Transferability of data-driven, many-body models for CO_2 simulations in the vapor and liquid phases

Shuwen Yue,[1] Marc Riera,[2, a)] Raja Ghosh,[2, a)] Athanassios Z. Panagiotopoulos,[1, b)] and Francesco Paesani[2, 3, 4, c)]

[1)] *Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA*

[2)] *Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093, USA*

[3)] *Materials Science and Engineering, University of California San Diego, La Jolla, California 92093, USA*

[4)] *San Diego Supercomputer Center, University of California San Diego, La Jolla, California 92093, USA*

(Dated: 28 November 2021)

Extending on previous work by Riera et al. [J. Chem. Theory Comput. 16, 2246 (2020)], we introduce a second generation family of data-driven many-body MB-nrg models for CO_2 and systematically assess how the strength and anisotropy of the CO_2–CO_2 interactions affect the models' ability to predict vapor, liquid, and vapor-liquid equilibrium properties. Building upon the many-body expansion formalism, we construct a series of MB-nrg models by fitting 1-body and 2-body reference energies calculated at the coupled cluster level of theory for large monomer and dimer training sets. Advancing from the first generation models, we employ the Charge Model 5 scheme to determine the atomic charges and systematically scale the 2-body energies to obtain more accurate descriptions of vapor, liquid, and vapor-liquid equilibrium properties. Comparisons with the polarizable TTM-nrg model, which is constructed from the same training sets as the MB-nrg models but using a simpler representation of short-range interactions based on conventional Born-Mayer functions, showcase the necessity of high dimensional functional forms for an accurate description of the multidimensional energy landscape of liquid CO_2. These findings emphasize the key role played by the training set quality and flexibility of the fitting functions in the development of transferable, data-driven models which, accurately representing high-dimensional many-body effects, can enable predictive computer simulations of molecular fluids across the entire phase diagram.

[a)] These authors contributed equally to this work.

[b)] Electronic mail: azp@princeton.edu

[c)] Electronic mail: fpaesani@ucsd.edu

I. INTRODUCTION

CO_2 plays a critical role in driving global climate change as a greenhouse gas that traps heat in the atmosphere.[1,2] The recent rapid increase in atmospheric CO_2 has important repercussions on the environment, such as ocean acidification,[3,4] polar ice sheets melting,[5] and extreme climate events.[6] Significant attention has been placed on limiting CO_2 emissions from chemical and industrial processes, particularly through carbon capture and geological carbon sequestration in deep saline aquifers and hydrocarbon reservoirs.[7] Due to the broad range of thermodynamic conditions found in these processes, experimental studies can be challenging to perform.[8,9] Many equation of state models have been developed which represent properties at certain ranges of temperatures and pressures but lack transferability to other conditions.[9–17] Molecular simulations provide a promising alternative to not only model structural, thermodynamic, and transport properties of complex fluids but also to identify molecular-level driving forces for macroscopic phenomena.[18] The predictive capabilities of such simulations, however, ultimately depend on the accuracy of the underlying molecular models.

In the last 30 years, several models have been used in molecular simulations for the study of CO_2 and its mixtures.[19–30] Early models such as Harris and Yung's EPM/EPM2 models,[19] Potoff and Siepmann's TraPPE model,[20] and Potoff and Panagiotopoulos' exponential-6 model[21] were constructed by fitting to experimental vapor-liquid equilibrium properties. These models generally perform well for bulk thermodynamic and transport properties of CO_2. For mixtures such as CO_2 and H_2O, however, these simple empirical models are unable to represent adequately both liquid and vapor phase compositions.[31,32] More recently, many-body effects and polarizability have been incorporated in models to address these inadequacies.[24–27] A 3-body model developed by Persson et al. led to improved predictions of the second and third virial coefficients, but no additional vapor-liquid coexistence properties were probed, possibly due to the added computational cost.[24] Jiang et al. incorporated polarizablity in a model using Drude oscillators and showed that for pure CO_2 systems, polarizability is not significant due to the linear and symmetric nature of CO_2[25]. Although polarizability is likely more significant for mixture systems when CO_2 is paired with polar components, Jiang et al. showed that even with polarizable H_2O and CO_2 models, accurate phase compositions could only be achieved by empirically fitting to the cross second virial

coefficient between the two species.[31] The necessity of empirical parameterization in these models, even those including polarizability, in order to achieve accurate predictions of thermodynamic properties of CO_2 and its mixtures point to inadequacies in the fundamental physics of the models.

Empirical models parameterized to experimental data, while likely to reproduce properties targeted for fitting, often rely on implicit error cancellation in fundamental physical interactions which limit their transferability. Alternatively, several *ab initio*-based models of CO_2 have been developed without such constraints. Bukowski et al. fit a model to symmetry-adapted perturbation theory (SAPT) calculations of dimer and trimer configurations and predicted second virial coefficients in excellent with experiment.[22] Bock et al. developed a 5-site CO_2 model with a high dimensional functional form parameterized on MP2 calculations which similarly produced excellent predictions of the second virial coefficient.[23] However, when it comes to bulk thermodynamic properties, Bratschi et al. concluded that both of these *ab initio*-based models could not adequately reproduce the vapor-liquid phase diagram.[33] Yuan et al. applied adaptive force matching to the development of CO_2 models fitted to *ab initio* data calculated using local second-order Møller-Plesset perturbation theory (LMP2).[34] The models were shown to provide an overall accurate representation of various CO_2 properties although both critical point and vapor pressure were underestimated, and some differences were found between the experimental and simulated liquid structure.

Instead of fitting the total energy derived from *ab initio* calculations, Yu et al. constructed a polarizable CO_2 model parameterized from SAPT dimer energy calculations decomposed into individual contributions, such as electrostatics, dispersion, induction, and polarization, which were then fitted independently using physically motivated functional forms.[26] This model yielded excellent representations of macroscopic properties such as density, isobaric heat capacity, diffusion, and vapor-liquid coexistence densities, although only after scaling the dispersion coefficients derived from the SAPT calculations. A subsequent extension of this model including 3-body contributions reduced the magnitude of the scaling factor of the dispersion coefficients, nonetheless scaling was still necessary to reproduce experimental properties.[27] Beyond simple physically motivated functional forms, more complex permutationally invariant polynomials (PIPs) have also been applied towards the development of CO_2 models. Wang et al. developed a 2-body model using PIPs and showed that a high dimensional functional form was able to accurately predict intramolecular vibrational

frequencies of the CO_2/H_2O dimer as well as cluster properties.[28] Riera et al. developed a transferable many-body model that combines a data-driven PIP representation of short-range 2-body interactions with an implicit representation of many-body effects based on classical polarization. This model was shown to predict well both structure and energetics of small clusters as well as the second virial coefficient and structure of liquid CO_2 at two thermodynamic state points.[30]

In this study, we introduce data-driven many-body CO_2 models which build on the work by Riera et al. in Ref. 30, and follow the "Thole-type model energy" (TTM-nrg) and "Many-body energy" (MB-nrg) theoretical frameworks originally introduced to represent the interactions of halide[35,36] and alkali-metal ions[37,38] with water. The models are obtained with a different charge determination scheme, and used a scaling factor, α, on the training data to investigate how small variations in the 2-body interaction strength and anisotropy affect the ability of the models to predict various liquid and vapor-liquid equilibrium properties. In addition, we examine the advantages of the high dimensional fitting capabilities of PIPs adopted by the MB-nrg models in comparison to conventional Born-Mayer functional forms used by the TTM-nrg model.

The article is organized as follows. In Section II, we discuss the development of the MB-nrg and TTM-nrg CO_2 models, and provide specific details about the simulations and analysis methods. In Section III, we discuss the model fitting, and present the results for various vapor, liquid, and vapor-liquid properties calculated with the new MB-nrg and TTM-nrg models. Finally, in Section IV, we provide a summary of our study and highlight possible future research directions.

II. METHODS

A. Many-body potential energy surfaces

The TTM-nrg and MB-nrg models presented in this study are rigorously derived from the many-body expansion (MBE)[39] of the total energy which recasts the total energy of a system composed by N atomic or molecular monomers into a sum of individual n-body

energies:

$$
\begin{aligned}
E_N(1,\ldots,N) \quad &= \sum_{i=1}^{N} \epsilon^{1\mathrm{B}}(i) \\
&+ \sum_{i<j}^{N} \epsilon^{2\mathrm{B}}(i,j) \\
&+ \sum_{i<j<k}^{N} \epsilon^{3\mathrm{B}}(i,j,k) \\
&+ \ldots + \epsilon^{\mathrm{NB}}(1,\ldots,N).
\end{aligned}
\tag{1}
$$

Since the MBE converges quickly for nonmetallic systems, it has been used as a rigorous theoretical framework for the development of several models for various molecular systems.[28,30,36,38,40–47] Notably, these models have shown remarkable agreement with experimental data for both gas and liquid properties,[30,40,47–52] providing evidence for the transferable nature of the MBE framework.

In the present study of CO_2, $\epsilon^{1\mathrm{B}}(i)$ in Eq. 1 is the 1-body energy associated with any distortion of molecule i from its equilibrium geometry, $\epsilon^{2\mathrm{B}}(i,j)$ is the 2-body energy between molecules i and j, and $\epsilon^{\mathrm{NB}}(1,...,N)$ is the N-body polarization contributions described by the extended Thole-type model originally introduced with the TTM4-F water model.[53]

While both the TTM-nrg and MB-nrg models are constructed following the MBE, they differ in two key aspects. The first difference is related to the 1-body term for which the TTM-nrg model adopts simple harmonic bond and angle stretching functional forms similar to those found in common force fields while the MB-nrg model adopts significantly more flexible PIPs.[54] The second difference is related to the representation of the 2-body term that can be decomposed into 3 distinct contributions:

$$
\epsilon^{2\mathrm{B}} = \epsilon^{2\mathrm{B}}_{\mathrm{elec}} + \epsilon^{2\mathrm{B}}_{\mathrm{disp}} + \epsilon^{2\mathrm{B}}_{\mathrm{sr}} \tag{2}
$$

In both the TTM-nrg and MB-nrg models, $\epsilon^{2\mathrm{B}}_{\mathrm{elec}}$ describes pairwise Coulomb electrostatic interactions between atomic charges on each monomer which are determined using the "Charge Model 5" scheme,[55] and $\epsilon^{2\mathrm{B}}_{\mathrm{disp}}$ describes dispersion interactions. The TTM-nrg and MB-nrg models differ in the expressions for $\epsilon^{2\mathrm{B}}_{\mathrm{sr}}$ describing short-range interactions. In the TTM-nrg model, $\epsilon^{2\mathrm{B}}_{\mathrm{sr}}$ is represented by simple pairwise Born-Mayer functions[56] while it is represented by PIPs in the MB-nrg models. The MB-nrg PIPs have been shown to quantitatively represent quantum-mechanical short-range interactions such as Pauli repulsion, and charge

transfer and penetration, all of which arise due to the overlap of the monomer's electron densities.[57–59] By assessing the differences in the ability of the TTM-nrg and MB-nrg models to predict CO_2 properties for different thermodynamic conditions, we can investigate how the representation of short-range 2-body interactions affect the overall accuracy of the various models.

The appeal of machine-learning data-driven models is the fact that they can precisely reproduce the data that they are trained on. However, the downside is that high quality of training data is crucial. Within the MB-nrg and TTM-nrg frameworks, the reference energies are computed, when possible, with the current "gold standard" for electronic structure calculations – coupled cluster with single, double and perturbative triple excitations, i.e., CCSD(T), in the complete basis set (CBS) limit. In previous work,[30] Riera et al. introduced first generation CO_2 models fitted to CCSD(T)-F12b reference energies.[60,61] Specifically, the 1-body reference energies were obtained using a two-point extrapolation[62,63] between the CCSD(T)-F12b 1-body energies calculated with the aug-cc-pVTZ (AVTZ) and aug-cc-pVQZ (AVQZ) basis sets, while, due to the associated computational cost, the 2-body reference energies were obtained using a two-point extrapolation between the CCSD(T)-F12b 2-body energies calculated with the aug-cc-pVDZ (AVDZ) and aug-cc-pVTZ basis sets.[64–68] The use of relatively smaller basis sets in the calculations of the CCSD(T)-F12b 2-body reference energies implies that these energies are not obtained in the CBS limit, which is usually achieved using a two-point extrapolation between energies calculated with the AVTZ and AVQZ basis sets.[43] In the following, the 2-body reference energies used in the training of the TTM-nrg and MB-nrg models are labeled as CCSD(T)/AV(DT)Z, where AV(DT)Z indicates that the energies were obtained from the two-point extrapolation between the values calculated with the AVDZ and AVTZ basis sets. In the first generation MB-nrg models, the charges were obtained from ChelpG[69] calculations carried out with Q-Chem 5.0[70] at the density functional theory (DFT) level using the ωB97M-V functional[71] in combination with the AVTZ basis set. The dipole polarizabilities were calculated using the exchange-dipole moment (XDM) model[72–74] as implemented in Q-Chem 5.0.[70] All models were parameterized following procedures described in Ref. 30.

The first generation MB-nrg model was shown to predict highly accurate gas-phase properties as well as the structure of liquid CO_2 at two thermodynamic state points, (300 K, 250 MPa) and (300 K, 470 MPa), for which X-ray diffraction data are available.[75] However, as

shown in the next section, the first generation MB-nrg model overestimates the density of liquid CO_2 for pressures below 100 MPa. After careful investigation, we deduced two possible causes for this discrepancy: (1) lack of accuracy in charge determination, and (2) lack of accuracy in the training set reference energies. The latter is caused, in part, by the relatively small magnitude of the interaction energies between CO_2 molecules, which poses a difficult challenge to data-driven models, such as MB-nrg, which aim to precisely reproduce the overall multidimensional reference energy landscape. Larger basis sets could, in principle, aid in improving the accuracy of the reference energies, but the associated computational costs effectively make these calculations impractical.

To address the limitations of the first generation MB-nrg model, we introduce here a second generation of CO_2 models that are developed by: (1) applying the Charge Model 5 charge determination scheme which derives partial atomic charges from the Hirshfeld population analysis,[55] and (2) systematically scaling the 2-body reference energies which allows for a more accurate description of vapor, vapor-liquid, and liquid equilibrium properties. All other components of the second generation of MB-nrg models remain the same as in the original MB-nrg model introduced in Ref. 30. The scaling scheme of the 2-body reference energies adopted by the new MB-nrg models is given by the following expression:

$$\epsilon^{\mathrm{2B}}_{\mathrm{scaled}} = \epsilon^{\mathrm{2B}} + \alpha|\epsilon^{\mathrm{2B}}| \tag{3}$$

where $\epsilon^{\mathrm{2B}}_{scaled}$ and ϵ^{2B} are the scaled and original (unscaled) 2-body reference energies, respectively. In the second generation MB-nrg models, the PIPs representing $\epsilon^{\mathrm{2B}}_{\mathrm{sr}}$ are then fitted to $\epsilon^{\mathrm{2B}}_{scaled}$, which is $(\alpha * 100)\%$ less attractive than the original value for $\epsilon^{\mathrm{2B}} < 0$ and $(\alpha * 100)\%$ more repulsive for $\epsilon^{\mathrm{2B}} > 0$. Using this scaling scheme, we derived four different scaled MB-nrg models that we will henceforth refer to as MB-nrg 100%, MB-nrg 95%, MB-nrg 90%, and MB-nrg 85% for $\alpha = 0, 0.05, 0.1, 0.15$, respectively.

As for the first generation model, the 2-body reference energies for the second generation MB-nrg models were obtained at the CCSD(T)-F12b level of theory, using a two-point extrapolation between the values obtained with the AVDZ and AVTZ basis sets. The training set, generated using the standard protocol defined in MB-Fit,[76] is composed of 40,000 configurations combined from three different sets. The first set is composed of 14,000 configurations of two CO_2 molecules at their optimized geometry, randomly rotated and separated by a distance r for each configuration. The values of r range between 1.5 Å and 9.0 Å in a

logarithmic distribution that ensures a denser grid at shorter distances and a more sparse grid at longer distances. The second set is composed of 20,000 configurations sampled the same way as the previous set but using distorted geometries of CO_2 instead of the optimized geometry, obtained by combining normal mode displacements with the piecewise distribution of a CO_2 molecule (see Ref. 76 for details). The last 6,000 configurations were generated analogously to the monomer distortions but using the dimer normal modes of the global minimum energy structure. All the fits were performed with the MB-Fit software.[76]

B. Calculation of thermodynamic properties

In order to determine the optimal scaling factor α for the MB-nrg models, we performed molecular dynamics (MD) simulations to calculate various thermodynamic properties. All MD simulations were carried out using the LAMMPS package[77] interfaced with the MBX library.[78] In the analyses presented in Sec. III, the statistical uncertainties of the thermodynamic properties are reported as standard error obtained at the 95% confidence level.

For each model, MD simulations in the isothermal-isobaric (NPT) ensemble were performed with periodic boundary conditions for a cubic box containing 256 CO_2 molecules. These NPT simulations were used to calculate the density and structure of liquid CO_2 at temperatures ranging from 240 K to 300 K and pressures ranging from the experimental vapor pressure to 100 MPa. Temperature and pressure were maintained using a global Nosé-Hoover thermostat and barostat[79] with relaxation times of 0.05 ps and 0.5 ps, respectively. Classical equations of motion were numerically integrated with a time step of 0.5 fs. These NPT simulations consisted of 3 ns of equilibration followed by 2 ns of sampling from which the bulk liquid density and structure were obtained. Short-range interactions were evaluated with a real-space cutoff of 9 Å, while long-range interactions (including electrostatic, dispersion, and polarization contributions) were calculated in reciprocal space using a particle-particle particle-mesh solver.

We performed direct coexistence interfacial simulations to calculate vapor-liquid coexistence density and surface tension at temperatures ranging from 240 K to 280 K for the MB-nrg models, and 120 K to 150 K for the TTM-nrg model. The upper limit of the temperature range sampled by each model was determined by the proximity to each of their respective critical points where stable and distinct liquid and vapor regions are able

to coexist. We generated an initial configuration of direct coexistence geometry by taking an equilibrated bulk liquid configuration in a cubic periodic box and expanding in the z-dimension to produce separated liquid and vapor regions in the x-y plane, with the z-dimension normal to the interface. Following the guidelines of Ref. 80, the final dimensions of the expanded box were set to be 20 Å × 20 Å × 200 Å, which allowed for appropriately sized z-axial homogeneous regions of both liquid and vapor phases that were unperturbed by the interfacial width. From these initial configurations, MD simulations in the canonical (NVT) ensemble were performed using the same parameters as those of the NPT simulations of liquid CO_2, excluding the barostat. For the MB-nrg 85% and 95% models, we sampled production periods ranging from 2–5 ns, after equilibration periods of 10 ns. Simulations of the MB-nrg 90% and 100% models were initiated from equilibrated configurations obtained from the aforementioned MB-nrg 85% and 95% trajectories, equilibrated for an additional 3 ns, and sampled over production periods of 3–4 ns.

To obtain the vapor and liquid region coexistence densities, we calculated the average density profile in the z-dimension which was then fitted to a hyperbolic tangent expression:

$$\rho(z) = \frac{\rho_l + \rho_v}{2} - \frac{\rho_l - \rho_v}{2}\tanh\left(\frac{z - z_0}{d}\right) \tag{4}$$

Here, ρ_l is the liquid-phase coexistence density and ρ_v is the vapor-phase coexistence density. The position of the interface, z_0, and the interfacial thickness, d, are also obtained from the fit. The vapor pressure was obtained from sampling the pressure tensor element normal to the interface, P_{zz}.

The critical parameters, ρ_c and T_c, were estimated for each model from the universal scaling of the coexistence densities near the critical point and the law of rectilinear diameters given by:

$$\rho_l - \rho_v = \Delta\rho_0(1 - T/T_c)^\beta \tag{5}$$

$$\frac{\rho_l + \rho_v}{2} = \rho_c + A(T_c - T) \tag{6}$$

Here, A and $\Delta\rho_0$ are system-specific parameters to be adjusted in the fitting, and the critical exponent $\beta \approx 0.326$, based on the three-dimensional Ising model universality class.[81] All data points from T = 240 K to T = 280 K used in this calculations are within ∼20% of the critical point, allowing for the use of Eq. 6 which is only valid close to the critical point. To estimate the uncertainties of the critical parameters ρ_c and T_c, Eqs. 5 and 6 were

fit to 10 block averages of the coexistence densities to obtain a distribution of ρ_c and T_c, from which the standard error was determined with a 95% confidence interval.

III. RESULTS AND DISCUSSION

We assess the accuracy of the five CO_2 models (TTM-nrg and four scaled MB-nrg models) through a systematic analysis of dimer energies, bulk liquid densities, and vapor-liquid coexistence properties. Juxtaposing the TTM-nrg model with the MB-nrg models showcases the necessity of a flexible and high dimensional functional form for the short-range interaction fitting towards the prediction of all properties of interest. The comparison of the four scaled MB-nrg models provides guidance toward selecting the most comprehensive model (and optimal scaling factor) for predictions of CO_2 properties across a wide range of thermodynamic phase space, while emphasizing how small differences in the 2-body training energies can be responsible for notable differences in condensed phase properties.

A. Model fitting

Fig. 1 shows the correlations between 2-body energies calculated with the TTM-nrg and four MB-nrg models and the reference CCSD(T)-F12b/AV(DT)Z values. Just as for other molecular systems, the TTM-nrg model lacks accuracy in predicting the 2-body energies, as shown by the wide range of deviations from the reference values. The MB-nrg models, on the other hand, yield remarkably high accuracy correlations. As expected, the deviations from

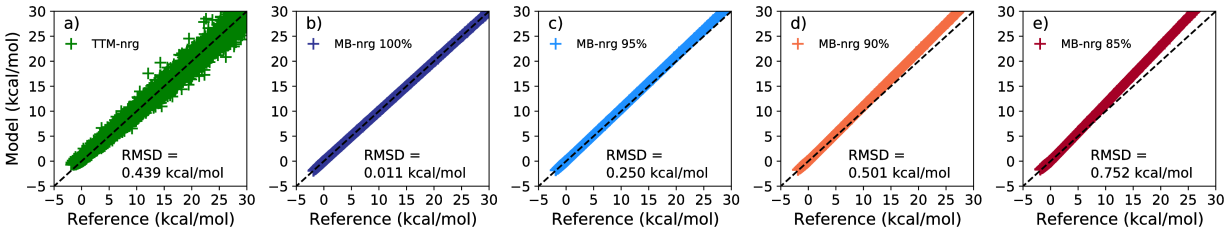

FIG. 1. Correlation between the TTM-nrg/MB-nrg 2-body energies and the corresponding reference values for dimer configurations in the training set with binding energies lower than 30 kcal/mol. Dashed line corresponds to the line $y = x$. The reference 2-body energies were calculated at the CCSD(T)-F12b/AV(DT)Z level of theory. See main text for details.

the reference energies and root-mean-square deviations (RMSDs) progressively increase as the scaling factor α is increased. However, the tight correlation of each of the four MB-nrg models remain, especially at low energy. As we will show in Sec. III D, a proper choice of the scaling factor α for the 2-body training energies appears to be necessary to correctly predict the properties of liquid CO_2 using the MB-nrg models.

B. Dimer energies

The interaction energies of the CO_2 dimer as a function of the carbon–carbon (C–C) distance are shown in Fig. 2 for five different geometric orientations. For each orientation, we show the CCSD(T)-F12b/AV(DT)Z interaction energies along with the corresponding values calculated with the TTM-nrg model and all four MB-nrg models. All MB-nrg models provide improved agreement with the reference energies compared to the TTM-nrg model, which predicts an overall more repulsive 2-body energy surface. As expected, the MB-nrg interaction energy becomes more repulsive for orientations that are generally attractive ("parallel shifted", "plus", "T") and more attractive for orientations that are repulsive ("parallel" and "line") as α increases. Hence, a single global scaling parameter on the dimer energies

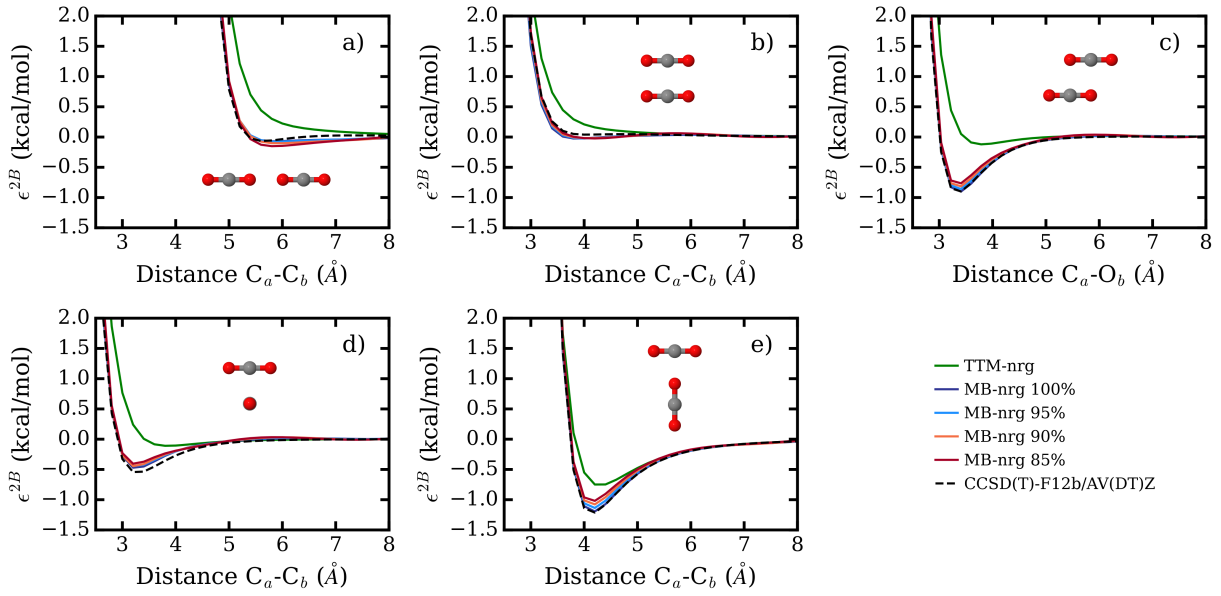

FIG. 2. Interaction energy as a function of distance between two CO_2 monomers for the a) "line", b) "parallel", c) "parallel shifted", d) "plus", and e) "T" orientations.

results in non-uniform shifts of the interaction strength for different dimer separations and orientations. The magnitude of energy changes for the attractive configurations are, however, significantly lower than that for the repulsive configurations. Although the deviations from the reference energies are smaller than 0.2 kcal/mol for all four scaled MB-nrg models, we will show in the next section that even such small differences have significant impact on the predictions of macroscopic thermodynamic properties of the liquid phase.

C. Second virial coefficient

The analysis of the deviations from the reference interaction energies provides a quantitative assessment of the accuracy of the different models. However, interaction energies for CO_2 dimers are not directly measurable experimentally. In order to provide greater insight on the predictive ability of the TTM-nrg and MB-nrg models, we have calculated the second virial coefficient for CO_2 following the same procedure as in Ref. 30. The results in Fig. 3

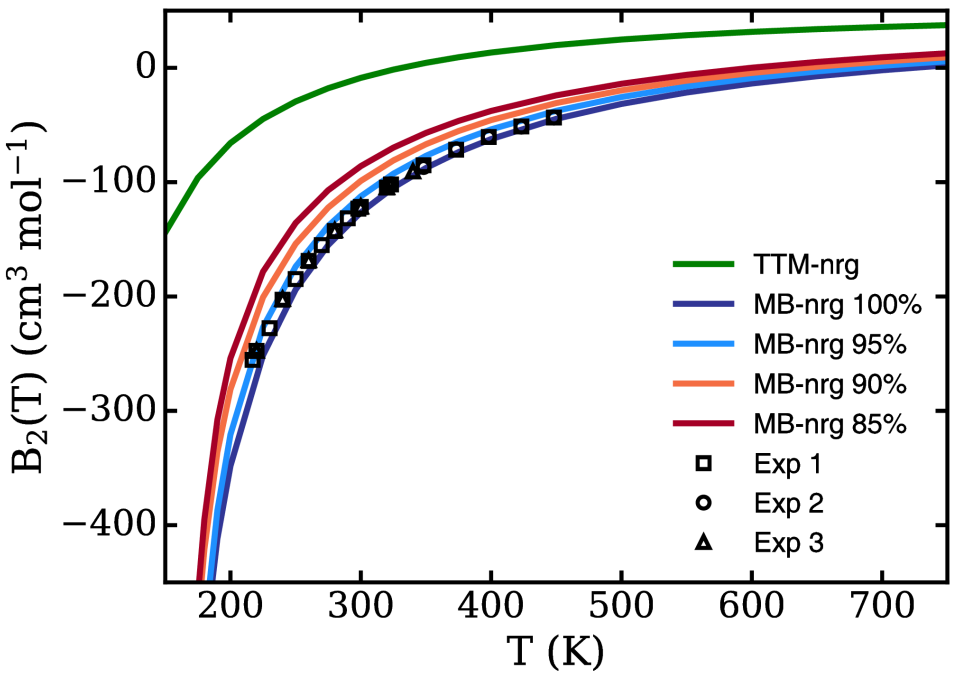

FIG. 3. Second virial coefficients, $B_2(T)$, predictions from the TTM-nrg model and MB-nrg models as a function of T. The experimental data is taken from Ref. 82 for Exp 1, Ref. 83 for Exp 2, and Ref. 84 for Exp 3.

show that even though the four MB-nrg models are fitted to different scaled 2-body energies, the second virial coefficient is in close agreement with the available experimental data.[82–84] By definition, the more repulsive the model is, the larger the virial coefficient. This is in agreement with the results reported in Fig. 3, where the virial coefficients calculated are larger as the 2-body energies are made more repulsive, with the TTM-nrg model predicting values for the virial coefficients that are largely overestimated. The MB-nrg 95% and 100% models display the best agreement with the experimental data, implying that very little or no scaling of the 2-body CCSD(T)/AV(DT)Z energies is necessary to represent the dimer configurations that primarily contribute to the second virial coefficient calculations. However, as we will see in the next section, greater scaling of the 2-body energies are necessary when extrapolating the models to predict the condensed liquid phase.

D. Thermodynamic properties

In order to investigate how the (small) differences in the 2-body energies seen in Fig. 2 affect the predictions of thermodynamic properties of liquid CO_2, we performed MD simulations with each scaled MB-nrg model over a wide range of pressures, from the saturation vapor pressure to 100 MPa, as shown in Fig. 4. Three temperatures (240 K, 270 K, and 300 K) were selected to traverse the liquid phase region. For the chosen thermodynamic conditions, the MB-nrg 95% and MB-nrg 100% scaled models overestimate the experimental densities across the entire range of pressures and temperatures. For the MB-nrg 85% and MB-nrg 90% scaled models, however, the differences from the experimental values are more subtle and dependent on the specific temperature and pressure (although all deviations are smaller than 5% for both models at all pressures excluding the saturation pressure). For higher pressure regimes corresponding to higher liquid densities, the MB-nrg 85% model predictions most closely approach the experimental values. However, at the saturation vapor pressure, the MB-nrg 90% model provides the closest agreement with experiment. While no clear "best" scaling factor emerges from this analysis, both the MB-nrg 85% and MB-nrg 90% models prove capable of providing a reasonably accurate representation of liquid CO_2.

Despite the appreciable differences in the overall liquid density between the four MB-nrg models each with different 2-body energy reference energy scaling, the local liquid structure is minimally affected, as shown by the radial distribution functions (RDFs) in Fig. 5.

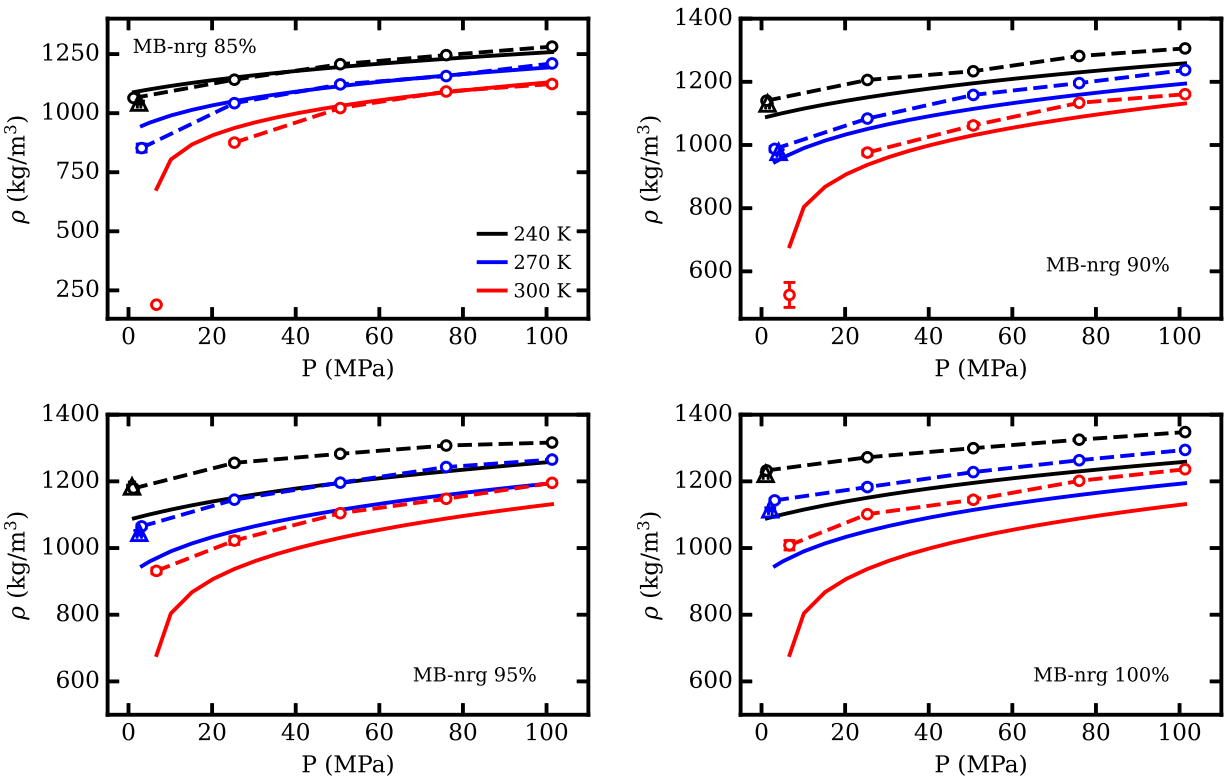

FIG. 4. Bulk liquid densities, represented by open circles connected by dotted lines, from the coexistence pressure to 100 MPa for the 4 MB-nrg models at 3 temperatures. At the coexistence pressure, some models predict a transformation to the vapor phase, as denoted by the open circles unconnected by dotted lines. Triangles represent liquid region densities obtained from vapor-liquid direct coexistence simulations. Experimental data, represented by solid lines, was obtained from NIST.[85]

Consistent with the overall bulk density predictions, an increasing scaling factor α generally reduces the local density, resulting in the gradually lower first peak heights in the RDF. Additionally, the magnitude of structural differences between the four MB-nrg models increases as the temperature increases, reflecting the greater influence of the scaled 2-body energies in lower density configurations at 300 K compared to higher density configurations at 240 K. These discrepancies demonstrate the challenges of representing accurate structural correlations across a range of thermodynamic conditions due to the weak and anisotropic nature of CO_2 interactions.

We then assess the accuracy of the scaled MB-nrg models with respect to the vapor-liquid coexistence behavior using direct coexistence interfacial simulations. Fig. 6a and 6b

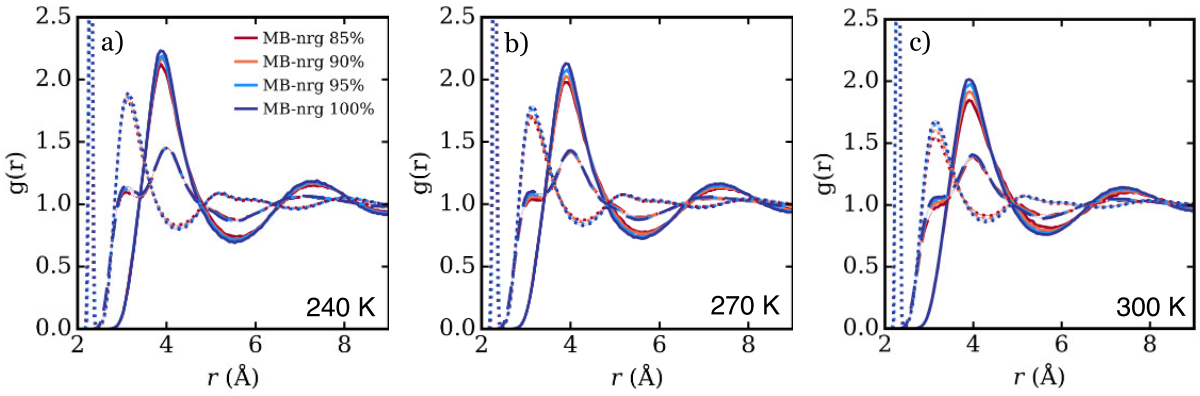

FIG. 5. Radial distribution functions, g(r), of all MB-nrg models at (a.) 240 K, (b.) 270 K, and (c.) 300 K, all at a pressure of 100 MPa. Solid lines represent the C-C correlation, dashed lines represent the C-O correlation, and dotted lines represent the O-O correlation.

show the coexistence densities for all four scaled MB-nrg models at temperatures 240 K $< T <$ 280 K, along with the estimated critical points from Eqs. 5 and 6. In the liquid region, the MB-nrg 95% and MB-nrg 100% models overestimate the coexistence density at all temperatures analyzed in this study, also consistent with our previous findings on the liquid density. On the other hand, the MB-nrg 85% and MB-nrg 90% models show significantly smaller differences with the experimental data, with the MB-nrg 90% model overestimating the coexistence density by $2-4\%$ and the MB-nrg 85% model underestimating it by $4-6\%$.

The vapor-phase representation, however, shows noticeable differences in comparison to the liquid-phase representation. Interestingly, in the vapor region, the MB-nrg 90% and MB-nrg 95% models agree most closely with the experimental coexistence densities across all temperatures sampled, which is also reflected in the most accurate prediction of the saturation vapor pressure, P_v, as shown in Fig. 6c. The surface tension, γ, is reflective of both the liquid and vapor representations. Fig. 6d shows the MB-nrg 90% model as the most accurate model for this property.

It appears from the assessments of the bulk liquid and vapor-liquid coexistence properties that different thermodynamic regimes call for different magnitudes of scaling of the 2-body energies. A clear trend emerges: more condensed systems require a larger scaling of the CCSD(T)/AV(DT)Z reference energies. Since the training set of dimer configurations are most closely representative of vapor-phase interactions, the smaller α necessary for the

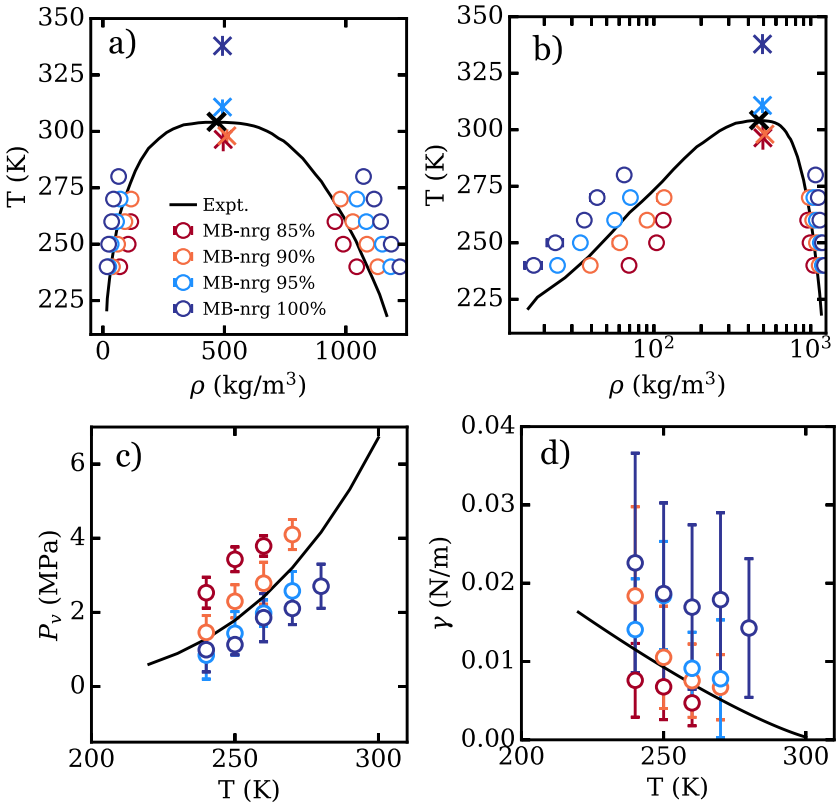

FIG. 6. a) Vapor-liquid phase diagram representing liquid and vapor coexistence densities ρ as a function of T for the four MB-nrg models. b) Same data as a) presented on a log-scale in density ρ to display the vapor phase densities. c) Saturation vapor pressure, P_v, as a function of temperature T predicted by all MB-nrg models. d) Surface tensions, γ, as a function of T predicted by all MB-nrg models. Experimental data were obtained from NIST.[85]

low density vapor systems (the MB-nrg 95% model with 5% scaling of the reference 2-body energies) is not surprising. By the same rationale, a larger α is necessary to achieve accuracy in higher density condensed liquid regimes (the MB-nrg 85% model with 15% scaling of the reference 2-body energies) which deviates further from the molecular environment of the original training set. Out of the four MB-nrg models investigated, we conclude that the MB-nrg 90% model provides the best compromise across the entire range of thermodynamic conditions. Additionally, it is notable that very small differences in the training set of 2-body energies, ∼0.2 kcal/mol, can result in up to 18% differences in liquid densities at moderate/high pressures and even larger differences from experiment when approaching the critical point. The extreme sensitivity of the models' performance can be traced back to the

nature of the interactions between CO_2 molecules which, being very weak and anisotropic, represent a difficult challenge for the development of accurate and transferable models capable of correctly predicting the properties of CO_2 across the entire phase diagram. The non-uniform scaling of different geometric molecular arrangement interactions as a result of a single global scaling factor on the 2-body energy training set (as discussed in Sec. III B) additionally contributes to the discrepancies in the MB-nrg models' predictions. The differences between experimental and simulated properties also suggest that achieving the CBS limit in the CCSD(T)-F12b calculations may be needed to improve the overall accuracy of the training set 2-body energies. Another avenue to explore in future work is the addition of an explicit 3-body term to the MB-nrg models, which was shown to be necessary to correctly reproduce the properties of liquid water[86] as well as the hydration structure of halide[52] and alkali-metal[49] ions. In this regard, it should be noted that the addition of a 3-body term was also found to improve the overall accuracy of the polarizable CO_2 model developed in Ref. 27.

In addition to the MB-nrg models, we also investigated the vapor-liquid coexistence behavior of the TTM-nrg model. Fig. 7 shows that the TTM-nrg model systematically underestimates the experimental coexistence temperature by $\sim$140 K. The saturation vapor pressure of the TTM-nrg model, reported in the Supporting Information, underestimates the experimental value by a similar temperature shift. From these findings, we conclude that the limitations of the simple Born-Mayer repulsive terms used in the current TTM-nrg model, in comparison to that of the PIP-fitted MB-nrg models, are responsible for the inability of the TTM-nrg model to correctly predict the properties of liquid and vapor phase CO_2. It should be noted that, while clearly performing poorly for the properties examined in this study, the present TTM-nrg model was developed only using electronic structure data with no input from other sources. It is, therefore, difficult to make comparisons with other polarizable models of CO_2[25–27] which have shown to correctly reproduce the vapor-liquid coexistence behavior but were empirically parameterized by fitting to experimental bulk properties. In this context, we expect that more accurate versions of the TTM-nrg model can be developed for specific applications by targeting a reduced range of 2-body dimer energies. At the same time, it should be emphasized that these more specialized TTM-nrg models are unlikely to be completely transferable across the entire phase diagram. The main conclusion drawn from the TTM-nrg model predictions is the importance and necessity of high dimensional short-

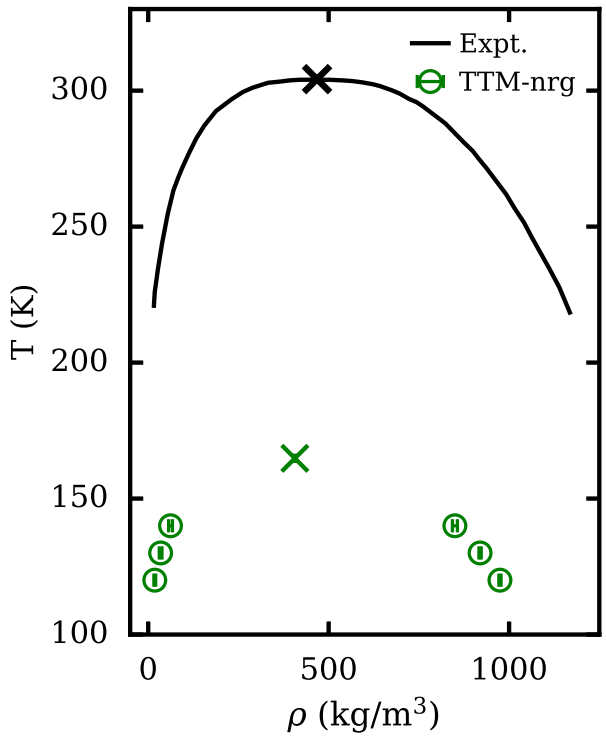

FIG. 7. Vapor-liquid phase diagram representing liquid and vapor coexistence densities ρ as a function of T for the TTM-nrg model. Experimental data was obtained from NIST.[85]

range functional forms (such as the PIPs employed by the MB-nrg models) to accurately represent both gas phase and liquid properties of CO_2 over a wide range of temperatures and pressures.

IV. CONCLUSION

In this work, we constructed second generation models of CO_2 using the MBE formalism and assessed their capabilities with respect to molecular dimer energies and macroscopic thermodynamic properties, namely liquid phase density and vapor-liquid coexistence behavior. We first evaluated four MB-nrg models each with a scaling factor applied to the training set of 2-body dimer interaction energies in an effort to determine the ideal scaling to reproduce thermodynamic properties. For all models surveyed, we note that although the 2-body fitting accuracy is excellent, even small deviations in the training energies of the potential energy surface can yield significant deviations in the prediction of thermodynamic properties. A larger scaling factor was found to be necessary to represent higher density

condensed liquid phase regimes, likely a result of the greater disparity of the molecular environment from the training set of dimer configurations. On the other hand, lower scaling factors were necessary for the predictions of lower density liquids and vapor systems, which more closely resemble the training set of dimer configurations. Out of all MB-nrg model studied in this work, we conclude that the MB-nrg 90% provides the best compromise across all thermodynamic regime investigated in this study. Nonetheless, a single global scaling factor of the 2-body dimer energies is likely unable to wholistically capture the weak and anisotropic interactions of CO_2 across a wide range of thermodynamic phase space, providing motivation for future work toward more targeted scaling procedures.

As for the TTM-nrg models, the low dimensional functional form of the short-range interaction terms severely limit both dimer and thermodynamic property predictions. For a model which is entirely first-principles derived, we conclude that high dimensional functional forms such as PIPs are necessary to achieve flexibility in fitting the potential energy surface of complex fluids such as CO_2 and its mixtures.

There remains a number of avenues to pursue towards the aim of improving the MB-nrg model beyond the capabilities of the present study. Incorporating trimer configurations and explicit 3-body interactions, increasing the level of electronic structure calculations, and fitting to higher level polynomials are all interesting and promising investigations, although all requiring additional computational expense. These topics will be the subjects of future work.

V. ACKNOWLEDGMENTS

The work of S.Y. and A.Z.P. was supported as part of the Computational Chemical Sciences Program funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award # DE-SC0002128, and Chemistry in Solution and at Interfaces (CSI) Center, Award # DE-SC001934. M.R., R.G., and F.P. were supported by the U.S. Department of Energy, through Award # DE-SC0019490. Computing resources were provided by the Princeton Institute for Computational Science and Engineering, National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract DE-AC02-05CH11231, and the Triton Shared Computing Cluster (TSCC) at the San Diego Supercomputer Center (SDSC).

REFERENCES

[1]S. Solomon, G.-K. Plattner, R. Knutti, and P. Friedlingstein, "Irreversible climate change due to carbon dioxide emissions," Proc. Natl. Acad. Sci. U.S.A. **106**, 1704–1709 (2009).

[2]J. H. Seinfeld and S. N. Pandis, *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 3rd ed. (Wiley, 2016) p. 1152.

[3]J. N. Butler, *Carbon Dioxide Equilibria and Their Applications* (Routledge, 2019).

[4]K. Caldeira and M. E. Wickett, "Oceanography: Anthropogenic carbon and ocean pH," Nature **425**, 365 (2003).

[5]J. L. Wadham, J. R. Hawkings, L. Tarasov, L. J. Gregoire, R. G. M. Spencer, M. Gutjahr, A. Ridgwell, and K. E. Kohfeld, "Ice sheets matter for the global carbon cycle," Nat. Commun. **10**, 3567 (2019).

[6]P. Shukla, J. Skea, E. C. Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. P. Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, and J. Malley, "IPCC, 2019: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems," Tech. Rep. (2019).

[7]K. Michael, A. Golab, V. Shulakova, J. Ennis-King, G. Allinson, S. Sharma, and T. Aiken, "Geological storage of CO_2 in saline aquifers—A review of the experience from existing storage operations," Int. J. Greenh. Gas Control **4**, 659–667 (2010).

[8]L. M. Anovitz, T. C. Labotka, J. G. Blencoe, and J. Horita, "Experimental determination of the activity-composition relations and phase equilibria of H_2O-CO_2-NaCl fluids at 500°C, 500 bars," Geochim. Cosmochim. Acta **68**, 3557–3567 (2004).

[9]Y. Ji, X. Ji, X. Feng, C. Liu, L. Lü, and X. Lu, "Progress in the study on the phase equilibria of the CO_2-H_2O and CO_2-H_2O-NaCl systems," Chin. J. Chem. Eng. **15**, 439–448 (2007).

[10]G. K. Jacobs and D. M. Kerrick, "Methane: An equation of state with application to the ternary system H_2O–CO_2–CH_4," Geochim. Cosmochim. Acta **45**, 607–614 (1981).

[11]T. S. Bowers and H. C. Helgeson, "Calculation of the thermodynamic and geochemical consequences of nonideal mixing in the system H_2O–CO_2–NaCl on phase relations in geologic systems: Equation of state for H_2O–CO_2–NaCl fluids at high pressures and temperatures,"

Geochim. Cosmochim. Acta **47**, 1247–1275 (1983).

[12]Z. Duan, N. Møller, and J. H. Weare, "An Equation of state for the CH_4–CO_2–H_2O system: I. Pure systems from 0 to 1000 ^{o}C and 0 to 8000 bar," Geochim. Cosmochim. Acta **56**, 2605–2617 (1992).

[13]Z. Duan, N. Møller, and J. H. Weare, "An equation of state for the CH_4–CO_2–H_2O system: II. Mixtures from 50 to 1000 ^{o}C and 0 to 1000 bar," Geochim. Cosmochim. Acta **56**, 2619–2631 (1992).

[14]K. S. Pitzer and S. M. Sterner, "Equations of state valid continuously from zero to extreme pressures for H_2O and CO_2," J. Chem. Phys. **101**, 3111–3116 (1994).

[15]Z. Duan, N. Møller, and J. H. Weare, "A general equation of state for supercritical fluid mixtures and molecular dynamics simulation of mixture PVTX properties," Geochim. Cosmochim. Acta **60**, 1209 – 1216 (1996).

[16]R. J. Bakker, "Adaptation of the Bowers and Helgeson (1983) equation of state to the H_2O–CO_2–CH_4–N_2–NaCl system," Chem. Geol. **154**, 225–236 (1999).

[17]Z. Duan and R. Sun, "An improved model calculating CO_2 solubility in pure water and aqueous NaCl solutions from 273 to 533 K and from 0 to 2000 bar," Chem. Geol. **193**, 257–271 (2003).

[18]Z. Tian, S. Dai, and D.-E. Jiang, "What can molecular simulation do for global warming?" Advanced Review **6**, 173–197 (2016).

[19]J. G. Harris and K. H. Yung, "Carbon dioxide's liquid–vapor coexistence curve and critical properties as predicted by a simple molecular model," J. Phys. Chem. **99**, 12021–12024 (1995).

[20]J. J. Potoff and J. I. Siepmann, "Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen," AIChE J. **47**, 1676–1682 (2001).

[21]J. Potoff, J. Errington, and A. Z. Panagiotopoulos, "Molecular simulation of phase equilibria for mixtures of polar and non-polar components," Mol. Phys. **97**, 1073–1083 (1999).

[22]R. Bukowski, J. Sadlej, B. Jeziorski, P. Jankowski, K. Szalewicz, S. A. Kucharski, H. L. Williams, and B. M. Rice, "Intermolecular potential of carbon dioxide dimer from symmetry-adapted perturbation theory," J. Chem. Phys. **110**, 3785 (1999).

[23]S. Bock, E. Bich, and E. Vogel, "A new intermolecular potential energy surface for carbon dioxide from ab initio calculations," Chem. Phys. **257**, 147–156 (2000).

[24]R. A. X. Persson, "Gaussian charge polarizable interaction potential for carbon dioxide," J. Chem. Phys. **134**, 034312 (2011).

[25]H. Jiang, O. A. Moultos, I. G. Economou, and A. Z. Panagiotopoulos, "Gaussian-charge polarizable and nonpolarizable models for CO_2," J. Phys. Chem. B **120**, 984–994 (2016).

[26]K. Yu, J. G. McDaniel, and J. Schmidt, "Physically motivated, robust, ab initio force fields for CO_2 and N_2," J. Phys. Chem. B **115**, 10054–10063 (2011).

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

[28]Q. Wang and J. M. Bowman, "Two-component, ab initio potential energy surface for CO_2–H_2O, extension to the hydrate clathrate, $CO_2@(H_2O)_{20}$, and VSCF/VCI vibrational analyses of both," J. Chem. Phys. **147**, 161714 (2017).

[29]O. Sode and J. N. Cherry, "Development of a flexible-monomer two-body carbon dioxide potential and its application to clusters up to $(CO_2)_{13}$," J. Comput. Chem. **38**, 2763–2774 (2017).

[30]M. Riera, E. P. Yeh, and F. Paesani, "Data-driven many-body models for molecular fluids: CO_2/H_2O mixtures as a case study," J. Chem. Theory Comput. **16**, 2246–2257 (2020).

[31]H. Jiang, I. G. Economou, and A. Z. Panagiotopoulos, "Molecular modeling of thermodynamic and transport properties for CO_2 and aqueous brines," Acc. Chem. Res. **50**, 751–758 (2017).

[32]G. A. Orozco, I. G. Economou, and A. Z. Panagiotopoulos, "Optimization of intermolecular potential parameters for the CO_2/H_2O mixture," J. Phys. Chem. B **118**, 11504–11511 (2014).

[33]C. Bratschi, H. Huber, and D. J. Searles, "Non-Hamiltonian molecular dynamics implementation of the Gibbs ensemble method. II. Molecular liquid-vapor results for carbon dioxide," J. Chem. Phys. **126**, 164105 (2007).

[34]Y. Yuan, Z. Ma, and F. Wang, "Leveraging local MP2 to reduce basis set superposition errors: An efficient first-principles based force-field for carbon dioxide," J. Chem. Phys. **151**, 184501 (2019).

[35]D. J. Arismendi-Arrieta, M. Riera, P. Bajaj, R. Prosmiti, and F. Paesani, "i-TTM model for ab initio-based ion–water interaction potentials. 1. Halide–water potential energy functions," J. Phys. Chem. B **120**, 1822–1832 (2015).

[36]P. Bajaj, A. W. Götz, and F. Paesani, "Toward chemical accuracy in the description of ion–water interactions through many-body representations.I. Halide–water dimer potential energy surfaces," J. Chem. Theory Comput. **12**, 2698–2705 (2016).

[37]M. Riera, A. W. Götz, and F. Paesani, "The i-TTM model for ab initio-based ion–water interaction potentials. II. Alkali metal ion–water potential energy functions," Phys. Chem. Chem. Phys. **18**, 30334–30343 (2016).

[38]M. Riera, N. Mardirossian, P. Bajaj, A. W. Götz, and F. Paesani, "Toward chemical accuracy in the description of ion–water interactions through many-body representations. Alkali-water dimer potential energy surfaces," J. Chem. Phys. **147**, 161715 (2017).

[39]D. Hankins, J. Moskowitz, and F. Stillinger, "Water molecule interactions," J. Chem. Phys. **53**, 4544–4554 (1970).

[40]R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. Van der Avoird, "Predictions of the properties of water from first principles," Science **315**, 1249–1252 (2007).

[41]Y. Wang, B. C. Shepler, B. J. Braams, and J. M. Bowman, "Full-dimensional, ab initio potential energy and dipole moment surfaces for water," J. Chem. Phys. **131**, 054511 (2009).

[42]Y. Wang, X. Huang, B. C. Shepler, B. J. Braams, and J. M. Bowman, "Flexible, ab initio potential, and dipole moment surfaces for water. I. Tests and applications for clusters up to the 22-mer," J. Chem. Phys. **134**, 094509 (2011).

[43]V. Babin, C. Leforestier, and F. Paesani, "Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, VRT spectrum, and second virial coefficient," J. Chem. Theory Comput. **9**, 5395–5403 (2013).

[44]V. Babin, G. R. Medders, and F. Paesani, "Development of a "first principles" water potential with flexible monomers. II: Trimer potential energy surface, third virial coefficient, and small clusters," J. Chem. Theory Comput. **10**, 1599–1607 (2014).

[45]G. R. Medders, V. Babin, and F. Paesani, "Development of a "first-principles" water potential with flexible monomers. III. Liquid phase properties," J. Chem. Theory Comput. **10**, 2906–2910 (2014).

[46]R. Conte, C. Qu, and J. M. Bowman, "Permutationally invariant fitting of many-body, non-covalent interactions with application to three-body methane–water–water," J. Chem. Theory Comput. **11**, 1631–1638 (2015).

[47]M. Riera, A. Hirales, R. Ghosh, and F. Paesani, "Data-driven many-body models with chemical accuracy for CH_4/H_2O mixtures," J. Chem. Phys. B **124**, 11207–11221 (2020).

[48]S. K. Reddy, S. C. Straight, P. Bajaj, C. Huy Pham, M. Riera, D. R. Moberg, M. A. Morales, C. Knight, A. W. Götz, and F. Paesani, "On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice," J. Chem. Phys. **145**, 194504 (2016).

[49]D. Zhuang, M. Riera, G. K. Schenter, J. L. Fulton, and F. Paesani, "Many-body effects determine the local hydration structure of Cs^+ in solution," J. Phys. Chem. Lett. **10**, 406–412 (2019).

[50]P. Bajaj, J. O. Richardson, and F. Paesani, "Ion-mediated hydrogen-bond rearrangement through tunnelling in the iodide–dihydrate complex," Nat. Chem. **11**, 367 (2019).

[51]M. C. Muniz, T. E. Gartner III, M. Riera, C. Knight, S. Yue, F. Paesani, and A. Z. Panagiotopoulos, "Vapor–liquid equilibrium of water with the MB-pol many-body potential," J. Chem. Phys. **154**, 211103 (2021).

[52]A. Caruso and F. Paesani, "Data-driven many-body models enable a quantitative description of chloride hydration from clusters to bulk," J. Chem. Phys. **155**, 064502 (2021).

[53]C. Burnham, D. Anick, P. Mankoo, and G. Reiter, "The vibrational proton potential in bulk liquid water and ice," J. Chem. Phys. **128**, 154519 (2008).

[54]B. J. Braams and J. M. Bowman, "Permutationally invariant potential energy surfaces in high dimensionality," Int. Rev. Phys. Chem. **28**, 577–606 (2009).

[55]A. V. Marenich, S. V. Jerome, C. J. Cramer, and D. G. Truhlar, "Charge model 5: An extension of hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases," J. Chem. Theory Comput. **8**, 527–541 (2012).

[56]A. J. Stone, *The Theory of Intermolecular Forces* (Oxford University Press, Oxford, 2013).

[57]F. Paesani, "Water: Many-body potential from first principles (from the gas to the liquid phase)," Handbook of Materials Modeling: Methods: Theory and Modeling , 635–660 (2020).

[58]B. B. Bizzarro, C. K. Egan, and F. Paesani, "Nature of halide–water interactions: Insights from many-body representations and density functional theory," J. Chem. Theory Comput. **15**, 2983–2995 (2019).

[59]C. K. Egan, B. B. Bizzarro, M. Riera, and F. Paesani, "Nature of alkali ion–water interactions: Insights from many-body representations and density functional theory. II," J. Chem. Theory Comput. **16**, 3055–3072 (2020).

[60]T. B. Adler, G. Knizia, and H. J. Werner, "A simple and efficient CCSD(T)-F12 approximation," J. Chem. Phys. **127**, 221106 (2007).

[61]G. Knizia, T. B. Adler, and H.-J. Werner, "Simplified CCSD(T)-F12 methods: Theory and benchmarks," J. Chem. Phys. **130**, 054104 (2009).

[62]J. G. Hill, K. A. Peterson, G. Knizia, and H.-J. Werner, "Extrapolating MP2 and CCSD explicitly correlated correlation energies to the complete basis set limit with first and second row correlation consistent basis sets," J. Chem. Phys. **131**, 194105 (2009).

[63]U. Góra, R. Podeszwa, W. Cencek, and K. Szalewicz, "Interaction energies of large clusters from many-body expansion," J. Chem. Phys. **135**, 224102 (2011).

[64]T. H. Dunning Jr, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," J. Chem. Phys. **90**, 1007–1023 (1989).

[65]R. A. Kendall, T. H. Dunning Jr, and R. J. Harrison, "Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions," J. Chem. Phys. **96**, 6796–6806 (1992).

[66]D. E. Woon and T. H. Dunning Jr, "Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon," J. Chem. Phys. **98**, 1358–1371 (1993).

[67]D. E. Woon and T. H. Dunning Jr, "Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties," J. Chem. Phys. **100**, 2975–2988 (1994).

[68]D. E. Woon and T. H. Dunning Jr, "Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon," J. Chem. Phys. **103**, 4572–4585 (1995).

[69]C. M. Breneman and K. B. Wiberg, "Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis," J. Comput. Chem. **11**, 361–373 (1990).

[70]E. Epifanovsky, A. T. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, *et al.*, "Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package," The Journal of

chemical physics **155**, 084801 (2021).

[71]N. Mardirossian and M. Head-Gordon, "ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation," J. Chem. Phys. **144**, 214110 (2016).

[72]A. D. Becke and E. R. Johnson, "Exchange-hole dipole moment and the dispersion interaction," J. Chem. Phys. **122**, 154104 (2005).

[73]E. R. Johnson and A. D. Becke, "A post-Hartree–Fock model of intermolecular interactions," J. Chem. Phys. **123**, 024101 (2005).

[74]E. R. Johnson and A. D. Becke, "A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections," J. Chem. Phys. **124**, 174104 (2006).

[75]F. Datchi, G. Weck, A. Saitta, Z. Raza, G. Garbarino, S. Ninet, D. Spaulding, J. Queyroux, and M. Mezouar, "Structure of liquid carbon dioxide at pressures up to 10 GPa," Phys. Rev. B **94**, 014201 (2016).

[76]E. F. Bull-Vulpe, M. Riera, A. W. Götz, and F. Paesani, "MB-Fit: Software infrastructure for data-driven many-body potential energy functions," J. Chem. Phys. **155**, 124801 (2021).

[77]S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," J. Comp. Phys. **117**, 1–19 (1995).

[78]"MBX: A many-body energy and force calculator," `http://paesanigroup.ucsd.edu/software/mbx.html`.

[79]H. Kamberaj, R. J. Low, and M. P. Neal, "Time reversible and symplectic integrators for molecular dynamics simulations of rigid molecules," J. Chem. Phys. **122**, 224114 (2005).

[80]E. A. Muller, Å. Ervik, and A. Mejía, "A guide to computing interfacial properties of fluids from molecular simulations [Article v1.0]," LiveCoMS **2**, 21385 (2021).

[81]J. Rowlinson and B. Widom, *Molecular Theory of Capillarity* (Dover Publications, Mineola, NY, 2002).

[82]J. Holste, K. Hall, P. Eubank, G. Esper, M. Watson, W. Warowny, D. Bailey, J. Young, and M. Bellomy, "Experimental (p,V_m,T) for pure CO_2 between 220 and 450 K," J. Chem. Thermodyn. **19**, 1233–1250 (1987).

[83]M. R. Patel, L. L. Joffrion, and P. T. Eubank, "A simple procedure for estimating virial coefficients from Burnett PVT data," AIChE J. **34**, 1229–1232 (1988).

[84]W. Duschek, R. Kleinrahm, and W. Wagner, "Measurement and correlation of the (pressure, density, temperature) relation of carbon dioxide I. The homogeneous gas and liquid

regions in the temperature range from 217 K to 340 K at pressures up to 9 MPa," J. Chem. Thermodyn. **22**, 827–840 (1990).

[85]P. Linstrom and W. Mallard, *NIST Chemistry WebBook, NIST Standard Reference Database Number 69* (National Institute of Standards and Technology, Gaithersburg MD) p. 20899.

[86]V. Babin, G. R. Medders, and F. Paesani, "Toward a universal water model: First principles simulations from the dimer to the liquid phase," J. Phys. Chem. Lett. **3**, 3765–3769 (2012).