



of things, knowledge graph, and recommender system, is particularly suitable for clusters and molecules. Compared to the commonly used artificial neural network (ANN) [14–21] that treats each atom independently, graph network is a natural choice to describe the geometric characteristic of a cluster or molecule, since it contains the information of interconnection between atoms [22–32]. Moreover, graph network is less sensitive to the choice of atomic descriptors. To date, several graph neural networks, namely, SchNet [23], CGCNN [26], MEGNet [28], MGCN [29], DimNet [30], DGANN [31] and DeepMolNet [32] have been developed for molecules and crystals. For instance, SchNet [23] introduces the continuous-filter convolutions (cfconv) and utilizes element type and atomic coordinates as input. DimNet [30] uses spherical Bessel functions and spherical harmonics to replace the Gaussian radial basis representations. CGCNN [26] allows multiple edges in crystal graph due to periodicity which can predict eight different properties of crystals. DGANN [31] does not require the complete graph and directly uses the Cartesian coordinates instead of the interatomic distance as input.

Compared to molecules with well-defined geometry, the structure of atomic cluster evolves with size (i.e., number of atoms) and searching the ground-state structure requires global exploration on the potential energy surface (PES). Due to the coexistence of numerous isomers on the PES, DL techniques, including graph network, have been rarely applied to cluster science yet. For a cluster, not only the tiny difference between similar isomers would confuse the neural network, but also the flexible bonding configuration is disadvantageous to build network.

Different from many other fields that need node information only, application of graph network on molecules, clusters, or crystals must consider information of both atoms and bonds. As one of the three leading authorities in deep learning, attention mechanism is proposed for machine translation by Bengio [33]. The effect of neighboring bonds on a center atom is suited to model by attention mechanism. Qian *et al.* [31] devised a directed graph attention neural network (DGANN) that encodes the local chemical environment by graph attention mechanism. The DeepMoleNet [32] developed by Ma's group adopts multilevel attention that can precisely predict 12 molecular properties integrating in one net. However, graph network associated with attention mechanism has not been adopted for structural optimization of clusters yet.

In this paper, we develop a cluster graph attention network (CGANet) to reproduce the binding energy, force, and molecular orbital energies of silver cluster from density functional theory (DFT) calculations. Combining with a home-made genetic algorithm code, we have performed global search for the ground state structures of  $\text{Ag}_n$  clusters with  $n = 14\text{--}26$  and unveiled several unprecedented lowest-energy structures. This

graph attention network, which is at least two orders of magnitude faster than DFT calculations, is a universal approach for structural optimization and property prediction of atomic clusters.

## 2 Network structure

### 2.1 Feature extraction

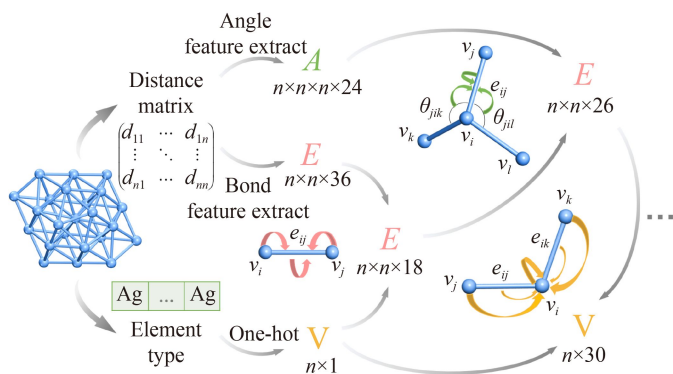
For an atomic cluster or molecule, element type, bond length and bond angle are the essential information that determines the physical and chemical properties of the system. We first extract this information from the element type and distance matrix as input. The element information should be transformed to a vector by one-hot encoding or embedding layer. For simplicity, herein we consider elemental clusters of silver atoms; thus, we do not have to distinguish the element type. The initial feature of atoms is a one-dimensional vector of constant 1. The Cartesian coordinates of atoms are not directly utilized as the input of network. Instead, we extract the structural information from Cartesian coordinates and put it into the network. The key operation of CGANet is schematically shown in Fig. 1. We start from the distance matrix  $\mathbf{D}$  because it is invariant under translation or rotation. For each pair of atom  $i$  and  $j$ , the distance  $d_{ij}$  in  $\mathbf{D}$  matrix is transformed to a 36-dimensional vector  $\mathbf{e}_{ij}^0$  by a Gaussian function [14]:

$$\mathbf{d}_{ij}^0 = \mathbf{e}^{-\xi_k(d_{ij}-\beta_k)^2}, \quad k = 1, 2, \dots, 36.$$

Here  $\xi_k$  and  $\beta_k$  are parameters to be trained. The bond angle  $\theta_{ijk}$  can be simply calculated from  $d_{ij}$ ,  $d_{ik}$  and  $d_{jk}$  using the cosine law. Then we introduce a 24-dimensional angle vector  $\mathbf{a}_{ijk}^0$  in a similar way:

$$\mathbf{a}_{ijk}^0 = \mathbf{e}^{-\eta_l(a_{ijk}-\alpha_l)^2}, \quad l = 1, 2, \dots, 24.$$

where  $\eta_l$  and  $\alpha_l$  are parameters to be trained.



**Fig. 1** Feature extraction and graph convolution operations.



## 2.2 Vertex convolutional layer

The classical graph convolutional operation is expressed as follows [34]:

$$\mathbf{X}^{l+1} = \sigma(\mathbf{A}\mathbf{X}^l\mathbf{W}),$$

where  $\mathbf{X}^l$  is the output of layer  $l$ ,  $\mathbf{A}$  is the adjacent matrix of graph,  $\mathbf{W}$  is the weight matrix of this layer with size  $n_l \times n_{l+1}$ . The product of  $\mathbf{X}^l\mathbf{W}$  transforms the feature from dimension  $n_l$  to dimension  $n_{l+1}$ .  $\mathbf{X}^l$  left multiplied by  $\mathbf{A}$  would sum all neighbor feature to the center atom. At last, an activation function  $\sigma$  is used to achieve the output of layer  $l+1$ . However, this graph convolutional operation has two deficiencies. First, different neighboring atoms have different effect on the center atom; therefore, a simple summation of them is unreasonable. The second issue is how to properly aggregate the feature of center atom itself. Hence, we introduce an improved graph attention layer. Let  $\mathbf{V}^l$  be the matrix of each atom's feature of layer  $l$ , and  $\mathbf{h} = \mathbf{V}^l\mathbf{W}$  be a new feature matrix of another dimension. Then, we aggregate neighbor features with different weights, and the weights are calculated by the center feature  $\mathbf{h}_i$ , the neighbor feature  $\mathbf{h}_j$  and the vector  $\mathbf{d}_{ij}$  [35]

$$\alpha_{ij} = \text{Softmax}(\text{LeakyReLU}(\boldsymbol{\alpha}^T[\mathbf{h}_i^{l+1} \parallel \mathbf{h}_j^{l+1} \parallel \mathbf{d}_{ij}])).$$

Here *Softmax* and *LeakyReLU* are activation functions,  $\alpha_{ij}$  is attention coefficient,  $\boldsymbol{\alpha}$  is a vector to be trained. The *Softmax* function can normalize the attention coefficient. For each center atom  $i$ , the feature of neighboring atom  $j$  is aggregated by the following formula:

$$\mathbf{v}_i^{l+1} = \text{Elu}(\mathbf{h}_i + \beta \sum_{j \in N(i)} \alpha_{ij} \mathbf{h}_j).$$

The aggregation includes two parts—the feature of atom  $i$  and the summation of features of all neighboring atoms  $j$  scaled with an attention coefficient  $\alpha_{ij}$  weighted by coefficient  $\beta$ . At last, the aggregation goes through an activation function *Elu* to yield the next output feature  $\mathbf{V}^{l+1}$ . Attention mechanism is permutation invariant because the attention coefficient  $\alpha_{ij}$  permutes with atom  $j$ .

## 2.3 Edge convolutional layer

The  $l+1^{\text{th}}$  layer edge feature is obtained from vertex features of the two ends and the edge feature itself in the previous  $l^{\text{th}}$  layer:

$$\mathbf{e}^{l+1} = \text{Elu}([\mathbf{e}_{ij}^l \parallel \mathbf{v}_i^l \parallel \mathbf{v}_j^l] \mathbf{w}).$$

## 2.4 Angle convolutional layer

Similar to the vertex convolutional layer, the new feature of edge  $\mathbf{e}_{ij}$  is aggregated by the edge  $\mathbf{e}_{ij}$  itself and all angle features  $\mathbf{a}_{ijk}$  using attentional mechanism:

$$\mathbf{h} = \mathbf{e}^l \mathbf{w},$$

$$\alpha_{ijk} = \text{Softmax}(\text{LeakyReLU}(\boldsymbol{\alpha}^T[\mathbf{h}_{ij} \parallel \mathbf{h}_{ik} \parallel \mathbf{a}_{ijk}])),$$

$$\mathbf{e}_{ij}^{l+1} = \text{Elu}(\mathbf{h}_{ij} + \beta \sum_{k \in N(i)} \alpha_{ijk} \mathbf{h}_{ik}).$$

## 2.5 Force convolutional layer

We combine the feature vectors  $v_i$ ,  $v_j$  and  $\mathbf{e}_{ij}$ , and map them into a single value as combination coefficient  $\alpha_{ij}$ . Then, the force of atom  $i$  can be related to the linear combination of vector  $\mathbf{r}_{ij}$  from atom  $i$  to atom  $j$ .

$$\alpha_{ij} = \boldsymbol{\alpha}^T[\mathbf{h}_i \parallel \mathbf{h}_j \parallel \mathbf{e}_{ij}],$$

$$\mathbf{f}_i = \sum_j \alpha_{ij} \mathbf{r}_{ij}'.$$

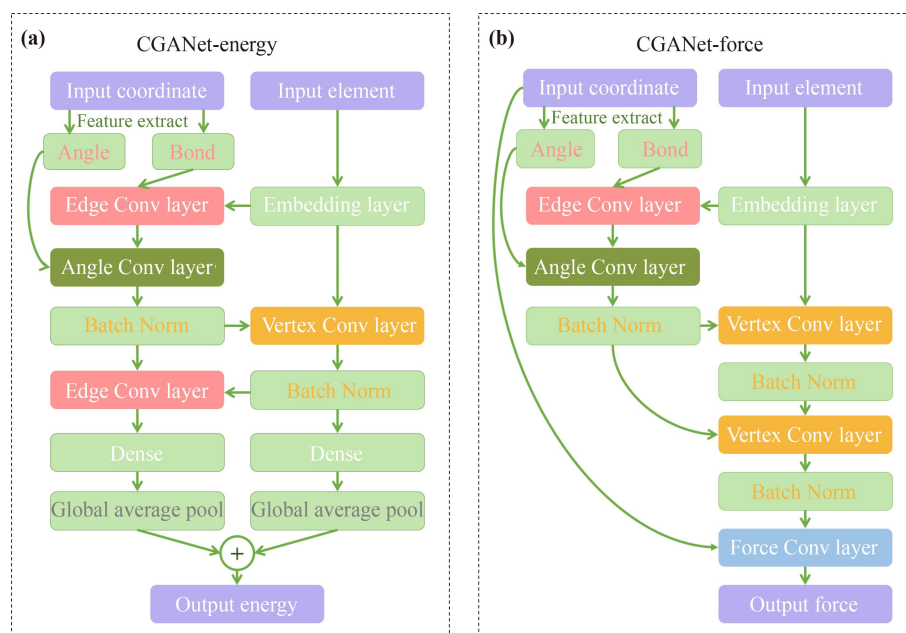
Here  $\mathbf{r}_{ij}'$  is the unit vector of  $\mathbf{r}_{ij}$ .

## 2.6 Energy prediction network

As shown in Fig. 2(a), the energy prediction network takes distance matrix  $\mathbf{D}$  as input. We use an edge convolutional layer to integrate the element information (embedding layer) and the distance feature extracted by distance matrix  $\mathbf{D}$ . Then we integrate the edge information and angle information to an angle convolutional layer. Next, the angle convolutional layer goes through a batch normalization and combines with the element information to enter a vertex convolutional layer. The resulting vertex information flows through a batch normalization layer and combines with the previous edge information to yield a new edge convolutional layer. In the dense layer, the obtained vertex feature and edge feature are mapped to a single number of  $E_v$  and  $E_e$ , respectively. In the final global average pool layer as the readout function, the average binding energy per atom is predicted by

$$E = \left(1 + \frac{c_1}{N} + \frac{c_2}{N^2}\right) \bar{E}_v + \left(1 + \frac{c_3}{N} + \frac{c_4}{N^2}\right) \bar{E}_e.$$

In fact, the binding energy of cluster usually increases and gradually approaches to the bulk limit as the size  $N$  increase, whose trend is similar to the curve of  $1 + c_1/N + c_2/N^2$ . Here we introduce four empirical parameters  $c_1$ ,  $c_2$ ,  $c_3$  and  $c_4$  to reflect the overall trend of size dependence of binding energy, and  $N$  is the cluster size. By definition, a positive binding energy means formation of a cluster from individual atoms is exothermic. The parameters  $c_1$ – $c_4$  are all negative, which successfully reflect the ascending tendency of binding energy as cluster size increases. Our test calculations show that inclusion of parameters for the size dependence of binding energy is beneficial for mixed-size train.



**Fig. 2** (a) Energy prediction network and (b) force prediction network.

### 2.7 Force prediction network

In a previous study, an energy-conserving force model was obtained by differentiating the energy model with respect to the atomic position [23]. Due to the relation of energy and force, the model has to be at least twice differentiable to allow for gradient descent of the force loss, the resulting force model is twice as deep and, hence, requires about twice amount of computational time. Therefore, the energy and force networks are trained separately in this work, which can avoid the consumption of differential operation. The force prediction network is depicted in Fig. 2(b). Generally, it resembles that of energy prediction but with little difference. Both networks share the same architecture until the vertex convolutional layer. Then we replace the last edge convolutional layer by a second vertex convolutional layer. Followed by a batch normalization layer, the force convolutional layer is connected as output. Moreover, similar CGANet architectures for energy, force and properties prediction with good performance can fully demonstrate the robustness of CGANet.

## 3 Train and test

### 3.1 Datasets

To demonstrate how to implement the CGANet, we consider medium-sized  $\text{Ag}_n$  clusters ( $n = 15\text{--}27$ ) to construct datasets. First, huge number of structures for  $\text{Ag}_{16}$ ,  $\text{Ag}_{20}$  and  $\text{Ag}_{24}$  clusters are obtained from unbiased search on the PES of clusters using a home-developed genetic algorithm program (namely CGA [36]). For each cluster, we conduct five independent CGA searches up

to 1000 iterations with symmetry constrain of  $C_1$ ,  $C_2$ ,  $C_3$ ,  $C_s$  and  $C_i$  point groups, respectively. All structures in CGA search are fully optimized with DFT, as implemented by the DMol<sup>3</sup> program [37], which utilizes the double numerical basis including d-polarization function (DND) with a global orbital cutoff of 6.0 Å (see Fig. S1 of the Supporting Information), the Perdew–Burke–Ernzerhof (PBE) functional for exchange–correlation interaction [38], and all-electron relativistic potential to treat the core. All intermediate structures during geometry optimization are collected to constitute the train set, unless the binding energy of the structure is negative (which means that the structure is too unstable). Finally, the total size of dataset for  $\text{Ag}_{16}$ ,  $\text{Ag}_{20}$  and  $\text{Ag}_{24}$  is 166 689, 136 165 and 145 014, respectively. We have also collected the reported isomeric structures of  $\text{Ag}_n$  clusters ( $n = 15, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27$ ) in literature [21, 39–45] and re-optimized them using the same PBE-DND scheme. Again, using the intermediate structures during geometry optimization, the resulting dataset contains 1825 cluster structures as the test dataset for mixed-size train.

### 3.2 Train

First, we train the CGANet on datasets of  $\text{Ag}_{16}$ ,  $\text{Ag}_{20}$  and  $\text{Ag}_{24}$  individually. For each cluster size, the data are split to train data and test data by 9:1. The loss function of force prediction net is mean absolute error (MAE) of force, while the loss function of energy prediction net is root mean square error (RMSE) of energy since RMSE is larger than MAE by definition and more sensitive to reflect the deviation between the data predicted by neural network and from DFT calculations. For

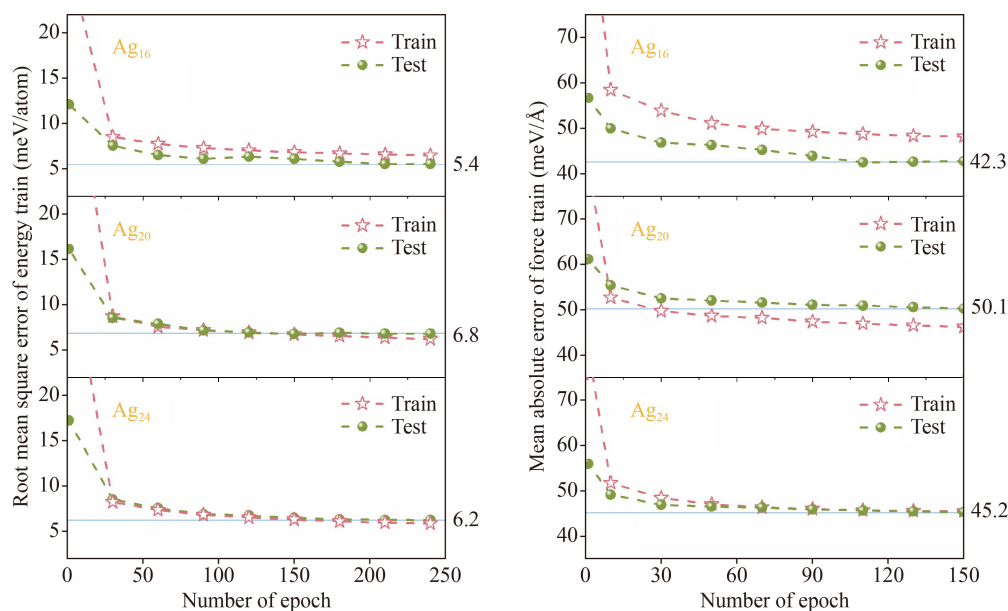
comparison, we have trained different networks for energy prediction on dataset of  $\text{Ag}_{20}$ . The ANN with two full connection layers after bond and angle feature extract layer has RMSE larger than 20 meV, while the graph convolutional network (GCN) also leads to test error larger than 20 meV. When we replace graph convolutional layer by graph attention layer, RMSE immediately decreases to 13 meV. Then, we improve the graph attention network by a series of testing. We introduce edge convolutional layer and angle convolutional layer to handle the original bond and angle information. Batch normalization layers are used to enhance the generalization ability and accelerate the training process. After testing various activation functions, the values of attention are gained by *LeakyReLU* function, the outputs of attention layer go through *Elu* function, and the activation function after full connection layer is chosen to be *LeakyReLU*. The numbers of initial bond and angle features are 36 and 24 respectively. The numbers of feature of first bond convolutional layer, angle convolutional layer, vertex convolutional layer and second bond convolutional layer are 18, 26, 30 and 32 respectively. It should be noticed that the number of these parameters is not the larger the better. At last, the test error for energy prediction net decreases to 6.8 meV and the architecture of CGANet is fixed. Using the same CGANet framework, the test losses of energy for  $\text{Ag}_{16}$  and  $\text{Ag}_{24}$  also approach about 6 meV/atom [Fig. 3(a)] after training of 100–150 epochs. Similarly, the network of force prediction is constructed and the test losses of force are 42.3, 50.1 and 45.2 meV/Å for  $n = 16, 20$  and 24, respectively [Fig. 3(b)].

It is crucial for a net to be transferable to different sizes. Therefore, we combine the datasets of  $\text{Ag}_{16}$ ,  $\text{Ag}_{20}$

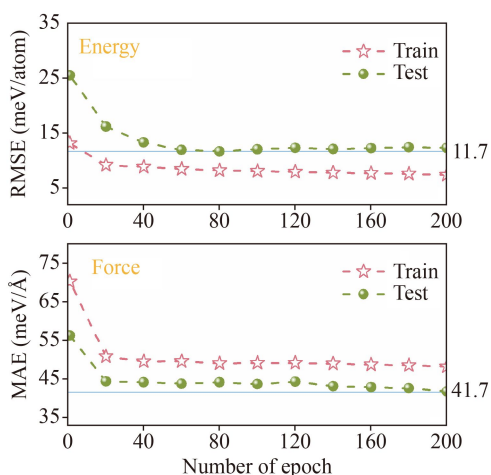
and  $\text{Ag}_{24}$  together as the train set and use the dataset of other sized  $\text{Ag}_n$  clusters ( $n = 15\text{--}27$ , excluding 16, 20, 24) as the test set for a mixed-size train. In this case, the test error of energy is obtained as 11.7 meV/atom (Fig. 4), which is larger than those from single-size trains, but still lower than those of ANI (26.0 meV) [16] and SchNet (23.8 meV) [23]. Differently, the test loss of force in the mixed-size train does not increase compared to single-size trains, but slightly reduces to 41.7 meV/Å (Fig. 4). This value is comparable to the reported RMSE values of 39 meV/Å in Cu–Zn binary cluster by HDNNP [46] and 43.4 meV/Å in organic molecules by GDML [47]. Overall speaking, the test errors of energy and force from mixed-size train are satisfactory and the trained CGANet should be able to describe the PES of  $\text{Ag}_n$  clusters of different sizes.

### 3.3 Structural prediction

Using the CGANet from mixed-sized train, we have performed global optimization of  $\text{Ag}_n$  clusters in size range of  $n = 14\text{--}26$  within the framework of genetic algorithm, as implemented by our CGA code. Briefly speaking, 128 configurations are generated from scratch as the initial population for each sized  $\text{Ag}_n$  cluster. Then, any two individuals in this population are chosen as parents to produce a child cluster via a “cut and splice” crossover operation, followed by an optional mutation operation of 35% probability. Each child cluster from CGA iteration is locally relaxed by a combined technique of quick molecular dynamic simulations of 20 steps at 300 K and numerical minimization with BFGS algorithm. Our test calculations demonstrate that this technique is rather efficient to avoid being trapped in shallow



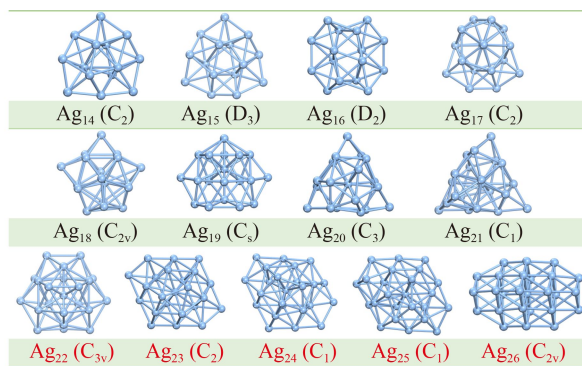
**Fig. 3** Root mean square error of energy (left) and mean absolute error of force (right) for  $\text{Ag}_{16}$ ,  $\text{Ag}_{20}$  and  $\text{Ag}_{24}$ .



**Fig. 4** Root mean square error (RMSE) of energy (upper) and mean absolute error (MAE) of force (lower) for mixed-size train of  $\text{Ag}_{16}$ ,  $\text{Ag}_{20}$  and  $\text{Ag}_{24}$ .

local minima on the PES. For each cluster size, the independent CGA search based on CGANet lasts for about 1000 iterations.

The global minima structures of  $\text{Ag}_n$  ( $n = 14\text{--}26$ ) are given in Fig. 5. To double check our finding, we have re-performed structural relaxation using Gaussian16 suite [48], with a metaGGA TPSS exchange-correlation functional [49] and the SDD basis set. As demonstrated by our test calculations on spin multiplicity (Supporting Information, Fig. S2), we choose the possibly lowest spin states for all  $\text{Ag}_n$  clusters, i.e., singlet for even-sized clusters and doublet for odd-sized ones. As presented in Table S1 of the Supporting Information, our test calculations demonstrate that the TPSS/SDD scheme is able to describe the binding energy, bond length, ionization potential, and vibrational frequency of  $\text{Ag}_2$  dimer well. First of all, we have reproduced the reported ground-state structures of silver clusters at  $n = 14\text{--}21$  from previous DFT studies [23, 41–47]. The ground state



**Fig. 5** Lowest-energy structures of  $\text{Ag}_n$  clusters ( $14 \leq n \leq 26$ ). The point group symmetry is given in parenthesis. Red color highlights the newly found lowest-energy structures, while black color denote those the same as in literature [21, 39–45].

structure of  $\text{Ag}_{14}$  is prolate with  $C_2$  symmetry, in agreement with previous works [39, 41, 43]. The ground state structure of  $\text{Ag}_{15}$  is similar to  $\text{Ag}_{14}$  but with  $D_3$  symmetry, and it can be viewed as the same structure reported by Tian *et al.* [39]. We find that the lowest-energy configuration of  $\text{Ag}_{16}$  also takes prolate shape with  $D_2$  symmetry. This structure is the same as Baishya’s result [41] but different from the other reports [39, 40, 42–44]. The lowest-energy configuration of  $\text{Ag}_{17}$  is identical to that reported by Baishya *et al.* [41], which can be obtained by adding 4 atoms on the waist of  $\text{Ag}_{13}$  icosahedron. The present lowest-energy structure  $\text{Ag}_{18}$  takes a distort star shape with  $C_s$  symmetry, and it has been reported by Baishya and McKee [44]. For  $\text{Ag}_{19}$ , We find a  $C_3$  structure that coincides with Yin and Du’s structure for  $\text{Ag}_{19}^-$  anion [45]. It is 0.13 eV and 0.11 eV lower than that reported by Chen [42] and Baishya [41], respectively, using DMol<sup>3</sup> calculation. However, at level of TPSS/SDD, the structure reported by Chen [42] is more energetically favorable. Similar to  $\text{Ag}_{19}$ , our lowest-energy structure of  $\text{Ag}_{20}$  is a twist tetrahedrons with  $C_3$  symmetry, which has been reported by Du [45] and McKee [44] respectively. The ground state structure of  $\text{Ag}_{21}$  found by us also takes twist tetrahedrons configuration and it can be considered as the same one in previous studies [39, 45].

More excitingly, unprecedented lowest-energy structures of  $\text{Ag}_n$  clusters at  $n = 22\text{--}26$  are obtained from our CGA search with CGANet. For  $n = 22\text{--}26$ , the lowest-energy structures from present work and those reported in recent studies [21, 39–45] are compared in Fig. S3 of the Supporting Information. The newly found structures energetically prevail those in literature by 43–960 meV and 70–702 meV at PBE/DND and TPSS/SDD level of theory, respectively. The ground state structure of  $\text{Ag}_{22}$  found here has  $C_3$  symmetry, which is different with all previous structures but similar with  $\text{Ag}_{23}$  structure reported by Chen [42]. It is also 0.35 eV lower than McKee’s structure [44] at TPSS/SDD level. All the lowest-energy structures of  $\text{Ag}_{23\text{--}25}$  find by us are 3-layer frustum with bottom face of 14 Ag atoms. We compare them with McKee’s structures [44] using TPSS/SDD calculations, and all the present structures are 0.2–1.2 eV lower in energy. Our lowest-energy structure of  $\text{Ag}_{26}$  adopts the configuration of a triangular bi-pyramid by removing 4 atoms at both two vertices and waist. This structure is 0.43 eV lower than the structure reported by Chen [42]. Based on the lowest-energy structures of  $\text{Ag}_n$  ( $n = 14\text{--}26$ ), the vertical ionization potentials are calculated and with experimental data [50] in Fig. S4 of the Supporting Information. The satisfactory agreement demonstrates the validity of our theoretical scheme.

Compared to the conventional DFT calculations using DMol<sup>3</sup>, CGANet is faster by at least two orders of magnitude but with comparable accuracy (11.7 meV/atom RMSE in energy). Combined with genetic algorithm or other global minimization approach, it can provide

robust description of the PES and allows more efficient search of the ground state structures. Our further investigation on large  $\text{Ag}_n$  clusters up to  $n = 60$  using CGANet is still underway and the results will be reported elsewhere.

### 3.4 Properties prediction

In addition to energy and force, it is desirable to predict the physical or chemical properties of a cluster using the developed graph neural network. As key electronic properties, here we consider the gap between highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) as well as the HOMO energy (whose negative value is approximately the vertical ionization energy according to the Koopman's theorem [51]). For either HOMO–LUMO gap or HOMO energy, the CGANet is the same as that for energy prediction described before, just by replacing the object function from energy to the desired property. Taking  $\text{Ag}_{20}$  as a representative, we train HOMO energy and HOMO–LUMO gap individually and the results are presented in Fig. 6. Most intermediate structures during geometry optimization are kept as the dataset, while about 18% of them with unreasonable binding energy are discarded. After 90 epochs train, the MAE of HOMO–LUMO gap for test set reaches 45.8 meV (Fig. 6), while the range of HOMO–LUMO gap values lies between 0.002 to 1.703 eV. For comparison, the MAE for HOMO–LUMO gap on QM9 dataset is 32.1 meV of DeepMoleNet and 34.8 meV of DimNet, respectively, and those for other models range between 56 to 69 meV. On the other hand, MAE of HOMO energy for test set is 50.0 meV, which is only about 1% of the absolute HOMO value (−4.217 eV in average). It is worthy to mention that the present CGANet is not limited to orbital energy, but can be extended to describe other properties such as polarizability, spectral

characteristics, NMR chemical shift, magnetic moment, magnetic anisotropic energy, adsorption energy and chemical reactivity. The capability of directly predicting the chemical or physical property from atomic structure is essential for rapid screening of functional clusters or molecules.

## 4 Conclusion

To summarize, a graph network for atomic clusters, namely CGANet, is built by aggregation layers of vertex, edge, bond angle and force information using attention mechanism. It takes only element type and coordinates as input and is able to predict the binding energy and force with satisfactory precision. Combining a genetic algorithm code for global search, CGANet can reproduce the DFT potential energy surface of  $\text{Ag}_n$  clusters with hundredfold acceleration. Despite our train set only includes data at selected sizes of  $n = 16, 20$  and  $24$ , it is successful for an extended size range of  $n = 14$ – $26$ . Especially, a number of unprecedented lowest-energy structures have been obtained for  $\text{Ag}_n$  clusters at  $n = 22$ – $26$ . Furthermore, CGANet can predict HOMO energy and HOMO–LUMO gap with MAE of about 50 meV, showing its capability of describing other physical and chemical properties. Although the specific way of constructing the aggregation layers of vertex, edge, bond angle and generating the dataset may vary with the system, the idea of using current graph attention network is universal and extendable to any atomic clusters and perhaps other materials. We anticipate that the developed CGANet can be widely used for rapidly searching the ground state structures of large clusters and reversal design of functional clusters with desired properties.

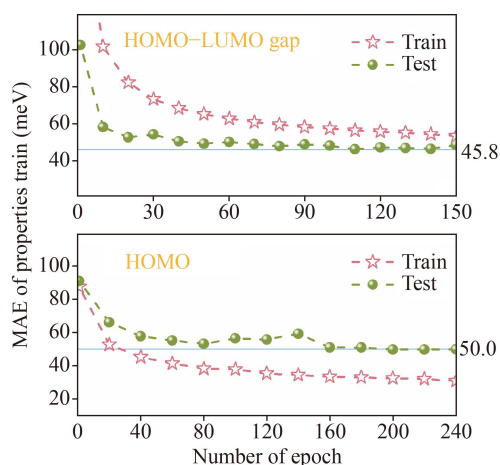
**Electronic supplementary material** Supplementary materials are available in the online version of this article at <https://doi.org/10.1007/s11467-022-1219-5> and <https://journal.hep.com.cn/fop/EN/10.1007/s11467-022-1219-5> and are accessible for authorized users. The details of the newly discovered lowest-energy structures and metastable structures of  $\text{Ag}_n$  clusters, including structural information and energy difference compared with the lowest-energy structures reported previously.

**Notes** The authors declare no competing financial interests.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (Grant Nos. 11804076 and 91961204), the Fundamental Research Funds for the Central Universities of China (No. B210202151), and the Changzhou Science and Technology Plan (No. CZ520012712).

## References

1. W. Gong and Q. Yan, Graph-based deep learning



**Fig. 6** Mean absolute error of HOMO energy (lower) and HOMO–LUMO gap (upper) trained for  $\text{Ag}_{20}$ .

- frameworks for molecules and solid-state materials, *Comput. Mater. Sci.* 195, 110332 (2021)
2. P. Friederich, F. Hase, J. Proppe, and A. Aspuru-Guzik, Machine-learned potentials for next-generation matter simulations, *Nat. Mater.* 20(6), 750 (2021)
  3. A. C. Mater and M. L. Coote, Deep learning in chemistry, *J. Chem. Inf. Model.* 59(6), 2545 (2019)
  4. L. Pattanaik, J. B. Ingraham, C. A. Grambow, and W. H. Green, Generating transition states of isomerization reactions with deep learning, *Phys. Chem. Chem. Phys.* 22(41), 23618 (2020)
  5. C. Coley, W. Jin, L. Rogers, T. Jamison, T. Jaakkola, W. Green, R. Barzilay, and K. Jensen, A graph-convolutional neural network model for the prediction of chemical reactivity, *Chem. Sci. (Camb.)* 10(2), 370 (2019)
  6. F. Nikitin, O. Isayev, and V. Strijov, DRACON: Disconnected graph neural network for atom mapping in chemical reactions, *Phys. Chem. Chem. Phys.* 22(45), 26478 (2020)
  7. Y. Ouyang, C. Yu, G. Yan, and J. Chen, Machine learning approach for the prediction and optimization of thermal transport properties, *Front. Phys.* 16(4), 43200 (2021)
  8. J. R. Kitchin, Machine learning in catalysis, *Nat. Catal.* 1(4), 230 (2018)
  9. C. McGill, M. Forsuelo, Y. Guan, and W. H. Green, Predicting infrared spectra with message passing neural networks, *J. Chem. Inf. Model.* 61(6), 2594 (2021)
  10. D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks, *Phys. Rev. Res.* 2(3), 033429 (2020)
  11. A. Khan, V. Ghorbanian, and D. Lowther, Deep learning for magnetic field estimation, *IEEE Trans. Magn.* 55(6), 1 (2019)
  12. B. Sanchez-Lengeling, J. N. Wei, B. K. Lee, R. C. Gerkin, A. Aspuru-Guzik, and A. B. Wiltschko, Machine learning for scent: Learning generalizable perceptual representations of small molecules, arXiv: 1910.10685 (2019)
  13. J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, Graph neural networks: A review of methods and applications, *AI Open* 1, 57 (2020)
  14. J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, *Phys. Rev. Lett.* 98(14), 146401 (2007)
  15. J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, *J. Chem. Phys.* 134(7), 074106 (2011)
  16. J. S. Smith, O. Isayev, and A. E. Roitberg, ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost, *Chem. Sci. (Camb.)* 8(4), 3192 (2017)
  17. X. Gao, F. Ramezanghorbani, O. Isayev, J. S. Smith, A. E. Roitberg, and A. N. I. Torch, A free and open source PyTorch-based deep learning implementation of the ANI neural network potentials, *J. Chem. Inf. Model.* 60(7), 3408 (2020)
  18. Z. L. Glick, D. P. Metcalf, A. Koutsoukas, S. A. Spronk, D. L. Cheney, and C. D. Sherrill, AP-Net: An atomic-pairwise neural network for smooth and transferable interaction potentials, *J. Chem. Phys.* 153(4), 044112 (2020)
  19. R. Lot, F. Pellegrini, Y. Shaidu, and E. Küçükbenli, PANNA: Properties from artificial neural network architectures, *Comput. Phys. Commun.* 256, 107402 (2020)
  20. R. Modee, S. Laghuvarapu, and U. D. Priyakumar, Benchmark study on deep neural network potentials for small organic molecules, *J. Comput. Chem.* 43(5), 308 (2022)
  21. L. Cao, P. Wang, L. Sai, J. Fu, and X. Duan, Artificial neural network potential for gold clusters, *Chin. Phys. B* 29(11), 117304 (2020)
  22. K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, Quantum-chemical insights from deep tensor neural networks, *Nat. Commun.* 8(1), 13890 (2017)
  23. K. T. Schütt, H. E. Sauceda, P. J. Kindermans, A. Tkatchenko, and K. R. Müller, SchNet – A deep learning architecture for molecules and materials, *J. Chem. Phys.* 148(24), 241722 (2018)
  24. J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Neural message passing for quantum chemistry, in: International Conference on Machine Learning (2017)
  25. N. Lubbers, J. S. Smith, and K. Barros, Hierarchical modeling of molecular energies using a deep neural network, *J. Chem. Phys.* 148(24), 241715 (2018)
  26. T. Xie and J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, *Phys. Rev. Lett.* 120(14), 145301 (2018)
  27. O. T. Unke, and M. Meuwly, PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges, *J. Chem. Theory Comput.* 15(6), 3678 (2019)
  28. C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, Graph networks as a universal machine learning framework for molecules and crystals, *Chem. Mater.* 31(9), 3564 (2019)
  29. C. Lu, Q. Liu, C. Wang, Z. Huang, P. Lin, and L. He, Molecular property prediction: A multilevel quantum interactions modeling perspective, in: Association for the Advancement of Artificial Intelligence (2019)
  30. J. Klicpera, J. Groß, and S. Günnemann, Directional message passing for molecular graphs, in: International Conference on Learning Representations (2020)
  31. C. Qian, Y. Xiong, and X. Chen, Directed graph attention neural network utilizing 3D coordinates for molecular property prediction, *Comput. Mater. Sci.* 200, 110761 (2021)
  32. Z. Liu, L. Lin, Q. Jia, Z. Cheng, Y. Jiang, Y. Guo, and J. Ma, Transferable multilevel attention neural network for accurate prediction of quantum chemistry properties via multitask learning, *J. Chem. Inf. Model.* 61(3), 1066 (2021)
  33. D. Bahdanau, K. Cho, and Y. Bengio, Neural machine translation by jointly learning to align and translate, in: International Conference on Learning Representations



- (2014)
34. T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, arXiv: 1609.02907 (2016)
  35. P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, Graph attention networks, arXiv: 1710.10903 (2017)
  36. J. Zhao, R. Shi, L. Sai, X. Huang, and Y. Su, Comprehensive genetic algorithm Forab intioglobal optimisation of clusters, *Mol. Simul.* 42(10), 809 (2016)
  37. B. Delley, From molecules to solids with the DMol<sup>3</sup> approach, *J. Chem. Phys.* 113(18), 7756 (2000)
  38. J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* 77(18), 3865 (1996)
  39. D. Tian, H. Zhang, and J. Zhao, Structure and structural evolution of Ag<sub>n</sub> ( $n = 3$ –22) clusters using a genetic algorithm and density functional theory method, *Solid State Commun.* 144(3–4), 174 (2007)
  40. M. Harb, F. Rabilloud, D. Simon, A. Rydlo, S. Lecoul-tre, F. Conus, V. Rodrigues, and C. Felix, Optical absorption of small silver clusters: Ag<sub>n</sub> ( $n = 4$ –22), *J. Chem. Phys.* 129(19), 194108 (2008)
  41. K. Baishya, J. C. Idrobo, S. Öğüt, M. Yang, K. Jackson, and J. Jellinek, Optical absorption spectra of intermediate-size silver clusters from first principles, *Phys. Rev. B* 78(7), 075439 (2008)
  42. M. Chen, J. E. Dyer, K. Li, and D. A. Dixon, Prediction of structures and atomization energies of small silver clusters, (Ag)<sub>n</sub>,  $n < 100$ , *J. Phys. Chem. A* 117(34), 8298 (2013)
  43. M. Liao, J. D. Watts, and M. Huang, Theoretical comparative study of oxygen adsorption on neutral and anionic Ag<sub>n</sub> and Au<sub>n</sub> clusters ( $n = 2$ –25), *J. Phys. Chem. C* 118(38), 21911 (2014)
  44. M. L. McKee and A. Samokhvalov, Density functional study of neutral and charged silver clusters Ag<sub>n</sub> with  $n = 2$ –22: Evolution of properties and structure, *J. Phys. Chem. A* 121(26), 5018 (2017)
  45. B. Yin, Q. Du, L. Geng, H. Zhang, Z. Luo, S. Zhou, and J. Zhao, Superatomic signature and reactivity of silver clusters with oxygen: double magic Ag<sub>17</sub><sup>−</sup> with geometric and electronic shell closure, *CCS Chemistry* 3(12), 219 (2021)
  46. J. Weinreich, A. Römer, M. L. Paleico, and J. Behler, Properties of  $\alpha$ -brass nanoparticles (1): Neural network potential energy surface, *J. Phys. Chem. C* 124(23), 12682 (2020)
  47. S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K. R. Müller, Machine learning of accurate energy-conserving molecular force fields, *Sci. Adv.* 3(5), e1603015 (2017)
  48. M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, General atomic and molecular electronic structure system, *J. Comput. Chem.* 14(11), 1347 (1993)
  49. J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids, *Phys. Rev. Lett.* 91(14), 146401 (2003)
  50. G. Alameddine, J. Hunter, D. Cameron, and M. M. Kappes, Electronic and geometric structure in silver clusters, *Chem. Phys. Lett.* 192(1), 122 (1992)
  51. T. Koopmans, Ordering of wave functions and eigenenergies to the individual electrons of an atom, *Physica* 1, 104 (1933)