Supplementary information to "Understanding the ML black box with simple descriptors to predict cluster-adsorbate interaction energy"

Sheena Agarwal,[†,‡] Shweta Mehta,[†,‡] and Kavita Joshi[*,†,‡]

† *Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pashan, Pune-411008, India.*

‡*Academy of Scientific and Innovative Research (AcSIR)*

E-mail: k.joshi@ncl.res.in;kavita.p.joshi@gmail.com

1. Machine Learning

1.1 Architecture of ML algorithm

Gradient boosting regression is a Tree based classifier technique. As the name suggests gradient descent along with boosting methods is applied. Gradient Descent is an iterative optimization algorithm to find the minimum of a function. And boosting is a method by which the performance of the model is improved by iteratively training weak learners on the errors made by its predecessors. Gradient boosting algorithms mainly involves three elements,

1. A loss function to be optimized,
2. A weak learner to make predictions, and
3. An additive model to add the weak learns to minimize the loss function.

After choosing a differentiable loss function $L(y, F(x))$, to be optimized a weak learner is to be selected. The weak learners are combined using an additive model of the form:

$$F(x) = \sum_{m=1}^{M} \gamma_m h_m(x)$$

where, $(x_i, y_i)_{i=1}^{n}$ is the training dataset, γ_m is the step length, $h_m(x)$ are the basis functions (aka weak learners) and M are the number of iterations. In case of gradient boosting, decision trees are used as weak learners. After calculating the loss, trees that reduce the loss are added to the model.

$$h_m = argmin_h \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + h(x_i))$$

The output for the new tree is then added to the output of the existing sequence of trees in an effort to correct or improve the final output of the model

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

To constrain the weak learners hyper parameters, such as a maximum number of layers, nodes, splits or leaf nodes are used. Finally, a gradient descent procedure is used as an additive model to minimize the loss while adding the trees. The steepest descent direction is the negative gradient of the loss function L evaluated at the current model F_{m-1} which can be calculated for any differentiable loss function:

$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^{n} \nabla_F L(y_i, F_{m-1}(x_i))$$

where the step length γ_m is chosen using line search:

$$\gamma_m = argmin_\gamma \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i)) - \gamma \frac{\partial L(y_i, F_m - 1(x_i))}{F_m - 1(x_i)}$$

A fixed number of trees are added or training stops once loss reaches an acceptable level or no longer improves on an external validation dataset.

1.2 MAE vs Interatomic correlation

The correlation of variation of Mean Absolute errors (MAE) as a function of nearest neighbor distances, as explained in detail for Al_{13} in the paper, is shown here for a few other representative cases like Al_5 , Al_7, and Al_9. We see the same correlation for all cases, i.e. reducing MAE with increasing system representation. Also, jumps in error plots are observed at places wherein differentiating information regarding the atomic arrangement is added in the model. Demonstrating the correlation between MAE and nearest neighbor distances.

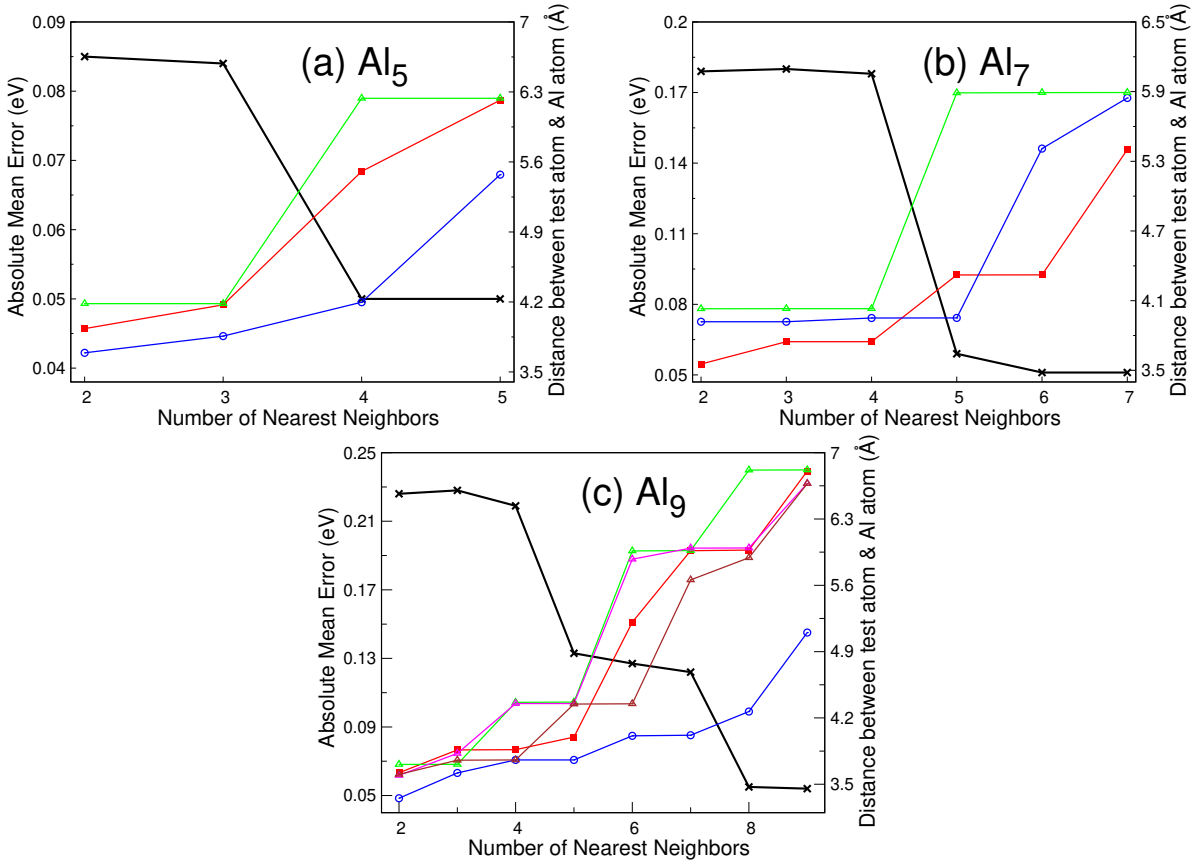

Figure 1: Mean absolute errors (on left y1 axis) is plotted in black line as a function of Number of Nearest neighbors included a descriptors. The nearest neighbor distrobution lines are shown in colored lines with distances between adsorbate and Al atom on the left y2 axis.

2. DFT

'Orderedness' of clusters: As evident from the figure, the interatomic distances could also be used asindicators of an 'ordered' cluster. By 'ordered' cluster we mean a cluster with many identical atoms in terms of the chemical environment that they experience. For example, Al_{13} , Al_{36} , and Al_{75} are 'ordered' clusters because all the surface atoms are grouped into 2 (for Al_{13}),6 (for Al_{36}), and 13 (for Al_{75}) classes whereas for all the disordered clusters, more than half of surface atoms experience unique environment. For example, in case of Al_{67} there are 34 distinct classes/atoms on the surface which also reflects into the interaction energy patterns of that cluster towards an adsorbate.

For clarity we show enlarged pictures of one disordered (Al_{55}) and one ordered (Al_{75}) cluster. Interaction energy curves and Nearest neighbor distribution plots for larger clusters i.e. Al_{55} and Al_{75} are shown. The one to one correlation between nearest neighbor distances and IE holds.

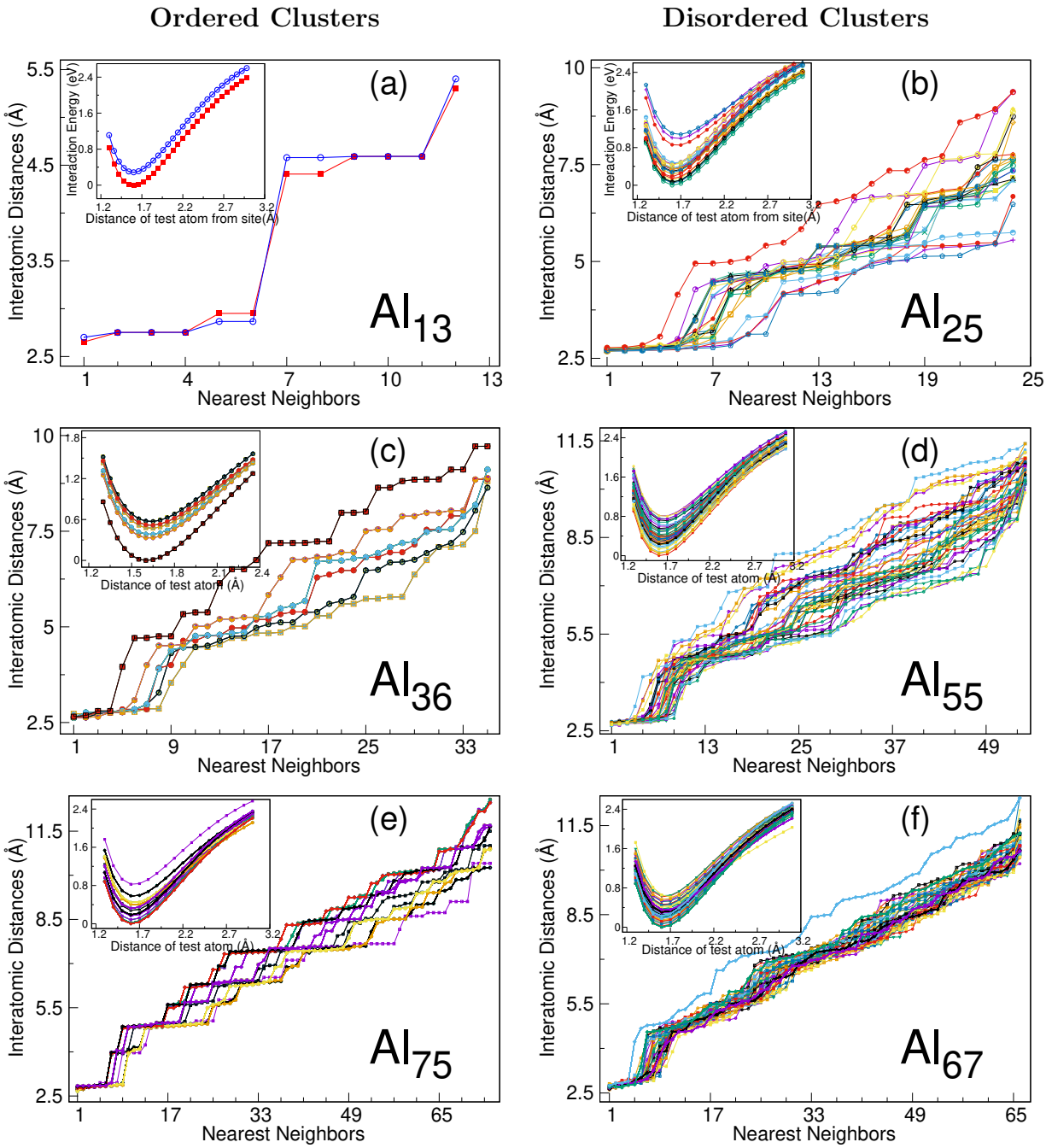

Figure 2: The figure shows variation in the interatomic distances as function of nearest neighbors for all surface atoms of few representative 'ordered' and 'disordered' clusters. The inset figure shows distance dependent interaction energy of all surface atoms of these clusters. Atoms with identical nearest neighbor distribution also exhibit identical interaction energy pattern towards a test atom

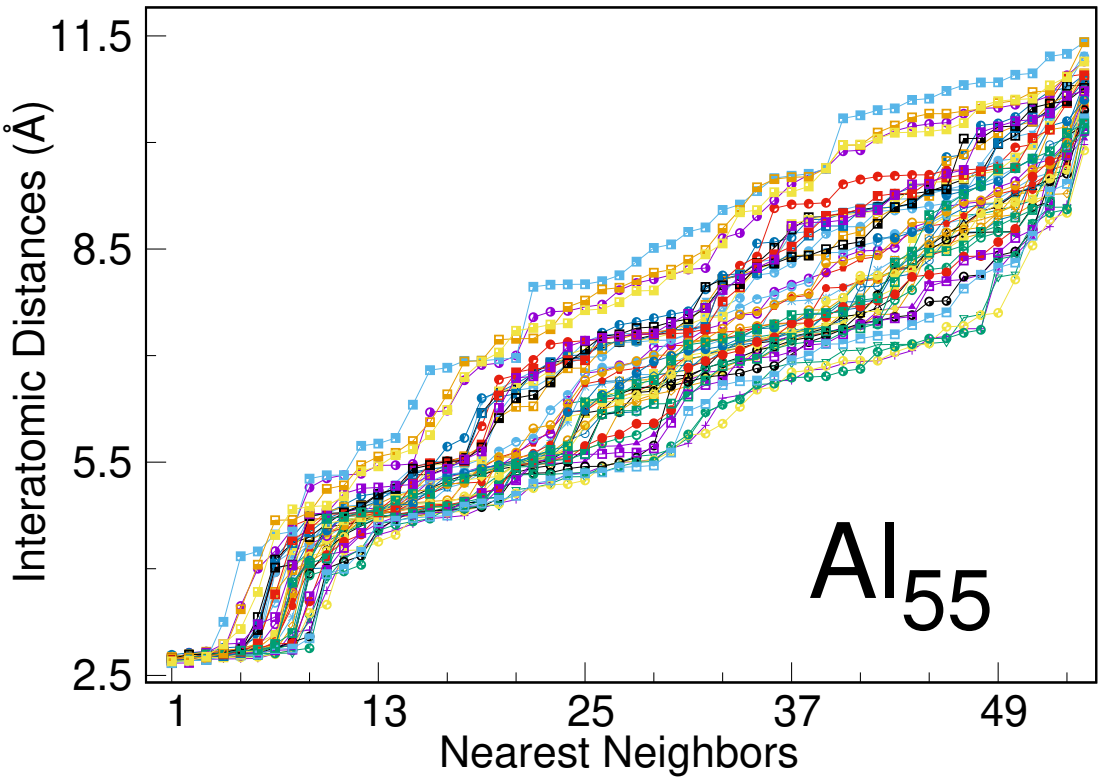

Figure 3: Nearest neighbor vs Interatomic distance plot for Al_{55}

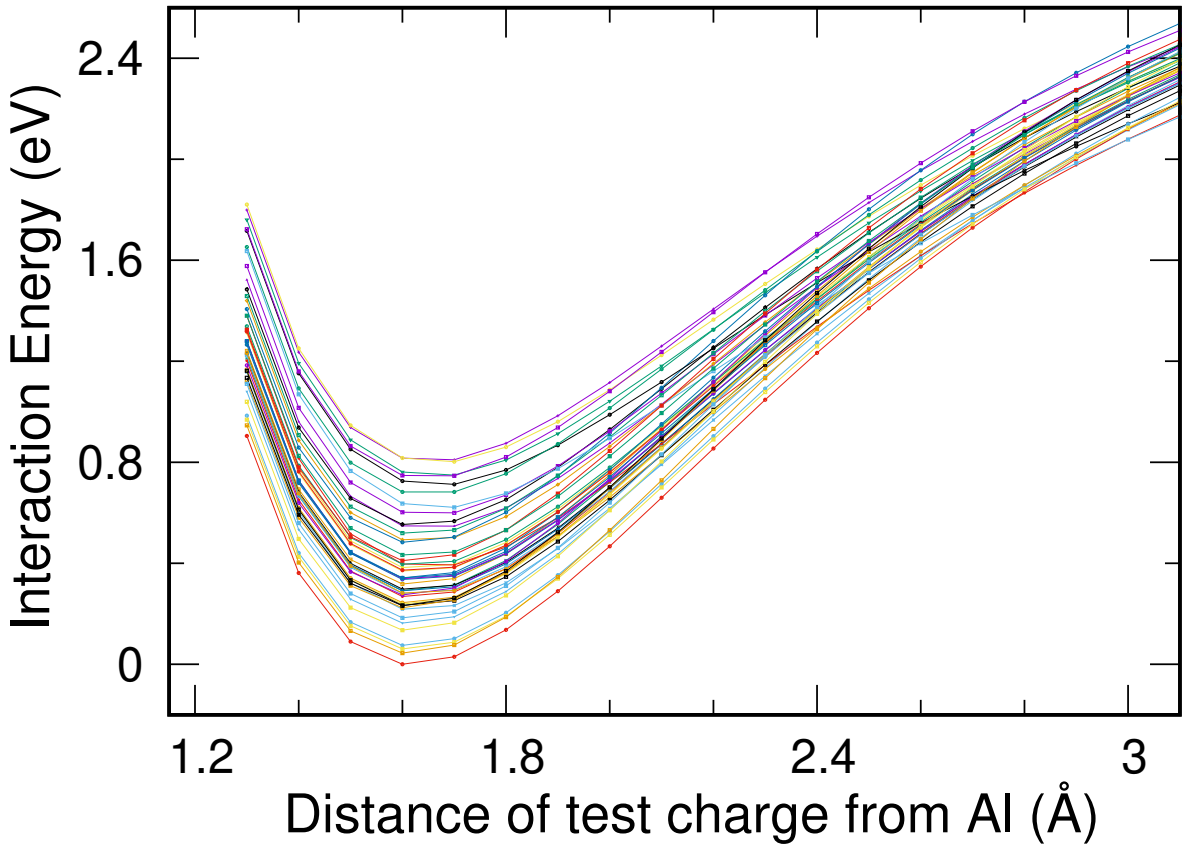

Figure 4: Interaction energy curves for Al_{55}

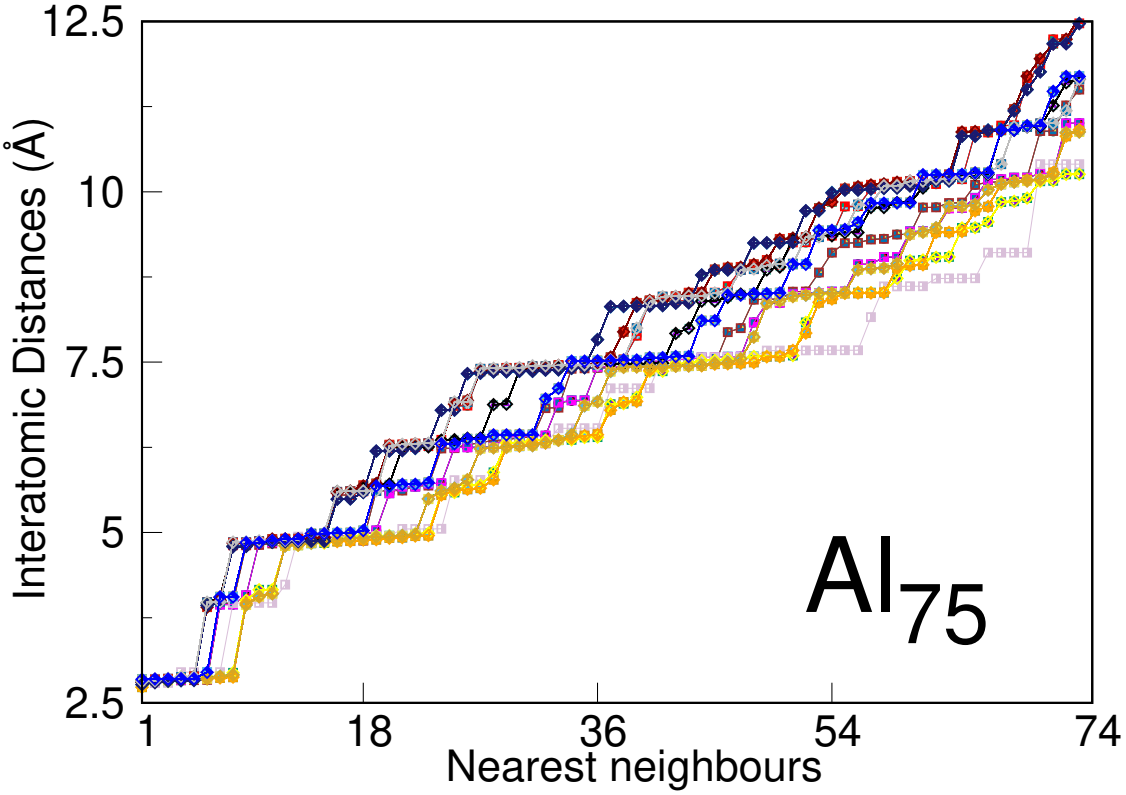

Figure 5: Nearest neighbor vs Interatomic distance plot for Al_{75}

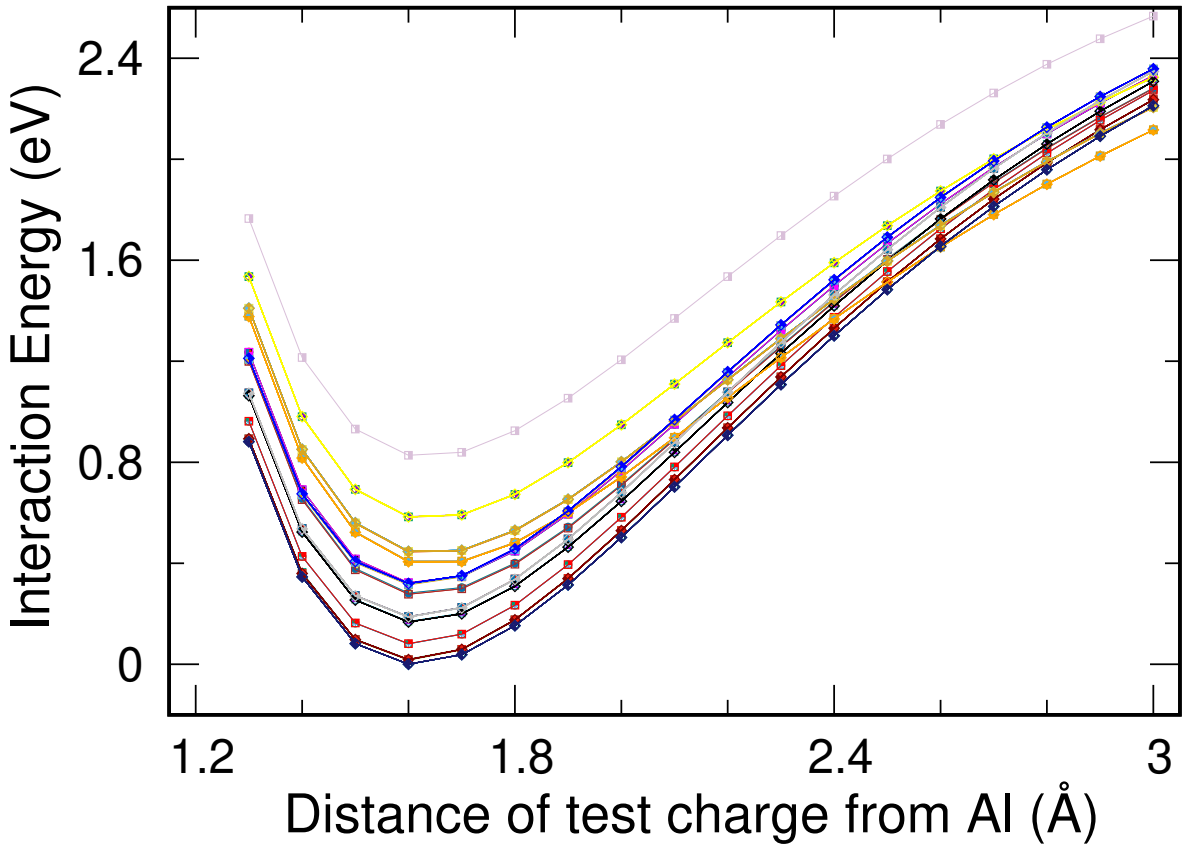

Figure 6: Interaction energy curves for Al_{75}