Cluster integrals and virial coefficients for realistic molecular models

Richard J. Wheatley,[1] Andrew J. Schultz,[2] Hainam
Do,[3] Navneeth Gokul,[2] and David A. Kofke[2]

[1]*School of Chemistry, University of Nottingham,*
University Park NG7 2RD, United Kingdom
[2]*Department of Chemical and Biological Engineering, University at Buffalo,*
The State University of New York, Buffalo, NY 14260-4200 USA
[3]*Department of Chemical and Environmental Engineering,*
University of Nottingham Ningbo China,
199 Taikang East Road, Ningbo 315100, China.

Abstract

We present a concise, general and efficient procedure for calculating the cluster integrals that relate thermodynamic virial coefficients to molecular interactions. The approach encompasses non-pairwise intermolecular potentials generated from quantum chemistry or other sources; a simple extension permits efficient evaluation of temperature and other derivatives of the virial coefficients. We demonstrate with a polarizable model of water. We argue that cluster-integral methods are a potent yet underutilized instrument for the development and application of first-principles molecular models and methods.

More than a century of effort has been put toward bridging nanoscopic and macroscopic behavior, with the aim of predicting and understanding thermophysical properties quantitatively from molecular considerations. Such a capability can have immense value: reliable thermophysical models are essential to engineering design, optimization, and control applications [1]; conversely well-structured thermophysical models provide a conduit for macroscopic experiments to inform molecular modeling, guiding the development of ab initio methods and semi-empirical force fields. Only two general approaches are available to bridge these scales rigorously. The first, molecular simulation, is very widely used. It is a type of surrogate for experiment, with many of the good and bad features this entails. It is computationally expensive, and hence cannot be employed on-the-fly as part of a larger calculation. Also, despite the detailed molecular information it can yield, it is still in some respects a blunt instrument—when simulation disagrees with experiment, it is difficult to know what features of the molecular model led to this failure.

The second rigorous nano/macro bridge is much less commonly applied. *Cluster integrals* reformulate the partition function into computationally tractable pieces. These integrals enter as parameters in models for the thermodynamic behavior, such as the equation of state, in which the integrals appear as the familiar virial coefficients [2, 3]. Cluster-integral approaches are (as yet) inapplicable to condensed or ordered phases, hence they apply primarily to vapor and supercritical states, but otherwise they are distinct from all other thermophysical models in being explicit, rigorous, and applicable to arbitrarily complex molecular species. Moreover, their hierarchical structure, in which 2-, 3-, and generally n-body interactions can be studied independently, coupled with the information obtainable from their temperature dependence, provides a rich source of focused information for assessing the strengths and

2

weaknesses of molecular models.

Cluster-integral methods received considerable attention in the decades following their introduction by Mayer and Mayer in the 1930's and 40's [4, 5], but in subsequent years interest waned due to the difficulty in computing coefficients for even the simplest models. However, the past decade has brought renewed attention, with considerable advances seen in theory, methods and applications [6–12]. While the approach remains inapplicable to condensed phases, there is an enormous range of non-trivial and technologically important behavior encompassed by the supercritical region to which it can apply [13–21]. Considering that first-principles calculations of ideal-gas properties now routinely outperform experiment in accuracy and cost [22–26], it should be self-evident that cluster integrals—which proceed naturally and methodically from the ideal-gas starting point—form the next natural frontier to advance ab initio computational chemistry. Indeed, in a recent study of helium [27], virial coefficients computed from first-principles force fields yielded supercritical property data with accuracy and precision that rivals—and probably exceeds—the best experimental measurements.

Recently, one of the co-authors of this Communication presented a recursive algorithm for calculation of the integrand of the cluster integrals for the virial coefficient [11]. The algorithm drastically reduces the time required to compute the integrand for a given configuration, and consequently it has made possible calculation of virial coefficients to higher order than previously thought possible. Following convention, the treatment was developed and derived for pairwise-additive models. Here, we present a fully general algorithm for calculating arbitrary virial coefficients for non-additive potential energy surfaces. This algorithm provides a better basis for framing the development of cluster integrals—it is not only more general, it is simpler too. The approach is developed in the form of a recursion, and it makes no mention of pair bonds or biconnected graphs.

Non-reliance on pairwise additivity is a crucial feature of this scheme. Although the assumption of pairwise additivity is broadly practiced in molecular simulation, it is wholly incompatible with a framework that purports to provide first-principles properties with accuracy that rivals experiment (see, e.g. [27, 28]). Virial coefficients have been computed for non-pairwise potentials previously [29–34], but no methods for multibody potentials exploit the recursion developed in [11]. Instead they rely on explicit enumeration of the clusters that sum to yield the coefficient; while progress has been made in economizing the

3

process [35], the approach quickly becomes unwieldy to apply for increasing order of the virial coefficient. Moreover, important related quantities, and in particular temperature derivatives of the integrals, are not handled at all.

Each coefficient appearing in density series for thermodynamic properties (see Supporting Information [36]) is expressible as an integral over configurations of N molecules:

$$Q_N = \frac{1}{N!\Lambda^{3N}} \int f_Q(N) d\mathbf{r}^N \tag{1a}$$

$$C_N = \frac{1}{N!} \int f_C(N) d\mathbf{r}^{N-1} \tag{1b}$$

$$B_N = \frac{1-N}{N!} \int f_B(N) d\mathbf{r}^{N-1} \tag{1c}$$

where Λ is the de Broglie wavelength. We introduce here a set of functions defined on the positions (N) of N molecules. $f_Q(N)$ is the Boltzmann factor of the potential energy for N molecules: $f_Q(N) = \exp(-\beta E(N))$, where $E(N) \equiv E(\mathbf{r}^N)$ is the potential energy of the set of N molecules in the configuration $\mathbf{r}^N$. This energy function is completely arbitrary in form—it may be additive, or explicitly or implicitly non-additive. The functions $f_C(N)$ and $f_B(N)$ are known as the Ursell and Husimi functions, respectively [37, 38]; their integrals hold the position of one molecule fixed at the origin.

The integrals labeled B_N are the conventional virial coefficients, and for a given molecular model they are computed by evaluating the integral in (1c). Monte Carlo methods can be used to sample configurations and collect averages [6–8]. The difficulty lies in the evaluation of f_B for each configuration, which depends on f_Q and thereby the molecular model as manifested in $E(N)$. This can in principle be accomplished in two steps—first evaluating the f_C in terms of the f_Q, and then the f_B from these f_C. In practice this is not easy to do; both f_C and f_B are defined for a given $\mathbf{r}^N$ as a complicated sum of products involving functions on subsets of the coordinates, with a number of terms that grows faster than exponentially with N. Here we present a means to accomplish these evaluations in a way that is efficient and easily implemented.

The inversion giving the f_C in terms of the f_Q is [39]

$$f_C(N) = \sum_{\lambda \vdash N} (-1)^{p-1}(p-1)! \prod_i^p f_Q(\lambda_i) \tag{2}$$

where the sum is over all partitions λ of the set N, with p the number of blocks λ_i in each
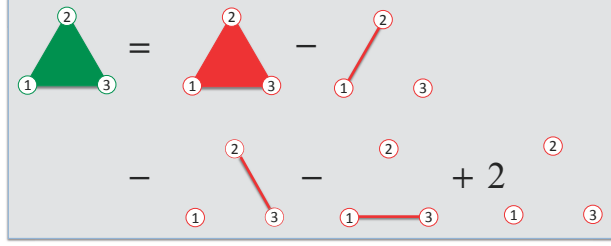
FIG. 1: Graphical representation of (3c). Green lines and faces (shaded polygons, left-hand side of equation) represent f_C, and red (seen on right-hand side of equation) represents f_Q.

partition. The first three formulas are:

$$f_C(\mathbf{r}_1) = f_Q(\mathbf{r}_1) \tag{3a}$$

$$f_C(\mathbf{r}_1, \mathbf{r}_2) = f_Q(\mathbf{r}_1, \mathbf{r}_2) - f_Q(\mathbf{r}_1)f_Q(\mathbf{r}_2) \tag{3b}$$

$$f_C(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = f_Q(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) - f_Q(\mathbf{r}_1, \mathbf{r}_2)f_Q(\mathbf{r}_3) \tag{3c}$$
$$- f_Q(\mathbf{r}_2, \mathbf{r}_3)f_Q(\mathbf{r}_1) - f_Q(\mathbf{r}_1, \mathbf{r}_3)f_Q(\mathbf{r}_2)$$
$$+ 2f_Q(\mathbf{r}_1)f_Q(\mathbf{r}_2)f_Q(\mathbf{r}_3)$$

It is useful to discuss these functions in reference to a representation as graphs. Equation (3c) is depicted in Fig. 1 as an example. The faces (and lines) join vertices that represent the labeled coordinates, with color used to indicate the function.

A computationally expedient means to obtain the f_C from the f_Q is via the recursion:

$$f_C(N) = f_Q(N) - \sum_S f_C(S)f_Q(S^*) \tag{4}$$

where the sum over S includes all possible labelled subsets of sizes 1 to $N-1$, from the set of N molecules, such that every subset S contains a chosen particle (labeled 1), and S^* is the complement of S. Since the f_Q are known, this expresses $f_C(N)$ entirely in terms of f_C's for sets of size smaller than N, so the recursion can be started from trivially small subsets, for example $f_C = f_Q = 1$ (Eq. (3a)) for one particle, and the f_C's then calculated for subsets of increasingly large size up to the required size N. The proof that (4) is consistent with (2) is given in the Supporting Information [36].

Stell [40] and Sherman [41] independently derived the "cluster-star inversion" and showed that $f_B(N)$ could be written as a weighted sum over connected tree graphs formed from f_C faces. While in principle this result provides the connection needed to complete the sequence
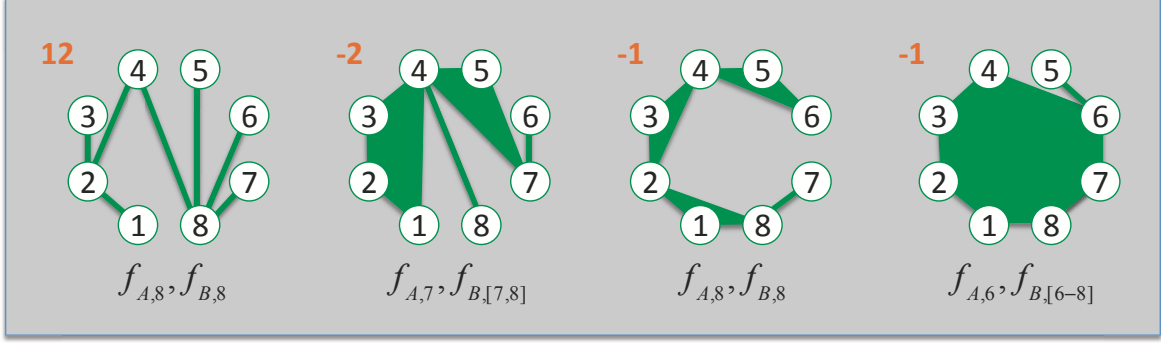
FIG. 2: Graphs representing some of the terms that sum to give $f_B(8)$. Green regions and lines represent f_C faces. Orange number above and left of each graph is its weight as given by (6). Beneath each graph is shown the functions to which it contributes (with v ranges shown in square brackets, when needed).

$E(N) \rightarrow f_Q(N) \rightarrow f_C(N) \rightarrow f_B(N) \rightarrow B_N \rightarrow$ properties, their result does not suggest an efficient computational route to realize this connection. We develop this result here, again using recursion.

The cluster-star inversion may be written as:

$$f_B(N) = \sum_{G \in \text{trees}} W(G) \prod_{\lambda_i \in G} f_C(\lambda_i) \tag{5}$$

Examples are shown in Fig. 2. Pairs of f_C faces may be joined at a single vertex they have in common, which then represents an articulation point, and the tree structure requires that no chain of joined faces can form a ring. More than two f_C faces can be joined by the same vertex, and we define the multiplicity $p(k)$ as the number of faces sharing (joined at) vertex k; $p(k) = 1$ if k is not an articulation point. The weight associated with each graph G formed this way is:

$$W(G) = \prod_{k=1}^{N} (-1)^{p(k)-1}(p(k) - 1)! \tag{6}$$

6

Given this result, we propose the following recursion to compute the f_B from the f_C:

$$f_B(N) = f_C(N) + \sum_{v=1}^{N} f_{A,v}(N) \tag{7a}$$

$$f_{A,v}(N) = -\sum_{S} f_{B,v}(S) f_{B,v-1}(S^* \cup v) \tag{7b}$$

$$f_{B,v}(N) = f_{B,v-1}(N) + f_{A,v}(N) \tag{7c}$$

$$f_{B,0}(N) = f_C(N) \tag{7d}$$

Here we have introduced the auxiliary functions $f_{A,v}$ and $f_{B,v}$ to support the recursion. These are both $W(G)$-weighted sums of f_C trees, just as f_B is, but they have restrictions on the graphs included in their sum: they both disallow graphs in which any vertex $k > v$ is an articulation point, and $f_{A,v}$ furthermore includes only graphs in which v itself *is* an articulation point ($f_{B,v}$ has no similar restriction on vertex v). The sum in (7b) is over all subsets S having between 2 and $N-1$ nodes that include both node v and the lowest-numbered node (node 1, or node 2 if $v = 1$), and $v-1$ is the node with highest number below v, or 0 if v is the lowest-numbered node. The recursion for each N involves calculating $f_{A,v}$ for each node v from information already calculated for smaller subsets, then calculating $f_{B,v}$ for increasing v from f_C, which is already known from (4), and the f_A's. The computation time taken to calculate the cluster integrand for N particles scales exponentially with N using this new algorithm. This is close to the theoretical maximum efficiency, since the number of input (sub-)cluster energies itself scales exponentially with N.

Proof that the recursion (7) yields f_B, timings for $N = 4$ to 12, comparison with the direct sum over graphs, and links to an implementation in software, are presented in the Supporting Information [36].

Temperature derivatives of the virial coefficients are needed to describe thermal properties, such as the heat capacity, the Joule-Thomson coefficient, and the speed of sound [12]. They are also useful in providing an accurate representation of the temperature dependence of the coefficients, and to aid in interpolation and extrapolation [42]. The conventional pairwise-additive treatment requires a tedious accounting for the unique irreducible graphs that can be generated from Mayer bonds and derivative bonds, and the complexity increases rapidly with derivative order (second derivatives are needed, for example, for the heat capacity). In contrast, these derivatives are given almost trivially from the recursion formulation, and moreover they remain applicable to the general case involving multibody potentials.

From (1c),

$$B_N^{(n)} = \frac{1-N}{N!} \int f_B^{(n)}(N) d\mathbf{r}^{N-1} \tag{8}$$

where the superscript in parentheses represents the nth derivative with respect to the inverse temperature β. For a given configuration, $f_B^{(n)}(N)$ is determined via straightforward propagation of the derivative via (4) and (7). The sum over products introduces another layer of recursion. For example, from (4):

$$f_C^{(n)}(N) = f_Q^{(n)}(N) - \sum_{k=0}^{n} \sum_S \binom{n}{k} f_C^{(n-k)}(S) f_Q^{(k)}(S^*) \tag{9}$$

(7b) is treated similarly. This culminates in the f_Q derivative, which we write first for the general case of a temperature-dependent potential:

$$f_Q^{(n)}(N) = \sum_{k=1}^{n} \binom{n-1}{k-1} (-\beta E(N))^{(k)} f_Q^{(n-k)}(N) \tag{10a}$$

and in the more typical case where E is independent of β:

$$f_Q^{(n)}(N) = (-E(N))^n e^{-\beta E(N)} \tag{10b}$$

Many other derivatives of the virial coefficients are of practical interest, and can be evaluated using an approach analogous to that outlined here for temperature derivatives. Derivatives with respect to an electric field, for example, are needed to develop virial series for the dielectric constant [43, 44]. Alternatively, derivatives with respect to intermolecular-potential parameters can be computed, providing an effective means to adjust molecular models to better match experimental data [42].

As an example, we consider an application to a model for water. The most widely used molecular models for water have been formulated to describe behavior in the liquid phase at ambient conditions, largely due to its biological importance. Almost all such models are pairwise additive with fixed electrostatic features. Unsurprisingly, such models perform poorly when describing the vapor or supercritical phases, and are unable to reproduce the temperature dependence of the lower-order virial coefficients [8, 29, 45]. Explicit two- and three-body potentials for rigid water molecules have been developed through fits to ab initio data [46] for water dimers and trimers, and these do very well in describing experimental second and third virial coefficients [47].

An alternative approach to non-additive interactions is the use of polarizable electrostatics. The Gaussian-charge polarizable model (GCPM) of Pericaud et al. [48] was developed

using data from both ab initio calculation and from experiment (notably, not experimental virial coefficients), so its parameterization implicitly includes nuclear quantum effects and 3-body dispersion, induction, and exchange (for water, it has been shown that these effects contribute at least 50% of the overall 3-body interaction [49], so if they were not already implicit to the model they would need to be handled explicitly). Virial coefficients up to B_5 (but not temperature derivatives) have been computed for GCPM water previously [29–31], and while these data are not given with the same precision as obtained here, they can serve to partially validate our methods and calculations.

We computed virial coefficients $B_N(T)$, $N = 2, \ldots, 6$ for this model for several temperatures T from 270 K to 1500 K. At each temperature, we computed the coefficient itself and its first three temperature derivatives. These data were fit to a polynomial form that provides a convenient analytic expression for $B_N(T)$ for all temperatures in the range (and beyond, given that extrapolation using a form fit to many derivatives can be very reliable [42]). Two independent sets of B_N values were computed, one based on Mayer sampling [6–8] and the other using nested sampling [50], both computing f_B for sampled configurations as described above. Details are given in the Supporting Information [36].

We used the coefficients to compute the pressure and the isochoric heat capacity c_V as a function of temperature and density. The pressure can be given via the usual virial equation of state, and in addition we examine a "parametric approximant" [51]. This is an equation of state that is formulated to capture the non-analytic behavior known to prevail in the vicinity of a critical point, while also matching the zero-density limiting behavior as quantified by the virial series. The model takes the critical temperature T_c and density ρ_c as input parameters, and provides a much better and more regular convergence of the virial series; we use "VN" and "AN" respectively to indicate the virial series and the approximant based on virial coefficients up to B_N. As input parameters, we use $T_c = 642.21$ K and $\rho_c = 0.3344$ g/cm^3 previously determined [48] for GCPM; the critical pressure is obtained as an output of the approximant, and the value from A6 (24.3 ± 0.4 MPa, with error indicating difference from A5) is in good agreement with molecular simulation for GCPM (24.56 MPa [48]). The estimated GCPM critical temperature, density, and pressure reported in [48] differ from experimental values [52] by −0.7%, +3.8%, and +11.3%, respectively.

Figure 3 shows the performance of the VN and AN series for three supercritical isotherms (viz., 650 K, 700 K, and 800 K, which are, respectively, approximately 1%, 10%, and 25%

above T_c); data are plotted up to about twice the critical density. Models are compared to data taken via isobaric Monte Carlo simulations of GCPM water, and also to data for real water [52, 53]. Uncertainties in the isotherms were determined by propagating the uncertainties in the coefficients, either analytically (for VN) or via bootstrapping (for AN).

One of the most valuable features of the series-based approaches is their ability to self-assess their accuracy through comparison of isotherms for successive orders of the series [12]. This feature is demonstrated for the 800 K isotherm, which presents the VEOS estimates for different orders. We see that each VN curve peels away from those at higher order at about the point where it deviates from molecular simulation data. Uncertainties in V6 prevent V5 from being assessed this way, highlighting the need to know confidence limits on the VN isotherms when evaluating series convergence in this manner.

The parametric approximant has a single parameter (labeled $\tilde{a}$) that is independent of T and series order N, and we adjust this parameter to a value that gives the best overall convergence of the approximant series. Notably, $\tilde{a}$ is not adjusted via comparison to simulation or experimental data; rather it is given by considerations completely internal to the parametric-approximant model. We observe that the value of $\tilde{a}$ that yields the best convergence is also the value that results in the best agreement with molecular simulation data, for the given values of T_c and ρ_c. As seen in Fig. 3, both the convergence of the series and its agreement with simulation data are remarkable. Still, the systematic disagreement at higher density when using simulation-based critical properties—even when the approximant apparently is converged—is interesting, and this behavior is worthy of further study.

Isochoric heat capacity is examined in the Supporting Information [36].

The methods presented in this Communication open a new frontier in computational quantum chemistry, moving beyond single-molecule properties to those that are based in multi-molecular interactions. Cluster-integral treatments not only provide a full spectrum of thermodynamic properties for the vapor and supercritical-fluid phases from intermolecular energies, but they can also support systematic development of intermolecular-energy methods and models via comparison to experimental temperature-dependent virial coefficients. Condensed-phase properties are still out of reach for the conventional virial series, but attention is being given to this issue [10, 54, 55]. In the meantime, a reliable way to predict properties of just supercritical fluids and their mixtures — ultimately to compute them with accuracy exceeding experiment — would be a major advance for science and
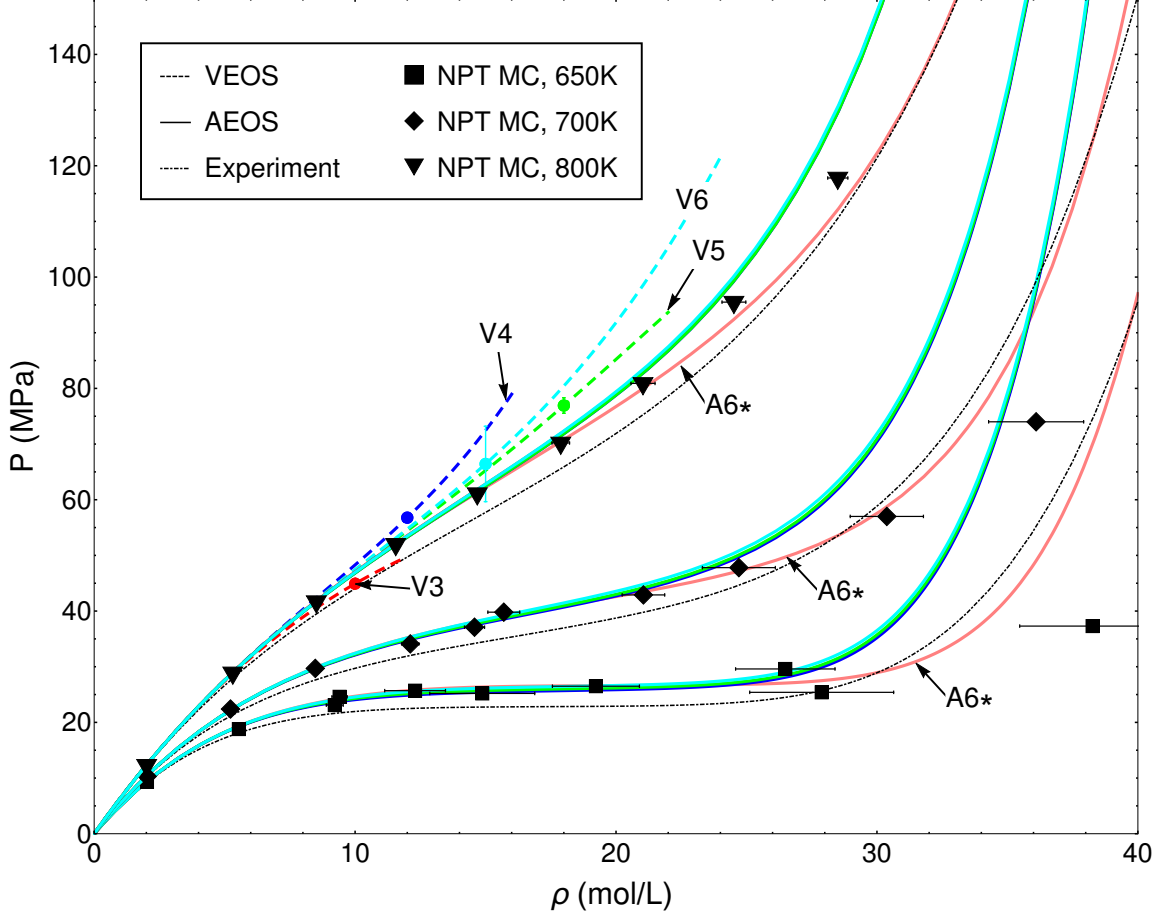
FIG. 3: Pressure-density isotherms of water for three supercritical temperatures. AEOS lines represent the parametric approximant [51] using values of T_c and ρ_c from [48]; curves are shown for A3 to A6, but are not distinguishable on the plot. Pink lines labeled A6* show the A6 parametric approximant using $T_c = 647$ K and $\rho_c = 0.37$ g/cm^3. VEOS lines are shown only for the $T = 800$ K isotherm, and are truncated where they start to diverge; this is representative of behavior at other temperatures. Experimentally derived data are as represented by an accurate correlation [52, 53] (black dot-dashed lines), and filled symbols are results of NPT Monte Carlo simulations for GCPM water. Error bars indicate uncertainties at 68% confidence level, and for AEOS are smaller than the line thickness.

engineering. Supercritical water (as studied here) is itself important to diverse areas such as geophysics [13] and gasification of biomass [14], while applications of supercritical fluids more generally encompass a broad range of science and technology, including separations [15], reaction engineering [16], nanotechnology [17], crystallization [18], power generation [19], carbon capture [20], and more [21]. These fields could more rapidly advance with

access to predictive (ideally first-principles) methods for computing properties of arbitrary compounds and their mixtures.

[1] Wilhelmsen, Ø, Aasen, A, Skaugen, G, Aursand, P, Austegard, A, Aursand, E, Gjennestad, M. A, Lund, H, Linga, G, & Hammer, M. (2017) Thermodynamic modeling with equations of state: Present challenges with established methods. *Ind. Eng. Chem. Res.* **56**, 3503–3515.

[2] McQuarrie, D. (1976) *Statistical Mechanics.* (Harper & Row, New York).

[3] Hansen, J.-P & McDonald, I. (2006) *Theory of Simple Liquids.* (Academic Press, London), 3rd edition.

[4] Mayer, J. E. (1937) The Statistical Mechanics of Condensing Systems. I. *J. Chem. Phys.* **5**, 67–73.

[5] Mayer, J & Mayer, M. (1977) *Statistical Mechanics.* (Wiley, New York).

[6] Singh, J. K & Kofke, D. A. (2004) Mayer sampling: Calculation of cluster integrals using free-energy perturbation methods. *Phys. Rev. Lett.* **92**, 220601.

[7] Benjamin, K. M, Schultz, A. J, & Kofke, D. A. (2006) Gas-phase molecular clustering of tip4p and spc/e water models from higher-order virial coefficients. *Ind. Eng. Chem. Res.* **45**, 5566–5573.

[8] Benjamin, K. M, Singh, J. K, Schultz, A. J, & Kofke, D. A. (2007) Higher-order virial coefficients of water models. *J. Phys. Chem. B* **111**, 11463–11473.

[9] Masters, A. J. (2008) Virial expansions. *J. Phys.: Condens. Matter* **20**, 283102.

[10] Ushcats, M. (2012) Equation of state beyond the radius of convergence of the virial expansion. *Phys. Rev. Lett.* **109**, 040601.

[11] Wheatley, R. J. (2013) Calculation of high-order virial coefficients with applications to hard and soft spheres. *Phys. Rev. Lett.* **110**, 200601.

[12] Yang, S, Schultz, A. J, & Kofke, D. A. (2016) Thermodynamic Properties of Supercritical CO_2/CH_4 Mixtures from the Virial Equation of State. *J. Chem. Eng. Data* **61**, 4296–4312.

[13] Scott, S, Driesner, T, & Weis, P. (2015) Geologic controls on supercritical geothermal resources above magmatic intrusions. *Nat. Commun.* **6**.

[14] Peng, G, Vogel, F, Refardt, D, & Ludwig, C. (2017) Catalytic supercritical water gasification: Continuous methanization of chlorella vulgaris. *Ind. Eng. Chem. Res.* **56**, 6256–6265.

[15] Marcus, Y. (2018) Extraction by Subcritical and Supercritical Water, Methanol, Ethanol and Their Mixtures. *Separations* **5**, 4.

[16] Nasir, N, Daud, W, Kamarudin, S, & Yaakob, Z. (2013) Process system engineering in biodiesel production: A review. *Renew. Sust. Energ. Rev.* **22**, 631 – 639.

[17] Philippot, G, Elissalde, C, Maglione, M, & Aymonier, C. (2014) Supercritical fluid technology: A reliable process for high quality BaTiO3 based nanomaterials. *Adv. Powder Technol.* **25**, 1415–1429.

[18] Badens, E, Masmoudi, Y, Mouahid, A, & Crampon, C. (2018) Current situation and perspectives in drug formulation by using supercritical fluid technology. *J. Supercrit. Fluids* **134**, 274–283. 12th International Symposium on Supercritical Fluids (ISSF), Antibes, FRANCE, 2018.

[19] Pizzarelli, M. (2018) The status of the research on the heat transfer deterioration in supercritical fluids: A review. *Int. Commun. Heat Mass* **95**, 132–138.

[20] Burant, A, Lowry, G. V, & Karamalidis, A. K. (2013) Partitioning Behavior of Organic Contaminants in Carbon Storage Environments: A Critical Review. *Environ. Sci. Technol.* **47**, 37–54.

[21] Knez, Z, Markocic, E, Leitgeb, M, Primozic, M, Hrncic, M. K, & Skerget, M. (2014) Industrial applications of supercritical fluids: A review. *Energy* **77**, 235–243.

[22] Fabian, W. M. F. (2008) Accurate thermochemistry from quantum chemical calculations? *Monatsch. Chem.* **139**, 309–318.

[23] Ghahremanpour, M. M, van Maaren, P. J, Ditz, J. C, Lindh, R, & van der Spoel, D. (2016) Large-scale calculations of gas phase thermochemistry: Enthalpy of formation, standard entropy, and heat capacity. *J. Chem. Phys.* **145**, 114305.

[24] Karton, A, Daon, S, & Martin, J. M. (2011) W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem. Phys. Lett.* **510**, 165 – 178.

[25] Ganyecz, Á, Kállay, M, & Csontos, J. (2017) Moderate-cost ab initio thermochemistry with chemical accuracy. *J. Chem. Theory Comput.* **13**, 4193–4204. PMID: 28777574.

[26] Ganyecz, Á, Kállay, M, & Csontos, J. (2018) High accuracy quantum chemical and thermochemical network data for the heats of formation of fluorinated and chlorinated methanes and ethanes. *J. Phys. Chem. A* **122**, 5993–6006. PMID: 29939026.

[27] Schultz, A. J & Kofke, D. A. (2019) Virial coefficients of helium-4 from *ab initio*-based molecular models. *J. Chem. Eng. Data* **64**, 3742.

[28] Yu, K & Schmidt, J. R. (2012) Many-body effects are essential in a physically motivated CO_2 force field. *J. Chem. Phys.* **136**, 034503.

[29] Benjamin, K. M, Schultz, A. J, & Kofke, D. A. (2007) Virial coefficients of polarizable water: Applications to thermodynamic properties and molecular clustering. *J. Phys. Chem. C* **111**, 16021–16027.

[30] Benjamin, K. M, Schultz, A. J, & Kofke, D. A. (2009) Fourth and fifth virial coefficients of polarizable water. *J. Phys. Chem. B* **113**, 7810–7815.

[31] Kim, H. M, Schultz, A. J, & Kofke, D. A. (2012) Virial equation of state of water based on wertheim's association theory. *J. Phys. Chem. B* **116**, 14078–14088.

[32] Jäger, B, Hellmann, R, Bich, E, & Vogel, E. (2011) Ab initio virial equation of state for argon using a new nonadditive three-body potential. *J. Chem. Phys.* **135**, 084308.

[33] Rouha, M, Nezbeda, I, Hrubý, J, & Moučka, F. (2017) Higher virial coefficients of water. *Journal of Molecular Liquids.*

[34] Hellmann, R. (2017) Nonadditive three-body potential and third to eighth virial coefficients of carbon dioxide. *J. Chem. Phys.* **146**, 054302.

[35] Hellmann, R & Bich, E. (2011) A systematic formulation of the virial expansion for nonadditive interaction potentials. *J. Chem. Phys.* **135**, 084117.

[36] Wheatley, R. J, Schultz, A. J, Do, H, Gokul, N, & Kofke, D. A. (2020) See supplemental material at [url will be inserted by publisher] for details of derivations of formulas and additional computational results.

[37] Husimi, K. (1950) Note on Mayers' Theory of Cluster Integrals. *J. Chem. Phys.* **18**, 682–684.

[38] Stell, G. (1976) in *Phase Transitions and Critical Phenomena*, eds. Domb, C & Green, M. S. (Academic Press, London) Vol. 5B, pp. 205–258.

[39] Uhlenbeck, G & Ford, G. (1962) in *Studies in Statistical Mechanics*, eds. de Boer, J & Uhlenbeck, G. (North-Holland) Vol. 1, pp. 119–211.

[40] Stell, G. (1965) Cluster-?Star Inversion by Means of Generating Functionals. *J. Math. Phys.* **6**, 1193–1200.

[41] Sherman, S. (1965) Cluster-?Star Inversion and the Möbius Formula. *J. Math. Phys.* **6**, 1189–1192.

[42] Hatch, H. W, Jiao, S, Mahynski, N. A, Blanco, M. A, & Shen, V. K. (2017) Communication: Predicting virial coefficients and alchemical transformations by extrapolating Mayer-sampling

Monte Carlo simulations. *J. Chem. Phys.* **147**, 231102.

[43] Neumann, M. (1983) Dipole moment fluctuation formulas in computer simulations of polar systems. *Mol. Phys.* **50**, 841–858.

[44] Yang, S, Schultz, A, & Kofke, D. A. (2017) Evaluation of second and third dielectric virial coefficients for non-polarizable molecular models. *Mol. Phys.* p. submitted.

[45] Kusalik, P. G, Linden, & Svishchev, I. M. (1995) Calculation of the third virial coefficient for water. *J. Chem. Phys.* **103**, 10169–10175.

[46] Góra, U, Cencek, W, Podeszwa, R, van der Avoird, A, & Szalewicz, K. (2014) Predictions for water clusters from a first-principles two- and three-body force field. *J. Chem. Phys.* **140**, 194101.

[47] Schultz, A. J, Kofke, D. A, & Harvey, A. H. (2015) Molecular-based virial coefficients of CO_2-H_2O mixtures. *AIChE J.* **61**, 3029–3037.

[48] Paricaud, P, Předota, M, Chialvo, A. A, & Cummings, P. T. (2005) From dimer to condensed phases at extreme conditions: Accurate predictions of the properties of water by a Gaussian charge polarizable model. *J. Chem. Phys.* **122**, 244511.

[49] Akin-Ojo, O & Szalewicz, K. (2013) How well can polarization models of pairwise nonadditive forces describe liquid water? *J. Chem. Phys.* **138**, 024316.

[50] Nielsen, S. O. (2013) Nested sampling in the canonical ensemble: Direct calculation of the partition function from NVT trajectories. *J. Chem. Phys.* **139**, 124104.

[51] Barlow, N. S, Schultz, A. J, Weinstein, S. J, & Kofke, D. A. (2015) Communication: Analytic continuation of the virial series through the critical point using parametric approximants. *J. Chem. Phys.* **143**, 071103.

[52] Wagner, W & Pruß, A. (2002) The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *J. Phys. Chem. Ref. Data* **31**, 387.

[53] Lemmon, E, Linden, M, & Friend, D. (2013) in *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, eds. Linstrom, P & Mallard, W. (National Institute of Standards and Technology).

[54] Schultz, A. J & Kofke, D. A. (2016) Vapor-phase metastability and condensation via the virial equation of state with extrapolated coefficients . *Fluid Phase Equilib.* **409**, 12–18.

[55] Ushcats, M. V, Bulavin, L. A, Sysoev, V. M, & Ushcats, S. Y. (2017) Divergence of activity

expansions: Is it actually a problem? *Phys. Rev. E* **96**, 062115.