2
32	0.416506137	3	6	2
33	0.492487787	4	6	2
34	−0.007594133	5	6	2
35	0.123791118	3	8	2
36	0.027697626	4	8	2
37	−0.042799502	5	8	2
38	−0.005719867	3	10	2
39	0.022433309	4	10	2
40	0.002604667	5	10	2