

Virial equation of state as a new frontier for computational chemistry ^F

Cite as: J. Chem. Phys. **157**, 190901 (2022); <https://doi.org/10.1063/5.0113730>

Submitted: 24 July 2022 • Accepted: 21 August 2022 • Accepted Manuscript Online: 08 September 2022 • Published Online: 15 November 2022

 Andrew J. Schultz and  David A. Kofke

COLLECTIONS

Paper published as part of the special topic on [JCP Editors' Choice 2022](#)

 This paper was selected as Featured



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Computing chemical potentials of solutions from structure factors](#)

The Journal of Chemical Physics **157**, 121101 (2022); <https://doi.org/10.1063/5.0107059>

[Phase diagrams—Why they matter and how to predict them](#)

The Journal of Chemical Physics **158**, 030902 (2023); <https://doi.org/10.1063/5.0131028>

[The Asakura–Oosawa theory: Entropic forces in physics, biology, and soft matter](#)

The Journal of Chemical Physics **156**, 080401 (2022); <https://doi.org/10.1063/5.0085965>



Time to get excited.

Lock-in Amplifiers – from DC to 8.5 GHz



[Find out more](#)


Zurich
Instruments

Virial equation of state as a new frontier for computational chemistry

Cite as: J. Chem. Phys. 157, 190901 (2022); doi: 10.1063/5.0113730

Submitted: 24 July 2022 • Accepted: 21 August 2022 •

Published Online: 15 November 2022



Andrew J. Schultz  and David A. Kofke^{a)} 

AFFILIATIONS

Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, New York 14260-4200, USA

^{a)} Author to whom correspondence should be addressed: kofke@buffalo.edu

ABSTRACT

The virial equation of state (VEOS) provides a rigorous bridge between molecular interactions and thermodynamic properties. The past decade has seen renewed interest in the VEOS due to advances in theory, algorithms, computing power, and quality of molecular models. Now, with the emergence of increasingly accurate first-principles computational chemistry methods, and machine-learning techniques to generate potential-energy surfaces from them, VEOS is poised to play a larger role in modeling and computing properties. Its scope of application is limited to where the density series converges, but this still admits a useful range of conditions and applications, and there is potential to expand this range further. Recent applications have shown that for simple molecules, VEOS can provide first-principles thermodynamic property data that are competitive in quality with experiment. Moreover, VEOS provides a focused and actionable test of molecular models and first-principles calculations via comparison to experiment. This Perspective presents an overview of recent advances and suggests areas of focus for further progress.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0113730>

I. INTRODUCTION

Ab initio and related methods for computing physical properties from first-principles considerations are advancing rapidly. Evaluation of single-molecule properties (e.g., charge distribution, spectroscopic behavior) for small- to medium-sized molecules is routine,¹ applications of density functional theory and its extensions² to crystalline phases are competitive with experiment, and first-principles techniques are increasingly finding their way into fluid-phase simulations.³ A logical progression from the current situation to the next level of application should advance through the non-condensed vapor and supercritical fluid phases, which provides a clear path from single-molecule, ideal-gas properties to nontrivial, dense gases. Properties of interest in this context include those relating to volumetric behavior—the equation of state in particular—as well as thermal properties, such as the heat capacity; a number of other properties of interest, e.g., the Joule–Thomson (JT) coefficient, can be derived from these. However, there has to date been no concerted effort to develop and apply first-principles methods in this direction. The value of doing so is twofold: (1) It provides a systematic approach to rigorous and precise testing of the molecular models and *ab initio* methods, through comparison to available

experimental data; and (2) it can yield useful computational-sourced data for properties of important systems for which no experimental data are available. The conventional approach for bridging molecular models to thermodynamic properties is molecular simulation, but, at least for gases and supercritical fluid phases, an alternative path exists that has some advantages; see Fig. 1.

More specifically, *ab initio* methods and molecular models can be connected rigorously to bulk fluid properties via calculation of “cluster integrals,” which are integrals over the configurations of a small number of molecules (often fewer than ten). These methods typically provide fluid properties in the form of a power series in density, with coefficients expressed in terms of the cluster integrals. As a modeling tool, cluster-integral methods hold a distinct position between molecular simulation on the one hand and traditional molecularly inspired thermodynamic models on the other. Specifically,

- like simulation, (1) for the state conditions where they are applicable, they provide an exact thermodynamic description for any system that is defined in terms of its molecular interactions; (2) they include the theoretically correct dependence on state variables, including mole fractions when

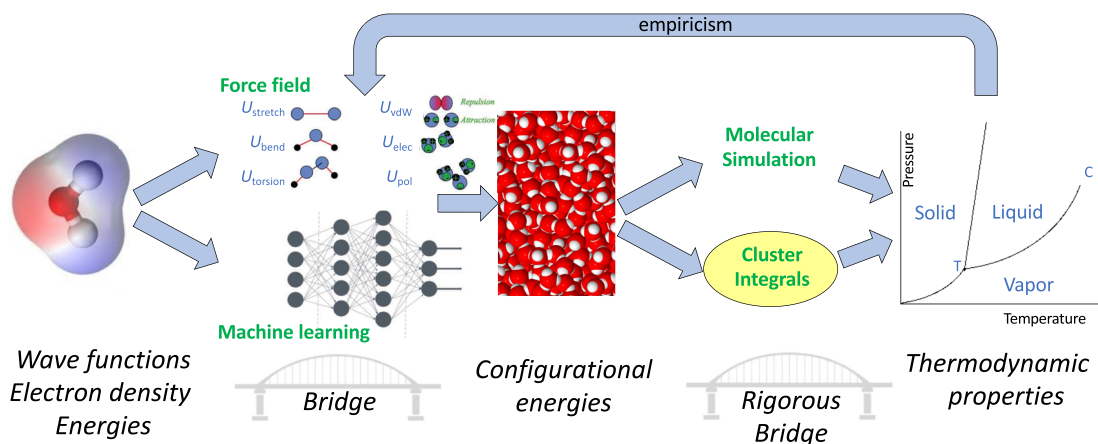


FIG. 1. Illustration of paths from first-principles methods to thermodynamic properties. *Ab initio* calculations feed into either force-field models (perhaps supplemented empirically with experimental data) or symmetry-aware machine-learning methods for interpolating the multidimensional potential-energy surface (PES). Knowledge of configurational energies is used via ensemble averaging to compute thermodynamic properties for the input molecular model; this can be accomplished rigorously using molecular simulation (most common path) or via calculation of cluster integrals (virial coefficients), which is the focus of this Perspective.

applied to mixtures; (3) they can accommodate any level of complexity in the model, all with full statistical-mechanical rigor; (4) they can provide information for a broad range of thermodynamic properties, all consistent with a given molecular model.

- like thermodynamic models, (5) they provide a model in the form of an equation (rather than discrete data), which can be manipulated, analyzed, and used in larger calculations (e.g., process design) as needed; (6) cluster integrals (i.e., thermodynamic model parameters) developed for one substance enter as part of the description of mixtures containing that substance, so effort made to characterize it is not discarded when it appears elsewhere.
- and, unique to cluster methods, (7) unfortunately, their applicability is (at present) limited to non-condensed fluid phases and, thus, excludes, e.g., liquids, which are separated from the low-density gas by a first-order phase transition; moreover, they become more difficult to apply with increasing density, even at supercritical temperatures; (8) however, unlike any other thermodynamic model, when failure occurs, it is detectable—variation of the property with addition of higher-order terms indicates that it is not accurate at a given order; (9) related to this, cluster methods are based on a well-defined approximation and they can be systematically improved; finally, (10) they can be interfaced with other known features of the fluid-property behavior to form approximants that are more effective over a broader range of conditions, compared to the basic series.

Thus, the approach as a whole really is in a class by itself, wherein the simultaneous character of being both a detailed molecular treatment and a rigorous thermodynamic model makes it a methodology with unique appeal. The small community of researchers working on the virial framework has made truly remarkable progress in recent years toward turning it into a practical tool to connect molecular and bulk properties, one that may come

to rival molecular simulation in some respects. Major advances have been made in methods for calculating the cluster integrals, understanding their convergence, and extending their range of application via interfacing with approximants. A particularly important development, which makes the proposed activity possible now for the first time, is the formulation of very efficient algorithms for computing cluster integrals and their derivatives for multibody potentials.

II. FRAMEWORK AND RECENT ADVANCES

A. Virial equation of state

1. Virial coefficients

The virial equation of state (VEOS) relates pressure p , temperature T , and number density ρ as a series expansion in density,^{4–6}

$$\frac{p}{k_B T} = \rho + B_2 \rho^2 + B_3 \rho^3 + \cdots + B_n \rho^n, \quad (1)$$

where k_B is the Boltzmann constant and B_k is the k th virial coefficient. We label the series truncated at order n (shown here) as VEOS $_n$. The series in terms of the activity $z = \exp(\mu/k_B T)$, where μ is the chemical potential, is sometimes considered in developing, understanding, or extending the overall framework,

$$\frac{p}{k_B T} = z + b_2 z^2 + b_3 z^3 + \cdots + b_n z^n. \quad (2)$$

The B_n and b_n coefficients can be computed from each other in a straightforward way.⁷

The virial coefficients are independent of density and depend only on temperature; coefficients for mixtures are given rigorously as a polynomial in mole fractions (see Sec. II A 2). The virial coefficients

are related to the intermolecular potential via the configuration integral,

$$B_n = \frac{1-n}{n!} \int f_B(\mathbf{r}^n) d\mathbf{r}_{12} d\mathbf{r}_{13} \cdots d\mathbf{r}_{1n}, \quad (3)$$

where molecule 1 defines the origin and positions of molecules 2, 3, ..., n are defined relative to it (integration over orientations or intramolecular coordinates would be performed as well, if needed). Here, f_B is the Husumi function, and it is given as a complicated sum of products of terms involving the energy U_k of groupings of $k \leq n$ molecules, formed from partitions of the entire set of n molecules in configuration $\mathbf{r}^n$. For pairwise-additive models, such that $U_k = \sum_{i < j} u_{ij}$, f_B can be represented as a sum of doubly connected graphs, with bonds given by the Mayer function, $f_{ij} = \exp(-u_{ij}/k_B T)$. Significantly, the approximation of pairwise additivity is not a requirement, and the VEOS can be applied with virial coefficients computed using potentials that include multibody contributions. It has been shown that such contributions are essential to providing property values that are competitive in accuracy to experiment.^{8–13}

2. Mixtures

The treatment of mixtures by the VEOS is handled rigorously. The coefficient B_n for a mixture of c components is given as a polynomial in mole fractions of the species, with coefficients given by cluster integrals computed for the corresponding set of molecules. For example, B_3 for a three-component mixture is

$$B_3 = B_{300}x_1^3 + 3B_{210}x_1^2x_2 + 3B_{120}x_1x_2^2 + B_{030}x_2^3 + 3B_{201}x_1^2x_3 + 6B_{111}x_1x_2x_3 + 3B_{021}x_2^2x_3 + 3B_{102}x_1x_3^2 + 3B_{012}x_2x_3^2 + B_{003}x_3^3, \quad (4)$$

where B_{ijk} is the cluster integral for i molecules of species 1, j molecules of species 2, and k molecules of species 3, with mole fractions x_1 , x_2 , and x_3 , respectively. The generalization to other orders n and number of species c is straightforward.

Often, the weak link in the treatment of mixtures is evaluating the cross-species configurational energy needed for coefficients that involve a heterogeneous set of molecules. Potentials are typically developed and refined for pure species, and when needed, unlike-molecule interactions are estimated using semiempirical combining rules, such as Lorentz–Berthelot. However, in principle, it is possible to formulate energy surfaces for heterogeneous sets of molecules, and this is sometimes done, but it obviously involves more computational effort. Nevertheless, calculations that aim for high accuracy relative to experiment should develop bespoke potential models for each set of molecule types and cannot rely on approximate combining rules. If the potential is modeled by summing only two- and three-body contributions, the effort required can be manageable.

3. Other properties

When supplemented by ideal-gas thermal properties, standard thermodynamic manipulations allow derivation of formulas¹⁴ for any thermodynamic property of interest, for any temperature, density, and composition within the region of convergence of the

series (and where coefficient values are available). This includes thermal expansion coefficient, heat capacities, Joule–Thomson inversion curve, fugacity coefficients, etc. Cluster integrals cannot be performed to compute transport coefficients, and attempts to develop an appropriate density expansion for these properties have been problematic,¹⁵ but extensions of the VEOS approach have been developed for calculation of other properties, such as those related to inhomogeneous fluids and surfaces,¹⁶ and electrostatic phenomena¹⁷ (e.g., dielectric constant).

The VEOS is a rigorous description of the equation of state for the molecular model used to compute the virial coefficients. Where it is converged, it provides exact values of the pressure for a given density, temperature, and mixture composition. In particular, it describes behavior in the thermodynamic limit, so it is not subject to finite-size errors that arise in molecular simulation.

B. Calculation of virial coefficients

1. Mayer-sampling Monte Carlo

The integral in Eq. (3) must be evaluated numerically using Monte Carlo (MC) methods. Mayer-sampling Monte Carlo (MSMC)¹⁸ is formulated by extending single-stage and multistage free-energy methods, and it has opened up many opportunities for evaluating cluster integrals. In MSMC, configurations are generated via a Markov process, sampled in proportion to a distribution $\pi(\mathbf{r}^n)$. For each configuration, f_B is evaluated, and the integral is given as the average

$$B_n = B_n^{\text{ref}} \frac{\langle f_B/\pi \rangle_\pi}{\langle f_B^{\text{ref}}/\pi \rangle_\pi}, \quad (5)$$

where $B_n^{\text{ref}} \equiv \frac{n-1}{n!} \int f_B^{\text{ref}} d\mathbf{r}^{n-1}$ is the virial coefficient (or some cluster integral) for a reference system (typically based on the hard-sphere model) and is taken as known. Importance sampling suggests that π be given by f_B , but because f_B can be negative, an absolute value is used. With this choice, it is likely that the reference system is incompletely sampled, so we routinely employ a multistage method^{19,20} (overlap sampling, a variant of Bennett's method²¹) in which configurations are generated in parallel by sampling on $\pi^{\text{ref}} = |f_B^{\text{ref}}|$. Averages in both samples are then taken for an overlap function that is defined in terms of f_B and f_B^{ref} , with a parameter [and allocation of central processing unit (CPU) time between the two samples] that is tuned to obtain precise B_n most efficiently.

Mayer-sampling calculations are performed in an open volume, with no containing walls and no periodic boundaries—importance sampling keeps the molecules together. Configurations with large steric overlap are included in the samples because f_B is nonzero there (Fig. 2). No correction for truncation of the potential is needed, even for systems with electrostatic interactions (i.e., no Ewald sum or similar methods are invoked). However, it is sometimes necessary to couple offsetting orientations when computing a weight for a configuration of molecules that have electrostatic interactions. This rough average can be as simple as including the direct configuration as well as others where each molecule is flipped to reverse the direction of its dipole moment. Electrostatic interactions decay more rapidly upon Boltzmann-weighted orientation averaging, relative to fixed orientation, and this flipping process prevents molecules from oversampling distant separations.

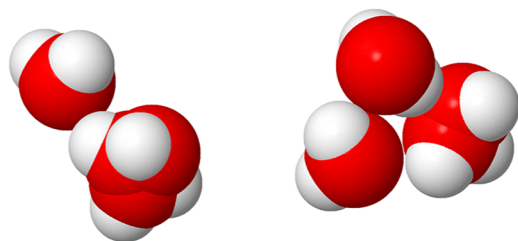


FIG. 2. Configurations from MSMC calculations to compute B_4 for water, using overlap sampling. Image on the left is from sampling of a hard-sphere reference system and on the right is based on sampling of the target system (water). The hard-sphere diameters in the reference are larger than the displayed oxygen atoms, and they possess orientations that do not affect sampling but are needed for the quantity being averaged.

2. Wheatley's recursion

Each configuration generated in a MSMC process requires evaluation of the integrand f_B . This calculation is increasingly complicated and computationally expensive with increasing order of the coefficient n . This is the major barrier to evaluation of virial coefficients to arbitrarily high order n . For example, for pairwise-additive potentials, f_B requires evaluation of a sum of terms that grows in number as 2^{n^2} . Table I lists the total number of graphs that can be formed from n distinguishable points with zero or one bonds joining each pair. The fraction of these that are doubly connected must be summed to evaluate f_B for each configuration. By just $n = 13$, the total number of graphs is of order 10^{23} , of which 96% are biconnected (also referred to as “irreducible”). Clearly, direct evaluation of such a sum, even once, is wholly infeasible.

Ten years ago, Wheatley²² proposed a very clever method to break through this barrier, or at least push it back. He presented a recursive method for evaluating f_B with computational effort that grows (only) exponentially with n , specifically as 3^n . This rate of growth is illustrated in the table as well, showing that calculation

TABLE I. Demonstration of the increase in computational effort required for direct sum of graphs (second column) vs Wheatley's recursion (fourth column).

Order n	Number of graphs $2^{n(n-1)/2}$	Fraction biconnected (%)	3^n
2	2	50	9
3	8	13	27
4	64	16	81
5	1024	23	243
6	32 768	34	729
7	2 097 152	48	2187
8	268 435 456	62	6561
9	68 719 476 736	74	19 683
10	35 184 372 088 832	83	59 049
11	36 028 797 018 963 968	89	177 147
12	73 786 976 294 838 206 464	93	531 441
13	0.501 mol	96	1 594 323

of B_n for $n = 13$ is only of order 10^5 times more computationally expensive than $n = 2$. The calculation still becomes untenable for not-very-large n , but the advance extends our capabilities by a useful amount. Apart from the expense of evaluating f_B , larger n implies a larger space of configurations to sample, and the presence of offsetting positive and negative contributions to the average can further slow convergence.

Arguably, even more important than the computational savings provided by Wheatley's algorithm is the relative simplicity it presents for writing a general-purpose code that allows calculation of f_B for any n . Previously, this required enumeration and coding of the relevant doubly connected graphs, which becomes increasingly tedious for increasing n and increasing number of species c . Moreover, without added effort, the enumeration will not handle permutations of molecule labels (yielding graphs that are equivalent upon integration but differ for a given configuration); Wheatley's algorithm handles these permutations automatically, reaping the most information from each configuration.

Even more, Wheatley and co-workers²³ recently presented a detailed proof showing that his algorithm yields the correct f_B for nonadditive potentials. The complexity of including non-pairwise terms in a direct sum significantly compounds the tedium already inherent in formulating the pairwise-additive sum. This realization opens the door to facile implementation of calculations of virial coefficients for highly realistic molecular models, suitable for high-accuracy property calculation, where three-body interactions (at a minimum) are needed to compete with state-of-the-art experimental data.

Finally, Wheatley's recursion can be extended in a straightforward manner to allow calculation of virial-coefficient derivatives, with added cost that is comparable to the cost of computing the coefficient value itself (in excess of the cost of the energy calculations, which do not need to be repeated for derivatives).²³ The temperature derivative in particular is important, as the first and second temperature derivatives are required for evaluation of thermal properties, such as the heat capacity and Joule–Thomson coefficient.

As a result of these advances, calculation of virial coefficients to seventh or maybe eighth order, for sufficiently simple molecular models, can be accomplished without extraordinary computational expense. The limit of what can be reached if extraordinary effort is applied is demonstrated by some recent publications, reporting coefficients up to B_{12} for the hard-sphere model,^{22,24} up to B_9 with full temperature dependence for the square-well model,²⁵ and up to B_{16} for a single low temperature for the Lennard-Jones model.²⁶

3. Flexible correction

The formulas typically presented for the virial coefficients in terms of doubly connected graphs are applicable only to molecules that are rigid, having no bend, torsion, stretch, or other intramolecular coordinate degrees of freedom. The mathematical manipulations that convert the activity series to the density series involve cancellation of graphs that are equivalent only if the molecules are rigid; see Fig. 3. Hence, to compute B_n correctly, it is necessary (for $n > 2$) to apply a well-defined “flexible correction.”²⁷ As of now, no general-purpose algorithm has yet been developed to compute the correction for arbitrary n and c , e.g., as an extension to Wheatley's recursion, so the correction has to be coded specifically for each case.



FIG. 3. Example of additional graphs that contribute to virial coefficients for flexible molecules. In deriving the density series for B_3 , the singly connected graph on the left and the disconnected graph on the right are differenced. For rigid molecules, this difference is zero. For flexible molecules, the interactions of molecule 2 with molecule 1 affects the latter's conformations, which, in turn, affects its overall interactions with molecule 3 (and vice versa). In the graph on the right, the interactions of 2 and 3 with 1 and 1' (respectively) are independent. The net result is that the difference is nonzero. A solid "field" point represents an integral over a molecule's position and conformation, while an open "root" point represents an integral over only conformation. Bonds represent Mayer f -function.⁵

The correction is small to the extent that the molecule is rigid, and an algorithm has been proposed to exploit this feature,²⁸ but in general it can be difficult to calculate the correction with good precision, and often it contributes the most to the overall uncertainty in B_n . Wheatley's algorithm, by the way, computes the b_n coefficients in the process of evaluating B_n , and the activity-series coefficients evaluated this way are correct even for flexible molecules—it is in the process of deriving B_n that the need for the correction is introduced.

It should be noted that "flexibility" in the intermolecular potential, e.g., electrostatic polarization, is not a concern in this context; it is only nuclear-coordinate flexibility that is at issue.

4. Nuclear quantum effects

Light atoms at low temperatures exhibit behavior that is not accurately described using a classical treatment, and it becomes necessary to introduce corrections or methods that capture nuclear quantum effects. Semiempirical intermolecular models incorporate these effects implicitly when the model parameters are fit to experimental data. In contrast, calculations based on models that are derived from high-accuracy first-principles electronic structure calculations must instead handle the effects explicitly. The main effect is due to the quantum nature of the nucleus, which does not allow it to be treated as a mathematical point particle but rather as a diffuse (albeit highly localized) density cloud. Another possible complication is nuclear exchange, which can further give rise to peculiar isotope effects; these considerations are important only at very low temperatures (less than 10 K) for very light atoms (hydrogen, helium).

The simplest approach to handle quantum diffuseness is to invoke a semiclassical approximation,^{29,30} which is derived as a series expansion in Planck's constant, $\hbar$, usually truncated at $\hbar^2$. The correction requires evaluation of first and second derivatives of the intermolecular potential with respect to the atom coordinates. Semiclassical methods can extend the range of temperature where accurate results can be obtained, but they eventually become inaccurate at sufficiently low temperatures. In these cases, path-integral methods should be applied.^{12,31,32} Curiously, introduction of path-integral techniques effectively makes rigid (or even monatomic)

molecules into flexible molecules, and the treatments discussed in Sec. II B 3 come into play for coefficients beyond B_2 .

C. Convergence of the VEOS

1. Performance, and assessment of convergence

The VEOS is a power series in density, and as such, it can provide accurate values for properties only for conditions where the series is converged. One can assess whether the series truncated at a particular order is converged by examining contributions to the properties made by successive terms in the series. At a given state condition, the magnitude of the last term in the sum gives a sense of the error made in neglecting the remaining terms. Of course, it is possible that the highest-order coefficient happens to be small where subsequent coefficients are not, so ideally one will have enough coefficients to observe a consistent trend of diminishing contributions. In several applications,^{14,33} we have performed a test of the reliability of the approach in the following manner. Say we have n coefficients' values available to us. We set a tolerance for accuracy, e.g., 1%, and we evaluate the order of the VEOS required to obtain the property value to this accuracy, gauging the truncated series VEOS_k against the full available series VEOS_n , $k < n$. In these tests, we also have available the correct property value, given either by experiment or molecular simulation for the same model. We perform the same evaluation, finding the VEOS_j that provides agreement of the series and the known property value within the tolerance. We typically observe that the self-assessment of convergence is highly consistent with the externally validated assessment, i.e., $j = k$. Thus, in contrast with other thermodynamic models, VEOS is unusual in its ability to provide a reliable self-assessment of its accuracy (for a given molecular model, whose accuracy itself is a different issue).

Condensation represents a singularity in the free-energy surface. Accordingly, VEOS, as a power series about $\rho = 0$, cannot converge in the liquid region. In fact, it was proved by Mayer and Mayer⁷ that the density series is not convergent for densities exceeding the spinodal density, i.e., where $(\partial p / \partial \rho)_T = 0$. In practice, we find that convergence is good up to the binodal density, sometimes even on approach to the critical point. The VEOS in principle can provide the location of the critical point, but in practice, we have not been able to do this reliably, such that increasing the series order leads to a smooth convergence to the critical conditions. Above the critical temperature, there is *a priori* no clear barrier to convergence with increasing density. However, we observe an apparent barrier when convergence is mapped in the pressure–temperature plane. Self-assessment of convergence viewed this way suggests that convergence becomes difficult upon approach to conditions that extrapolate the coexistence line in a Clausius–Clapeyron representation; this behavior is illustrated in Fig. 4. We suspect this barrier may have some relation to a Widom line,³⁴ which is connected to maxima in the thermodynamic response functions³⁵ (e.g., compressibility). The effect attenuates as one moves away from the critical point, toward higher temperatures, such that a series using about six coefficients is sufficient to reach the freezing point, at least for simple, nonpolar fluids.

2. Approximants

The VEOS is just one way to represent the equation of state as a power series. The series in activity [Eq. (2)] is an alternative,

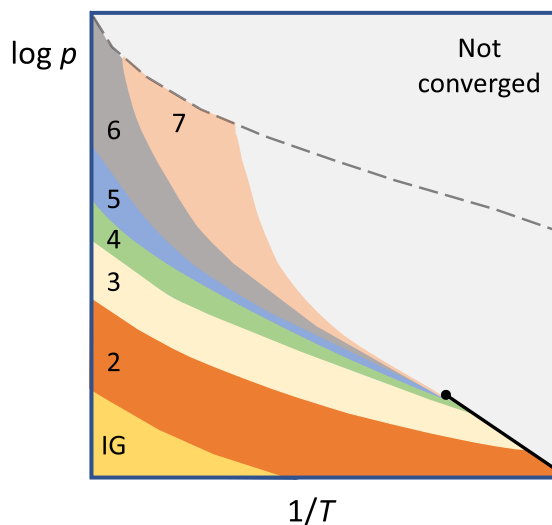


FIG. 4. Schematic illustration of observed convergence behavior of VEOS for a simple fluid. Labeled regions show where ideal gas, VEOS2, VEOS3, etc., are sufficient to describe behavior within a given tolerance, and light gray region at the top and right (extending both above and below the dashed line) shows conditions where VEOS7 is found insufficient. Black solid line on the right represents vapor–liquid coexistence, ending at the critical point, and dashed line represents fluid–solid coexistence.

and the reversion of Eq. (1), giving density as a series in pressure, is another. The convergence properties of these various approaches differ from each other. Studies have found³⁶ that the conventional VEOS [Eq. (1)] exhibits the best behavior of these alternatives.

It is worthwhile to consider more sophisticated ways to recast the VEOS as a different series. In doing this, we want to take advantage of any independent physical information that may be available about the system and build that into the treatment. Padé approximants, for example, are formed as a ratio of two series with coefficients selected to be consistent with the original, direct series. They can accelerate convergence significantly, but one must be careful in choosing how they are formulated, in particular, how many terms to include in the numerator series and how many in the denominator (the total number of independent terms cannot exceed the number of terms in the original series). In application to the soft-sphere model, we showed that convergence of the approximant depended critically on whether the approximant captured an estimated form for the pressure at high density.³⁷

Another example of a physically motivated approximant considers behavior near the critical point. Here, all fluids in a given universality class exhibit behavior having a common mathematical form, and a strategy for enhancing convergence of the VEOS in the vicinity of criticality is to formulate an approximant with this universal behavior built in. The simplest such example attempts to capture the critical singularity along a path following the critical isotherm. In particular, the following form is imposed³⁸ (subscript c indicates a value at the critical point):

$$p = p_c - (A_0 + A_1\rho + A_2\rho^2 + \cdots + A_n\rho^n) \left(1 - \frac{\rho}{\rho_c}\right)^\delta. \quad (6)$$

The coefficients A_k are determined by expanding this form about $\rho = 0$ and matching the VEOS. The exponent δ is universal, with the value 4.789. Extension of this approximant to approach the critical point from any direction (not just along T_c) must be handled carefully. We developed an approximant³⁹ based on the Schofield's parametric equation of state⁴⁰ and found significantly better performance over the basic VEOS. However, there is still much room (and need) for improvement.

3. Condensation and the liquid state

It is interesting to consider the prospect of extending the VEOS into the liquid phase. Mayer and Mayer⁷ proved that the VEOS strictly fails at the point of condensation, the binodal density. Ushcats⁴¹ developed a recursive formulation of the partition function (and from it, an equation of state, which we will designate UEOS) in terms of the same irreducible cluster integrals appearing in the VEOS. The formulation depends on the number of molecules N and the volume V independently (rather than only as their ratio $\rho = N/V$). In the limit $N \rightarrow \infty$, UEOS and VEOS are equivalent for ρ less than the binodal density, but for densities beyond the point where $\partial p / \partial \rho = 0$, UEOS follows a flat line, mimicking the true coexistence behavior. However, this line does not terminate at the coexisting liquid density but instead extends to arbitrarily large density without change in pressure. UEOS fails to identify the coexisting liquid and the subsequent increase in pressure with density in the liquid region. This failure is attributed to the cluster integrals' lack of volume dependence, as the effects of constraining walls (or periodic boundaries) are removed in the development of the cluster-integral framework.

Ushcats and co-workers have done much to advance thinking on this topic over the past decade.⁴² They raise issues regarding the relative importance of reducible vs irreducible cluster integrals in determining the needed volume dependence and how this dependence might be represented in practice. They argue that volume dependence must be manifested in “macroscopic” cluster integrals, where n is of order ρV , where V is large enough to support a liquid phase. Indeed, given that any such volume dependence must scale in a way that ensures that in the thermodynamic limit, properties depend on N/V and not N and V independently, one might consider whether there exists a universal form for the volume dependence of large cluster integrals, independent of microscopic details in the same manner as critical-point universality. Regardless, the topic seems ripe for some type of breakthrough and deserves additional attention from the statistical physics community.

III. CAN ACCURACY OF VEOS PROPERTIES EXCEED EXPERIMENT?

Attempts to generate VEOS-based experiment-quality property data from *ab initio* computational chemistry calculations are showing significant progress. The natural starting point for such efforts is helium, for which a history of virial-coefficient calculations through 2021 may be found in Ref. 33. A high-quality first-principles pair potential has been presented and refined several times over the past decade,⁴³ and a good three-body potential has been available since 2009.⁴⁴ We used these models to compute virial coefficients up to B_7 , and in Fig. 5, we compare the errors in the pressure

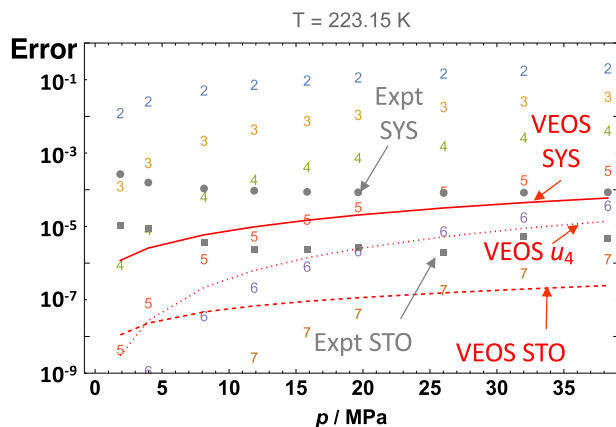


FIG. 5. Comparisons of errors in the compressibility factor $Z = p/\rho k_B T$ of helium-4 at 223.15 K. Numerals k indicate the contribution to Z of the B_k term in VEOS. Estimated stochastic and systematic errors in experimental measurements^{45,46} are given by gray squares and circles, respectively. Estimated error in VEOS7 due to stochastic and systematic errors in virial coefficients¹⁰ is given by dashed and solid red lines, respectively. Additional error estimated due to neglect of four-body contributions¹² to B_4 is given by red dotted line.

(expressed as the compressibility factor, Z) given via VEOS to those estimated for state-of-the-art experimental measurements.^{45,46} The figure shows that the contribution to Z from each additional term in the series steadily decreases such that the error from truncating the series is safely less than 10^{-5} . At the same time, stochastic error in the VEOS pressure, propagated from uncertainties in the coefficients, is negligible; estimated systematic error in VEOS, propagated from bounds on errors published with the pair and three-body potentials, and estimated error due to neglect of four-body interactions¹² are the limiting (largest) errors for VEOS, but they do not exceed 10^{-4} . Meanwhile, estimated stochastic errors published with the experiments are small, but estimated systematic errors do exceed 10^{-4} for all pressures. Hence, based on best estimates of all errors contributing to VEOS and experiment, we could conclude that the *ab initio*-based VEOS calculations exceed the accuracy and precision of the experimental measurements for helium at this temperature.

Propagation of estimated systematic error in the intermolecular potential into estimated error in a virial coefficient is typically accomplished by perturbing the entire potential-energy surface in one direction by its estimated error and computing how this changes the coefficient. Although imperfect, this provides a reasonable gauge of the effect of errors in the potential. The process can be completed efficiently without computing the coefficient twice, by focusing on calculating the difference itself³³ or by evaluating functional derivatives.¹²

Going beyond helium, greater challenge (and interest) is found with multiatomic molecules. Hellmann generated *ab initio* pair⁴⁷ and three-body⁴⁸ potential functions for rigid CO₂ molecules, and he used these potentials to compute first-principles virial coefficients for CO₂ up to B_8 . We used his coefficients to compute the Joule–Thomson (JT) inversion curve for CO₂ and compare the result with experimental data in Fig. 6. The JT coefficient requires

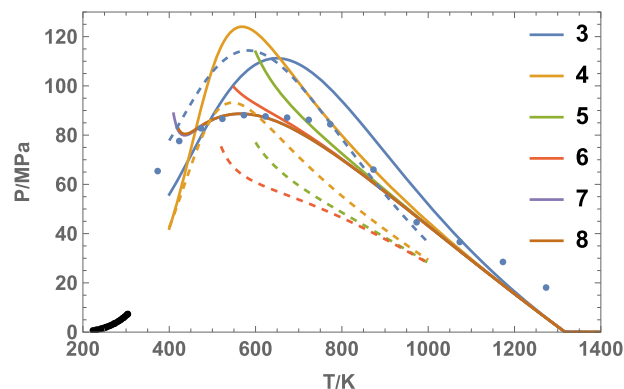


FIG. 6. Joule–Thomson inversion curve for CO₂. Lines are based on VEOS n to the order n indicated by the caption, and points are experimental data from Price.⁴⁹ Solid lines are based on *ab initio* virial coefficients reported by Hellmann,⁴⁸ and dashed lines are based on coefficients (known only up to 1000 K) determined using the TraPPE model.⁵⁰ Black points at bottom left show the vapor–liquid coexistence line, ending at the critical point. Lines for $n = 3$ –8 follow from top to bottom (to aid interpretation in the absence of color).

first and second temperature derivatives of the virial coefficients, and JT inversion is generally considered to be a challenging property to capture accurately with any equation of state; it is also a difficult property to evaluate experimentally. The figure demonstrates the convergence of the VEOS and the very good agreement with experimental data. The figure also shows the performance of a semiempirical pair potential (TraPPE), which does well in fitting CO₂ properties overall. In this application, however, its limits are apparent, and they highlight the success of the first-principles coefficients in capturing this behavior.

Finally, we point out a very recent paper from Graham and Wheatley,¹¹ who computed properties from first-principles VEOS for the mixture CO₂ + Ar. This work is notable for its use of the Gaussian Processes (GP) machine-learning method to describe the potential-energy surface (PES) based on inputs from two- and three-molecule *ab initio* calculation. Separate *ab initio* calculations and GP representations were determined for each set of pairs (AA, AB, BB, where A and B represent CO₂ and Ar molecules, respectively) and triplets (AAA, AAB, ABB, and BBB) needed to compute the PES for any set of species. Then from these, all mixture virial coefficients up to B_5 were computed from 150 to 370 K (with some pure-species *ab initio* coefficients taken from the literature). Graham and Wheatley report excellent agreement with experiment at conditions where VEOS is converged, at all compositions of the mixture. Tested properties include the pressure, speed of sound, and Joule–Thomson inversion.

We would suggest that these recent studies argue for the feasibility of computing properties from first principles via the VEOS, over regions where it is converged, with results that are competitive to experiment in both precision and accuracy.

IV. MOVING FORWARD

The steady accumulation of advances in disconnected fields—first-principles molecular modeling, cluster-integral theory

and methods, machine-learning tools for parameterizing force fields, and computer hardware and software—has now set the stage for a concerted, systematic effort to generate highly accurate first-principles equations of state for bulk fluids. We can now for the first time calculate properties of technological importance for a huge variety of supercritical fluid mixtures from molecular considerations. We can generate data for far more systems than can be measured in the laboratory, and we can present these data as a reliable equation of state suitable for manipulation in design and optimization or in other investigations (e.g., estimating critical lines). In doing so, we generate information that can be used to identify weaknesses in *ab initio* methods, thereby guiding efforts in that arena.

For semiempirical pairwise-additive (or polarizable) potentials that are formulated in part by fitting to experiment, the virial coefficients are just another property that may or may not be reproduced well. These potentials necessarily make compromises that enable them to cover a broad range of properties adequately over a range of state conditions and to do so while being computationally inexpensive and thereby suitable for molecular simulations. Often, these compromises are manifest in systematic but compensating differences from experimental virial coefficients, such that, for example, B_2 is overestimated while B_3 is underestimated.¹⁴

Deviations of this sort should not be tolerated for virial coefficients computed from potential-energy functions that purport to provide the true pair, three-body, etc., interactions. Moving forward, we would advocate that the presentation of any *ab initio* potential-energy function be accompanied by a demonstration that it is consistent with any available experimental virial-coefficient data. Such a comparison should cover a range of temperatures when possible, because different temperatures will give emphasis to different parts of the PES. Indeed, often this comparison is dutifully performed for new potentials, but in some cases, it is not. Potentials that are validated this way are more valuable, because they provide more reliable probes of behavior and can serve as a more reliable basis for property calculations, whether via VEOS or molecular simulation.

Related to this, a quantitative estimate of the magnitude of any systematic error that may be present in the potential is needed to generate error bounds on the properties computed from them. Such errors could have origins in the limitations of the *ab initio* theory or basis sets or in the means used to represent the PES from the limited *ab initio* data. It would be helpful to developers and users to have standards and procedures in place for estimating these errors and an expectation that they will be applied and their results reported with any new model or data computed from it.

Methods for calculation of virial coefficients given a function for the PES are adequate for simple systems, but some advances in application to flexible molecules will be needed to allow for application to a broad range of molecule types.

The utility of high-accuracy virial coefficients can be greatly extended if we can improve the reach of the VEOS. This may be accomplished by the development of more powerful approximants, which can, for example, be built upon treatments of critical universality other than previously attempted or perhaps with accommodation of the anomalies associated with the Widom region; approximants for mixtures have not yet been attempted. The prospect of reaching condensed states is alluring, but success in

that direction is uncertain despite significant advances over the past decade. All in all, the number of researchers who have reported attempts to extend the reach of the VEOS is small, and continued advances may arrive more quickly if a broader community of researchers can be persuaded to look at this problem.

Finally, the ability of VEOS to advance first-principles and machine-learning molecular models hinges on the availability of high-quality experimental data that can be used to assess and guide the model development. We have advocated that an effective form of this interplay is through comparison of temperature-dependent virial coefficients. However, to extract meaningful virial coefficients from experimental data (e.g., compressibility factor, speed of sound), one should extrapolate to zero density and evaluate the limiting intercept, slope, curvature, etc. This is difficult to do with good precision, and consequently experimental data for coefficients of order B_3 and (especially) higher are rare; the situation is even worse for mixtures. A variety of experimental methods exist for generating data that can yield virial coefficients,⁴ and others may yet be invented. It is our hope that an emergence of the VEOS as a new frontier for computational chemistry would spur a concomitant renewal of interest in experimental gas-phase thermodynamics as a means to inform and support these advances.

ACKNOWLEDGMENTS

The authors are grateful for support from the U.S. National Science Foundation, Grant No. CBET-2152946. D.A.K. is grateful for the support of the Walter E. Schmid Family Foundation.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Andrew J. Schultz: Conceptualization (equal); Writing – review & editing (equal). **David A. Kofke:** Conceptualization (equal); Writing – original draft (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- ¹A. Karton, “A computational chemist’s guide to accurate thermochemistry for organic molecules,” *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **6**, 292–310 (2016).
- ²A. Paul and T. Birol, “Applications of DFT + DMFT in materials science,” *Annu. Rev. Mater. Res.* **49**, 31–52 (2019).
- ³A. A. Hassanali, J. Cuny, V. Verdolino, and M. Parrinello, “Aqueous solutions: State of the art in *ab initio* molecular dynamics,” *Philos. Trans. R. Soc., A* **372**, 20120482 (2014).
- ⁴E. Mason and T. Spurling, *The Virial Equation of State* (Pergamon Press, Oxford, 1969).
- ⁵J.-P. Hansen and I. McDonald, *Theory of Simple Liquids*, 4th ed. (Academic Press, London, 2013).

- ⁶A. J. Masters, "Virial expansions," *J. Phys.: Condens. Matter* **20**, 283102 (2008).
- ⁷J. Mayer and M. Mayer, *Statistical Mechanics* (Wiley, New York, 1977).
- ⁸K. M. Benjamin, A. J. Schultz, and D. A. Kofke, "Virial coefficients of polarizable water: Applications to thermodynamic properties and molecular clustering," *J. Phys. Chem. C* **111**, 16021–16027 (2007).
- ⁹W. Cencek, G. Garberoglio, A. H. Harvey, M. O. McLinden, and K. Szalewicz, "Three-body nonadditive potential for argon with estimated uncertainties and third virial coefficient," *J. Phys. Chem. A* **117**, 7542–7552 (2013).
- ¹⁰A. J. Schultz and D. A. Kofke, "Virial coefficients of helium-4 from *ab initio* based molecular models," *J. Chem. Eng. Data* **64**, 3742–3754 (2019).
- ¹¹R. S. Graham and R. J. Wheatley, "Machine learning for non-additive intermolecular potentials: Quantum chemistry to first-principles predictions," *Chem. Commun.* **58**, 6898–6901 (2022).
- ¹²G. Garberoglio and A. H. Harvey, "Path-integral calculation of the fourth virial coefficient of helium isotopes," *J. Chem. Phys.* **154**, 104107 (2021).
- ¹³R. Hellmann, "Eighth-order virial equation of state for methane from accurate two-body and nonadditive three-body intermolecular potentials," *J. Phys. Chem. B* **126**, 3920–3930 (2022).
- ¹⁴S. Yang, A. J. Schultz, and D. A. Kofke, "Thermodynamic properties of supercritical CO₂/CH₄ mixtures from the virial equation of state," *J. Chem. Eng. Data* **61**, 4296–4312 (2016).
- ¹⁵E. G. D. Cohen, "Fifty years of kinetic theory," *Physica A* **194**, 229–257 (1993).
- ¹⁶A. Bellemans, "Statistical mechanics of surface phenomena: I. A cluster expansion for the surface tension," *Physica* **28**, 493–510 (1962).
- ¹⁷C. G. Gray, K. E. Gubbins, and C. G. Joslin, *Theory of Molecular Fluids, Vol. 2: Applications* (Oxford University Press, New York, 2011).
- ¹⁸J. K. Singh and D. A. Kofke, "Mayer sampling: Calculation of cluster integrals using free-energy perturbation methods," *Phys. Rev. Lett.* **92**, 220601 (2004).
- ¹⁹K. M. Benjamin, A. J. Schultz, and D. A. Kofke, "Gas-phase molecular clustering of TIP4P and SPC/E water models from higher-order virial coefficients," *Ind. Eng. Chem. Res.* **45**, 5566–5573 (2006).
- ²⁰K. M. Benjamin, J. K. Singh, A. J. Schultz, and D. A. Kofke, "Higher-order virial coefficients of water models," *J. Phys. Chem. B* **111**, 11463–11473 (2007).
- ²¹C. H. Bennett, "Efficiency estimation of free energy differences from Monte Carlo data," *J. Comput. Phys.* **22**, 245 (1976).
- ²²R. J. Wheatley, "Calculation of high-order virial coefficients with applications to hard and soft spheres," *Phys. Rev. Lett.* **110**, 200601 (2013).
- ²³R. J. Wheatley, A. J. Schultz, H. Do, N. Gokul, and D. A. Kofke, "Cluster integrals and virial coefficients for realistic molecular models," *Phys. Rev. E* **101**, 051301 (2020).
- ²⁴A. J. Schultz and D. A. Kofke, "Fifth to eleventh virial coefficients of hard spheres," *Phys. Rev. E* **90**, 023301 (2014).
- ²⁵H. Do, C. Feng, A. J. Schultz, D. A. Kofke, and R. J. Wheatley, "Calculation of high-order virial coefficients for the square-well potential," *Phys. Rev. E* **94**, 013301 (2016).
- ²⁶C. Feng, A. J. Schultz, V. Chaudhary, and D. A. Kofke, "Eighth to sixteenth virial coefficients of the Lennard-Jones model," *J. Chem. Phys.* **143**, 044504 (2015).
- ²⁷S. Caracciolo, B. M. Mognetti, and A. Pelissetto, "Virial coefficients and osmotic pressure in polymer solutions in good-solvent conditions," *J. Chem. Phys.* **125**, 094903 (2006).
- ²⁸K. R. S. Shaul, A. J. Schultz, and D. A. Kofke, "Mayer-sampling Monte Carlo calculations of uniquely flexible contributions to virial coefficients," *J. Chem. Phys.* **135**, 124101 (2011).
- ²⁹B. Guillot and Y. Guissani, "Quantum effects in simulated water by the Feynman-Hibbs approach," *J. Chem. Phys.* **108**, 10162 (1998).
- ³⁰K. R. S. Shaul, A. J. Schultz, D. A. Kofke, and M. R. Moldover, "Semiclassical fifth virial coefficients for improved *ab initio* helium-4 standards," *Chem. Phys. Lett.* **531**, 11–17 (2012).
- ³¹P. Diep and J. K. Johnson, "An accurate H₂-H₂ interaction potential from first principles," *J. Chem. Phys.* **112**, 4465 (2000).
- ³²K. R. S. Shaul, A. J. Schultz, and D. A. Kofke, "Path-integral Mayer-sampling calculations of the quantum Boltzmann contribution to virial coefficients of helium-4," *J. Chem. Phys.* **137**, 184101–184113 (2012).
- ³³N. Gokul, A. J. Schultz, and D. A. Kofke, "Speed of sound in helium-4 from *ab initio* acoustic virial coefficients," *J. Chem. Eng. Data* **66**, 3258–3281 (2021).
- ³⁴B. Widom, *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, 1972), Vol. 2.
- ³⁵H.-O. May and P. Mausbach, "Riemannian geometry study of vapor-liquid phase equilibria and supercritical behavior of the Lennard-Jones fluid," *Phys. Rev. E* **85**, 031201 (2012).
- ³⁶J. G. Briano and E. D. Glandt, "The rate of convergence of the virial and related series," *Fluid Phase Equilib.* **5**, 207–223 (1981).
- ³⁷N. S. Barlow, A. J. Schultz, S. J. Weinstein, and D. A. Kofke, "An asymptotically consistent approximant method with application to soft- and hard-sphere fluids," *J. Chem. Phys.* **137**, 204102 (2012).
- ³⁸N. S. Barlow, A. J. Schultz, D. A. Kofke, and S. J. Weinstein, "Critical isotherms from virial series using asymptotically consistent approximants," *AIChE J.* **60**, 3336–3349 (2014).
- ³⁹N. S. Barlow, A. J. Schultz, S. J. Weinstein, and D. A. Kofke, "Communication: Analytic continuation of the virial series through the critical point using parametric approximants," *J. Chem. Phys.* **143**, 071103 (2015).
- ⁴⁰P. Schofield, "Parametric representation of the equation of state near a critical point," *Phys. Rev. Lett.* **22**, 606–608 (1969).
- ⁴¹M. V. Ushcats, "Equation of state beyond the radius of convergence of the virial expansion," *Phys. Rev. Lett.* **109**, 040601 (2012).
- ⁴²M. V. Ushcats, S. Y. Ushcats, L. A. Bulavin, and V. M. Sysoev, "Equation of state for all regimes of a fluid: From gas to liquid," *Phys. Rev. E* **98**, 032135 (2018).
- ⁴³P. Czachorowski, M. Przybytek, M. Lesiuk, M. Puchalski, and B. Jeziorski, "Second virial coefficients for ⁴He and ³He from an accurate relativistic interaction potential," *Phys. Rev. A* **102**, 042810 (2020).
- ⁴⁴W. Cencek, K. Patkowski, and K. Szalewicz, "Full-configuration-interaction calculation of three-body nonadditive contribution to helium interaction potential," *J. Chem. Phys.* **131**, 064105 (2009).
- ⁴⁵M. O. McLinden and C. Lösch-Will, "Apparatus for wide-ranging, high-accuracy fluid (*p, ρ, T*) measurements based on a compact two-sinker densimeter," *J. Chem. Thermodyn.* **39**, 507–530 (2007).
- ⁴⁶M. R. Moldover and M. O. McLinden, "Using *ab initio* 'data' to accurately determine the fourth density virial coefficient of helium," *J. Chem. Thermodyn.* **42**, 1193–1203 (2010).
- ⁴⁷R. Hellmann, "*Ab initio* potential energy surface for the carbon dioxide molecule pair and thermophysical properties of dilute carbon dioxide gas," *Chem. Phys. Lett.* **613**, 133–138 (2014).
- ⁴⁸R. Hellmann, "Nonadditive three-body potential and third to eighth virial coefficients of carbon dioxide," *J. Chem. Phys.* **146**, 054302 (2017).
- ⁴⁹D. Price, "Thermodynamic functions of carbon dioxide. Joule-Thomson coefficient, isochoric heat capacity, and isentropic behavior at 100° to 1000°C and 50 to 1400 bars," *Ind. Eng. Chem. Chem. Eng. Data Ser.* **1**, 83–86 (1956).
- ⁵⁰N. Gokul, A. J. Schultz, and D. A. Kofke, "Properties of supercritical N₂, O₂, CO₂, and NH₃ mixtures from the virial equation of state," *AIChE J.* **67**, e17072 (2021).