

Machine learning for non-additive intermolecular potentials: from quantum chemistry to first principles predictions - Electronic supplementary information

Richard S. Graham^{a*} and Richard J. Wheatley^{b†}

^aSchool of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD.

^bSchool of Chemistry, University of Nottingham, Nottingham NG7 2RD.

May 11, 2022

1 Introduction

Uteva *et al.* (2017, 2018) presented a method to use Gaussian Processes (GPs) to produce accurate potential energy surfaces (PES) from a limited number of ab-initio calculations of the interaction potential of two molecules. Herein we generalise this method to non-additive three-body interactions and use resulting PES to make first-principles predictions for the thermophysical properties of CO₂–Ar mixtures.

2 Angular and distance coordinates

We begin with the coordinates used to describe the relative position (geometry) of three molecules. A geometry contains either two or three molecules, each of which can be either a CO₂ molecule or an Ar atom. We describe a geometry via the angular and centre-to-centre distance coordinates shown in figure S1. Here r_{ij} denotes the centre-to-centre distances between molecules i and j and $\tilde{\theta}$ is the angle between the direction of r_{12} and the line between the midpoint of r_{12} and the centre of molecule 3. The C–O bond was assumed to be fixed at 1.1632 Å for all calculations. When assigning molecule numbers we assign the lowest slots to the CO₂ molecules and then fill the remaining with Ar atoms. The relative positions of the centre of each molecule can be specified via all three centre-to-centre distances. Furthermore, β_i and α_i denote, respectively, the angle between r_{12} and the CO₂ axis and the torsional angle of the CO₂ axis, for the i^{th} molecule. The relevant ranges of these coordinates for each system are shown in tables S1 and S2.

*Richard.Graham@nottingham.ac.uk

†Richard.Wheatley@nottingham.ac.uk

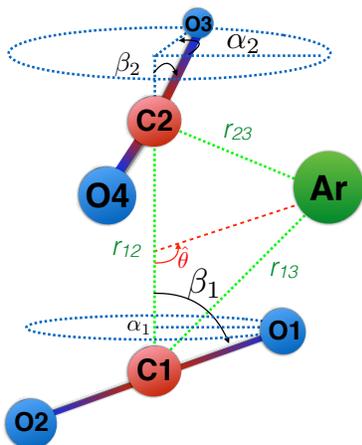


Figure S1: An illustration of the angular and distance co-ordinates for the trimolecular complex $(\text{CO}_2)_2\text{-Ar}$

3 Latin hypercube design

3.1 The two-body algorithm of Uteva *et. al*

Our GPs are trained and tested against sets of ab initio interaction calculations. The data sets comprise molecular geometries and their corresponding interaction energy, which are computed later by ab initio calculation once the design is chosen. All data sets are initially produced by a Latin hyper cube (LHC) design, which is intended to evenly fill the GP region of the space. For two-body systems we followed the algorithm of Uteva *et al.* (2017, 2018), which we briefly summarise herein. For full details see the original papers. We then generalise this algorithm to three-body interactions.

To avoid very widely separated geometries and very high energies, a geometric constraint and a high energy cut off are imposed. A geometry is rejected if $\min(\{r_i\}) < r_{\min}$ or $\min(\{r_i\}) > r_{\max}$ or $E(\{r_i\}) > E_{\text{cut}}$, where $\{r_i\}$ is the set of all interatomic distances in a geometry and E is the energy of a geometry. The value of r_{\min} depends on the atoms involved in the interatomic distance: the values are $r_{\min} = 1.5\text{\AA}$ when neither atom is Ar, 2.0\AA when just one atom is Ar and 2.87\AA when both atoms are Ar. Both r_{\max} and E_{cut} are constants with values of 9\AA and $0.005E_h$, respectively.

The LHC is generated as follows: generate an initial LHC in the relevant distance and angle coordinates (see table S1); reject geometries outside the geometric constraint, and repeat over many candidate LHC designs; retain the LHC candidate with the best space-filling (defined as largest minimum distance between geometries); perform interaction calculations for all geometries in the chosen LHC; and reject any geometries that exceed the high energy cut off.

	$r_{12}[\text{\AA}]$	$\cos(\beta_1)$	α_2	$\cos(\beta_2)$	Other constraints
(Ar) ₂	1.5-9	-	-	-	-
CO ₂ -Ar	1.5-10	0-1	-	-	-
(CO ₂) ₂	1.5-11	0-1	0- π	0-1	$\cos(\beta_1) < \cos(\beta_2)$

Table S1: The co-ordinate ranges used to generate LHC data for the two-body systems in this work.

3.2 Three-body ranges of angles and distances

	$r_{12}[\text{\AA}]$	$r_{13}[\text{\AA}]$	$r_{23}[\text{\AA}]$	Other constraints
(Ar) ₃	2.87-13.5	2.87-13.5	2.87-13.5	$r_{12} < r_{13} < r_{23}$

	$r_{12}[\text{\AA}]$	$r_{23}[\text{\AA}]$	$\tilde{\theta}$	α_1	$\cos(\beta_1)$	Other constraints
CO ₂ -(Ar) ₂ : region A	2.0-5.0	2.87-9.3	0- $\pi/2$	0- π	0-1	$r_{12} < r_{13} < r_{23}$
CO ₂ -(Ar) ₂ : region B	2.0-5.0	2.87-5.5	$\pi/2$ - π	0- π	0-1	$r_{12} < r_{23} < r_{13}$
CO ₂ -(Ar) ₂ : region C	2.0-6.65	2.87-5.0	$\pi/2$ - π	0- π	0-1	$r_{23} < r_{12} < r_{13}$

	$r_{12}[\text{\AA}]$	$r_{23}[\text{\AA}]$	$\tilde{\theta}$	α_1	$\cos(\beta_1)$	α_2	$\cos(\beta_2)$	Other constraints
(CO ₂) ₂ -Ar: region A	1.5-7.0	2.0-9.7	0- $\pi/2$	0- π	0-1	0-2 π	0-1	$r_{12} < r_{13} < r_{23}$
(CO ₂) ₂ -Ar: region D	1.5-8.0	2.0-9.7	0- $\pi/2$	0- π	0-1	0-2 π	0-1	$r_{13} < r_{12} < r_{23}$
(CO ₂) ₂ -Ar: region E	1.5-11.0	2.0-6.7	0- $\pi/2$	0- π	0-1	0-2 π	0-1	$r_{13} < r_{23} < r_{12}$

Table S2: The co-ordinate ranges used to generate LHC data for the three-body systems in this work. These ranges are used to generate candidate geometries within a LHC.

We subdivide the large three-body systems into smaller regions, to limit the number of training points required for each GP evaluation. We base the regions on the order of the centre-to-centre distances. Thus there are six regions, from the six possible orderings of the three centre-to-centre distances. However, for the systems herein the constraints due to interchange of identical molecules mean we need to consider, at most, three regions per system (see table S2).

The relative centre-to-centre distances can be specified by specifying r_{12} , $\hat{\theta}$ and r_{23} , for all regions in table S2. The ranges of angles and interatomic distances used for three-body LHCs are shown in table S2. The range of angles are determined by the symmetry distinct region. The symmetries due to interchange of identical molecules also introduced constraints on the ordering of the centre-to-centre distances. For example, for (Ar)₃, all atoms are identical so, without loss of generality, we have $r_{12} \leq r_{13} \leq r_{23}$.

3.3 Three-body geometric constraint

The geometric constraint is based on minimum and maximum thresholds for the interatomic distances and we generalise this for three-body interactions as follows. The condition at short range is identical to the two-body case: namely if *any* distance is below the relevant r_{\min} we reject the entire geometry, with the values for r_{\min} given in section 3.1. For the maximum distances, we define three maximum distances, r_{12}^{\max} , r_{13}^{\max} and r_{22}^{\max} . The value of these maximum distances

depends on the system and on the region (see table S3). For $(\text{Ar})_3$ and $\text{CO}_2\text{-(Ar)}_2$ we reject the entire geometry if $r_{12} > r_{12}^{\max}$ or $r_{13} > r_{13}^{\max}$ or $r_{23} > r_{23}^{\max}$. For $(\text{CO}_2)_2\text{-Ar}$ we reject the entire geometry if $\min(\{r_{12}\}) > r_{12}^{\max}$ or $\min(\{r_{13}\}) > r_{13}^{\max}$ or $\min(\{r_{23}\}) > r_{23}^{\max}$, where $\{r_{ij}\}$ denotes the set of all interatomic distances between atoms on molecule i and atoms on molecule j .

	$r_{12}^{\max}[\text{\AA}]$	$r_{13}^{\max}[\text{\AA}]$	$r_{23}^{\max}[\text{\AA}]$
$(\text{Ar})_3$	9.0	9.0	9.0
$\text{CO}_2\text{-(Ar)}_2$: region A	5.0	5.5	9.3
$\text{CO}_2\text{-(Ar)}_2$: region B	5.0	9.3	5.5
$\text{CO}_2\text{-(Ar)}_2$: region C	6.65	10.5	5.0
$(\text{CO}_2)_2\text{-Ar}$: region A	4.5	5.5	8.5
$(\text{CO}_2)_2\text{-Ar}$: region D	5.5	4.5	8.5
$(\text{CO}_2)_2\text{-Ar}$: region E	8.5	4.5	5.5

Table S3: Values for r_{\max} for the geometric constraint for the three-body systems in this work. These values are used to reject geometries from the LHC and to define outer perimeter of the GP region.

3.4 High energy cut off for three-body systems

The high energy cut-off is generalised for three-body interactions as follows. A geometry exceeds the cut off if any of its three binary interaction energies exceeds E_{cut} , which remains at $E_{cut} = 0.005E_h$. The geometry is also rejected if its total interaction energy exceeds $3E_{cut}$. We were able to implement the condition on the binary interactions during the LHC design algorithm, as accurate PES for all binary interactions were completed before the three-body calculations. This caught most geometries that exceed the high energy cut-off before expensive three-body interaction calculations were performed.

3.5 Ab-initio interaction calculations

Interactions calculations were carried out in Molpro [Werner *et al.* (2012)] and were performed at two levels: moderate accuracy and high accuracy. Both levels used coupled-cluster theory with single, double and non-iterative triple excitations [CCSD(T)]. Moderate accuracy calculations used the augmented correlation-consistent triple-zeta (aug-cc-pVTZ) basis set, whereas high accuracy calculations utilised complete basis set (CBS) extrapolation of the interaction energies from the augmented correlation-consistent quadruple-zeta (aug-cc-pVQZ) and aug-cc-aVTZ basis sets. In all calculations, basis set superposition errors are corrected using the full counterpoise correction procedure. For two-body interactions we return the total interaction energy between the two molecules. For three-body interactions we return all three binary interaction energies, the total energy and the non-additive energy.

3.6 LHC generation for three-body interactions

We wish to generate a dataset of model evaluations, $\{\mathbf{x}_i, E(\mathbf{x}_i)\}_{i=1}^N$, that can be used to train the Gaussian process, where the \mathbf{x}_i represent N distinct molecular geometries and E is the non-

additive energy. Each element of \mathbf{x}_i is the inverse distance between two atoms, running over all pairs of nuclei on different molecules. We generate multiple candidate designs and use a maximin criterion to evaluate these designs. Thus we seek designs which maximise the minimum distance between any two design points. We define the effective distance between points \mathbf{x}_i and \mathbf{x}_j in the design to be

$$|\mathbf{x}|_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j). \quad (1)$$

The LHC algorithm for three-body interactions can be summarised as follows

1. Generate a LHC the in centre-to-centre coordinates (either $1/r_{12}$, $1/r_{13}$ and $1/r_{23}$ or $1/r_{12}$, $1/r_{23}$ and $\tilde{\theta}$) and the rigid-body rotation angles, with the ranges specified in table S2
2. Convert the LHC data to atomic positions and compute all interatomic distances for pairs of atoms on separate molecules.
3. Reject the geometries that do not obey the geometric constraint or other constraints in table S2.
4. Use the previously computed two-body PES to calculate all two-body interactions and reject any geometry where any two-body energy exceeds E_{cut} ($0.005E_h$).
5. Reject the entire LHC if it does not contain at least the target number of geometries.
6. Find the minimum $|\mathbf{x}|_{ij}^2$ within the current LHC.
7. Repeat for as many new LHCs as desired and return the LHC with the largest minimum $|\mathbf{x}|_{ij}^2$.
8. For the chosen LHC design, compute the interaction energies at moderate accuracy, as detailed in section 3.5.
9. Reject any geometries whose total energy exceeds $3E_{cut}$.

4 Gaussian Process training

4.1 Applying Gaussian Processes to potential energy surfaces

We use the non-parametric machine learning technique, Gaussian Processes (GP) [Rasmussen and Williams (2006)], to interpolate our ab-initio interaction calculations, to produce a potential energy surface (PES). The general method for training a GP to produce a PES is described by Uteva *et al.* (2017). Significantly, the method does not change when the chemistry or number of molecules change. Here, we summarise the approach of Uteva *et al.* (2017).

GPs are non-parametric models which have proved successful in creating theory-free models of complex datasets. GPs are mathematically tractable and interpretable, and allow prior information to be built into the model. The prior specification of a GP consists of a mean function (often taken as zero) and a covariance function $k(\mathbf{x}, \mathbf{x}')$, expressing the covariance between $f(\mathbf{x})$ and $f(\mathbf{x}')$, where f is the function being interpolated. Training data, consisting of observations of the value of f at various locations, update the mean and covariance functions to give a posterior model which predicts the function at any location.

The intermolecular energy is a non-stationary function of distance, as it varies rapidly at small interatomic separations, but more gently at larger separation. To achieve approximate stationarity Uteva *et al.* (2017) used the inverse interatomic distances as covariates in the GP. Thus the GP coordinates were $\mathbf{x} = (1/r_1, \dots, 1/r_{N_D})$ where r_i is the interatomic distance, running over all pairs of nuclei on different molecules. Their GP has a zero mean function and a squared-exponential covariance function

$$\kappa(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \prod_{i=1}^{N_D} \exp \left[-\frac{(x_i - x'_i)^2}{2l_i^2} \right],$$

where σ_f^2 is the signal variance and l_i is the correlation length for each dimension. The potential energy surfaces should be invariant under various permutations of \mathbf{x} , due to flips of the head-to-tail symmetric CO₂ molecules and interchange of two molecules of the same species. To ensure the GP reflects this, Uteva *et al.* (2017) let G represent the permutation group containing permutations of elements of \mathbf{x} under which the energy surface is unchanged. If $l_i = l_j$ when coordinates x_i and x_j swap for some permutation in G , then a covariance function of the form

$$k_{\text{sym}}(\mathbf{x}, \mathbf{x}') = \sum_{g \in G} \kappa(g\mathbf{x}, \mathbf{x}').$$

resulted in a GP which shares the symmetries of the energy surface.

GP training was performed with the GPy package [GPy (2015)]. Zero-mean Gaussian observation error [Rasmussen and William (2006)] was assumed on the function outputs, with standard deviation σ_n . Thus the model’s hyperparameters were σ_f , σ_n and $\{l_i\}$ and these hyperparameters were estimated by optimising the log-likelihood. Broad *et al.* (2021) demonstrated that more reliable optimisation is achieved by imposing a weak prior distribution on all hyperparameters. Hence, like Broad *et al.* (2021), we used a gamma distribution with an expectation of one and a variance of two as the prior distribution for all hyperparameters. Typically, 20 – 30 random restarts were sufficient to find the optimal hyperparameters multiple times.

4.2 LHC learning

Uteva *et al.* (2017) used LHC design as their strategy to distribute geometries within their training sets. In this method a sequence of training LHCs of increasing sizes is produced, a GP is trained to each LHC and the root-mean-square error (RMSE) is computed against a very large, independent LHC test set. Once an acceptable RMSE is achieved, the calculations for this training LHC are upgraded to a higher level of theory (a process known as transfer learning) to produce an accurate PES.

4.3 Sequential learning

In a later publication, Uteva *et al.* (2018) described a series of active and sequential learning techniques, which aim to improve the placement of training points compared to LHC learning. For a given training set size, these methods give a smaller RMSE than LHC learning. Herein, we adopt the sequential learning technique from Uteva *et al.* (2018), known as the highest error search. This method begins with an empty training set and a large reference set. First, the

largest energy geometry is moved from the reference set, into the training set and a GP is trained to this set. For this GP, the RMSE is computed against the reference set and the geometry in the reference set with the largest absolute error is located. This largest error point is then moved from the reference set to the training set. This process of adding a geometry to the training set, retraining the GP and computing the RMSE is repeated until a sufficiently small RMSE is achieved.

5 Results

We require a PES for each interaction that is relevant to CO₂-Ar mixtures, up to three-body interactions. An additive (Ar)₂ PES and a non-additive PES for (CO₂)₃ are available in the literature [Patkowski and Szalewicz (2010); Hellmann (2017)], so we produced GP force fields for the remaining two and three-body interactions, as shown in tables S4 and S5. For the smallest four systems LHC learning was sufficient to give a very good RMSE for a reasonable training set size (see table S4). However, for the more challenging three-body systems, we found that sequential learning via the largest error method was necessary to produce a sufficiently accurate RMSE within a reasonable training set size. For example for CO₂-(Ar)₂ in region A our best LHC result was $4.1 \times 10^{-7} E_h$, which is over an order of magnitude larger than the RMSE from sequential learning with the same training set size (see figure S2a). In all sequential learning the training set remained significantly smaller than the reference set throughout. We also monitored continuously the root mean square value (RMS) of the reference set to ensure that changes in this RMS due to depletion of the reference set were minimal.

	Test set size	Training / transfer set size	RMSE of transfer set [E_h]	Percentage error
(Ar) ₂	1398	72	2.1×10^{-8}	0.004%
CO ₂ -Ar	1337	221	7.3×10^{-8}	0.007%
(CO ₂) ₂	2190	146	3.9×10^{-6}	0.35%
(Ar) ₃	5398	337	4.1×10^{-8}	0.1%

Table S4: The test and training set sizes, along with the resulting RMSE for the systems that were modelled by LHC training. The percentage errors are expressed as a percentage of the root mean square value of the initial reference set.

Plots of the reducing RMSE with increased training set size from sequential learning are shown in figure S2 for CO₂-(Ar)₂ (a) and (CO₂)₂-Ar (b). For each region of CO₂-(Ar)₂ we obtained an RMSE of $6 - 11 \times 10^{-8} E_h$ ($\sim 0.5\%$ of the mean square value of the reference set) within 1000 sequentially learnt data points (see table S3). (CO₂)₂-Ar is more challenging because of the two additional angular degrees of freedom. Nevertheless we obtained a usable RMSE (1 - 2% error) within ~ 1000 training points (see table S5).

6 PES behaviour outside the geometric constraint

Outside the geometric constraint we require alternative forms to complement the GP. For two-body interactions, suitable solutions are a strongly repulsive function at short range and an

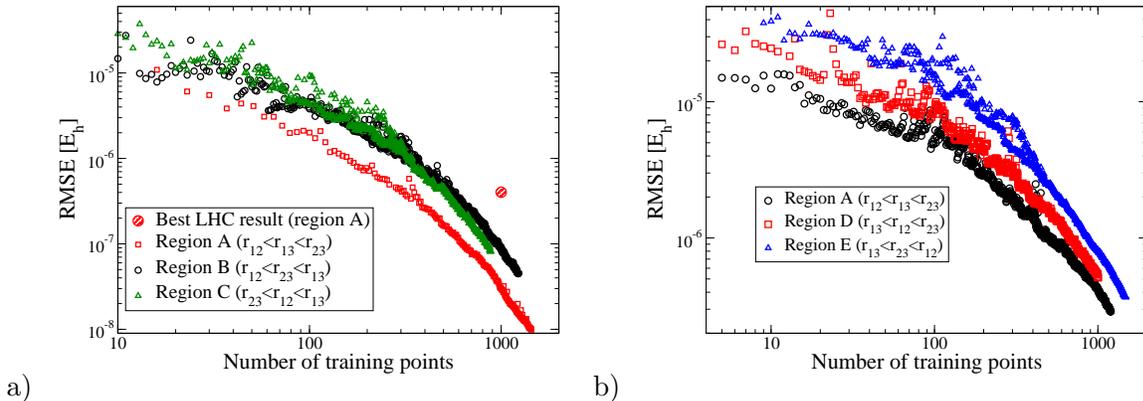


Figure S2: The reduction in RMSE with increasing training set size from sequential learning for $\text{CO}_2-(\text{Ar})_2$ (a) and $(\text{CO}_2)_2-\text{Ar}$ (b). The single point in fig (a) is the best RMSE obtained from LHC learning in region A.

	Reference set size	Transfer learning set size	RMSE of transfer set [E_h]	Percentage error
$\text{CO}_2-(\text{Ar})_2$: region A	5044	786	6.5×10^{-8}	0.5%
$\text{CO}_2-(\text{Ar})_2$: region B	8058	1069	6.8×10^{-8}	0.45%
$\text{CO}_2-(\text{Ar})_2$: region C	5056	909	1.1×10^{-7}	0.48%
$(\text{CO}_2)_2-\text{Ar}$: region A	6206	1000	4.2×10^{-7}	1.2%
$(\text{CO}_2)_2-\text{Ar}$: region D	5115	1000	5.1×10^{-7}	2.3%
$(\text{CO}_2)_2-\text{Ar}$: region E	5000	1400	4.2×10^{-7}	1.1%

Table S5: The reference and training set sizes, along with the resulting RMSE for the systems that were modelled by sequential learning. The percentage errors are expressed as a percentage of the root mean square value of the initial reference set.

asymptotic function at short range [Uteva *et al.* (2017)]. At short-range the high interaction energy of the binary interactions means these configurations are very rare in a thermal ensemble and contribute only weakly to equilibrium properties. For long-range interactions Uteva *et al.* (2017) and Broad *et al.* (2021) used a truncated multipole expansion of the interaction energy from intermolecular perturbation theory for two-body interactions. However, this approach is not readily generalised to three-body interactions, particular when two molecules are close, but the third is distant from both others. Instead, we use carefully chosen interaction calculations to characterise an empirical form for the asymptotic expansion, by generalising the empirical approach of Broad *et al.* (2021).

6.1 Two-body systems

For interactions where atoms are closer than the geometric constraint we use the same form as Uteva *et al.* (2017), namely

$$E = E_{\max} \frac{1}{N} \sum_{i=1}^N (r_{\min}/r_i)^{12}, \quad (2)$$

where E_{\max} is the largest energy in the test set before applying the high energy cut-off and r_{\min} is the shortest distance allowed by the geometric constraint. The values of E_{\max} for $(\text{Ar})_2$, $\text{CO}_2\text{-Ar}$ and $(\text{CO}_2)_2$, respectively, are 0.027, 0.12 and $0.52E_h$. For the long range asymptotic function, multipole series were employed for all systems. The contributions included for $(\text{Ar})_2$ and $\text{CO}_2\text{-Ar}$ were dipole and polarisability. For $(\text{CO}_2)_2$ we used the asymptotic function developed by Broad *et al.* (2021).

6.2 Three-body systems

For a three-body system, short range configurations are already made sufficiently rare by the repulsive part of the two-body interactions. Hence, for configurations below the three-body geometric cut-off we set the non-additive potential to zero.

	$r_{\min}[\text{\AA}]$	$r_{\text{short}}^{\max}[\text{\AA}]$	$r_{\text{medium}}^{\max}[\text{\AA}]$	$r_{\text{long}}^{\max}[\text{\AA}]$
$(\text{Ar})_3$	2.87	5.5	5.5	8.5
$\text{CO}_2\text{-(Ar)}_2$: region A	2.0	4.5	5.5	9.3
$\text{CO}_2\text{-(Ar)}_2$: region B	2.0	4.5	5.5	9.3
$\text{CO}_2\text{-(Ar)}_2$: region C	2.0	4.5	5.5	8.5
$(\text{CO}_2)_2\text{-Ar}$: region A	2.0	4.5	5.0	9.3
$(\text{CO}_2)_2\text{-Ar}$: region D	2.0	4.5	5.0	8.5
$(\text{CO}_2)_2\text{-Ar}$: region E	2.0	4.5	5.0	8.5

Table S6: Parameters to define the slide and scale values used to generate the empirical long range asymptotic function for each of the three-body systems.

For the long range asymptotic behaviour of the non-additive potential we combine our existing calculations with an empirical power-law behaviour. This approach involves scaling the overall configuration and/or sliding the most distant molecule into the GP region to compute the potential, before extrapolating to the required long-range potential via an empirical power law. The threshold values for this scale and slide are shown in table S6. Although the same general concepts are used for each system, as the systems became increasingly complex we needed to refine the method. Hence we describe each system in turn below. In each case we verified the asymptotic function against new test data outside the geometric constraint and the errors were comparable to or smaller than the corresponding GP RMSE values in tables S4 and S5.

6.2.1 $(\text{Ar})_3$

For a given long range configuration we begin by permuting the distances so that $r_{12} < r_{13} < r_{23}$. Next we consider the line connecting the midpoint of the Ar1-Ar2 line and Ar3, which makes the angle $\tilde{\theta}$ with the Ar1-Ar2 line (as shown in figure S3). We assume that, for sufficiently large atomic separations, the non-additive potential, E , obeys

$$E = \frac{A(\tilde{\theta})}{r_{12}^3 r_{13}^3 r_{23}^3}. \quad (3)$$

We obtain A for the particular configuration by mapping to the edge of the GP region, without changing $\tilde{\theta}$. This is achieved through the following sequence of operations:

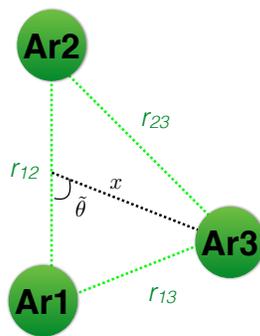


Figure S3: Slide geometry

1. If $r_{12} > r_{\text{short}}^{\text{max}}$, scale down all three distances so that $r_{12} = r_{\text{short}}^{\text{max}}$.
2. If $r_{13} > r_{\text{medium}}^{\text{max}}$, slide Ar3 towards Ar1 and 2 at fixed $\tilde{\theta}$ until $r_{13} = r_{\text{medium}}^{\text{max}}$.
3. If $r_{23} > r_{\text{long}}^{\text{max}}$ scale down all distances until $r_{23} = r_{\text{long}}^{\text{max}}$.

Evaluation of the GP at this transformed point provides the potential, from which $A(\tilde{\theta})$ can be obtained from equation (3). With $A(\tilde{\theta})$ determined, the potential at the original point can be computed via equation (3).

6.2.2 $\text{CO}_2-(\text{Ar})_2$

The long-range behaviour for $\text{CO}_2-(\text{Ar})_2$ uses a similar scale and slide approach to $(\text{Ar})_3$, but with the power-law exponent for the slide determined from the main GP. When performing a slide we always slide the most distant molecule towards the mid-point of the other two molecules, as shown in figure S3. We define $r_{\text{short/medium/large}}$ to be the shortest, middle and largest centre-to-centre distances, according to the corresponding region. We denote $\tilde{\theta}$ as the angle between the line joining the centres of the two closest molecules and the line joining the centre of the furthest molecule to the midpoint of the other two molecules. All scaling and sliding is performed with fixed α_1 , β_1 and $\tilde{\theta}$.

For sufficiently widely separated configurations, we assume that after a scale of all three interatomic distances then $E_0 = \lambda^9 E'$, where λ is the scaling ratio, E_0 is the original potential energy and E' is the potential after the scale. We also assume a slide changes the non-additive potential according to

$$\tilde{E} = E \left(\frac{r_{\text{medium2}}}{r_{\text{medium1}}} \right)^\gamma, \quad (4)$$

where E and \tilde{E} are the energies before and after the slide, γ is an exponent, which depends on $\tilde{\theta}$, r_{12} , α_1 and β_1 , and r_{medium1} and r_{medium2} are the middle centre-to-centre distances before and after the slide, respectively.

For a given configuration beyond the geometric constraint, \mathbf{r} , we map to two configurations within the geometric constraint and compute both energies using the GP model. The first configuration is obtained via:

- If $r_{\text{short}} > r_{\text{short}}^{\text{max}}$, scale down all three centre-to-centre distances until $r_{\text{short}} = r_{\text{short}}^{\text{max}}$.
- If, after this scaling, $r_{\text{medium}} > r_{\text{medium}}^{\text{max}}$ then propose a slide until $r_{\text{medium}} = r_{\text{medium}}^{\text{max}}$. If, after this slide, $r_{\text{large}} < r_{\text{large}}^{\text{max}}$ then accept this slide.
- Otherwise define $\Delta r = r_{\text{large}}^{\text{proposed}} - r_{\text{large}}^{\text{max}}$; reject the scale and slide above; scale down until $r_{\text{short}} = r_{\text{short}}^{\text{proposed}} - \Delta r/2$; then slide until $r_{\text{long}} = r_{\text{long}}^{\text{max}}$.

Thus the scale produces the set of interatomic distances \mathbf{r}' , which have a scaling of size λ and an interaction E' . Furthermore, the slide produces the interatomic distances \mathbf{r}'' and a non-additive potential of E'' , which can be evaluated through the GP.

The second GP configuration results from further slide of \mathbf{r}'' , at fixed $\tilde{\theta}$, such that the distance x (see figure S3) is reduced by a further 0.5\AA . This results in interatomic distances \mathbf{r}''' and interaction potential E''' . The non-additive potentials from the two slides, E'' and E''' , can be computed from the GP and then used to obtain the exponent γ by substituting into equation (4) and rearranging for γ . If the resulting γ is smaller than 4.5 then we assume $\gamma = 4.5$, to guarantee that the non-additive potential falls off sufficiently quickly at wide atomic separations to ensure that the virial integrals converge. We now have that $E_0 = \lambda^{-9}E'$ and $E' = E'' \left(\frac{r''_{\text{medium}}}{r'_{\text{medium}}} \right)^\gamma$. Combining these gives

$$E_0 = E'' \lambda^{-9} \left(\frac{r''_{\text{medium}}}{r'_{\text{medium}}} \right)^\gamma \quad (5)$$

6.2.3 $(\text{CO}_2)_2\text{-Ar}$

For $(\text{CO}_2)_2\text{-Ar}$ we follow the same scaling and sliding algorithm as used for $\text{CO}_2\text{-(Ar)}_2$ for the first transformation to produce \mathbf{r}'' and E'' . However, in this case the target distances for the scales and slides are defined as the shortest interatomic distance between the pair of molecules that comprise the distance being targeted. As with $\text{CO}_2\text{-(Ar)}_2$, the scale produces interatomic distances, \mathbf{r}' and scale factor λ , while the slide produces interatomic distances \mathbf{r}'' , with the interaction potential E'' .

To characterise the slide behaviour, instead of making a second deeper slide into the GP region (as with $\text{CO}_2\text{-(Ar)}_2$) we employ an auxiliary GP specifically for wide separations, in each region. These auxiliary GPs were produced and verified in the same way as the main GPs, but have r_{medium} held fixed at a reference value, $r_{\text{reference}}$, with all other geometric constraints being dropped. Details of the auxiliary GP for each region are given in table S7. Having obtained a configuration at the edge of the main GP by following the slide/scale algorithm for $\text{CO}_2\text{-(Ar)}_2$ (with scale factor λ , interatomic distances \mathbf{r}'' and non-additive potential E''), we then produce a configuration in the auxiliary GP region, \mathbf{r}''' , by a further slide so that $r_{\text{medium}} = r_{\text{reference}}$; the non-additive potential of this configuration, E''' , can then be computed from the auxiliary GP.

We now assume that, at wide separations, the non-additive potential obeys

$$E(r) = Ar_m^\alpha + Br_m^\beta \quad (6)$$

where A and B are constant coefficients, α and β are constant exponents, and r_m is the shortest interatomic distance between atoms in the pair of molecules with the medium separation. We

	Region A	Region D	Region E
$r_{\text{reference}} [\text{\AA}]$	7.4	8.5	7.4
$N_{\text{reference}}$	1000	750	800
N_{transfer}	250	300	400
RMSE [E_h]	5.9×10^{-9}	1.5×10^{-8}	2.3×10^{-8}

Table S7: Characteristics of the auxiliary GP for $(\text{CO}_2)_2\text{-Ar}$: $r_{\text{reference}}$ is the fixed reference value of the centre-to-centre distance of the pair of molecules with the medium separation; $N_{\text{reference}}$ is the size of the LHC reference set used for active learning and to compute the RMSE; N_{transfer} is the number of training points used in transfer learning to produce the auxiliary GP; and RMSE is the error obtained from the GP from active learning with N_{transfer} training points when predicting the reference set.

take values of $\alpha = -4$ and $\beta = -6$, except for region A where $\alpha = -6$ and $\beta = -7$. We have a pair of simultaneous equations $E(r''_m) = E''$ and $E(r'''_m) = E'''$, which is linear in the coefficients A and B , and so can be solved for these coefficients. Hence E_0 can be obtained by evaluating equation (6) and taking $E_0 = \lambda^9 E(r'_m)$.

7 Total computational costs

Table S8 breaks down the total CPU time to construct our three new non-additive PES, for calculations run on the University of Nottingham HPC cluster. This includes the calculations at both moderate accuracy and accuracy calculations. In each case the CPU effort was divided roughly evenly between a large number of moderate accuracy calculations required for the reference set and a smaller number of high accuracy calculations for the final PES. If it was necessary to upgrade all moderate accuracy calculations to high accuracy, as would be required for traditional parametric fitting, the costs increase by a factor of 3.5-6. The memory requirement even for the most expensive $(\text{CO}_2)_2\text{-Ar}$ calculations is relatively low (< 0.5 GB).

8 First principles predictions

8.1 Virial EoS

The pressure-density behaviour of a fluid can be written as an expansion in powers of the density

$$p = RT(\rho_m + B_2(T)\rho_m^2 + B_3(T)\rho_m^3 + \dots) \quad (7)$$

where p is pressure, T is temperature, R is the universal gas constant, ρ_m is the molar density and $B_i(T)$ is the i^{th} virial coefficient. For binary mixtures the virial coefficients are given by

$$B_n(T) = \sum_{j=0}^n {}^n C_j \phi_X^j \phi_Y^{n-j} B_{jX/(n-j)Y}^n, \quad (8)$$

where ${}^n C_j$ is the binomial coefficient, $\phi_{A/B}$ is the mol fraction of species A/B and $B_{jX/(n-j)Y}^n$ is the n th virial coefficient involving j molecules of type X and $n - j$ molecules of type Y.

(Ar)₃

	Total number of calculations	CPU time per calculation [hrs]	Total CPU time [hrs]
Moderate accuracy	5398	0.12	648
High accuracy	337	1	337
Total			985

CO₂-(Ar)₂

	Total number of calculations	CPU time per calculation [hrs]	Total CPU time [hrs]
Moderate accuracy	18158	2	36,316
High accuracy	2764	17	46,988
Total			83,304

(CO₂)₂-Ar

	Total number of calculations	CPU time per calculation [hrs]	Total CPU time [hrs]
Moderate accuracy	18871	10	188,710
High accuracy	4350	115	500,250
Total			688,960

Table S8: The total cost of ab-initio calculations for each non-additive PES created in this work.

We computed virial coefficients and their uncertainties by integrating over the PES, using the Monte Carlo methods of Wheatley *et al.* (2020). This also allows us to compute temperature gradients analytically, rather than via finite differences. For the (Ar)₂ pair potential and non-additive potential for (CO₂)₃ we used a potential from the literature [Patkowski and Szalewicz (2010); Hellmann (2017)] and used our newly obtained GP PES everywhere else. Thus we compute all cross virial coefficients, $B_{jX/(n-j)Y}^n$, required for CO₂-Ar mixtures, up to and including B_5 , along with B_{2-4} for pure CO₂. We took the remaining coefficients for pure Ar and pure CO₂ from the literature [Jäger *et al.* (2011); Hellmann (2017)]. The first and second temperature derivatives of each coefficient also are required for the speed of sound and Joule-Thompson predictions below. We computed these, and their uncertainties, by analytic differentiation of the virial coefficient integrand, before integrating as above. Our new results are tabulated in tables S9-S18. Uncertainties are generally low for the lower virials, but grow for higher virials, particularly at lower temperatures.

Virial coefficients have previously been computed, by Hellmann and coworkers, up to and including B_7 for both pure Ar [Jäger *et al.* (2011)] and pure CO₂ [Hellmann (2017)] using non-additive three-body potentials. We obtain temperature derivatives from these by differentiating the fitting formulae given by Hellman *et al.* Uncertainties for the coefficients are given by Hellman *et al.* and we estimate uncertainties in the temperature derivatives by using the finite difference expressions for the derivatives with $\Delta T = 3K$ and assuming uncorrelated uncertainties in the coefficient for this ΔT . We confirmed that, for pure CO₂, our calculations are consistent with results from Hellmann (2017), within uncertainties, for the coefficient and their temperature

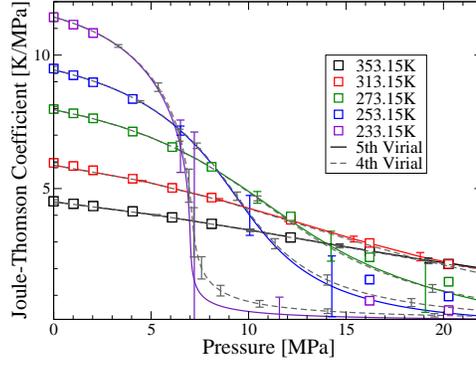


Figure S4: First-principles predictions for Joule-Thomson coefficient over an extended pressure range, compared to measurements Strakey *et al.* (1974) at $\phi_{\text{CO}_2} = 0.464$.

derivatives. We also confirmed that the temperature gradient uncertainties we estimated from Hellmann’s results are comparable to or somewhat larger than those obtained from our direct integration. We did not repeat the pure Ar virial calculations as the uncertainties quoted by Jäger *et al.* (2011) are extremely small ($< 0.1\%$). Thus we used our own calculations for pure CO_2 virials B_{2-4} and Hellmann *et al.*’s results for the CO_2 B_5 and for Ar B_{2-5} . All results presented herein are essentially unchanged if we swap Hellmann’s pure CO_2 virial results for our own integrals as the values are consistent and the pure CO_2 contribution to the uncertainties is small compared to that from the cross virials.

8.2 Calculating the speed of sound and Joule-Thomson coefficient

The virial EoS can be used to make first principle predictions of the speed of sound and Joule-Thomson coefficient. The speed of sound, c , can be rewritten as [El Hawary *et al.* (2019)]

$$c^2 = \frac{1}{M} \left(\frac{\partial p}{\partial \rho_m} \right)_T + \frac{T}{\rho^2 c_\nu} \left(\frac{\partial p}{\partial T} \right)_\rho^2, \quad (9)$$

where M is the molar mass of the mixture and c_ν is the isochoric heat capacity, given by

$$c_\nu = c_\nu^0 - \frac{T}{M} \int_0^{\rho_m} \frac{1}{\rho_m^2} \left(\frac{\partial^2 p}{\partial T^2} \right)_{\rho_m} d\rho_m, \quad (10)$$

where c_ν^0 is the ideal gas isochoric heat capacity.

The Joule-Thomson coefficient, μ measures rate of temperature loss with reduction in pressure. It is given by [Hirose *et al.* (1990)]

$$\mu = \frac{1}{c_p \rho_m} \left[\frac{T \left(\frac{\partial p}{\partial T} \right)_{\rho_m}}{\rho_m \left(\frac{\partial p}{\partial \rho_m} \right)_T} - 1 \right], \quad (11)$$

where c_p is the isobaric specific heat capacity, which can be obtained from

$$c_p - c_\nu = \frac{T}{\rho_m^2} \left(\frac{\partial p}{\partial T} \right)_{\rho_m}^2 / \left(\frac{\partial p}{\partial \rho_m} \right)_T. \quad (12)$$

Temperature derivatives of the pressure were computed by differentiating equation (7) and using the temperature derivatives of the virial coefficients obtained from direct integration of the PES (see tables S9-S18).

8.3 Extended Joule-Thomson plot

Figure S4 plots the Joule-Thomson coefficient prediction across the full pressure range for both the 4th and 5th order virial equation of state in fig B below. Firstly, the uncertainties are much lower for the 4th order virial model, confirming that the large uncertainties are due to uncertainties in the 5th virial coefficients. Secondly, uncertainties in the liquid region are too large to allow meaningful prediction here, confirming the expectation that the virial model is unsuitable for the liquid region.

9 Tabulated virial calculations

CO ₂ -Ar						
Temp	B ₂	Uncertainty	$\frac{dB_2}{d\beta}$	Uncertainty	$\frac{d^2B_2}{d\beta^2}$	Uncertainty
K	cm ³ /mol		cm ³ K/mol		cm ³ K ² /mol	
150	-1.73611e+02	1.8e-02	-5.0711e+04	3.3e+00	-3.7680e+06	2.9e+02
160	-1.53119e+02	1.6e-02	-4.7688e+04	3.1e+00	-3.4918e+06	2.6e+02
170	-1.36049e+02	1.5e-02	-4.5203e+04	3.0e+00	-3.2699e+06	2.4e+02
180	-1.21619e+02	1.4e-02	-4.3126e+04	2.9e+00	-3.0888e+06	2.2e+02
190	-1.09269e+02	1.4e-02	-4.1364e+04	2.8e+00	-2.9393e+06	2.1e+02
210	-8.92541e+01	1.2e-02	-3.8535e+04	2.6e+00	-2.7098e+06	1.9e+02
230	-7.37526e+01	1.1e-02	-3.6360e+04	2.5e+00	-2.5460e+06	1.7e+02
250	-6.14086e+01	1.0e-02	-3.4631e+04	2.4e+00	-2.4276e+06	1.6e+02
270	-5.13580e+01	9.7e-03	-3.3219e+04	2.4e+00	-2.3421e+06	1.5e+02
288	-4.37929e+01	9.2e-03	-3.2148e+04	2.3e+00	-2.2864e+06	1.5e+02
290	-4.30244e+01	9.1e-03	-3.2038e+04	2.3e+00	-2.2813e+06	1.5e+02
293	-4.18961e+01	9.1e-03	-3.1878e+04	2.3e+00	-2.2740e+06	1.4e+02
298	-4.00780e+01	8.9e-03	-3.1618e+04	2.3e+00	-2.2627e+06	1.4e+02
303	-3.83341e+01	8.8e-03	-3.1368e+04	2.3e+00	-2.2525e+06	1.4e+02
308	-3.66600e+01	8.7e-03	-3.1127e+04	2.3e+00	-2.2433e+06	1.4e+02
310	-3.60090e+01	8.6e-03	-3.1033e+04	2.3e+00	-2.2399e+06	1.4e+02
313	-3.50516e+01	8.6e-03	-3.0895e+04	2.3e+00	-2.2351e+06	1.4e+02
318	-3.35053e+01	8.5e-03	-3.0670e+04	2.3e+00	-2.2279e+06	1.4e+02
323	-3.20176e+01	8.4e-03	-3.0454e+04	2.2e+00	-2.2216e+06	1.4e+02
330	-3.00272e+01	8.2e-03	-3.0163e+04	2.2e+00	-2.2141e+06	1.4e+02
350	-2.48705e+01	7.8e-03	-2.9398e+04	2.2e+00	-2.2012e+06	1.3e+02
370	-2.03827e+01	7.5e-03	-2.8719e+04	2.2e+00	-2.1993e+06	1.3e+02

Table S9: 2nd virial values and temperature gradients from direct integration of our PES; $\beta = 1/T$.

CO ₂ -2Ar						
Temp	B ₃	Uncertainty	$\frac{dB_3}{d\beta}$	Uncertainty	$\frac{d^2B_3}{d\beta^2}$	Uncertainty
K	cm ⁶ /mol ²		cm ⁶ K/mol ²		cm ⁶ K ² /mol ²	
150	3.30881e+03	3.2e+00	-2.5090e+05	2.0e+03	-6.2379e+08	6.1e+05
160	3.31786e+03	2.6e+00	1.7868e+05	1.5e+03	-4.1637e+08	4.8e+05
170	3.20267e+03	2.1e+00	4.3057e+05	1.2e+03	-2.7449e+08	3.8e+05
180	3.03641e+03	1.7e+00	5.7613e+05	1.0e+03	-1.7464e+08	3.2e+05
190	2.85517e+03	1.5e+00	6.5648e+05	8.6e+02	-1.0268e+08	2.7e+05
210	2.50879e+03	1.1e+00	7.1016e+05	6.4e+02	-1.0276e+07	2.0e+05
230	2.21625e+03	9.2e-01	6.9553e+05	5.1e+02	4.2519e+07	1.5e+05
250	1.98085e+03	7.7e-01	6.5443e+05	4.2e+02	7.3852e+07	1.3e+05
270	1.79403e+03	6.7e-01	6.0474e+05	3.6e+02	9.2771e+07	1.1e+05
288	1.65922e+03	6.0e-01	5.5923e+05	3.2e+02	1.0324e+08	9.3e+04
290	1.64588e+03	5.9e-01	5.5426e+05	3.1e+02	1.0412e+08	9.2e+04
293	1.62644e+03	5.8e-01	5.4687e+05	3.1e+02	1.0536e+08	9.0e+04
298	1.59548e+03	5.6e-01	5.3469e+05	3.0e+02	1.0720e+08	8.8e+04
303	1.56620e+03	5.5e-01	5.2273e+05	2.9e+02	1.0879e+08	8.5e+04
308	1.53851e+03	5.4e-01	5.1100e+05	2.8e+02	1.1015e+08	8.2e+04
310	1.52785e+03	5.3e-01	5.0638e+05	2.8e+02	1.1064e+08	8.1e+04
313	1.51230e+03	5.2e-01	4.9951e+05	2.7e+02	1.1131e+08	8.0e+04
318	1.48749e+03	5.1e-01	4.8828e+05	2.7e+02	1.1228e+08	7.8e+04
323	1.46399e+03	5.0e-01	4.7731e+05	2.6e+02	1.1309e+08	7.6e+04
330	1.43313e+03	4.9e-01	4.6240e+05	2.5e+02	1.1396e+08	7.3e+04
350	1.35649e+03	4.5e-01	4.2268e+05	2.3e+02	1.1509e+08	6.6e+04
370	1.29396e+03	4.2e-01	3.8717e+05	2.1e+02	1.1467e+08	6.1e+04

Table S10: 3rd virial values and temperature gradients from direct integration of our PES; $\beta = 1/T$.

2CO ₂ -Ar						
Temp	B ₃	Uncertainty	$\frac{dB_3}{d\beta}$	Uncertainty	$\frac{d^2B_3}{d\beta^2}$	Uncertainty
K	cm ⁶ /mol ²		cm ⁶ K/mol ²		cm ⁶ K ² /mol ²	
150	-5.02098e+03	3.1e+01	-1.6339e+07	3.1e+04	-1.1160e+10	1.6e+07
160	1.16424e+02	2.1e+01	-8.8979e+06	2.0e+04	-6.9914e+09	1.0e+07
170	2.56417e+03	1.5e+01	-4.7188e+06	1.4e+04	-4.5303e+09	7.0e+06
180	3.68139e+03	1.1e+01	-2.2844e+06	1.0e+04	-3.0058e+09	5.1e+06
190	4.12255e+03	8.2e+00	-8.2884e+05	7.8e+03	-2.0231e+09	3.8e+06
210	4.13509e+03	5.2e+00	5.9520e+05	4.8e+03	-9.1911e+08	2.3e+06
230	3.76575e+03	3.6e+00	1.1143e+06	3.3e+03	-3.7790e+08	1.6e+06
250	3.34527e+03	2.6e+00	1.2703e+06	2.4e+03	-9.1433e+07	1.1e+06
270	2.96602e+03	2.1e+00	1.2739e+06	1.8e+03	6.8605e+07	8.5e+05
288	2.67648e+03	1.7e+00	1.2212e+06	1.5e+03	1.5401e+08	6.8e+05
290	2.64732e+03	1.7e+00	1.2136e+06	1.4e+03	1.6121e+08	6.7e+05
293	2.60468e+03	1.6e+00	1.2019e+06	1.4e+03	1.7133e+08	6.5e+05
298	2.53643e+03	1.6e+00	1.1814e+06	1.3e+03	1.8651e+08	6.2e+05
303	2.47160e+03	1.5e+00	1.1600e+06	1.3e+03	1.9981e+08	5.8e+05
308	2.41003e+03	1.4e+00	1.1379e+06	1.2e+03	2.1148e+08	5.6e+05
310	2.38629e+03	1.4e+00	1.1290e+06	1.2e+03	2.1573e+08	5.5e+05
313	2.35159e+03	1.4e+00	1.1155e+06	1.2e+03	2.2169e+08	5.3e+05
318	2.29612e+03	1.3e+00	1.0927e+06	1.1e+03	2.3063e+08	5.1e+05
323	2.24348e+03	1.3e+00	1.0699e+06	1.1e+03	2.3843e+08	4.9e+05
330	2.17426e+03	1.2e+00	1.0380e+06	1.0e+03	2.4770e+08	4.6e+05
350	2.00216e+03	1.1e+00	9.4879e+05	8.7e+02	2.6582e+08	3.9e+05
370	1.86206e+03	9.6e-01	8.6509e+05	7.6e+02	2.7514e+08	3.4e+05

Table S11: 3rd virial values and temperature gradients from direct integration of our PES; $\beta = 1/T$.

CO ₂ -3 Ar						
Temp	B ₄	Uncertainty	$\frac{dB_4}{d\beta}$	Uncertainty	$\frac{d^2B_4}{d\beta^2}$	Uncertainty
K	cm ⁹ /mol ³		cm ⁹ K/mol ³		cm ⁹ K ² /mol ³	
150	7.98431e+04	2.0e+03	-6.8765e+06	1.9e+06	-7.3262e+10	9.5e+08
160	7.30294e+04	1.3e+03	3.3240e+07	1.3e+06	-2.7393e+10	6.4e+08
170	5.82790e+04	9.5e+02	4.4230e+07	9.2e+05	-4.6818e+09	4.5e+08
180	4.37822e+04	7.0e+02	4.3283e+07	6.8e+05	6.4447e+09	3.3e+08
190	3.18505e+04	5.3e+02	3.7830e+07	5.1e+05	1.1596e+10	2.5e+08
210	1.61335e+04	3.3e+02	2.4498e+07	3.2e+05	1.3965e+10	1.5e+08
230	8.33109e+03	2.3e+02	1.3388e+07	2.2e+05	1.2539e+10	1.1e+08
250	5.10602e+03	1.7e+02	5.4211e+06	1.6e+05	1.0265e+10	7.6e+07
270	4.33649e+03	1.3e+02	-5.3918e+03	1.2e+05	8.0274e+09	5.8e+07
288	4.73596e+03	1.0e+02	-3.3084e+06	9.8e+04	6.2446e+09	4.7e+07
290	4.81873e+03	1.0e+02	-3.6031e+06	9.6e+04	6.0619e+09	4.6e+07
293	4.95339e+03	9.8e+01	-4.0217e+06	9.3e+04	5.7937e+09	4.4e+07
298	5.20222e+03	9.3e+01	-4.6604e+06	8.9e+04	5.3618e+09	4.2e+07
303	5.47630e+03	8.9e+01	-5.2313e+06	8.4e+04	4.9484e+09	4.0e+07
308	5.77040e+03	8.5e+01	-5.7403e+06	8.0e+04	4.5530e+09	3.8e+07
310	5.89262e+03	8.3e+01	-5.9278e+06	7.9e+04	4.3997e+09	3.7e+07
313	6.08003e+03	8.1e+01	-6.1930e+06	7.7e+04	4.1749e+09	3.6e+07
318	6.40136e+03	7.8e+01	-6.5942e+06	7.3e+04	3.8134e+09	3.5e+07
323	6.73112e+03	7.4e+01	-6.9486e+06	7.0e+04	3.4679e+09	3.3e+07
330	7.20175e+03	7.0e+01	-7.3739e+06	6.6e+04	3.0094e+09	3.1e+07
350	8.55705e+03	6.1e+01	-8.2127e+06	5.7e+04	1.8460e+09	2.7e+07
370	9.86156e+03	5.4e+01	-8.6302e+06	5.0e+04	8.6672e+08	2.3e+07

Table S12: 4th virial values and temperature gradients from direct integration of our PES; $\beta = 1/T$

2 CO ₂ –2 Ar						
Temp K	B ₄ cm ⁹ /mol ³	Uncertainty	$\frac{dB_4}{d\beta}$ cm ⁹ K/mol ³	Uncertainty	$\frac{d^2B_4}{d\beta^2}$ cm ⁹ K ² /mol ³	Uncertainty
150	7.83418e+04	1.9e+04	-8.7702e+08	2.7e+07	-1.2716e+12	2.0e+10
160	2.73720e+05	1.1e+04	-1.6321e+08	1.5e+07	-5.2857e+11	1.1e+10
170	2.79550e+05	6.5e+03	9.1653e+07	9.0e+06	-2.0127e+11	6.4e+09
180	2.34279e+05	4.2e+03	1.6933e+08	5.7e+06	-5.2820e+10	4.0e+09
190	1.82507e+05	2.9e+03	1.7827e+08	3.9e+06	1.4523e+10	2.7e+09
210	1.01578e+05	1.5e+03	1.3815e+08	2.0e+06	5.4444e+10	1.3e+09
230	5.39379e+04	9.0e+02	9.2051e+07	1.2e+06	5.4038e+10	7.7e+08
250	2.81424e+04	5.8e+02	5.7314e+07	7.5e+05	4.5149e+10	4.9e+08
270	1.48538e+04	4.0e+02	3.3314e+07	5.1e+05	3.5767e+10	3.3e+08
288	8.92708e+03	3.1e+02	1.8457e+07	3.9e+05	2.8468e+10	2.5e+08
290	8.50129e+03	3.0e+02	1.7111e+07	3.8e+05	2.7737e+10	2.4e+08
293	7.93128e+03	2.8e+02	1.5190e+07	3.6e+05	2.6669e+10	2.3e+08
298	7.14699e+03	2.7e+02	1.2234e+07	3.3e+05	2.4967e+10	2.1e+08
303	6.54444e+03	2.5e+02	9.5583e+06	3.1e+05	2.3357e+10	2.0e+08
308	6.09792e+03	2.3e+02	7.1372e+06	2.9e+05	2.1837e+10	1.9e+08
310	5.95791e+03	2.3e+02	6.2347e+06	2.9e+05	2.1253e+10	1.8e+08
313	5.78519e+03	2.2e+02	4.9468e+06	2.7e+05	2.0402e+10	1.8e+08
318	5.58703e+03	2.1e+02	2.9653e+06	2.6e+05	1.9048e+10	1.6e+08
323	5.48681e+03	1.9e+02	1.1733e+06	2.4e+05	1.7770e+10	1.5e+08
330	5.48397e+03	1.8e+02	-1.0503e+06	2.3e+05	1.6101e+10	1.4e+08
350	6.10656e+03	1.5e+02	-5.9043e+06	1.8e+05	1.2005e+10	1.2e+08
370	7.27801e+03	1.2e+02	-9.0972e+06	1.5e+05	8.7276e+09	9.5e+07

Table S13: 4th virial values and temperature gradients from direct integration of our PES; $\beta = 1/T$

3 CO ₂ –Ar						
Temp	B ₄	Uncertainty	$\frac{dB_4}{d\beta}$	Uncertainty	$\frac{d^2B_4}{d\beta^2}$	Uncertainty
K	cm ⁹ /mol ³		cm ⁹ K/mol ³		cm ⁹ K ² /mol ³	
150	-1.79496e+07	7.3e+05	-4.6266e+10	1.5e+09	-5.2294e+13	1.5e+12
160	-5.57863e+06	3.2e+05	-1.7356e+10	6.4e+08	-2.1402e+13	6.5e+11
170	-1.44005e+06	1.5e+05	-6.6248e+09	3.0e+08	-9.3195e+12	3.1e+11
180	-5.57544e+04	7.9e+04	-2.3996e+09	1.6e+08	-4.2079e+12	1.6e+11
190	3.62561e+05	4.4e+04	-6.8402e+08	8.8e+07	-1.9115e+12	8.8e+10
210	4.01121e+05	1.7e+04	2.6907e+08	3.3e+07	-3.1374e+11	3.3e+10
230	2.64171e+05	7.7e+03	3.4180e+08	1.5e+07	6.0419e+10	1.4e+10
250	1.56901e+05	4.0e+03	2.6659e+08	7.6e+06	1.3491e+11	7.3e+09
270	8.99361e+04	2.3e+03	1.8577e+08	4.3e+06	1.3203e+11	4.1e+09
288	5.37159e+04	1.5e+03	1.2861e+08	2.8e+06	1.1362e+11	2.7e+09
290	5.07009e+04	1.5e+03	1.2322e+08	2.7e+06	1.1136e+11	2.5e+09
293	4.64878e+04	1.4e+03	1.1548e+08	2.5e+06	1.0798e+11	2.4e+09
298	4.02230e+04	1.3e+03	1.0343e+08	2.3e+06	1.0238e+11	2.1e+09
303	3.48040e+04	1.1e+03	9.2396e+07	2.1e+06	9.6876e+10	1.9e+09
308	3.01266e+04	1.0e+03	8.2301e+07	1.9e+06	9.1535e+10	1.7e+09
310	2.84426e+04	1.0e+03	7.8510e+07	1.8e+06	8.9450e+10	1.7e+09
313	2.60997e+04	9.4e+02	7.3074e+07	1.7e+06	8.6383e+10	1.6e+09
318	2.26427e+04	8.6e+02	6.4644e+07	1.6e+06	8.1442e+10	1.4e+09
323	1.96851e+04	7.9e+02	5.6945e+07	1.4e+06	7.6724e+10	1.3e+09
330	1.62673e+04	7.1e+02	4.7279e+07	1.3e+06	7.0499e+10	1.2e+09
350	1.00364e+04	5.2e+02	2.5578e+07	9.4e+05	5.5089e+10	8.5e+08
370	7.29950e+03	4.0e+02	1.0497e+07	7.2e+05	4.2804e+10	6.4e+08

Table S14: 4th virial values and temperature gradients from direct integration of our PES; $\beta = 1/T$

CO ₂ -4 Ar						
Temp K	B ₅ cm ¹² /mol ⁴	Uncertainty	$\frac{dB_5}{d\beta}$ cm ¹² K/mol ⁴	Uncertainty	$\frac{d^2B_5}{d\beta^2}$ cm ¹² K ² /mol ⁴	Uncertainty
210	-1.11887e+06	1.4e+05	-5.5004e+08	1.9e+08	7.2246e+11	1.3e+11
230	-7.97884e+05	7.7e+04	-9.3150e+08	1.1e+08	2.2493e+11	7.4e+10
250	-4.59412e+05	4.7e+04	-9.8177e+08	6.8e+07	-5.9364e+10	4.7e+10
270	-1.78652e+05	3.1e+04	-8.9873e+08	4.6e+07	-2.0769e+11	3.2e+10
288	1.69091e+04	2.2e+04	-7.8585e+08	3.3e+07	-2.7347e+11	2.4e+10
290	3.55696e+04	2.2e+04	-7.7264e+08	3.2e+07	-2.7816e+11	2.3e+10
293	6.24994e+04	2.1e+04	-7.5277e+08	3.0e+07	-2.8441e+11	2.2e+10
298	1.04664e+05	1.9e+04	-7.1969e+08	2.8e+07	-2.9296e+11	2.0e+10
303	1.43611e+05	1.8e+04	-6.8687e+08	2.6e+07	-2.9948e+11	1.9e+10
308	1.79547e+05	1.6e+04	-6.5451e+08	2.4e+07	-3.0425e+11	1.7e+10
310	1.93123e+05	1.6e+04	-6.4173e+08	2.3e+07	-3.0572e+11	1.7e+10
313	2.12672e+05	1.5e+04	-6.2277e+08	2.2e+07	-3.0752e+11	1.6e+10
318	2.43178e+05	1.4e+04	-5.9176e+08	2.1e+07	-3.0953e+11	1.5e+10
323	2.71250e+05	1.3e+04	-5.6158e+08	1.9e+07	-3.1045e+11	1.4e+10
330	3.06791e+05	1.2e+04	-5.2080e+08	1.8e+07	-3.1023e+11	1.3e+10
350	3.87721e+05	9.8e+03	-4.1438e+08	1.4e+07	-3.0278e+11	1.0e+10
370	4.44592e+05	8.1e+03	-3.2277e+08	1.1e+07	-2.8954e+11	8.3e+09

Table S15: 5th virial values and temperature gradients from direct integration of our PES; $\beta = 1/T$.

2 CO ₂ -3 Ar						
Temp K	B ₅ cm ¹² /mol ⁴	Uncertainty	$\frac{dB_5}{d\beta}$ cm ¹² K/mol ⁴	Uncertainty	$\frac{d^2B_5}{d\beta^2}$ cm ¹² K ² /mol ⁴	Uncertainty
210	-1.96799e+06	6.7e+05	2.1294e+09	1.1e+09	4.7538e+12	9.9e+11
230	-2.18943e+06	3.4e+05	-7.0717e+08	5.7e+08	2.2114e+12	4.8e+11
250	-1.74190e+06	1.9e+05	-1.6958e+09	3.2e+08	7.4119e+11	2.7e+11
270	-1.19976e+06	1.2e+05	-1.8882e+09	2.0e+08	-2.4802e+10	1.6e+11
288	-7.71109e+05	8.3e+04	-1.7884e+09	1.4e+08	-3.7528e+11	1.1e+11
290	-7.28504e+05	8.0e+04	-1.7698e+09	1.3e+08	-4.0145e+11	1.1e+11
293	-6.66536e+05	7.5e+04	-1.7401e+09	1.3e+08	-4.3702e+11	1.0e+11
298	-5.68378e+05	6.9e+04	-1.6871e+09	1.1e+08	-4.8742e+11	9.3e+10
303	-4.76493e+05	6.3e+04	-1.6308e+09	1.0e+08	-5.2817e+11	8.5e+10
308	-3.90668e+05	5.8e+04	-1.5724e+09	9.6e+07	-5.6073e+11	7.8e+10
310	-3.57979e+05	5.6e+04	-1.5487e+09	9.3e+07	-5.7174e+11	7.6e+10
313	-3.10647e+05	5.3e+04	-1.5129e+09	8.8e+07	-5.8633e+11	7.2e+10
318	-2.36147e+05	4.9e+04	-1.4529e+09	8.2e+07	-6.0602e+11	6.7e+10
323	-1.66868e+05	4.5e+04	-1.3932e+09	7.6e+07	-6.2070e+11	6.2e+10
330	-7.80731e+04	4.1e+04	-1.3107e+09	6.8e+07	-6.3430e+11	5.6e+10
350	1.29701e+05	3.1e+04	-1.0887e+09	5.2e+07	-6.4155e+11	4.2e+10
370	2.82655e+05	2.4e+04	-8.9311e+08	4.1e+07	-6.2138e+11	3.3e+10

Table S16: 5th virial values and temperature gradients from direct integration of our PES; $\beta = 1/T$.

3 CO ₂ –2 Ar						
Temp	B ₅	Uncertainty	$\frac{dB_5}{d\beta}$	Uncertainty	$\frac{d^2B_5}{d\beta^2}$	Uncertainty
K	cm ¹² /mol ⁴		cm ¹² K/mol ⁴		cm ¹² K ² /mol ⁴	
150	-6.31681e+08	7.5e+08	-3.0939e+12	1.8e+12	-5.4312e+15	2.3e+15
160	1.72742e+07	2.7e+08	-5.6658e+11	6.6e+11	-1.4164e+15	8.1e+14
170	9.73989e+07	1.1e+08	1.7713e+08	2.7e+11	-3.3014e+14	3.3e+14
180	7.59159e+07	5.2e+07	9.9423e+10	1.2e+11	-3.1759e+13	1.5e+14
190	4.69822e+07	2.6e+07	9.1490e+10	6.2e+10	4.1695e+13	7.4e+13
210	1.32440e+07	8.2e+06	4.3380e+10	1.9e+10	4.2746e+13	2.2e+13
230	1.48910e+06	3.2e+06	1.6097e+10	7.3e+09	2.3432e+13	8.5e+12
250	-1.81319e+06	1.5e+06	4.2858e+09	3.3e+09	1.1405e+13	3.8e+12
270	-2.29057e+06	7.5e+05	-4.4097e+08	1.7e+09	5.0865e+12	1.9e+12
288	-1.97564e+06	4.5e+05	-2.0507e+09	1.0e+09	2.1069e+12	1.1e+12
290	-1.92537e+06	4.3e+05	-2.1460e+09	9.5e+08	1.8753e+12	1.1e+12
293	-1.84740e+06	3.9e+05	-2.2670e+09	8.8e+08	1.5565e+12	1.0e+12
298	-1.71300e+06	3.5e+05	-2.4182e+09	7.8e+08	1.0933e+12	8.8e+11
303	-1.57615e+06	3.1e+05	-2.5172e+09	6.9e+08	7.0370e+11	7.7e+11
308	-1.43960e+06	2.7e+05	-2.5746e+09	6.1e+08	3.7645e+11	6.9e+11
310	-1.38552e+06	2.6e+05	-2.5879e+09	5.8e+08	2.6082e+11	6.6e+11
313	-1.30531e+06	2.4e+05	-2.5991e+09	5.4e+08	1.0199e+11	6.1e+11
318	-1.17469e+06	2.2e+05	-2.5975e+09	4.9e+08	-1.2774e+11	5.5e+11
323	-1.04870e+06	2.0e+05	-2.5755e+09	4.4e+08	-3.1954e+11	4.9e+11
330	-8.81269e+05	1.7e+05	-2.5189e+09	3.8e+08	-5.3510e+11	4.2e+11
350	-4.65511e+05	1.2e+05	-2.2617e+09	2.6e+08	-9.0732e+11	2.9e+11
370	-1.39331e+05	8.5e+04	-1.9545e+09	1.9e+08	-1.0587e+12	2.1e+11

Table S17: 5th virial values and temperature gradients from direct integration of our PES; $\beta = 1/T$.

4 CO ₂ –Ar						
Temp	B ₅	Uncertainty	$\frac{dB_5}{d\beta}$	Uncertainty	$\frac{d^2B_5}{d\beta^2}$	Uncertainty
K	cm ¹² /mol ⁴		cm ¹² K/mol ⁴		cm ¹² K ² /mol ⁴	
150	-3.44887e+10	1.4e+10	-1.2306e+14	4.3e+13	-2.0089e+17	6.3e+16
160	-6.83925e+09	4.2e+09	-2.9623e+13	1.2e+13	-5.3510e+16	1.8e+16
170	-9.47008e+08	1.4e+09	-7.0263e+12	4.2e+12	-1.5213e+16	6.2e+15
180	2.20433e+08	5.4e+08	-1.2833e+12	1.6e+12	-4.3209e+15	2.4e+15
190	3.44577e+08	2.3e+08	1.2187e+11	6.8e+11	-1.0617e+15	1.0e+15
210	1.79461e+08	5.4e+07	3.4491e+11	1.6e+11	1.6871e+14	2.3e+14
230	7.00796e+07	1.7e+07	1.8327e+11	4.7e+10	1.7867e+14	6.9e+13
250	2.50219e+07	6.3e+06	8.4360e+10	1.8e+10	1.0594e+14	2.5e+13
270	7.76750e+06	2.8e+06	3.6901e+10	7.7e+09	5.7423e+13	1.1e+13
288	1.80942e+06	1.5e+06	1.6500e+10	4.1e+09	3.2457e+13	5.6e+12
290	1.43253e+06	1.4e+06	1.4994e+10	3.8e+09	3.0442e+13	5.3e+12
293	9.39908e+05	1.3e+06	1.2944e+10	3.5e+09	2.7645e+13	4.8e+12
298	2.84672e+05	1.1e+06	1.0019e+10	3.0e+09	2.3523e+13	4.1e+12
303	-2.01710e+05	9.5e+05	7.6133e+09	2.5e+09	1.9991e+13	3.5e+12
308	-5.55208e+05	8.3e+05	5.6367e+09	2.2e+09	1.6962e+13	3.0e+12
310	-6.66001e+05	7.8e+05	4.9491e+09	2.1e+09	1.5875e+13	2.8e+12
313	-8.04328e+05	7.2e+05	4.0147e+09	1.9e+09	1.4364e+13	2.6e+12
318	-9.71686e+05	6.3e+05	2.6858e+09	1.7e+09	1.2134e+13	2.3e+12
323	-1.07524e+06	5.6e+05	1.5997e+09	1.5e+09	1.0218e+13	2.0e+12
330	-1.13958e+06	4.7e+05	4.0888e+08	1.2e+09	7.9777e+12	1.7e+12
350	-1.01963e+06	3.0e+05	-1.5453e+09	7.9e+08	3.6508e+12	1.0e+12
370	-7.15187e+05	2.0e+05	-2.2747e+09	5.3e+08	1.2626e+12	6.9e+11

Table S18: 5th virial values and temperature gradients from direct integration of our PES; $\beta = 1/T$.

References

- Uteva, E., R. S. Graham, R. D. Wilkinson and R. J. Wheatley, “Interpolation of intermolecular potentials using Gaussian processes,” *J. Chem. Phys.* **147**, 161706 (2017).
- Uteva, E., R. S. Graham, R. D. Wilkinson and R. J. Wheatley, “Active learning in Gaussian process interpolation of potential energy surfaces,” *J. Chem. Phys.* **149**, 174114 (2018).
- Werner *et al.*, H. J., “MOLPRO version 2012.1,” <http://www.molpro.net> (2012).
- Rasmussen, C. and C. K. I. William, *Gaussian Processes for Machine Learning*, MIT Press (2006).
- GPY, “GPY: A Gaussian Process framework in python,” <http://github.com/SheffieldML/GPY> (2012–2015).
- Broad, J., S. Preston, R. J. Wheatley and R. S. Graham, “Gaussian Process models of potential energy surfaces with boundary optimization,” *J. Chem. Phys.* **155**, 144106 (2021).
- Patkowski, K. and K. Szalewicz, “Argon pair potential at basis set and excitation limits,” *J. Chem. Phys.* **133**, 094304 (2010).
- Hellmann, R., “Nonadditive three-body potential and third to eighth virial coefficients of carbon dioxide,” *J. Chem. Phys.* **146**, 054302 (2017).
- Wheatley, R. J., A. J. Schultz, H. Do, N. Gokul and D. A. Kofke, “Cluster integrals and virial coefficients for realistic molecular models,” *Phys. Rev. E* **101**, 051301 (2020).
- Jäger, B., R. Hellmann, E. Bich and E. Vogel, “Ab initio virial equation of state for argon using a new nonadditive three-body potential,” *J. Chem. Phys.* **135**, 084308 (2011).
- El Hawary, A., R. Hellmann, K. Meier and H. Busemann, “Eighth-order virial equation of state and speed-of-sound measurements for krypton,” *J. Chem. Phys.* **151**, 154303 (2019).
- Hirose, Y., T. Kitazawa and T. Yoshida, “The Joule-Thomson expansion coefficient by formula manipulation,” *Ind. Eng. Chem. Res.* **29**, 1555–1558 (1990).
- Strakey, J. P., C. O. Bennett and B. F. Dodge, “Joule-Thomson coefficients of argon-carbon dioxide mixtures,” *AIChE J.* **20**, 803–814 (1974).