Computational Materials Science 183 (2020) 109803

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Vacancy formation energy and its connection with bonding environment in solid: A high-throughput calculation and machine learning study

YingXing Cheng[a,b], Linggang Zhu[a,b,*], Guanjie Wang[a,b], Jian Zhou[a], Stephen R. Elliott[c], Zhimei Sun[a,b,*]

[a] School of Materials Science and Engineering, Beihang University, Beijing 100191, China
[b] Center for Integrated Computational Materials Engineering, International Research Institute for Multidisciplinary Science, Beihang University, Beijing 100191, China
[c] Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

ARTICLE INFO

Keywords:
High-throughput calculation
Machine learning
Vacancy
Chemical bonding
First-principles calculation

ABSTRACT

The generation of the vacancy involving the bond breaking/re-formation occurs naturally in the material. Here, we present a framework for automatically computing the vacancy-formation energy (E_f) and for analyzing the bonding environment concealed in the E_f by using an artificial neural network (ANN). The 'effective' bonding that determines the energy of the system and the E_f will be clarified. The phase-change memory material GeTe is used as a case study. Firstly, 791 Ge-vacancy containing GeTe structures are studied and a large data set of the formation energy of the Ge-vacancy is obtained, which is helpful to understand the vacancy-induced issue of the amorphous GeTe including the resistance drift, etc. By using the ANN fitting based on the large energy data set, a bonding picture that is applicable to both the crystalline and the amorphous state of GeTe is predicted. In terms of the contribution to the formation energy of the vacancy, the weight ratio of the bond with length of 3.0–3.6 Å and 3.6–4.5 Å can be approximated as 6:1. The bonding information is further confirmed by using the first-principles electronic structure analysis on the randomly chosen samples. The bonding analysis using the ANN method based on a large vacancy-formation-energy data set is demonstrated to be a novel alternative technique to understand the bonding in the material. The proposed framework can be applied to a wide range of materials.

1. Introduction

As a ubiquitous defect, vacancy can significantly change the properties of the material. In practice, the concentration of the vacancies can be tuned by design to engineer the performance of the material [1–3]. The formation energy of a vacancy is defined as the energy change on breaking the bonds of one atom with its ligands in the parent material and forming new bonds with ligands in the reference system. The formation energy of a vacancy is widely used to evaluate the probability of vacancy formation as well as its concentration. Moreover, as suggested by its definition, the formation energy of a vacancy can also be used as a probe to study the bonding environment in the materials. However, in the literature, the in-depth analysis of extensive vacancy-formation energy data are less focused than on the energy data themselves. Here, we have implemented a high-throughput, first-principles calculation and the machine-learning analysis to investigate the vacancy in the material. In addition to the extensive data on the formation energy that can be quite easily obtained, physical information which is computationally expensive to obtain, such as the bonding environment in the material, can be investigated by using the machine-learning study.

Pseudo-binary compounds on the compositional tie-line between GeTe and Sb_2Te_3, are widely used in the phase-change random-access memory (PCRAM) [4]. For the phase-change material, the crystalline and amorphous state representing logic states 1 and 0, respectively, can be rapidly and reversibly switched under an external field, which is utilized for the information storage [5–9]. The study of the bonding in the phase-change material, which is believed to be the underlying mechanism responsible for the rapid phase-change, has attracted tremendous efforts of the researchers [10–15]. On the other hand, the vacancy in the phase-change material has been demonstrated to exist at a quite high concentration in the crystalline state and also the vacancy plays a vital role in the fast phase transition from the crystalline to the amorphous state [16,17]. Here, a prototype phase-change material GeTe is used as a case study for the proposed framework. Crystalline GeTe can be regarded as having a distorted rock-salt geometry with an elongation along the [111] direction, and for the six coordinated Te(Ge) atoms of a central Ge(Te) atom, three of them are closer to the Ge(Te) than the other three [18]. For the amorphous state of GeTe, the

* Corresponding authors at: School of Materials Science and Engineering, Beihang University, Beijing 100191, China.
E-mail addresses: lgzhu7@buaa.edu.cn (L. Zhu), zmsun@buaa.edu.cn (Z. Sun).

https://doi.org/10.1016/j.commatsci.2020.109803
Received 21 March 2020; Received in revised form 11 May 2020; Accepted 14 May 2020
Available online 01 June 2020

defective octahedral and tetrahedral coordination of Ge commonly exist, with the coordination number reducing to 3–5. The Ge-Ge homopolar bond, not existing in the crystalline state and thus named as 'wrong' bonds, is found in the tetrahedral fragment of the amorphous structure [11,19,20]. Breaking of the Ge-Ge bond during the structural relaxation or ageing is believed to account for the time-dependent resistance drift of the amorphous GeTe [11,21,22]. Ge vacancy (V_{Ge}) is found to have a concentration around $8\% - 10\%$ in the crystalline GeTe [23,24]. For the amorphous state, it has been shown that the removal of the tetrahedrally-coordinated Ge leading to the breaking of the Ge-Ge wrong bond, can actually stabilize the amorphous system [22], indicating a vital role of V_{Ge} in the amorphous GeTe considering the phase stability and the resistance drift.

The most straightforward way to investigate the bonding is by studying the pair-correlation function and the coordination number. However, as recently pointed out by Lee et al. [25], this conventional method ignores the electronic charge distribution, and is less reliable than an electron-density-based analysis using electron-localization functions or maximally localized Wannier functions, etc. Compared to all the previous studies, probing the bonding environment in the phase-change material by using the vacancy-formation energy, as proposed in the present work, can be considered as another different strategy. And we believe that such a vacancy-formation-energy based analysis takes into account all the probable interactions, including the elastic and chemical interactions between atoms.

2. Theoretical framework and method

Fig. 1 presents the work flow, as proposed in the present study. The framework consists of three parts: the automatic generation of the vacancy-containing structure; the high-throughput DFT (Density Functional Theory) calculation, including the structure relaxation and the vacancy-formation energy (E_f) calculation etc.; finally, the data analysis based on an artificial neural network (ANN). By using the ANN method, firstly fitting the relationship between the fingerprint and known energetic data (from the DFT calculation), the energy of the unknown structure can be predicted and the E_f by ANN can be obtained. Moreover, by tuning the fingerprint and comparing the resultant fitted energy data, the bonding environment in the material can be analyzed. Each of the three components of the framework will be introduced in detail in the following parts.

2.1. Generation of the vacancy-containing structure

First of all, it is worth noting that, for the amorphous structure, the word "void" is used in several papers to describe the space left by a missing atom, while here, for brevity, the word "vacancy" is used for both the crystalline and amorphous structures. To generate vacancy-containing structures, the key step is to identify an inequivalent atom with a different surrounding environment in a crystalline or amorphous system. Here, we employed a simplified strategy: within a specified cutoff radius, using the atomic type of the neighbor and the corresponding distance (denoted as "atomic-type: atomic-distance" pair) near an atom/vacancy to identify the atomic structure. Two equivalent atom/vacancy sites in the structure should have identical "atomic-type: atomic-distance" pairs within error limits. It is worth mentioning that the strategy described here can also be applied to generate substitutional defects, in which case the inequivalent site needs to be

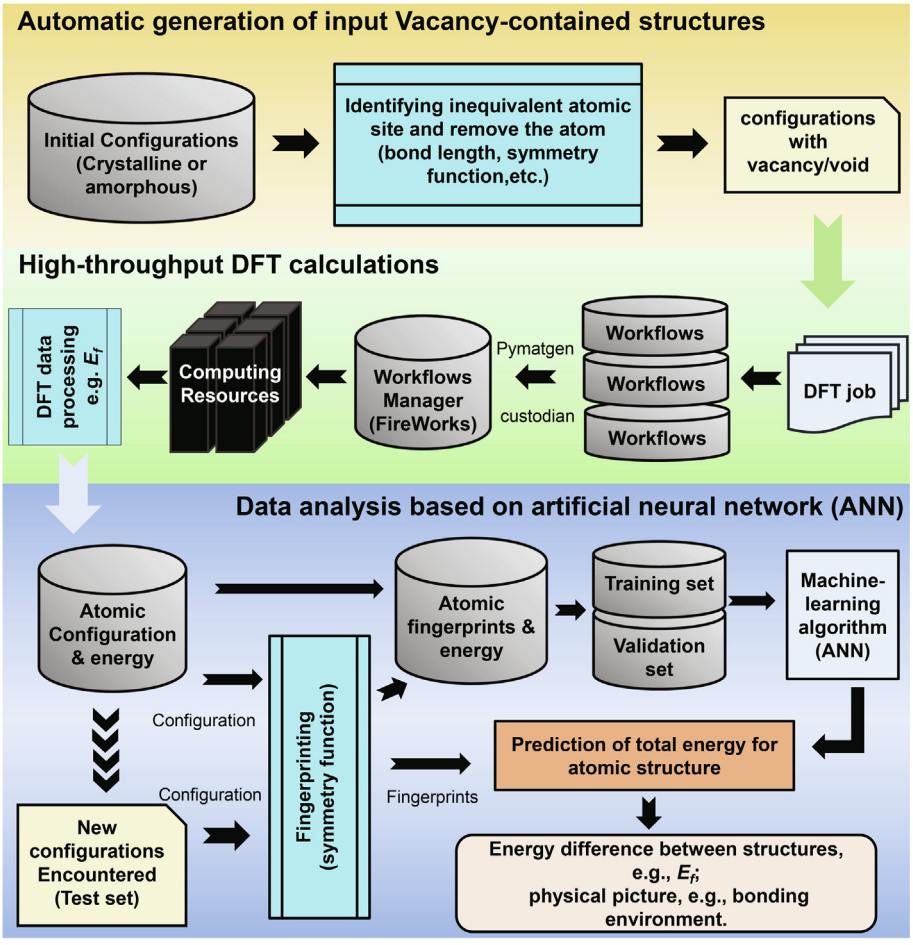

Fig. 1. Work flow implemented in the present study. E_f represents the formation energy of a vacancy.

determined first.

2.2. High-throughput calculation of the vacancy-formation energy

For the high-throughput calculation, we have incorporated the open-source package Atomate [26]. Atomate provides a high-level interface to generate the workflow based on Pymatgen [27], Fireworks [28], and Custodian, for the input structure generation, jobs management and error handling, respectively. The workflow for the evaluation of the vacancy-formation energy, including the structure optimization and energy calculation, using the first-principles calculation package VASP [29], can be easily customized. The formation energy of a vacancy (E_f) is calculated by using:

$$E_f = [E_V - E_0] + \mu_i \quad (1)$$

where E_V and E_0 are the total energy of the vacancy-containing and the initial configurations, respectively. μ_i is the chemical potential of the atom i removed from the initial structure that generate a vacancy, and it depends on the chemical environment where the atom is replaced [30]. For the first-principles calculation using VASP [29], the generalized-gradient approximation (GGA-PBE) [31] is employed for the exchange-correlation interaction between electrons. The cutoff energy for the plane wave is 400 eV. A $1 \times 1 \times 1$ k-point grid is employed for the optimization and electronic-structure calculation of the GeTe supercell containing 300 atoms.

2.3. Data analysis based on ANN

The artificial neural network (ANN), vaguely inspired by the biological neural network, is composed of interconnected nodes (neurons). A node will process the signal it receives, and then pass the result to the next node. From the input to the output of the network, the signal will pass through layers of nodes. ANN is able to fit any real-valued function with arbitrary accuracy [32]. As for the present study, in order to fit the relationship between the vacancy-formation energy and the atomic structure, an accurate evaluation of the total energy of the system (especially for the vacancy-containing structure) is a prerequisite. ANN combining specific symmetry function describing the atomic environments has shown the potential to reproduce the potential-energy surface (PES) of GeTe [33] and other materials [34,35]. The details for the specific symmetry function (input layer for ANN), proposed by Behler, can be found in their original paper [34]. Simply put, three radial symmetry functions describing the radial (coordination) environment of the central atom and two angular symmetry functions, which are a summation of cosine functions of the angles centered at the atom, are defined. Each symmetry function is multiplied by at least one cutoff function to ensure that the total symmetry function becomes zero in value and in slope at the cutoff radius. The cutoff radius needs to be carefully chosen to obtain an accurate fitting. Interestingly, later we will show that the bonding information in the material can be analyzed using the value of the cutoff radius. In our framework shown in Fig. 1, firstly, with the ANN method and the symmetry function as the fingerprint of the structure, the relationship between the atomic structure and its energy is fitted. In the present ANN implementation, all the samples are divided into three independent sets, namely, training set, validation set and test set. The training set is mainly for adjusting the weights of the connections between the neurons in the ANN; the validation set in the energy-fitting process makes it possible to detect the overfitting and select the best hyper-parameter, ANN architecture and the cutoff radius; the test set is used to estimate the performance of the model on unseen data and to confirm the actual predictive power of the network. Typically, the root mean-squared error (RMSE) and the mean absolute error (MAE) are used to evaluate the fitting performance [35]. Given the total energy of the vacancy-containing and initial structure, the formation energy of a defect can be calculated using Eq. (1). In the ANN analysis part of our workflow, in addition to the energy prediction, by playing around the cutoff radius, the effect of the bond with certain length on the energy can be analyzed. Thus finally a physical picture such as the bonding environment of the material can be obtained. The ANN analysis in the present work is performed using the Atomic Energy Network (ænet) software package [35,36], normally used for the construction of the atomic-interaction potential.

3. An application to the phase-change memory material GeTe

In the following parts, the above-mentioned general framework will be applied to the study of the phase-change memory material, GeTe. We will present an unprecedentedly comprehensive evaluation of the formation energy of V_{Ge} in GeTe, and analyze the bonding environment based on the energy data.

3.1. Generation of the GeTe data set

For the ANN study, a large reference data set for the system of interest is needed. Here, in order to obtain various GeTe structures in which the surrounding environment of the atoms is different, we adopt the structures generated during the crystallization process of the amorphous GeTe, as shown in our previous work [37], which mimics the transition from a memory state “0” to “1” of the PCRAM. The crystallization process is simulated using the ab initio molecular dynamics (AIMD) in a NVT ensemble maintained at 470 K (close to the experimentally measured crystallization temperature) [37]. The simulation starts from the structures of amorphous GeTe bonded to a “crystalline seed”, which are ordered GeTe layers sitting at the opposing faces of the supercell. The 300-atom supercell is firstly annealed for a few picoseconds to relax the sharp interface between the amorphous and crystalline GeTe [37]. Here, one atomic structure is picked out every 30 ps as the crystallization process proceeds, denoted as the annealing time of 0, 30, 60, 90, 120, 150, and 180 ps. Then, all the chosen structures are relaxed at 0K before being used as the initial structures for the V_{Ge} generation. The GeTe structure selected from the phase-transformation process contain a crystalline part, an amorphous part and the boundary between the two, which we believe [37] is helpful to obtain a general bonding picture applicable for both the crystalline and amorphous GeTe. Given the initial GeTe structure, a V_{Ge} is generated by removing an inequivalent Ge atom identified by comparing the “atomic-type: atomic-distance” pair in the surrounding environment as introduced in section 2.1. In total, 791 single-vacancy-containing structures are generated including the perfectly crystalline GeTe (with one inequivalent V_{Ge}).

By using the first-principles calculation, the formation energy of V_{Ge} (E_f), computed by using Eq. (1), is shown in Fig. 2(a), and here the chemical potential of Ge is set as the energy per atom in diamond Ge, -4.47 eV. The vacancy-formation energy, E_f, for crystalline GeTe is depicted as the dashed line, marked as $E_{vac} = 0.01$ eV. Energy values of $E_{vac} + 0.20$ and $E_{vac} - 0.40$ eV are also denoted to guide the eye. It can be seen that E_f spans over a wide range from -1.82 to 0.86 eV. Interestingly, the formation of V_{Ge} can be much easier in the disordered system than in the crystalline structure. The existence of V_{Ge} with a small formation energy in the amorphous-crystal mixed structure may account for the structural evolution or time-dependent resistance drift of the amorphous state of the phase-change material [11,24,22]. As the annealing time elapses and the crystallization of the amorphous GeTe proceeds, the range of the formation-energy data slightly shrinks, as seen from the overall decreasing trend of the standard deviation, σ, of the Gaussian fitting over time (Fig. 2(b)). The formation energy of V_{Ge} in the supercell cannot converge to E_{vac} in the crystal since a disordered region is inevitable due to the formation of the grain boundaries in the phase transformation process [37]. The surrounding atomic environment of the representative Ge vacancy with large or small E_f values is summarized in Fig. 2(c). In Fig. 2(c) the neighboring atom and its distance to the vacancy is shown, while the corresponding atomic

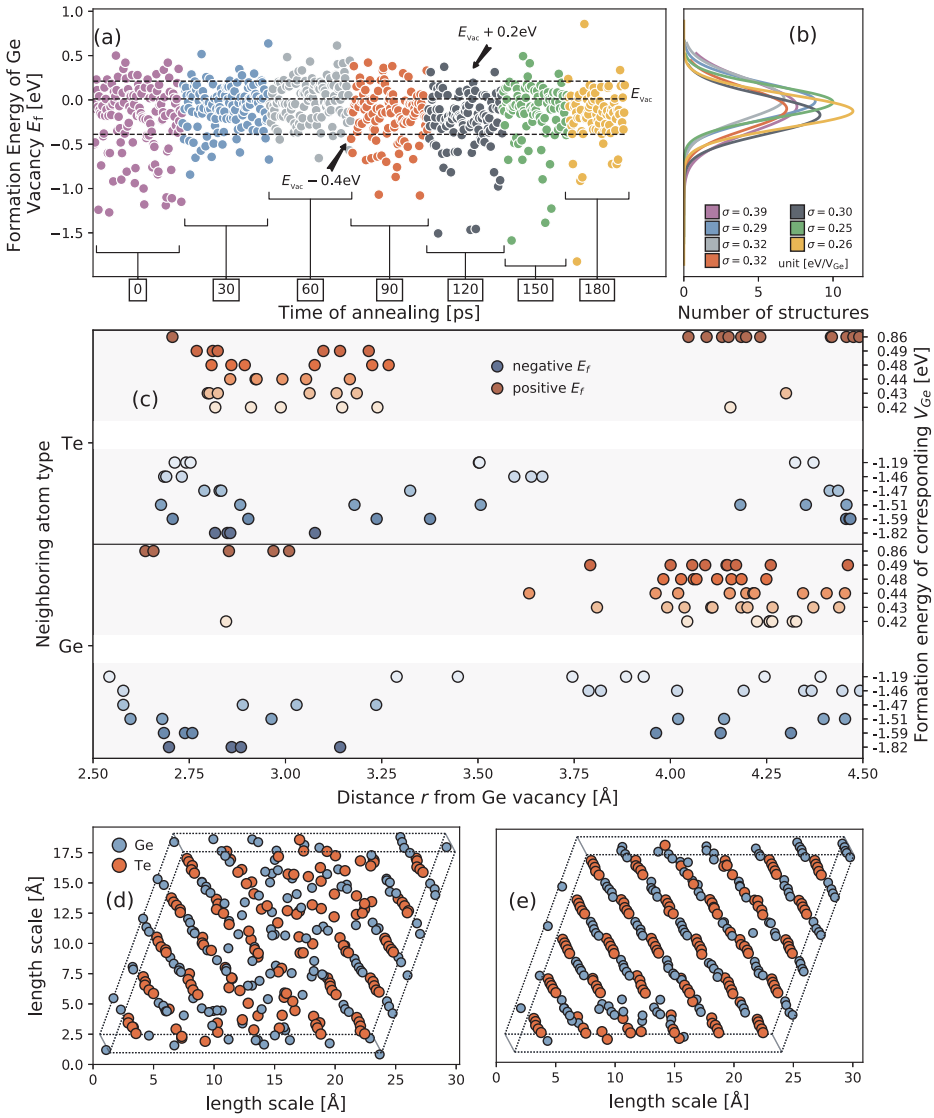

Fig. 2. All 791 formation energies of Ge vacancies in GeTe. (a) Formation energy of V_{Ge} in the GeTe structures under different annealing time, E_{vac} represents the formation energy of V_{Ge} in perfect crystalline GeTe; (b) Fitted Gaussian distribution of the formation energy of V_{Ge} and the standard deviation, σ, denoted by the same color scheme as in (a); (c) Neighboring atom and its distance to the Ge vacancy with positive (red circle) and negative (blue circle) value of E_f. The darker color indicates a larger absolute value of E_f. The corresponding atomic geometry where the vacancy is located can be found in Figs. S1 and S2 in the Supplemental material. (d) and (e) are the GeTe structures at annealing times of 0 ps and 180 ps, respectively. The initial structure is annealed for a few picoseconds firstly to relax the sharp interface, and then the time count is started.

geometry where the vacancy is located can be found in Figs. S1 and S2 in the Supplemental material. As shown in Fig. 2(c), in the case of a negative value of E_f (blue circle in the figure), Ge vacancy has more neighboring Ge atoms, especially in the distance range within 3.50 Å, i.e., Ge-Ge bonding is quite significant. Ge-Ge bond, so-called 'wrong bond' in GeTe, can destabilize the system. On the other hand, as demonstrated by the ANN fitting later (Fig. 4(b)), bond with length smaller than 3.6 Å (some Ge-Ge bonds locate within this range) can significantly affect the energetics of the system. Thus, both the vacancy-formation energy data and the ANN fitting highlights the critical role of the Ge-Ge bonding for the easy formation of the Ge vacancy. Vacancy in the phase-change material has quite a large impact on the electronic property of the material, and can even lead to an insulator-metal transition when the concentration of the vacancy is high [38]. Therefore, it can be concluded that the existence of the Ge-Ge contacts within 3.6 Å can lead to the easy formation of the Ge vacancy and finally may result in the resistance drift of the phase-change material. For a positive value of E_f (red circle in the figure), the corresponding Ge vacancy mainly has Te atoms as neighbors within a distance range of 3.25 Å. From 3.25 to 4.50 Å, the Ge vacancy is neighbored by the Ge atoms, like the coordination of Ge in the crystalline phase. From the atomic configuration where the vacancy is created, it is noticed that, in general, the negative value of E_f corresponds to a Ge atom in the interstitial region or in the complex defect, such as $Ge_{interstitial}$-vacancy-$Ge_{interstitial}$, as shown in Fig. S1 in the Supplemental material. In this case, significant local strain (expansion) may exist, and thus the formation of a vacancy, which can release the strain, will be energetically favorable. While for the positive value of E_f, the corresponding Ge vacancy has a similar environment to that of the crystal (Fig. S2 in the Supplemental material), with small local distortions in some cases. As the antisite defect exists (such as in the Te-rich region), the formation of a Ge vacancy will be more difficult, meaning an even larger value of E_f. Two representative structures at annealing time of 0 ps and 180 ps are shown in Figs. 2(d) and (e), respefctively, and such structures are utilized for the vacancy generation, as mentioned above.

The extensive structures and their DFT-calculated energies are then

used for the ANN study. k-fold cross validation is employed: the whole data set is divided into k groups, among which $k-1$ sets are for the training of the model and the other set is used for the test. In k-fold cross validation, any one of the k sets can be the test set; thus, in total, the learning/test process will be repeated k times, and the estimation is based on the average of the data in each time. In this work, 10-fold cross-validation is employed. We include a validation set in the fitting process, thus the ratio of the number of the samples for the training set, validation set and test set is 8: 1: 1. It is worth noting that, for the sake of fitting accuracy, here ANN is only used to fit the energy of the V_{Ge}-containing structure, given its large population (791 structures). While for the initial structure before the generation of the vacancy (only 8 structures are involved), the energy from the DFT calculation is adopted as the reference value. Then the formation energy of V_{Ge} by ANN is computed using Eq. (1).

3.2. Fingerprint selection for the GeTe system

In the present work, the fingerprint in the machine learning study is the parameter or hyper-parameter that can well describe the atomic environment. The symmetry function described in Section 2.2 is used as the fingerprint. The structural fingerprint constructed here comprises 8 radial symmetry functions for the interactions of Ge (Te) and 18 angular symmetry functions for the interactions of Ge-Te, Ge-Ge, and Te-Te atom pairs. Thus, the dimension of the input layer of the ANN network is $70(2\times 8+3\times 18)$. Here, the radial basis function of type G^2 and the angular basis function of type G^5, following Behler's original notation [34], are chosen. The radial function centered at atom i is given by:

$$G_i^2=\sum_{j\neq i} e^{-\eta(R_{ij}-R_s)^2}\cdot f_c(R_{ij}) \tag{2}$$

where η and R_s are adjustable parameters, R_{ij} is the distance between atom i and j. Only atoms within a certain distance R_c are taken into consideration, which is realized through a cosine cutoff function given by:

$$f_c(R_{ij})=\begin{cases}0.5\left[cos\left(\frac{\pi R_{ij}}{R_c}\right)+1\right], & R_{ij}\leqslant R_c\\ 0, & R_{ij}>R_c.\end{cases} \tag{3}$$

The angular three-body function, centered at atom i, is defined as:

$$G_i^5=2^{1-\zeta}\sum_{j\neq i}^{all}(1+\lambda cos\theta_{ijk})^{\zeta}\cdot e^{-\eta(R_{ij}^2+R_{ik}^2)}\cdot f_c(R_{ij})\cdot f_c(R_{ik}), \tag{4}$$

where R_{ij}, R_{ik} represent the atomic distances between atom i, j and k;θ_{ijk} is the angle defined by the three atoms, and ζ, λ and η are adjustable parameters. For the present work, all the parameters in the above-mentioned symmetry function are listed in Table S1 and Table S2 in the Supplemental material, which have been successfully applied to various polymorphs of the binary compound TiO_2 [35] and Cu-Au nanoalloys [39]. These parameters are to control the accuracy of the chosen symmetry functions (70 functions in total in this work), determining the input layer of the ANN. It seems that the fitting result may not be sensitive to these parameters, since the same parameters can be used for the two distinct binary systems TiO_2 and Cu-Au.

3.3. ANN-architecture construction

To construct an ANN, its architecture determined by the number of layers in the network and the number of nodes per layer should be designed first. An example of structure of ANN is shown in Fig. 3, in which the first layer is input layer that receive symmetry function values as input and these input are transported to hidden layer with several unknown so-called weights and bias to hidden layers. There are also unknown weights and bias between different hidden layers to be determined and finally these hidden layers transmit to generate output. The main task of ANN is to optimize these unknown weights and bias. An ANN with too few model parameters cannot well describe all the reference structures and the fitting might be inaccurate, while too many model parameters may lead to an overfitting and normally demand more computer time [35]. Thus, tests are needed to construct a suitable architecture. With the present ANN, the total energy of the V_{Ge}-containing GeTe is fitted. To estimate the accuracy of the ANN, the root-mean-squared error (RMSE) of the ANN energy with respect to the energy from the DFT calculation is computed. It is worth emphasizing that, for the computation of the formation energy of V_{Ge} using Eq. (1), the energy of the initial structure E_0 (8 structures in total in the present calculation) and the chemical potential of Ge atom μ_i are from the DFT calculation. Thus, the RMSE of the fitted energy for the V_{Ge}-containing GeTe structure equals the RMSE of E_f. Fig. 4(a) shows the RMSE of E_f with respect to the DFT energy of the structure in the validation set after 1000 batch L-BFGS (limited-memory Broyden-Fletcher-Goldfard-Shanno) training iterations. For a clearer presentation, the figure only shows data for the 70-M-N-1 ANN architecture, i.e., two hidden layers, M and N, are included, with M-N = 2–2, 5–5, 10–10, 15–15, 20–20, 20–10, 20–15, 25–25 and 30–30. As shown in Fig. 4(a), RMSE decreases rapidly when M-N increases from 2–2 to 5–5, while for $N>5$, the improvement is less significant. For a too large ANN size, the standard deviation of the RMSE (error bars in Fig. 4(a)) increases, which might be due to the overfitting [35]. Based on the results of the parameter test for the validation sets in Fig. 4 and a consideration of the computer time, we finally employed a 70-15-15-1 ANN architecture (including 1321 weight parameters).

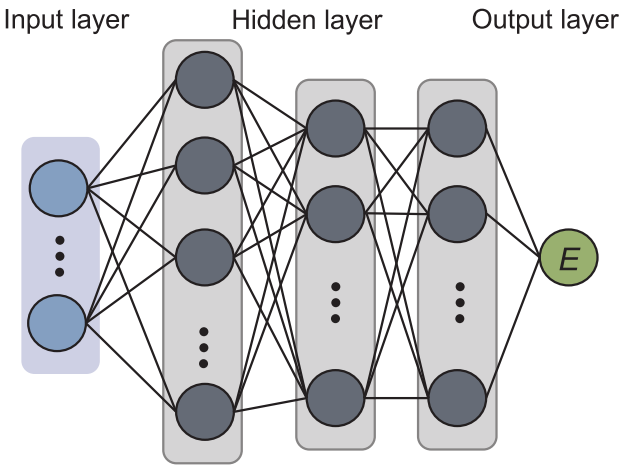

Fig. 3. Graph representation of a multilayer artificial neural network, in which blue nodes construct an input layer; three grey blocks including several grey nodes represent three different hidden layers and the output layer has only one node as energy output, i.e., green node.

3.4. Cutoff-radius test for the symmetry function

In addition to the symmetry-function type and the ANN architecture, the cutoff radius is another vital parameter to be determined. Clearly, the cutoff radius should be large enough to include all the effective interactions between an atom and its neighbors. However, if the radius is too large, more symmetry functions or a more complex ANN architecture should be needed, which will complicate the fitting process and require more computer resources. With the chosen symmetry function and the ANN architecture as discussed above, different cutoff radius is tested and the resultant RMSE is shown in Fig. 4(b). It can be seen that the RMSE of the validation set firstly exhibits a sharp decrease as the cutoff radius increases from 3.0 to 3.6 Å, and then the increasing slope becomes much smaller in the region between 3.6 and 4.5 Å. The

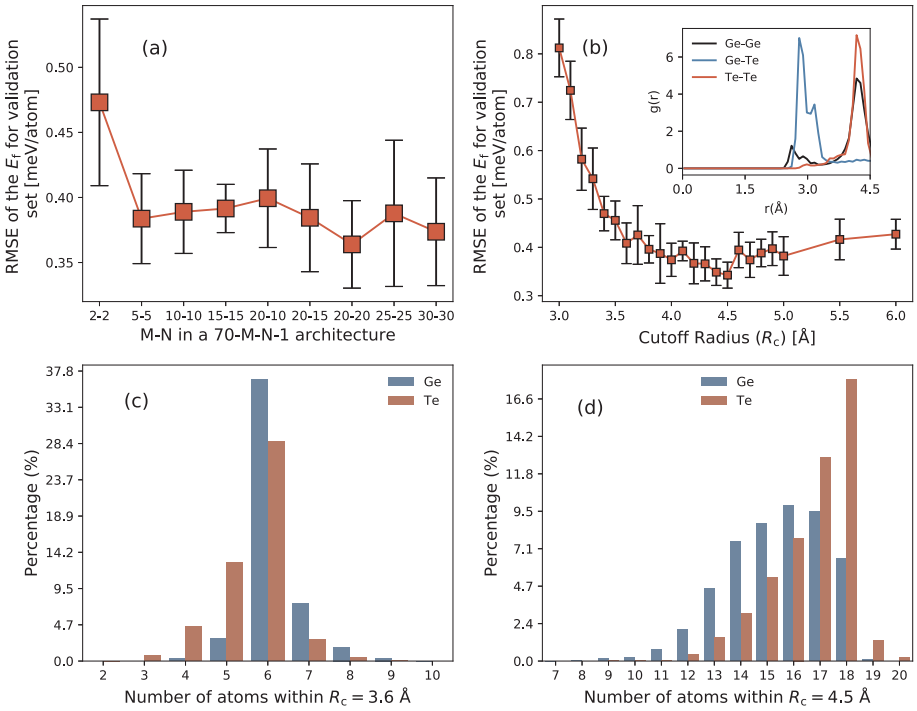

Fig. 4. (a) RMSE (root mean squared error) of the ANN energy for the structures in the validation set after 1000 L-BFGS iterations, as the function of the ANN architecture. The data is the mean value of 10-folds training runs, and the standard deviation is shown as the error bar. (b) RMSE of the ANN energy for the structures in the validation set after 1000 L-BFGS iterations, as function of the cutoff radius. The data is the mean value of 10-folds training runs, and the standard deviation is shown as the error bar. The inset shows the radial distribution function (g(r)) of the seven initial atomic structures in which the vacancy is created. (c)–(d) Statistic of the number of the neighboring atoms of central Ge atom/vacancy within the cutoff radius of 3.6 Å and 4.5 Å, respectively, based on all the training configurations.

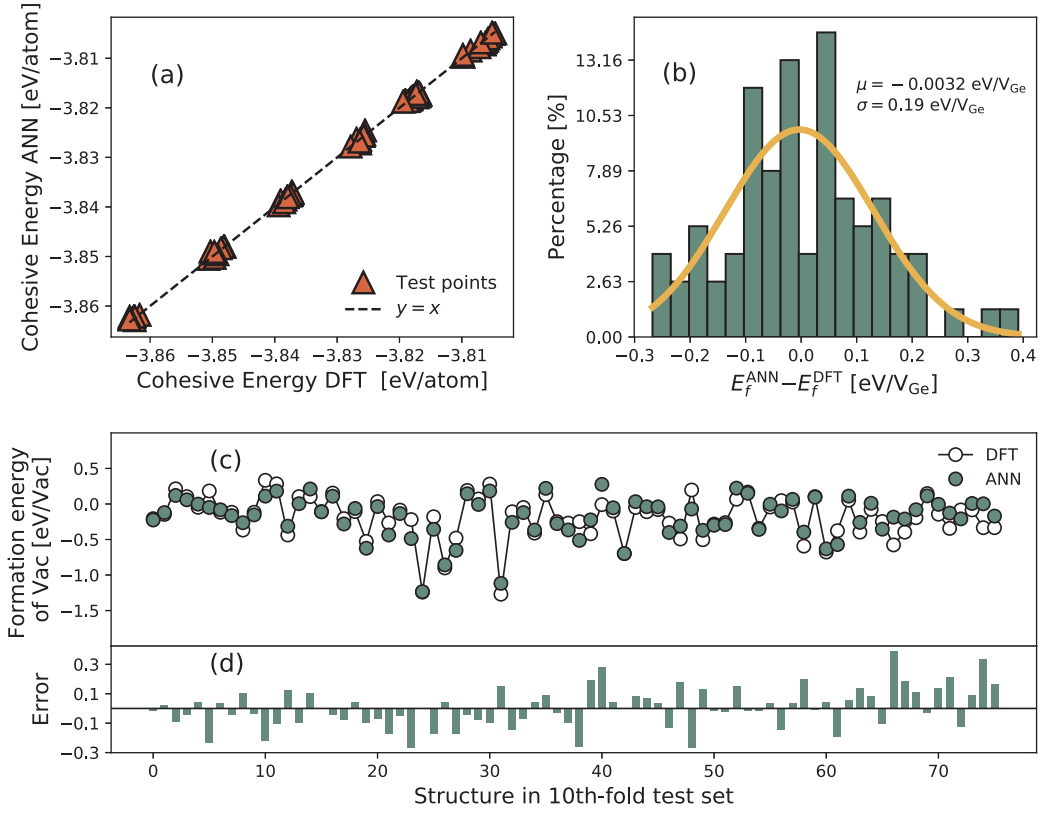

Fig. 5. (a) Cohesive energy for the test set by ANN and DFT method, and $y = x$ function is drawn to guide the eye, (b) Statistics of the E_f difference between ANN and DFT method, result of the Gaussian fitting is shown using a yellow line, with expectation value $\mu = -0.0032$ eV/V_{Ge} and standard deviation $\sigma = 0.19$ eV/V_{Ge}, (c) and (d) shows the concrete data of E_f by ANN and DFT method and the error (energy difference) for different structures in the test set.

RMSE almost converges at a cutoff radius of 4.5 Å, and further increase of the radius leads to an increase of the RMSE, which might be due to the fact that such a large radius may require a more complicated ANN architecture or due to the overfitting. Considering that the RMSE is stabilized at a very low value for a cutoff radius of 4.5 Å, we believe a larger cutoff radius, or a more complex ANN architecture, is unnecessary for the present study. The changing of the RMSE over the cutoff radius (first decreasing then increasing) is rather 'typical' for the

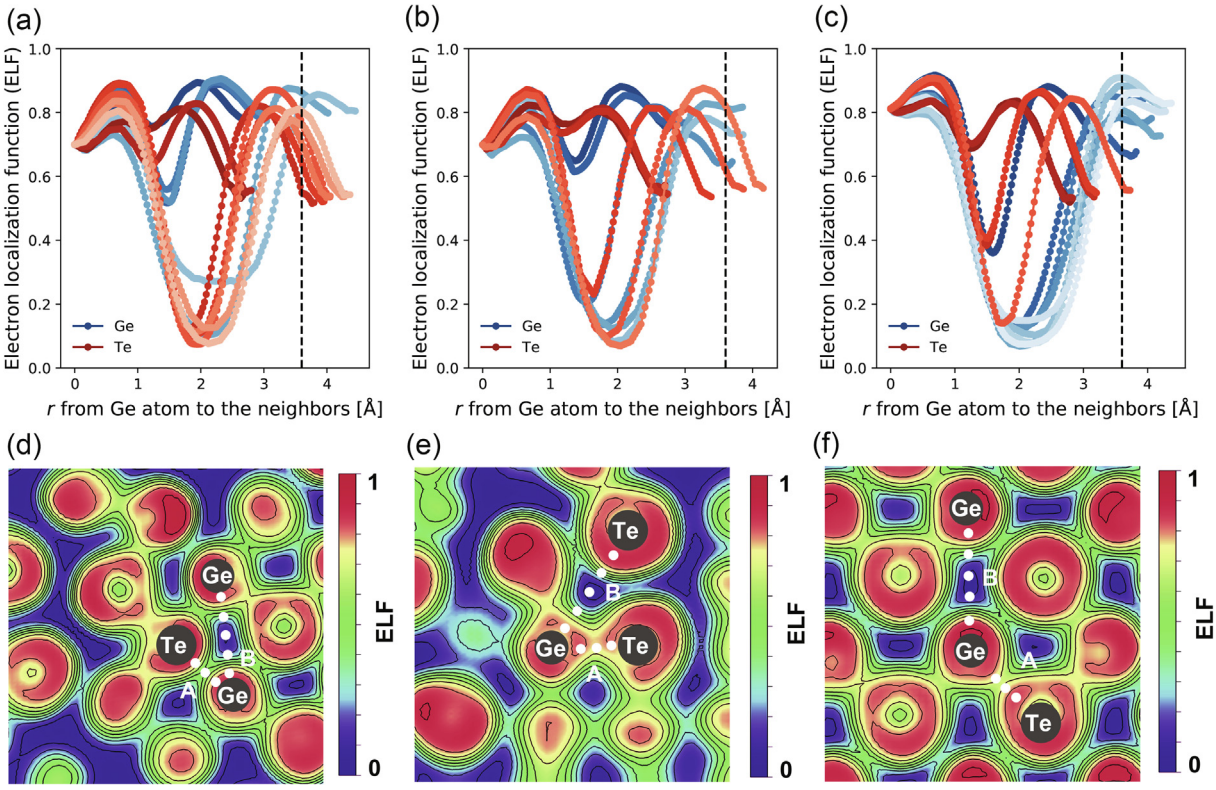

Fig. 6. The result of the electron-localization function (ELF) calculation. (a)–(c): Line profiles of the ELF between Ge and its neighbors, with the vertical dashed lines representing a cutoff distance of r = 3.6 Å; (d)–(f): the corresponding 2D contour plot in which the bonding path between Ge wand its nearest neighbor (A) or the farthest neighbor within a distance range of 4.5 Å is drawn.

ANN fitting, however, here we will show that by the combination with the pair distribution function of the material, more information can be dig out from the 'typical' curve shown in Fig. 4(b). As seen from the partial pair distribution function shown in the inset of Fig. 4(b), the bonds with the length between 3.0 and 3.6 Å are mainly composed of Ge-Te bonds and a small proportion of Ge-Ge bonds, while between 3.6 and 4.5 Å, there are homopolar Ge-Ge and Te-Te correlations. Given the variation of the RMSE of ANN fitting over the cutoff radius in Fig. 4(b), it can be deduced that the energy of the system and the E_f of the Ge vacancy is mainly determined by the Ge-Te bond, with a small contribution from the Ge-Ge bond. Moreover, in terms of the contribution to the formation energy of the vacancy, the weight ratio of the bond with length of 3.0–3.6 Å and 3.6–4.5 Å can be approximated as 6:1. The atoms distributed near a central Ge atom within the cutoff radius of 3.6 and 4.5 Å are analyzed in Fig. 4(c) and (d), respectively. It can be seen that the coordination number (CN) within the radius of 3.6 Å is 4–8, and CN equals 6 in most cases. Thus, the 4 to 8 neighbors of a Ge atom/vacancy make a significant contribution to the formation energy of V_{Ge}. It is known that, in crystalline GeTe, each Ge atom has 6 neighboring Te atoms if a cutoff radius of 3.6 Å is taken. While for amorphous GeTe, CN equals 4, corresponding to a tetrahedral or defective octahedral coordination [11]. Thus, our conclusion on the coordination environment based on the study of the material which is a mixture of the crystalline and amorphous GeTe, agrees with the previous studies. For a large cutoff radius of 4.5 Å, the CN of the Ge atom is mainly in the range of 12–18.

3.5. ANN energy vs DFT energy

After the fitting process based on all the optimized parameters/models, the energy of the unseen structure (test set) can be predicted by using ANN. Fig. 5 summarizes the fitting results for the 10th-fold test set which has a RMSE value of 0.45 meV/atom and a MAE value of 0.36 meV/atom, and the results of the other ninefold (test) sets can be found in Table S3 in the Supplemental material. Firstly, a comparison of the cohesive energies of all the structures in the test set predicted by the ANN with their DFT reference values can be seen in Fig. 4(a), exhibiting an excellent agreement. With the predicted energy of the vacancy-containing system by using ANN, the formation energy of V_{Ge} can be further calculated using Eq. (1), which is compared with the data based on the DFT energy, as shown in Fig. 5(b)–(d). A Gaussian fit (yellow line in Fig. 5(b)) is employed to analyze the difference between the DFT and ANN energy, and it shows that the ANN underestimates E_f by 3 meV/V_{Ge}, with a standard deviation of 0.19 eV/V_{Ge}. The result in Fig. 5(b) for the 10th-fold test set is quite close to the data for other test sets shown in Table S3 in the Supplemental material. The value of E_f obtained by the ANN and DFT method, and the discrepency, can be seen in Fig. 5(c) and (d). In principle, the performance of the ANN could be improved by including more input structures and/or more complex ANN architecture. However, for the present ANN application, an accurate prediction of the vacancy-formation energy is not the main object. The contribution of different types of bonding to the energy or vacancy-formation energy is of more interests here, and we think the present accuracy of ANN is sufficient.

3.6. Bonding inside the cutoff radius

In the previous sections, by using an ANN analysis, it is concluded that the interactions between the atoms in GeTe (including the amorphous and crystalline state) are mainly confined within a cutoff radius (Rc) of 4.5 Å. In this section, we randomly pick three structures from the data set, and analyze the bonding nature within a distance range less than Rc = 4.5 Å by using the electron-localization function (ELF) [40,41]. The value of ELF lies in the range [0, 1]: ELF = 1 means a perfect electron localization, while ELF = 0.5 indicates a homogenous electron gas which can represent a metallic interaction. The ELF has

been widely applied to investigate the chemical bonding and the bonding strength in a range of material [42]. The minimum value of ELF along the bonding path can be used as an indicator of the bond strength [43,44].

As shown in Fig. 6(a)–(c), for the neighbor of a Ge atom within a distance of 3.6 Å (dashed line in the figure), the ELF value in between is apparently larger than that between neighbors with a larger distance. From the minimum value of the ELF, it can be seen that Ge-Te bond is stronger than the Ge-Ge bond, with both exhibiting a feature of covalent bonding. For the neighbors with an interatomic distance of 3.6–4.5 Å, the bond strength becomes much weaker, while in certain cases, such as in Fig. 6(a), the strength Ge-Ge bond cannot be neglected. The bonding information obtained by analyzing the ELF is in excellent agreement with the 'general' bonding picture from the ANN analysis in Fig. 4(b) based on extensive energy data. The 2D contour of the ELF between Ge and its nearest and farthest neighbor within Rc = 4.5 Å, is shown in Fig. 6(d)–(f). In consistence with the conclusions in Fig. 6(a)–(c), bonding is more significant with the nearest than with the farthest neighbors.

4. Conclusions

In this work, an automatic workflow for the high-throughput calculation of the vacancy-formation energy in the material and for the analysis of the bonding environment using ANN method based on the energy data, is presented. Firstly, the inequivalent atom is identified by comparing its surrounding atoms within a specific cutoff radius, including the element type and the distance of the neighbor. Then the inequivalent atom is removed to generate a vacancy. The first-principles evaluation of the vacancy-formation energy is performed via a high-throughput calculation platform. More importantly, we show that the bonding information can be derived from the extensive formation-energy data, by using an artificial neural-network (ANN) analysis. Using the phase-change memory material, GeTe, as a case study, the formation energy of the Ge vacancy (V_{Ge}) in GeTe undergoing a phase-change process from the amorphous to the crystalline state, is studied. It is found that the formation energy of V_{Ge} in GeTe is mainly located in the range between 0.86 and −1.82 eV, while in the perfect crystal, the value is 0.01 eV. The existence of V_{Ge} with a small formation energy in the amorphous/crystal mixed structure may account for the structural evolution or the time-dependent resistance drift of the amorphous state in PCRAM. Then, an applicable ANN architecture, as well as the symmetry function (input layer for the ANN), are constructed for GeTe. And the energy of the atomic structure can be predicted by ANN, in agreement with the DFT energy. By tracking the performance of the ANN prediction with the change of the cutoff radius in the symmetry function which represent an "interaction range" of the atom with its neighbors, the bonding environment in GeTe is unraveled. It is found that bond with a length of about 3.6 Å (mostly Ge-Te bond) makes a significant contribution to the energy of the system (including the formation energy of V_{Ge}); bond with the length between 3.6 Å and 4.5 Å (mostly Ge-Ge bond) makes a smaller contribution, while 'bond' with length larger than 4.5 Å has a negligible effect on the energy. Moreover, we show that the bonding information obtained by the ANN analysis agrees well with the result from electron-localization-function computation. We believe that our study of the bonding information as well as the relative contribution of different bonds to the energy of the system, by using the ANN analysis on the extensive vacancy-containing structures, provides a novel alternative perspective of the bonding picture of GeTe. More importantly, the whole framework can be applied to other phase-change memory material or to other complex system such as the amorphous material, where the atomic interactions are to be studied. Going much further beyond standard research method and obtaining "new/unexpected" physical quantity or input-output relationship that is impossible to study using traditional techniques in materials science using ANN is expected. To achieve this goal, feature engineering with more interpretable feature vectors containing knowledge in material science is absolutely required.

CRediT authorship contribution statement

Linggang Zhu contributes the conceptualization, formal analysis,Investigation, Writing-original draft, Writing-review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is financially supported by the National Key Research and Development Program of China (Grant No. 2017YFB0701700) and the National Natural Science Foundation of China (No. 51871009). The authors also acknowledge support of the "111" Project (B17002).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.commatsci.2020.109803.

References

[1] H.-S. Kim, J.B. Cook, H. Lin, J.S. Ko, S.H. Tolbert, V. Ozolins, B. Dunn, Oxygen vacancies enhance pseudocapacitive charge storage properties of MoO_{3-x}, Nat. Mater. 16 (4) (2016) 454–460, https://doi.org/10.1038/nmat4810.
[2] S.O.J. Long, A.V. Powell, P. Vaqueiro, S. Hull, High thermoelectric performance of bornite through control of the Cu(II) content and vacancy concentration, Chem. Mater. 30 (2) (2018) 456–464, https://doi.org/10.1021/acs.chemmater.7b04436.
[3] T. Schuler, C. Barouh, M. Nastar, C.-C. Fu, Equilibrium vacancy concentration driven by undetectable impurities, Phys. Rev. Lett. 115 (1).https://doi.org/10.1103/physrevlett.115.015501.
[4] M. Wuttig, N. Yamada, Phase-change materials for rewriteable data storage, Nat. Mater. 6 (11) (2007) 824–832, https://doi.org/10.1038/nmat2009.
[5] D. Loke, T.H. Lee, W.J. Wang, L.P. Shi, R. Zhao, Y.C. Yeo, T.C. Chong, S.R. Elliott, Breaking the speed limits of phase-change memory, Science 336 (6088) (2012) 1566–1569, https://doi.org/10.1126/science.1221561.
[6] D. Lencer, M. Salinga, B. Grabowski, T. Hickel, J. Neugebauer, M. Wuttig, A map for phase-change materials, Nat. Mater. 7 (12) (2008) 972–977, https://doi.org/10.1038/nmat2330.
[7] Z. Sun, J. Zhou, A. Blomqvist, B. Johansson, R. Ahuja, Formation of large voids in the amorphous phase-change memory $Ge_2Sb_2Te_5$ alloy, Phys. Rev. Lett. 102(7). https://doi.org/10.1103/physrevlett.102.075504.
[8] Z. Sun, J. Zhou, R. Ahuja, Structure of phase change materials for data storage, Phys. Rev. Lett. 96(5).https://doi.org/10.1103/physrevlett.96.055507.
[9] J.-J. Wang, J. Wang, H. Du, L. Lu, P.C. Schmitz, J. Reindl, A.M. Mio, C.-L. Jia, E. Ma, R. Mazzarello, M. Wuttig, W. Zhang, Genesis and effects of swapping bilayers in hexagonal $GeSb_2Te_4$, Chem. Mater. 30 (14) (2018) 4770–4777, https://doi.org/10.1021/acs.chemmater.8b01900.
[10] K. Shportko, S. Kremers, M. Woda, D. Lencer, J. Robertson, M. Wuttig, Resonant bonding in crystalline phase-change materials, Nat. Mater. 7 (8) (2008) 653–658, https://doi.org/10.1038/nmat2226.
[11] V.L. Deringer, W. Zhang, M. Lumeij, S. Maintz, M. Wuttig, R. Mazzarello, R. Dronskowski, Bonding nature of local structural motifs in amorphous GeTe, Angew. Chem. Int. Ed. 53 (40) (2014) 10817–10820, https://doi.org/10.1002/anie.201404223.
[12] P. Ma, H. Tong, T. Huang, M. Xu, N. Yu, X. Cheng, C.-J. Sun, X. Miao, Variations of local motifs around ge atoms in amorphous GeTe ultrathin films, J. Phys. Chem. C 121 (2) (2017) 1122–1128, https://doi.org/10.1021/acs.jpcc.6b09841.
[13] A. Kolobov, M. Krbal, P. Fons, J. Tominaga, T. Uruga, Distortion-triggered loss of long-range order in solids with bonding energy hierarchy, Nat. Chem. 3 (4) (2011) 311–316, https://doi.org/10.1038/nchem.1007.
[14] M. Zhu, O. Cojocaru-Mirédin, A.M. Mio, J. Keutgen, M. K + +pers, Y. Yu, J.-Y. Cho, R. Dronskowski, M. Wuttig, Unique bond breaking in crystalline phase change materials and the quest for metavalent bonding, Adv. Mater. 30 (18) (2018) 1706735.https://doi.org/10.1002/adma.201706735.
[15] Y. Saito, Y. Sutou, P. Fons, S. Shindo, X. Kozina, J.M. Skelton, A.V. Kolobov, K. Kobayashi, Electronic structure of transition-metal based Cu_2GeTe_3 phase change material: Revealing the key role of cu d electrons, Chem. Mater. 29 (17) (2017) 7440–7449, https://doi.org/10.1021/acs.chemmater.7b02436.
[16] Z. Sun, J. Zhou, Y. Pan, Z. Song, H.-K. Mao, R. Ahuja, Pressure-induced reversible amorphization and an amorphous-amorphous transition in $Ge_2Sb_2Te_5$ phase-change

memory material, PNAS 108 (26) (2011) 10410–10414, https://doi.org/10.1073/pnas.1107464108.
[17] S. He, L. Zhu, J. Zhou, Z. Sun, Metastable stacking-polymorphism in $Ge_2Sb_2Te_5$, Inorg. Chem. 56 (19) (2017) 11990–11997, https://doi.org/10.1021/acs.inorgchem.7b01970.
[18] J. Goldak, C.S. Barrett, D. Innes, W. Youdelis, Structure of alpha GeTe, J. Chem. Phys. 44 (9) (1966) 3323–3325, https://doi.org/10.1063/1.1727231.
[19] S. Gabardi, S. Caravati, G. C. Sosso, J. Behler, M. Bernasconi, Microscopic origin of resistance drift in the amorphous state of the phase-change compound GeTe, Phys. Rev. B 92(5).https://doi.org/10.1103/physrevb.92.054201.
[20] Z. Sun, Y. Pan, J. Zhou, B. Sa, R. Ahuja, Origin of p-type conductivity in layered $nGeTe{\cdot}mSb_2Te_3$ chalcogenide semiconductors, Phys. Rev. B 83(11).https://doi.org/10.1103/physrevb.83.113201.
[21] A.V. Kolobov, P. Fons, J. Tominaga, Local instability ofp-type bonding makes amorphous GeTe a lone-pair semiconductor, Phys. Rev. B 87(15).https://doi.org/10.1103/physrevb.87.155204.
[22] J.Y. Raty, W. Zhang, J. Luckas, C. Chen, R. Mazzarello, C. Bichara, M. Wuttig, Aging mechanisms in amorphous phase-change materials, Nat. Commun. 6(1).https://doi.org/10.1038/ncomms8467.
[23] F. Tong, X.S. Miao, Y. Wu, Z.P. Chen, H. Tong, X.M. Cheng, Effective method to identify the vacancies in crystalline GeTe, Appl. Phys. Lett. 97 (26) (2010), https://doi.org/10.1063/1.3531664 261904.
[24] A.V. Kolobov, J. Tominaga, P. Fons, T. Uruga, Local structure of crystallized GeTe films, Appl. Phys. Lett. 82 (3) (2003) 382–384, https://doi.org/10.1063/1.1539926.
[25] T.H. Lee, S.R. Elliott, The relation between chemical bonding and ultrafast crystal growth, Adv. Mater. 29 (24) (2017) 1700814, https://doi.org/10.1002/adma.201700814.
[26] K. Mathew, J.H. Montoya, A. Faghaninia, S. Dwarakanath, M. Aykol, H. Tang, I. heng Chu, T. Smidt, B. Bocklund, M. Horton, J. Dagdelen, B. Wood, Z.-K. Liu, J. Neaton, S.P. Ong, K. Persson, A. Jain, Atomate: a high-level interface to generate, execute, and analyze computational materials science workflows, Comput. Mater. Sci. 139 (2017) 140–152.https://doi.org/10.1016/j.commatsci.2017.07.030.
[27] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python materials genomics (pymatgen): a robust, open-source python library for materials analysis, Comput. Mater. Sci. 68 (2013) 314–319, https://doi.org/10.1016/j.commatsci.2012.10.028.
[28] A. Jain, S.P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, D. Gunter, K.A. Persson, FireWorks: a dynamic workflow system designed for high-throughput applications, Concurrency Comput.: Practice Exp. 27 (17) (2015) 5037–5059, https://doi.org/10.1002/cpe.3505.
[29] G. Kresse, J. Hafner, Ab initiomolecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium, Phys. Rev. B 49 (20) (1994) 14251–14269, https://doi.org/10.1103/physrevb.49.14251.
[30] A.H. Edwards, A.C. Pineda, P.A. Schultz, M.G. Martin, A.P. Thompson, H.P. Hjalmarson, C.J. Umrigar, Electronic structure of intrinsic defects in crystalline germanium telluride, Phys. Rev. B 73 (2006), https://doi.org/10.1103/PhysRevB.73.045210.
[31] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (18) (1996) 3865–3868, https://doi.org/10.1103/physrevlett.77.3865.
[32] J. Hertz, A. Krogh, R.G. Palmer, Introduction to the Theory of Neural Computation, CRC Press, 2018, https://doi.org/10.1201/9780429499661.
[33] G. C. Sosso, G. Miceli, S. Caravati, J. Behler, M. Bernasconi, Neural network interatomic potential for the phase change material GeTe, Phys. Rev. B 85(17). https://doi.org/10.1103/physrevb.85.174103.
[34] J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, J. Chem. Phys. 134 (7) (2011), https://doi.org/10.1063/1.3553717 074106.
[35] N. Artrith, A. Urban, An implementation of artificial neural-network potentials for atomistic materials simulations: performance for TiO_2, Comput. Mater. Sci. 114 (2016) 135–150, https://doi.org/10.1016/j.commatsci.2015.11.047.
[36] N. Artrith, A. Urban, G. Ceder, Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species, Phys. Rev. B 96(1). https://doi.org/10.1103/physrevb.96.014112.
[37] L. Zhu, Z. Li, J. Zhou, N. Miao, Z. Sun, Insight into the role of oxygen in the phase-change material GeTe, J. Mater. Chem. C 5 (14) (2017) 3592–3599, https://doi.org/10.1039/c7tc00127d.
[38] W. Zhang, A. Thiess, P. Zalden, R. Zeller, P.H. Dederichs, J.-Y. Raty, M. Wuttig, S. Blügel, R. Mazzarello, Role of vacancies in metal–insulator transitions of crystalline phase-change materials, Nat. Mater. 11 (11) (2012) 952–956, https://doi.org/10.1038/nmat3456.
[39] N. Artrith, A.M. Kolpak, Grand canonical molecular dynamics simulations of cu–au nanoalloys in thermal equilibrium using reactive ANN potentials, Comput. Mater. Sci. 110 (2015) 20–28, https://doi.org/10.1016/j.commatsci.2015.07.046.
[40] A.D. Becke, K.E. Edgecombe, A simple measure of electron localization in atomic and molecular systems, J. Chem. Phys. 92 (9) (1990) 5397–5403, https://doi.org/10.1063/1.458517.
[41] B. Silvi, A. Savin, Classification of chemical bonds based on topological analysis of electron localization functions, Nature 371 (6499) (1994) 683–686, https://doi.org/10.1038/371683a0.
[42] A. Savin, R. Nesper, S. Wengert, T.F. Fässler, ELF: the electron localization function, Angew. Chem. Int. Ed. Engl. 36 (17) (1997) 1808–1832, https://doi.org/10.1002/anie.199718081.
[43] M. Xu, Y.Q. Cheng, H.W. Sheng, E. Ma, Nature of atomic bonding and atomic structure in the phase-change $Ge_2Sb_2Te_5$ 103(19).https://doi.org/10.1103/physrevlett.103.195502.
[44] S. Shi, L. Zhu, H. Zhang, Z. Sun, R. Ahuja, Mapping the relationship among composition, stacking fault energy and ductility in Nb alloys: a first-principles study, Acta Mater. 144 (2018) 853–861, https://doi.org/10.1016/j.actamat.2017.11.029.