

DART: Deep Learning Enabled Topological Interaction Model for Energy Prediction of Metal Clusters and its Application in Identifying Unique Low Energy Isomers

Rohit Modee^a, Sheena Agarwal^{b,c}, Ashwini Verma^b, Kavita Joshi^{*b,c}, U. Deva Priyakumar^{*a}

^aCenter for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Hyderabad 500032, India

^bPhysical and Materials Chemistry Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune-411008, India.

^cAcademy of Scientific and Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh- 201002, India.

Abstract

Recently, Machine Learning (ML) has proven to yield fast and accurate predictions of chemical properties to accelerate the discovery of novel molecules and materials. The majority of the work is on organic molecules, and much more work needs to be done for inorganic molecules, especially clusters. In the present work, we introduce a simple Topological Atomic Descriptor called TAD, which encodes chemical environment information of each atom in the cluster. TAD is a simple and interpretable descriptor where each value represents the atom count in three shells. We also introduce the DART, Deep Learning Enabled Topological Interaction model, which uses TAD as a feature vector to predict energies of metal clusters, in our case Gallium clusters with size ranging from 31 to 70 atoms. DART model is designed based on the principle that energy is a function of atomic interactions and allows us to model these complex atomic interactions to predict the energy. We further introduce a new dataset called GNC_31-70, which comprises structures and DFT optimized energies of Gallium clusters with sizes ranging from 31 to 70 atoms. We show how DART can be used to accelerate the identification of ground-state structures without geometry optimization. Albeit using topological descriptor, DART achieves MAE of 3.59 kcal/mol (0.15 eV) on testset. We also show that our model can distinguish core and surface atoms in the

Email addresses: k.joshi@nccl.res.in (Kavita Joshi*), deva@iiit.ac.in (U. Deva Priyakumar*)

Ga-70 cluster, which the model has never encountered earlier. Finally, we demonstrate the transferability of DART model by predicting energies for about 6k unseen configurations picked up from Molecular Dynamics (MD) data for three cluster sizes (46, 57, and 60) within seconds. The DART model was able to reduce the load on DFT optimizations while identifying unique low energy structures from MD data.

Keywords: Machine Learning, topological descriptors, Gallium clusters,

1. Introduction

Clusters are a bridge between two very well-understood extremes, i.e., the atoms and bulk. [1, 2] Clusters have been researched to develop a fundamental understanding of systems at a length scale where addition or subtraction of even one atom affects their physicochemical properties. [3, 4, 5, 6, 7] Computationally, structure generation techniques like random sampling, [8] genetic algorithm, [9, 10] minima hopping, [11] etc., are combined with geometry optimization DFT algorithms to obtain low energy/stable structures. The hunt for such stable structures is greatly limited by the computation expense involved. And hence, combining traditional approaches of energy prediction with machine learning is one such recent trend. [12, 13, 14, 15, 16, 17, 18]

Machine Learning techniques are finding increasing application to problems of materials science, right from exploring suitable materials and structures for a desired property dependent application to digging out hidden patterns in the ML datasets.[19, 20, 21, 22, 23] The power of ML to assist domain experts with insights from vast datasets has proven to be of immense promise. [17, 24, 25, 26] An important factor that lies at the heart of any ML problem is accurate “representation of the data”. And hence, design of accurate “descriptors” or “features” for various class of materials/compounds is still an active area of research. Various approaches have been developed for engineering features that give a fair representation of cluster structures.[18, 27, 28, 29, 21, 30, 31, 32] The most direct and crude form of descriptor to represent a cluster structure are the Cartesian coordinates. But their direct application for energy prediction is severely restricted due to lack of rotational and translational invariance.

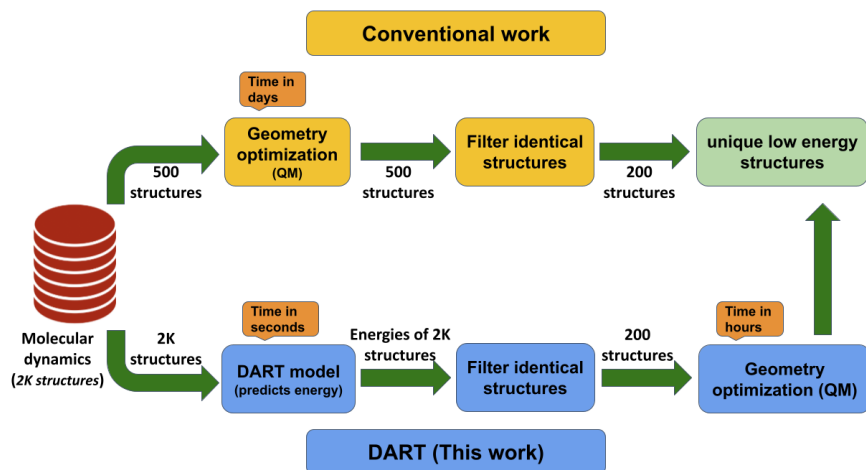


Figure 1: Work flow to identify ground state Gallium clusters using molecular dynamics simulations and Quantum Mechanics geometry optimization

Symmetry functions that depend on interatomic distances and angles are yet another successful class of descriptors applied to various systems like bulk, clusters, and molecules for prediction of formation energies.[33, 34, 35, 36, 37, 38] These approaches mainly involve accurate mapping of local atomic environment to describe the structure and hence predict their energies. Descriptors like the Coulomb Matrix (CM), [36] Smooth Overlap of Atomic Positions (SOAP), [37] Atom Centered Symmetry Function (ACSF) [33], Partial Radial Distribution Function (PRDF) [39], and so on have been developed and applied to problems of finding structural similarity and also structure energy predictions. These functions, while they describe the structure to a very reasonable extent and are accurate for energy predictions, are also complex in nature, require fitting of multiple parameters and need energy minimized structures for predicting energies[40].

Recently, graph-based representations for molecules and materials is also finding applications in property/energy prediction. [41, 42] Interatomic bonds are represented as edges and atoms as nodes of a graph. Jensen et. al., develop graph based features for atomic clusters for prediction of various molecular properties.[41] They showed that

they could predict octanol solubility of molecules with as low as 200 training examples
40 using appropriate molecular representation and MAE of $0.4 \log_{10}$ units. Further, the
author also acknowledges the lack of precise 3D positional information that usually
graph based descriptors encounter and hence limiting their prediction capacity. While,
graph based representation are very simple to understand, they come at the cost of
prediction accuracy as they miss out on structural details which may be of interest dur-
45 ing energy/property predictions. [41, 43] In another study by Wolverton et. al., they
demonstrate representation of crystalline compounds derived from Voronoi tessella-
tions of the structure. [42] Their model outperformed CM and PRDF based methods
with an MAE of 2.03 kcal/mol (88 meV/atom), at a training set size of 30,000 entries.
They design descriptors in a way such that they are insensitive to the choice of unit
50 cell or even the volume of unit cell. Their descriptors are a step ahead in the direction
of reducing independence of descriptors on atomic coordinates of a structure. Hence,
developing descriptors that are easy to design/understand yet accurate for energy pre-
dictions is a topic that is still under research.[44]

In this work, we propose a novel workflow for energy prediction of metallic clusters
55 as shown in Figure 1. For this purpose, we introduce, DART (Deep learning-enabled
topologicAl inteRacTion) model, which is designed based on the principle that energy
is a function of atomic interactions. To model these interactions, we developed an
interaction block which is inspired by SchNet.[45] Contrary to other models that use
structural information, DART uses features that capture topological/connectivity infor-
60 mation to predict metallic clusters energies. As the descriptor consists of just the atom
counts in each shell, it captures topological information. Other models such as High-
Dimensional Neural Network (HDNN)[46] uses Behler-Parinello Symmetry Function
(BPSF) as a feature vector that encodes structural information. DART with only topo-
logical information is able to reach Mean Absolute Error (MAE), which is very close
65 to HDNN, which requires structural information. Our model is able to learn physically
meaningful atomic interactions as well as distinguish between core and surface atoms.

2. Methods

This section is divided into four sub-sections. In the sub-section 2.1 we describe the generation of the new dataset called as Gallium Neutral Clusters, GNC_31-70 dataset. We introduce a novel feature vector, namely Topological Atomic Descriptor (TAD) in sub-section 2.2. TAD encodes the atomic environment information of a particular atom using the atom count information of its neighboring atoms. In sub-section 2.3, the building blocks of DART and its complete architecture which consists of five Multi-layered perceptron namely MLP_f , MLP_I , MLP_K , MLP_L , and MLP_{int} is described. DART uses TAD to predict the energy of gallium clusters. In sub-section 2.4 we describe the training procedure and mention other computational details. The code for this work, along with examples, is available at <https://github.com/devalab/DART>

2.1. Dataset

The dataset comprises of optimized, unique low energy isomers of gallium clusters with size ranging from 31 atoms to 70 atoms and their binding energies. All the calculations were carried out within the Kohn-Sham formulation of DFT. Projector Augmented Wave potential [47, 48] was used, with Perdew–Burke–Ehrzenhof (PBE) [49] approximation for the exchange-correlation and generalized gradient [50] approximation, as implemented in planewave, pseudopotential based code, VASP.[51, 52, 53] We begin by optimizing the previously published geometries for neutral Ga_n clusters (size $n = 31$ to 70). [54, 55] The dataset has total of 6851 structures and are shared as a Supplementary file and is also available at <https://figshare.com/s/9808d756f107e8fd6c69>. See Figure 2 for distribution of number of isomers across cluster size ranging from 31 to 69 atoms. We create train, validation and test dataset by randomly choosing structures from the cluster size from range 31 to 69. Train:valid:test split of 80:10:10 was used which gives 5265 structures in train set, 656 structures each in validation and test sets. We also created another data set called Ga-70 data set which has 285 structures all having 70 Gallium atom. Ga-70 dataset was created to demonstrate our models transferability in terms of predicting energy for Ga-70 clusters by training the model on Ga clusters of size 31 to 69.

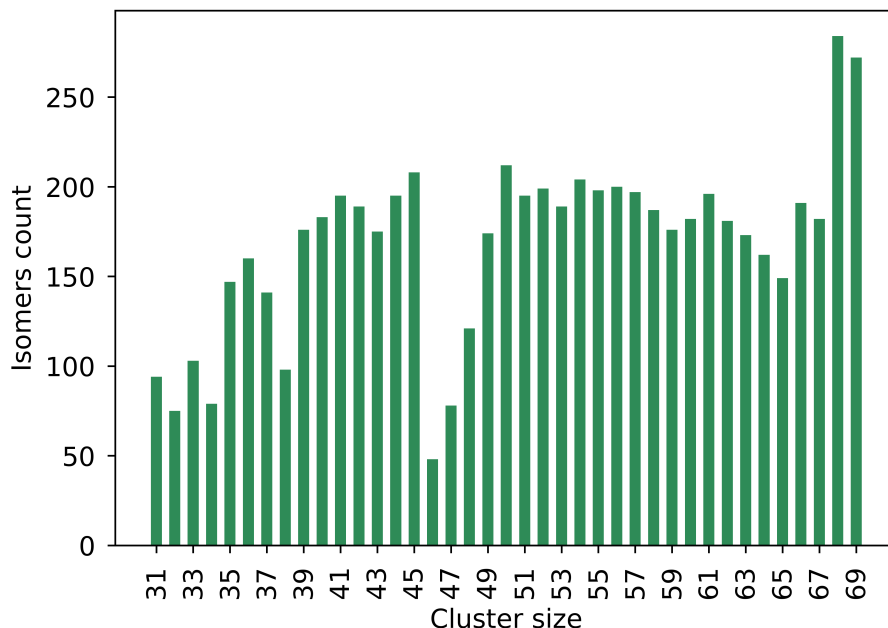


Figure 2: Shows distribution of number of isomer w.r.t Gallium cluster size.

2.2. Descriptor

We use TAD as the feature vector. As the name suggests, TAD encodes topological information of the cluster. TAD is designed to capture topological information and hence does not require exact structural information to identify low energy clusters; we show this in the results section. As seen in Figure 3, TAD encodes the atomic environment information of a particular i^{th} atom A by dividing the neighboring atoms into three different shells viz. J-shell, K-shell, and L-shell. These shells around i^{th} atom A are created using distance cutoffs. Here, we have used distance cutoff less than 3.49 Å for J-shell, 3.49 Å to 6.3 Å for K-shell, and beyond 6.3 Å is L-shell. The rationale behind choosing the distance cutoffs is based on the Ga-Ga pairwise distance distribution as shown in Figure 4. We count the number of atoms in each of these shells to create TAD. Hence, TAD is a very simple and elegant descriptor which in some sense tries to encode structural information by dividing the connectivity information using distance cutoffs.

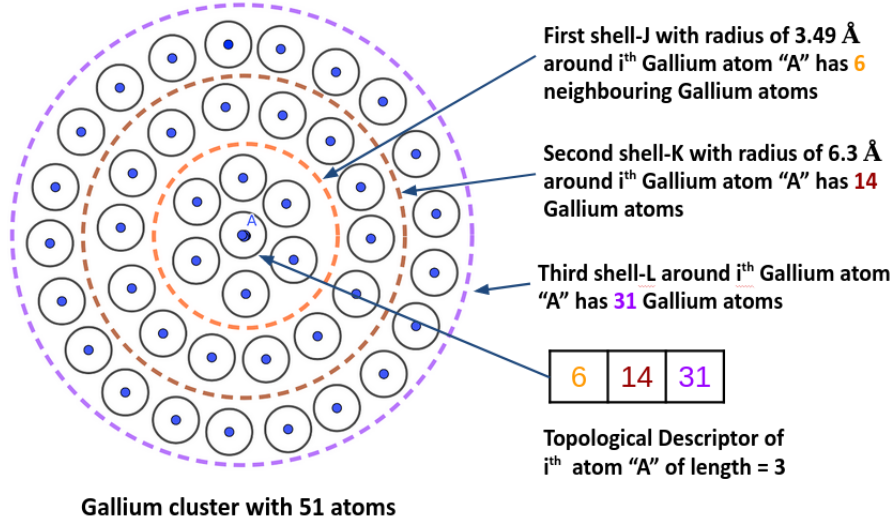


Figure 3: Shows topological descriptor of atom A in Cluster of 51 gallium atoms.

2.3. DART (Deep Learning Enabled Topological Interaction)

The architecture of DART is shown in Figure 5, it uses TAD as the feature vector to predict the energy of a single i^{th} atom. It sums the energy of all the atom's to obtain clusters energy as shown in Eq 1, where E_i is i^{th} atom energy and $i = 1..N$ where N is the total number of atoms in the cluster. Figure 5 shows TAD of i_1 atom in twelve atom representation of cluster. The remaining eleven atoms belong to either of the three shells based on the distance cutoff from the focal atom (i_1 in this case) whose energy is to be predicted. DART consists of five multilayered perceptrons (MLP), four for atomwise feature refinement, and one MLP is to model/refine atomwise interactions. As in equation 2 where W are the weights and b is the bias of MLP , the feature vector x_f of the focal atom i_1 is passed through MLP_f .

$$E_{Tot} = \sum_i^N E_i \quad (1)$$

$$Y_f = W_f x_f + b_f \quad (2)$$

The feature vector of each atom in the J^{th} -shell is feed to MLP_J as shown in equation 3 and element-wise summation of all the atoms provides us with J^{th} shell features.

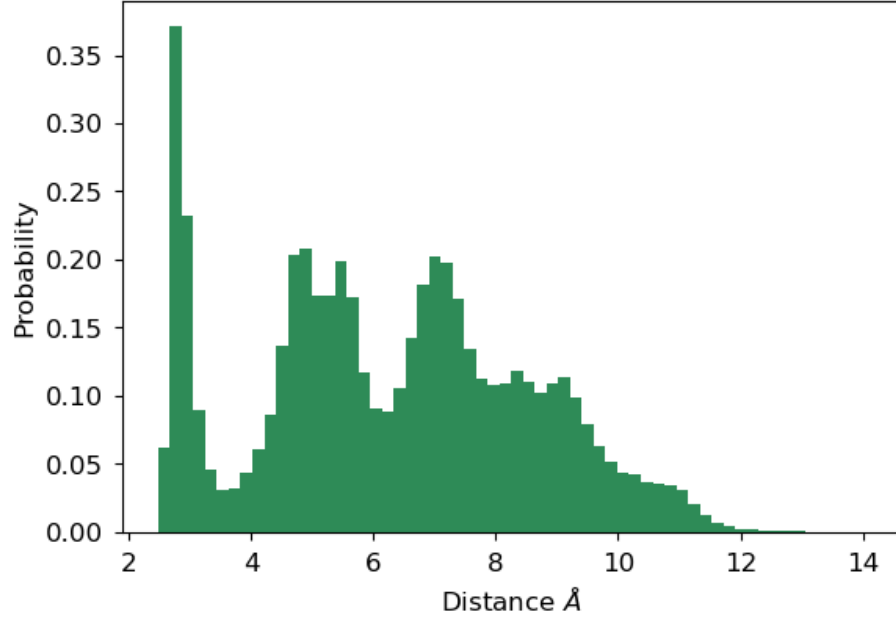


Figure 4: Shows Ga-Ga pairwise distance distribution.

Similarly, we get K^{th} -shell and L^{th} -shell features as given in equation 4 and 5.

$$Y_J = \sum_j W_J x_j + b_J \quad (3)$$

$$Y_K = \sum_k W_K x_k + b_K \quad (4)$$

$$Y_L = \sum_l W_L x_l + b_L \quad (5)$$

Further, we do an element-wise summation of features of the focal atom with the
 125 rest of the shells to get Y_{fTotal} as in equation 6.

$$Y_{fTotal} = Y_f + Y_J + Y_K + Y_L \quad (6)$$

Which is then feed to MLP_{int} , where int stands for interaction, which further refines these features to model atomic interactions. The final output of the MLP_{int} is the focal

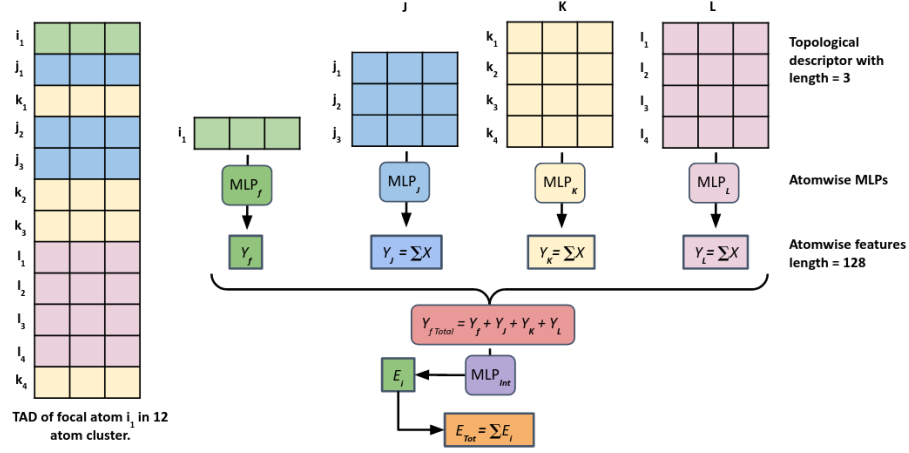


Figure 5: On the left is topological descriptor of 12 atom dummy cluster. On right is DART architecture, where energy of the i th atom of interest is predicted, with their corresponding atoms in J, K, and L shells, for $i = 1..N$ respectively, where N is total number of atoms in the cluster. There are five different types of Multilayered perceptron one each for focal atom, J-shell, K-shell, L-shell and interaction block are MLP_f , MLP_j , MLP_k , MLP_l , and MLP_{int} respectively. It should be noted that the number of MLP_j depends on number of atoms in J-shell

atom's energy, see equation 7.

$$E_f = WY_{fTotal} + b \quad (7)$$

2.4. Training the models

130 DART model is implemented in Pytorch[56]. Before training the model, all the learnable parameters are initialized using Kaiming initialization[57]. We have 2-hidden layers in MLP_f , MLP_j , MLP_k and MLP_l and 4-hidden layers in the interaction block MLP_{int} . The network architecture plays a major role in the performance of neural network potential. Too small of a network has reduced flexibility which causes poor
135 performance; on the other hand, larger networks tend to overfit the data leading to bad generalization, especially on small datasets. Table 1 gives the dimensions of DART model. After initialization, we train the DART model using a batch size of 32 to predict energy by minimizing the Mean Absolute Error (MAE) between predicted and actual energies using ADAM optimizer with an initial learning rate of 0.001, other ADAM

Module	Dimensions
MLP_f	3:128:128
MLP_J	3:128:128
MLP_K	3:128:128
MLP_L	3:128:128
MLP_{int}	128:256:128:32:1

Table 1: Number of hidden layers input size, and output size of each module is given. MLP_f is multi-layered perceptron of focal atom f , similarly MLP_J is for J-shell, MLP_K is for K-shell, MLP_L is for L-shell, and MLP_{int} is for interaction block. All these values are obtained after hyperparameter optimization

140 parameters set to $\beta_1 = 0.9$, $\beta_2 = 0.999$. Learning rate is multiplied by 0.1 after reaching a plateau with the patience of 25 epochs and $\text{eps} = 10^{-9}$. All the intermediate layers were activated using the Continuously Differentiable Exponential Linear Units (CeLU) activation function. We stop training by setting the early stopping learning rate to 10^{-8} to avoid overfitting. Training of HDNN-BPSF is similar to DART model with a
145 batch size of 32, $L1$ loss function except weights are updated using ADAMW, and bias is updated using SGD. All hyperparameter values are obtained after hyperparameter optimization for both models.

3. Results

3.1. BPSF-HDNN

150 Behler-Parrinello Symmetry Function (BPSF) descriptor encodes the atomic environment using 3D geometry. In contrast, we developed a topological descriptor that encodes the atomic environments using atom counts as described in section 2.2. The topological descriptor captures the topological information as compared to BPSF, which captures structural information. Hence, BPSF with the HDNN model is con-
155 sidered a hard limit that a model using topological descriptors cannot out perform. We compare the DART model, which uses TAD, with the BPSF-HDNN model, which uses 3D structures. BPSF-HDNN has an architecture of 640:128:128:64:1, and we used a radial cutoff of 15 Å and angular cutoff of 6.5 Å to generate atomic environment vector

(AEV). The radial and angular cutoffs dictate the distance neighboring environment
 160 should be probed. We used $\eta_r = 16.0$, $\zeta = 32.0$ and $\eta_a = 8.0$. Sixty-four evenly
 spaced radial shifting parameters were used for the radial part, and a total of twenty-
 four radial and twenty-four angular shifting parameters were used for the angular part.
 The total length of AEV is $64(radial) + 24 * 24(angular) = 640$. [46, 58]. We get MAE
 of 2.4 kcal/mol, RMSE of 3.03 kcal/mol on the test set. As expected, BPSF performs
 165 very well in predicting the binding energy of Gallium clusters.

3.2. Accuracy and Transferability of DART

Model	Testset (kcal/mol)	
	MAE	RMSE
BPSF-HDNN (structural)	2.4	3.03
DART (topological)	3.59	4.55
DART without MLP_K	4.06	5.16
DART without MLP_L	3.93	5.05
DART without MLP_K & MLP_L	4.29	5.45
DART without MLP_{int}	12.33	16.40

Table 2: Behler Parinello Symmetry Functions (BPSF) - HDNN uses structural information whereas DART uses topological information. Test set includes 655 clusters

DART model has been applied on Gallium clusters dataset introduced in section
 2.1 to predict binding energies by using topological information. After training DART
 on 80 % of data .i.e. 5256 clusters, we achieve MAE of 3.59 kcal/mol and Root Mean
 170 Squared Error (RMSE) of 4.55 kcal/mol on a test set that has 656 clusters with size
 ranging from 31 to 69 atom clusters. The largest Gallium clusters in the train set contain
 69 atoms; hence we also test the model on the Ga-70 cluster to check its transferability
 to larger clusters that the model has not seen during training. When trained on $N = 31$
 to $N = 69$ atom clusters and tested on $N + 1 = 70$ atom clusters we get MAE of 4.71
 175 kcal/mol and RMSE of 5.96 kcal/mol.

In another experiment, we validate our model on three test cases each having dif-
 ferent test-set to show the robustness of our model on different cluster sizes. In first test

case we train our model on Ga-31 to Ga-70 excluding Ga-46, 57 and 60. These sizes (46, 57 and 60) of Ga clusters will be used as test-set. In second test case our model
 180 is trained on Ga-31 to Ga-70 excluding Ga-46, 57, 66 and excluded sizes (46, 57, 66) are considered as test-set. Similarly for third test case we train model on Ga-31 to 70 excluding sizes (57, 60, 66) of Ga clusters. These excluded sizes (57, 60, 66) of Ga cluster are used as test-set. Performance of DART model on each of these test cases is summarized in table 3, it can be seen that DART models performance is consistent
 185 across all the test cases. TAD and DART can be extended to other metallic clusters.

Results for all the test cases for each Gallium cluster individually is given in Supplementary information table 1,2, and 3. As it is evident from the results that the model is robust, the RMSE values for Gallium-57, 60 and 66 are consistent across three experiments. There are small structural changes in the core-shell of Ga-46 cluster which
 190 leads to large change in energy whereas these small structural changes are not fully captured by our descriptor (TAD) hence we observe higher values of RMSE for Ga-46 which is a magic cluster.

Model	DART (values in kcal/mol)		
	MAE	RMSE	# structures
Test case 1 - Test set: (Ga-46, 57, and 60)	4.77	6.16	427
Test case 2 - Test set: (Ga-46, 57, and 66)	4.76	6.02	436
Test case 3 - Test set (Ga-57, 60, and 66)	4.34	5.58	570

Table 3: Performance of DART model on three different test cases each having different test-set to show the robustness of our model. last column gives the number of structures present in the test-set. All values are in (kcal/mol).

3.3. Importance of each shell and interaction block

To further show the importance of each of the modules in the DART model, we
 195 perform a few experiments (reported in table 3) . In these experiments, we train our

model by excluding or "switching off" each of these modules individually or in combination to see how the absence of each of these modules affects the performance of the DART model. It should be noted that all these experiments are compared with the DART model (topological), which has all-shells and interaction block "switched on".

200 In the first experiment, the MLP, which provides information about the L-shell, i.e., MLP_f , is switched off; we get MAE of 3.93 kcal/mol, RMSE of 5.05 kcal/mol on the test-set. For the second experiment, the MLP that provides information about the K-shell is switched off, which gives MAE of 4.06 kcal/mol, RMSE of 5.16 kcal/mol on the test-set. For the third experiment, the MLP_K and MLP_L , which provide information

205 about the K-shell and L-shell, respectively, are switched off, which gives MAE of 4.29 kcal/mol and RMSE of 5.45 kcal/mol on test-set. Finally, in the fourth experiment, the interaction module MLP_{int} , which learns the physically meaningful interaction between the atoms, is switched off, which gives MAE of 12.33 kcal/mol, RMSE of 16.40 kcal/mol on test-set. It is evident that when compared to DART (topological), which

210 achieves MAE of 3.59 kcal/mol and RMSE of 4.55 kcal/mol on test-set that switching off MLP_K gives a larger MAE value than the absence of MLP_L hence K_{th} -shell is more critical than L_{th} -shell. Switching off the interaction module has the most significant effect on the DART model's performance; MAE values go as high as 12.33 kcal/mol. Hence, we conclude that the interaction module MLP_{int} is critical in the DART model,

215 which helps learn the physically meaningful atomic interactions.

3.4. Model learns to distinguish between core and surface atoms

This exercise shows our model's ability to distinguish between the Gallium clusters' core and surface atom. The model learns this distinction from our descriptor TAD, which encodes the chemical environment information of each atom in the cluster.

220 Which further proves that our descriptor TAD encodes the chemical environment information of each atom within the cluster, which is sufficient for the model to distinguish between core and surface atoms. It should be noted that not all atoms contribute equally towards the binding energy of the cluster. The core atoms will have more contribution towards binding energy as compared to the surface atoms. In Ga-70 cluster,

225 atoms within 3.49 Å from the center of mass (CoM) of the cluster are considered

as core atoms, and atoms beyond 6.3 Å from CoM are considered as surface atoms. We extract the features from the third hidden layer of interaction block MLP_{int} . The third hidden layer of the interaction module MLP_{int} is just before MLP_{int} outputs the predicted atomic energy. We perform t-Distributed Stochastic Neighbor Embedding (t-SNE) on the extracted features. The t-SNE plot of the previously extracted feature
 230 in Figure 6 shows a clear distinction of core and surface atoms.

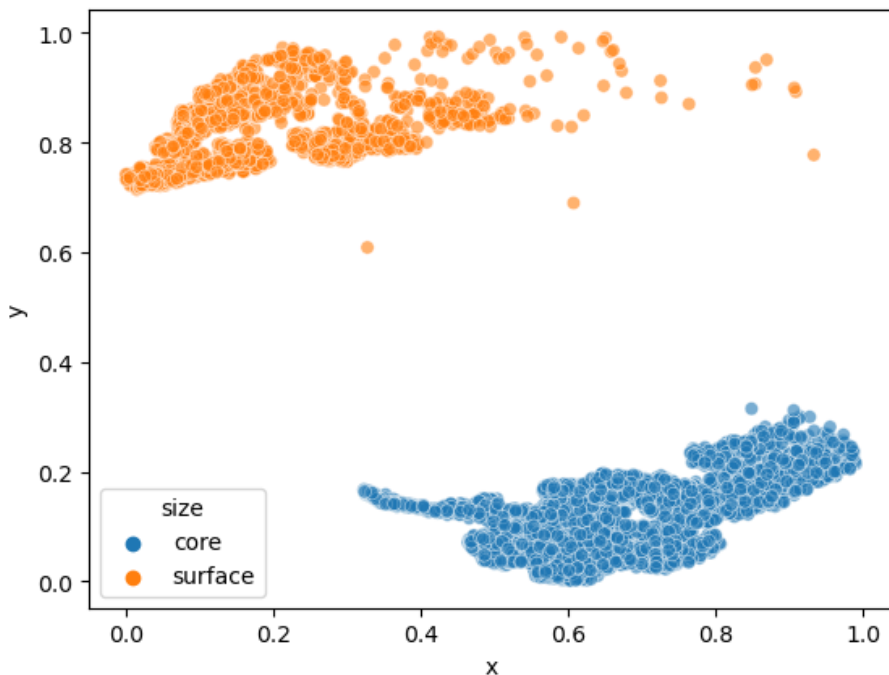


Figure 6: t-Distributed Stochastic Neighbor Embedding (t-SNE) of the feature extracted from the 3rd hidden layer of interaction module during testing of Ga-70 dataset.

3.5. Identification of stable isomers from molecular dynamics data

This section demonstrates the model’s ability to identify/filter unique low energy isomers from unseen structures taken from molecular dynamics simulation trajectories.
 235 Three randomly selected sizes (46, 57, 60) of Ga clusters were taken from molecular dynamics simulation. They were used as a test set for the DART model trained on geometry optimized data from size 31 to 70, excluding size (46, 57, and 60) of Ga

cluster. Molecular dynamics test set consists of 1919, 1051, and 2596 structures of Ga-46, Ga-57, and Ga-60, respectively. The model was able to predict energy values for all 5566 structures within seconds. Firstly, the predicted energies were sorted and 100 lowest energy structures were picked for DFT optimisation. Upon optimisation it was observed that the DART model had correctly identified the ground state geometry from the vast dataset on which it was tested. And hence, greatly reducing the load on DFT calculations. Further, the DART model was also used to identify the different low energy isomers for each size. Conventionally, about 25-30 % structures picked for each size from molecular dynamics would be chosen to perform further DFT optimizations, which would in turn yield roughly 100-150 unique structures. This search, i.e., the number of DFT optimizations, could be significantly reduced with the application of the DART model. Since the model was trained on optimization data, energy predictions were closer to optimized geometry values than MD. Moreover, it was possible to significantly reduce the number of DFT optimizations while maximizing the probability of finding unique low energy structures. There were two possible approaches to identify unique isomers out of the existing MD data; i.e. structures could either be segregated into bins of varying descriptors in the feature space or bins of varying ML predicted energy values. Optimisation of one structure from each bin would in turn identify the unique isomers. However, given that TAD counts the coordination of each atom in its neighborhood within three shells, classification based on descriptors would lead to very broad bins of structures. This in turn would lead to classifying similar structures of clusters belonging to different local minima's into one bin and missing out on many unique isomers. Hence, structures were segregated into bins of varying ML predicted energy values. About 100 structures were chosen (one from each energy bin) and optimized using DFT to identify unique low energy geometries. It was noted that out of 93 DFT optimizations performed for each of the three cluster sizes, nearly 97 % of the structures turned out to be unique. Hence, we note that DART not only predicts energy but can also identify unique low energy isomers from a large number of structures generated from molecular dynamics simulations.

4. Conclusion

In this study, a deep learning model, namely DART, developed using interaction block, capable of modeling atomic interactions using topological descriptor (TAD) is reported. DART is significantly more efficient than the conventional work-flow for identifying ground state isomer in terms of time and computational resources. A novel feature vector called Topological Atomic Descriptor (TAD), which effectively encodes the clusters topology, is used to train the DART model. A deep neural network DART model exhibits its ability to predict energy of Gallium cluster, demonstrating the trained network's robustness and TAD's adequacy as a feature vector compared to symmetry functions. The interaction module is crucial to mimic atomic interactions. Examining the features extracted from this module indicates that the model has learned the underlying chemistry by its ability to differentiate between core and surface atoms. Furthermore, DART model can be extended to any metal cluster and can be modified to be used on nanoalloys[59]. Also, DART model can identify unique low energy structures from a corpus of structures obtained from MD simulations, thereby reducing the number of QM calculations manifolds. This method successfully predicts clusters energy and can identify unique low energy isomers. Next objective is to develop a model which can learn and generate three-dimensional structures of the clusters than just identifying the low energy isomers. Future work in this direction is in progress.

Acknowledgment

KJ and AV would like to acknowledge CSIR-4PI for the computational facility.

References

- [1] P. Jena, A. W. Castleman, Clusters: A bridge across the disciplines of physics and chemistry, *Proceedings of the National Academy of Sciences* 103 (28) (2006) 10560–10569.
- [2] A. Castleman Jr, S. Khanna, Clusters, superatoms, and building blocks of new materials, *The Journal of Physical Chemistry C* 113 (7) (2009) 2664–2675.

- [3] K. D. Ball, R. S. Berry, R. E. Kunz, F.-Y. Li, A. Proykova, D. J. Wales, From topographies to dynamics on multidimensional potential energy surfaces of atomic clusters, *Science* 271 (5251) (1996) 963–966.
- [4] M. Schmidt, R. Kusche, B. von Issendorff, H. Haberland, Irregular variations in the melting point of size-selected atomic clusters, *Nature* 393 (6682) (1998) 238–240.
- [5] A. A. Shvartsburg, M. F. Jarrold, Solid clusters above the bulk melting point, *Physical Review Letters* 85 (12) (2000) 2530.
- [6] R. S. Berry, Potential surfaces and dynamics: What clusters tell us, *Chemical Reviews* 93 (7) (1993) 2379–2394.
- [7] K. Joshi, S. Krishnamurty, D. Kanhere, “magic melters” have geometrical origin, *Physical Review Letters* 96 (13) (2006) 135703.
- [8] C. J. Pickard, R. Needs, Ab initio random structure searching, *Journal of Physics: Condensed Matter* 23 (5) (2011) 053201.
- [9] D. M. Deaven, K.-M. Ho, Molecular geometry optimization with a genetic algorithm, *Physical Review Letters* 75 (2) (1995) 288.
- [10] B. Hartke, Global geometry optimization of clusters using genetic algorithms, *The Journal of Physical Chemistry* 97 (39) (1993) 9973–9976.
- [11] S. Goedecker, Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems, *The Journal of Chemical Physics* 120 (21) (2004) 9911–9917.
- [12] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, C. Wolverton, The open quantum materials database (oqmd): assessing the accuracy of dft formation energies, *npj Computational Materials* 1 (1) (2015) 1–15.
- [13] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in un-

- 320 constrained composition space with machine learning, *Physical Review B* 89 (9) (2014) 094104.
- [14] J. Jie, Z. Hu, G. Qian, M. Weng, S. Li, S. Li, M. Hu, D. Chen, W. Xiao, J. Zheng, et al., Discovering unusual structures from exception using big data and machine learning techniques, *Science Bulletin* 64 (9) (2019) 612–616.
- 325 [15] T. Mueller, A. G. Kusne, R. Ramprasad, Machine learning in materials science: Recent progress and emerging applications, *Reviews in Computational Chemistry* 29 (2016) 186–273.
- [16] R. Jinnouchi, H. Hirata, R. Asahi, Extrapolating energetics on clusters and single-crystal surfaces to nanoparticles by machine-learning scheme, *The Journal of*
 330 *Physical Chemistry C* 121 (47) (2017) 26397–26405.
- [17] Y. Pathak, K. S. Juneja, G. Varma, M. Ehara, U. D. Priyakumar, Deep learning enabled inorganic material generator, *Physical Chemistry Chemical Physics* 22 (46) (2020) 26935–26943.
- [18] A. M. Cooper, J. Kästner, A. Urban, N. Artrith, Efficient training of ANN potentials by including atomic forces via Taylor expansion and application to water and
 335 a transition-metal oxide, *npj Computational Materials* 6 (1) (2020) 1–14.
- [19] S. Agarwal, S. Mehta, K. Joshi, Understanding the ml black box with simple descriptors to predict cluster–adsorbate interaction energy, *New Journal of Chemistry* 44 (20) (2020) 8545–8553.
- 340 [20] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, *Nature* 559 (7715) (2018) 547–555.
- [21] S. A. Meldgaard, E. L. Kolsbjerg, B. Hammer, Machine learning enhanced global optimization by clustering local environments to enable bundled atomic energies, *The Journal of Chemical Physics* 149 (13) (2018) 134104.
- 345 [22] G. R. Schleder, A. C. Padilha, C. M. Acosta, M. Costa, A. Fazzio, From dft to machine learning: recent approaches to materials science—a review, *Journal of Physics: Materials* 2 (3) (2019) 032001.

- [23] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects, npj Computational Materials 3 (1) (2017) 1–13.
- [24] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, Accelerating materials property predictions using machine learning, Scientific Reports 3 (2013) 2810.
- [25] S. Kapse, S. Janwari, U. V. Waghmare, R. Thapa, Energy parameter and electronic descriptor for carbon based catalyst predicted using qm/ml, Applied Catalysis B: Environmental 286 (2021) 119866.
- [26] S. Saxena, T. S. Khan, F. Jalid, M. Ramteke, M. A. Haider, In silico high throughput screening of bimetallic and single atom alloys using machine learning and ab initio microkinetic modelling, Journal of Materials Chemistry A 8 (1) (2020) 107–123.
- [27] A. Seko, H. Hayashi, K. Nakayama, A. Takahashi, I. Tanaka, Representation of compounds for machine-learning prediction of physical properties, Physical Review B 95 (14) (2017) 144110.
- [28] X. Chen, D. Chen, M. Weng, Y. Jiang, G.-W. Wei, F. Pan, Topology-based machine learning strategy for cluster structure prediction, The Journal of Physical Chemistry Letters 11 (11) (2020) 4392–4401.
- [29] P. Pankajakshan, S. Sanyal, O. E. de Noord, I. Bhattacharya, A. Bhattacharyya, U. Waghmare, Machine learning and statistical analysis for materials science: stability and transferability of fingerprint descriptors and chemical insights, Chemistry of Materials 29 (10) (2017) 4190–4201.
- [30] A. Jain, T. Bligaard, Atomic-position independent descriptor for machine learning of material properties, Physical Review B 98 (21) (2018) 214112.
- [31] S. Laghuvarapu, Y. Pathak, U. D. Priyakumar, Band nn: A deep learning framework for energy prediction and geometry optimization of organic small molecules, Journal of Computational Chemistry 41 (8) (2020) 790–799.

- [32] F. Faber, A. Lindmaa, O. A. von Lilienfeld, R. Armiento, Crystal structure representations for machine learning models of formation energies, *International Journal of Quantum Chemistry* 115 (16) (2015) 1094–1101.
- [33] J. Behler, Atom-centered symmetry functions for constructing high-dimensional
380 neural network potentials, *The Journal of Chemical Physics* 134 (7) (2011) 074106.
- [34] H. Gassner, M. Probst, A. Lauenstein, K. Hermansson, Representation of intermolecular potential functions by neural networks, *The Journal of Physical Chemistry A* 102 (24) (1998) 4596–4605.
- [35] S. Lorenz, M. Scheffler, A. Gross, Descriptions of surface chemical reactions
385 using a neural network representation of the potential-energy surface, *Physical Review B* 73 (11) (2006) 115431.
- [36] M. Rupp, A. Tkatchenko, K.-R. Müller, O. A. Von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning, *Physical Review Letters* 108 (5) (2012) 058301.
390
- [37] A. P. Bartók, R. Kondor, G. Csányi, On representing chemical environments, *Physical Review B* 87 (18) (2013) 184115.
- [38] S. Jindal, S. Chiriki, S. S. Bulusu, Spherical harmonics based descriptor for neural network potentials: Structure and dynamics of Au_{147} nanocluster, *The Journal of Chemical Physics* 146 (20) (2017) 204301.
395
- [39] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K.-R. Müller, E. K. Gross, How to represent crystal structures for machine learning: Towards fast prediction of electronic properties, *Physical Review B* 89 (20) (2014) 205118.
- [40] N. Artrith, A. Urban, G. Ceder, Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species, *Physical Review B*
400 96 (1) (2017) 1–5.

- [41] C. W. Coley, R. Barzilay, W. H. Green, T. S. Jaakkola, K. F. Jensen, Convolutional embedding of attributed molecular graphs for physical property prediction, *Journal of Chemical Information and Modeling* 57 (8) (2017) 1757–1772.
- 405 [42] L. Ward, R. Liu, A. Krishna, V. I. Hegde, A. Agrawal, A. Choudhary, C. Wolverton, Including crystal structure attributes in machine learning models of formation energies via voronoi tessellations, *Physical Review B* 96 (2) (2017) 024104.
- [43] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, P. Riley, Molecular graph convolutions: moving beyond fingerprints, *Journal of Computer-aided Molecular Design* 30 (8) (2016) 595–608.
- 410 [44] A. M. Miksch, T. Morawietz, J. Kästner, A. Urban, N. Artrith, Strategies for the Construction of Machine-Learning Potentials for Accurate and Efficient Atomic-Scale Simulations, *Machine Learning: Science and Technology (ML)* (2021) 1–21.
- 415 [45] K. T. Schütt, H. E. Sauceda, P. J. Kindermans, A. Tkatchenko, K. R. Müller, SchNet - A deep learning architecture for molecules and materials, *Journal of Chemical Physics* 148 (24) (2018) 241722.
- [46] J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, *Physical Review Letters* 98 (14) (2007) 146401.
- 420 [47] P. E. Blöchl, Projector augmented-wave method, *Physical Review B* 50 (1994) 17953–17979.
- [48] G. Kresse, D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, *Physical Review B* 59 (1999) 1758–1775.
- 425 [49] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, *Physical Review Letters* 77 (1996) 3865–3868.
- [50] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, *Physical Review Letters* 78 (1997) 1396–1396.

- [51] G. Kresse, J. Hafner, *Ab initio* molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium, Physical Review B 49 (1994) 14251–14269.
- [52] G. Kresse, J. Furthmüller, Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set, Physical Review B 54 (1996) 11169–11186.
- [53] G. Kresse, J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Computational Material Science 6 (1) (1996) 15 – 50.
- [54] V. Kaware, K. Joshi, Scaling up the shape: A novel growth pattern of gallium clusters, The Journal of Chemical Physics 141 (5) (2014) 054308.
- [55] A. Susan, A. Kibey, V. Kaware, K. Joshi, Correlation between the variation in observed melting temperatures and structural motifs of the global minima of gallium clusters: An ab initio study, The Journal of Chemical Physics 138 (1) (2013) 014303.
- [56] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimsheine, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035.
- [57] K. He, X. Zhang, S. Ren, J. Sun, Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, Proceedings of the IEEE International Conference on Computer Vision 2015 Inter (2015) 1026–1034.
- [58] J. S. Smith, O. Isayev, A. E. Roitberg, ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost, Chemical Science 8 (4) (2017) 3192–3203.

- [59] R. N. Dhital, K. Nomura, Y. Sato, S. Haesuwannakij, M. Ehara, H. Sakurai, Pt-Pd nanoalloy for the unprecedented activation of carbon-fluorine bond at low temperature, *Bulletin of the Chemical Society of Japan* 93 (10) (2020) 1180–1185.