Using neural network potentials to study defect formation and phonon properties of nitrogen vacancies with multiple charge states in GaN

Koji Shimizu [1,*] Ying Dou,[2] Elvis F. Arguelles [1] Takumi Moriya,[1] Emi Minamitani,[3] and Satoshi Watanabe [1,†]

[1]*Department of Materials Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
[2]*Institute of Industrial Science (IIS), The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*
[3]*Institute for Molecular Science, 38 Nishigo-Naka, Myodaiji, Okazaki 444-8585, Japan*

Investigation of charged defects is necessary to understand the properties of semiconductors. Whereas density functional theory calculations can accurately describe the relevant physical quantities, these calculations increase the computational loads substantially, which often limits the application of this method to large-scale systems. In this paper, we propose a different scheme of neural network potential (NNP) to analyze the point defect behavior in multiple charge states. The proposed scheme necessitates only minimal modifications to the conventional scheme. We demonstrated the prediction performance of the proposed NNP using wurzite-GaN with a nitrogen vacancy with charge states of 0, 1+, 2+, and 3+. The proposed scheme accurately trained the total energies and atomic forces for all the charge states. Furthermore, it fairly reproduced the phonon band structures and thermodynamics properties of the defective structures. Based on the results of this paper, we expect that the proposed scheme can enable us to study more complicated defective systems and lead to breakthroughs in novel semiconductor applications.

I. INTRODUCTION

The presence of defects in materials, inevitable in most systems, alters their electronic and dynamic properties from their pristine forms [1]. In semiconductors, as-grown samples often contain certain amounts of native defects, resulting in high carrier concentrations [2,3]. Moreover, impurity atoms are often introduced intentionally as dopants into materials to control their conductivity. These dopants can be either *n* or *p* type [4]. Another problem, particularly, in nitride semiconductors, is the inevitability of dislocations during the growth processes of materials. These dislocations critically impact the mechanical and thermal properties of materials [5,6]. The above discussion highlights the need for research focused on studying defects in semiconductor materials.

First-principles calculations based on the density functional theory (DFT) have been indispensable in studying various defect properties of materials. The accuracy of such calculations depends on several factors such as the level of approximation in treating the electron-electron interaction, supercell size in modeling, and choice of the functional. In semiconductors, DFT calculations employing hybrid functionals are known to improve accuracy [7,8]; however, this approach increases the computational cost substantially. Furthermore, performing dynamical calculations, e.g., phonons and molecular dynamics on defective systems, typically requires significant configurational space and long computational times. These problems highlight the limitations of DFT calculations in treating defect properties even at the local density approximation or generalized gradient approximation (GGA) level.

Interatomic potentials using machine learning (ML) techniques, such as the high-dimensional neural network potential (NNP) [9], the Gaussian approximation potential [10], the moment tensor potential [11], and the spectral neighbor analysis potential [12], have been gaining increasing attention because of their low computational costs (by several orders of magnitude) and accuracy comparable with that of DFT calculations. Such ML potentials have been applied to various materials, e.g., Li_3PO_4 [13], GaN [14,15], and Au-Li [15,16] to successfully obtain the relevant dynamical quantities.

For NN-based ML potentials, several extensions have been proposed to achieve an improved expression of physical quantities. Interatomic potentials that use a cutoff radius to truncate the surrounding atomic environment in their descriptors generally possess a potential risk of inaccurate descriptions of long-range interactions, particularly, when applied to ionic materials. The initial approach to address this problem was to overlay another NN on the standard one, which was designed for predicting atomic charges and evaluating long-range electrostatic interaction [17]. More recently, an advanced method based on using the charge neutrality condition to predict atomic charges demonstrated accurate predictions of nonlocal charge transfer, which could not be achieved by conventional models [18].

The ML potentials have never been applied to defective systems to treat multiple charge states because of the difficulty in optimizing the fitting parameters owing to the nonunique output values [19] (total energies and atomic forces) for the same or similar inputs (structural descriptors). In this paper,

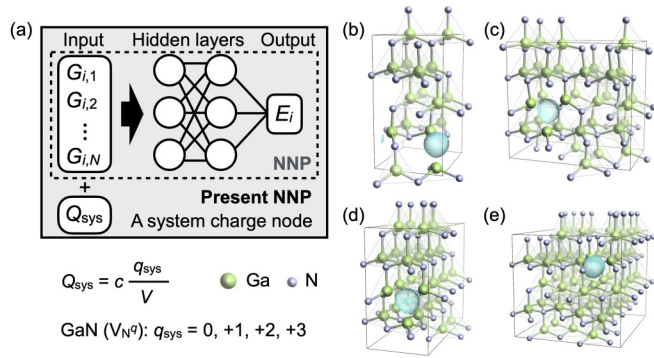*shimizu@cello.t.u-tokyo.ac.jp
†watanabe@cello.t.u-tokyo.ac.jp

FIG. 1. (a) Schematic of the proposed NNP architecture where the input layer contains the system charge node (Q_{sys}). (b)–(e) Snapshot of wurtzite-GaN structures containing one N-vacancy (V_N). (b) 31 atoms, (c) and (d) 63 atoms, and (e) 127 atoms per supercell in total. The localized defective charge densities are illustrated by the isosurfaces. Structures are visualized using the VESTA package [23].

we demonstrated the construction of an ML potential for such systems. We used wurtzite-GaN including a nitrogen vacancy (V_N^q) as a prototype case, where V_N^q takes various charge states q depending on the Fermi level (E_F). As mentioned earlier, the defect study of GaN is of significance for many applications, such as light-emitting diodes [20] and high-power devices [21].

II. METHODOLOGY

A. Neural network potential architecture

To develop our ML potentials, we employed the Behler-Parrinello-type NNP [9] as a base model with minimal modifications. Figure 1(a) depicts the schematic of the proposed NNP model. The local atomic features expressed by the radial and angular symmetry functions (SFs) were provided in the input layer. In addition, we added one system charge (Q_{sys}) node in the input layer. We defined Q_{sys} as constant value c times a charge state of the considered supercell q_{sys} divided by its volume V. For instance, in the cases of pristine and V_N^0, $Q_{\text{sys}} = 0$. In contrast, $Q_{\text{sys}} = c/V$, $2c/V$, and $3c/V$ for V_N^{1+}, V_N^{2+}, and V_N^{3+}, respectively. Although we used $c = 100$ for the ease of scaling, we emphasize that the present scaling method may not be the optimal one. All input nodes were fully connected to the first hidden layer. The atomic energy (E_i) for the two hidden layer case is expressed as

$$
\begin{aligned}
E_i = f_a^{\text{out}} \Bigg[& w_{0,1}^{\text{out}} + \sum_{k=1}^{k_0} w_{k,1}^{\text{out}} f_a^2 \Bigg\{ w_{0,k}^2 \\
& + \sum_{j=1}^{j_0} w_{j,k}^2 f_a^1 \left(w_{0,j}^1 + \sum_{\mu=1}^{\mu_0} w_{\mu,j}^1 G_i^\mu + w_{\mu_0+1,j}^1 Q_{\text{sys}} \right) \Bigg\} \Bigg],
\end{aligned}
\tag{1}
$$

where G_i^μ is the SFs of atom i with multiple input nodes μ. $w_{l,m}^n$ denotes a weight parameter connecting the nth and precedent layers of the lth and mth nodes, respectively. f_a^n corresponds to the activation function of the nth layer. We used the hyperbolic tangent for the first and second hidden

layers, whereas a linear function was used for the output layer. The sum of E_i over all constituent atoms $E = \sum_i E_i$ is the total energy. The forces along atomic coordinates $\alpha = x, y, z$ can be expressed as $F_{\alpha_i} = -\partial E/\partial \alpha_i = -\sum_\mu \partial E/\partial G^\mu \partial G^\mu/\partial \alpha_i$. Note that Q_{sys} has no explicit contribution to atomic forces.

B. Training dataset generation

To generate the training dataset of NNP, we first performed the molecular dynamics (MD) simulations using the Stillinger-Weber potential [22]. Using the pristine GaN (conventional cell) comprising 32 atoms, we performed canonical ensemble MD simulations at temperatures of 300–2700 K with 400-K intervals for 1 ns. From the obtained trajectories, we extracted the structures of every 10 ps. We used the optimal lattice constant as well as the compressed and expanded ±1% and ±2% and obtained 3500 structures. Next, we generated V_N structures by introducing a nitrogen vacancy into the above-mentioned pristine structures. Using the top two components of the principal component analysis based on the SFs (see Ref. [14] for the parameters of SFs), we performed clustering using a k-means algorithm to generate 3500 clusters. We chose one data point randomly from each cluster to obtain 3500 V_N structures. In addition, we performed MD simulations with the systems including 64 and 128 atoms in the same manner. The simulation time was 0.1 ns, and the snapshots of every 10 ps were extracted. In these cases, we obtained 700 and 350 V_N structures of 63- and 127-atom systems by randomly introducing a nitrogen vacancy. We then performed DFT calculations with 0, 1+, 2+, and 3+ charge states for the V_N structures. In this way, we generated 9100 ($3500 \times 2 + 700 \times 2 + 350 \times 2$) structures in total and the corresponding 22750 DFT data. Figures 1(b)–1(e) depicts one of these V_N structures. We used the Vienna *ab initio* simulation package [24,25] for all DFT calculations. We used the GGA with the Perdew-Burke-Ernzerhof functional [26], plane-wave basis set (550-eV cutoff energy), and the projector augmented-wave method [27]. Brillouin-zone integration was performed using the sampling technique of Monkhorst and Pack [28] for the training dataset ($4 \times 4 \times 2$ sampling mesh for 32-atom systems, $4 \times 2 \times 2$ and $2 \times 4 \times 2$ for 64-atom systems, and $2 \times 2 \times 2$ for 128-atom systems), whereas the Γ-centered $2 \times 2 \times 2$ was used for phonon calculations. The convergence criteria were 10^{-7} eV and 10^{-4} eV/Å for the self-consistent field and ionic relaxation, respectively. The number of electrons was varied for V_N^q by adding a jellium background charges q to neutralize the systems.

III. RESULTS AND DISCUSSION

A. NNP construction

Figure 2 depicts a comparison between DFT and the proposed NNP with respect to their total energies and atomic forces. For this comparison, we used randomly chosen 10% of the structures from the dataset as the test data. Both total energies and atomic forces were aligned along diagonal lines, suggesting that the constructed NNP accurately predicted all structures and charge states considered. The comparison plots using the conventional NNP are given as Fig. S1 in the Supplemental Material [29]. The root-mean-square errors (RMSEs)
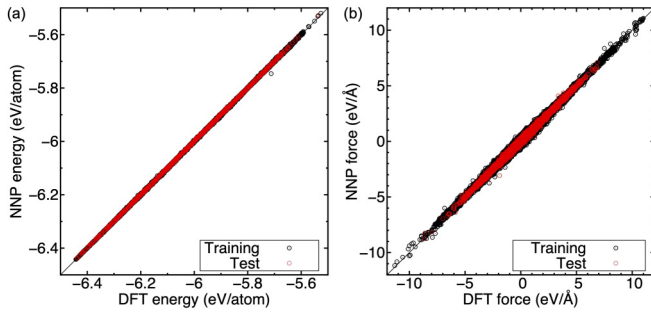
FIG. 2. Comparison between DFT and NNP in respect of (a) total energies and (b) atomic forces. The black and red circles represent the training and test data, respectively.

of the total energies and atomic forces of the proposed model were, respectively, 1.45 meV/atom and 63.8 meV/Å for the training dataset and 1.44 meV/atom and 64.6 meV/Å for the

test dataset. In this case, we used 8 and 24 types of radial and angular SFs, respectively, for each elemental combination ($8 \times 2 + 24 \times 3 = 88$) with the 7-Å cutoff distance. We used the NN consisting of two hidden layers with 30 and 20 nodes. Therefore, the NN architecture was [89-30-20-1]. Note that we used this setting for subsequent calculations based on several trial trainings (see Secs. 2–4 in the Supplemental Material for the details of the SF parameters, comparison between DFT and the proposed NNP separately depicted for each data type and transferability tests, respectively [29]).

B. Phonon properties

Figure 3 depicts the calculated phonon bands of the pristine and V_N^q structures. We used a $4 \times 4 \times 2$ supercell with 0.01-Å atomic displacement. The V_N^q structures contained one nitrogen vacancy in the supercell. All the phonon calculations were performed using PHONOPY software [30],
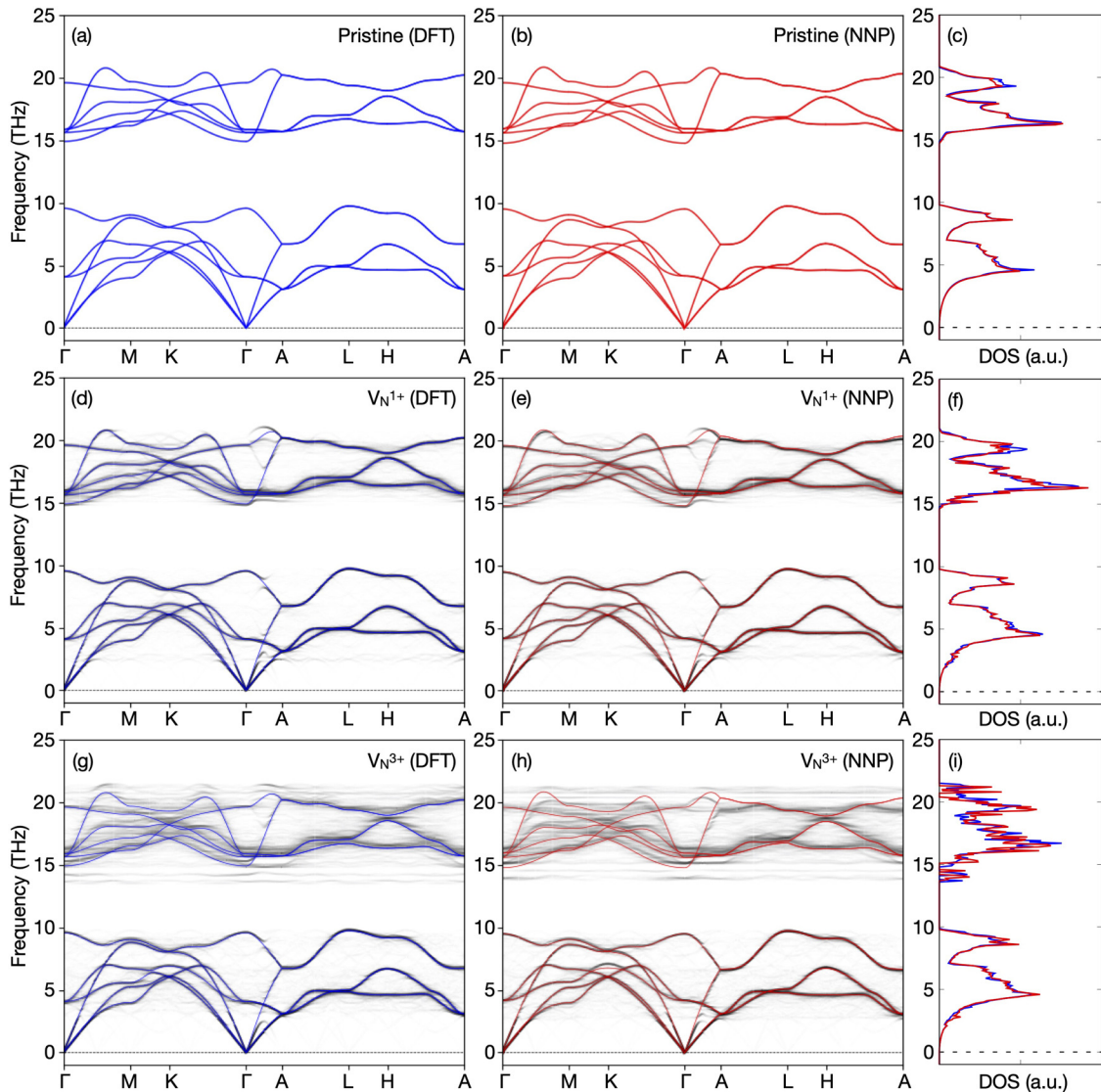

FIG. 3. Calculated phonon bands and densities of states of (a)–(c) pristine, (d)–(f) V_N^{1+}, and (g)–(i) V_N^{3+} structures using DFT and NNP. The densities of states obtained using DFT and NNP are plotted together. The phonon bands of the pristine case are superimposed on those of V_N^q.

and the band-unfolding package was used [31,32]. In the pristine case, the calculated phonon band structures and densities of states depicted in Figs. 3(a)–3(c) for DFT and NNP agreed well. These results were in agreement with experimental results [33]. Whereas some slight differences were seen at the higher frequency end, the detailed structures of both acoustic and optical modes matched considerably in the two methods. Note that we used the optimized structures obtained through DFT for the phonon calculations using NNP.

The calculated phonon band structures and densities of states of V_N^{1+} and V_N^{3+} are depicted in Figs. 3(d)–3(f) and 3(g)–3(i), respectively (see Fig. S6 in the Supplemental Material for phonon bands of other valence states [29]). In this case, the pristine phonon bands were superimposed on the defective bands for ease of comparison. Overall, the introduction of the defect caused band splitting at various regions. In addition, softening could be seen clearly for the V_N^{1+} (also V_N^0) case at around 2 to 3 THz (see Fig. S7 in the Supplemental Material for phonon bands showing the features of DFT and NNP and the differences between them [29]), whereas, the bandwidths of acoustic modes were comparable in all the cases. We found that the softened mode corresponded to the vibration around the defect position (see the Supplemental Material animation 1 for the softened vibration at the Γ point [29]). In the optimized structure of V_N^0, the Ga atoms nearest to the defect slightly moved toward the defect direction when compared with their pristine positions. In contrast, the Ga atoms moved away from the defect position in the $V_N^{1+,2+,3+}$ cases, and the shifts were larger for the higher charge states (see Fig. S8 and Table S5 in the Supplemental Material for the atomic configurations of V_N^q [29]). A previous study has also reported structural changes in V_N^{1+} and V_N^{3+} [34], which agree with the present results. This suggests that the potential-energy gradient became shallower for the lower charge states thereby causing the corresponding phonon softening.

The optical phonon modes, on the contrary, differ significantly clearly among the charge states. Particularly, we found flat bands at the bottom of optical modes in V_N^{3+}. The lowest frequency mode corresponded to nitrogen vibrations along the *ab* plane. The second lowest mode included nitrogen vibrations along the *c* direction in addition to the ones along the *ab* plane (see the Supplemental Material animations 2 and 3 for the vibrations of the flat bands at the Γ point [29]). The phonon band obtained using this NNP also exhibited the emergence of a flat band in V_N^{3+}. Furthermore, the proposed NNP reproduced the entire phonon band structures well although with some slight discrepancies.

Furthermore, we evaluated the thermodynamic properties of the defective GaN within the quasiharmonic approximation, aiming to see to what extent the physical quantities of actual materials could be altered from their bulk values by the presence of defects. We used the optimal, $\pm 1\%$, and $\pm 2\%$ lattice constants for these calculations. Figure 4 depicts the calculated Grüneisen parameters of the pristine and V_N^q systems as functions of temperature. The obtained values are close to the reported Grüneisen parameter of 0.87 at 300 K [35]. We found that the Grüneisen parameters of V_N^0 and V_N^{1+} were larger than that of the pristine at $T < 100$ K. The magnitudes of Grüneisen parameters reflect the size of phonon anharmonicity. Therefore, V_N^{1+} had the strongest
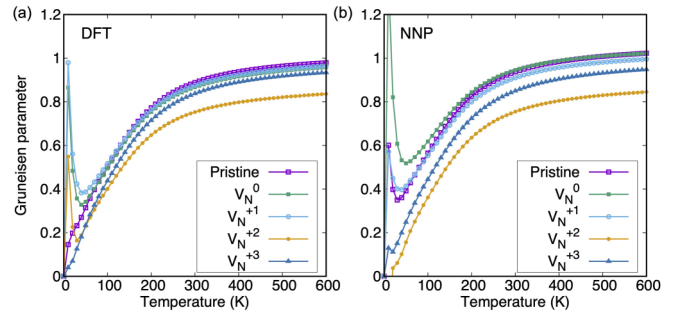

FIG. 4. Calculated Grüneisen parameters of pristine and V_N^q structures as functions of temperature: (a) DFT and (b) NNP.

anharmonicity among the considered systems. This result coincided with the observed phonon softening. In contrast, we found smaller Grüneisen parameters for V_N^{2+} and V_N^{3+} because V_N^{2+} had a smaller bulk modulus and thermal expansion coefficient. For V_N^{3+}, the thermal expansion coefficient was small; however, its bulk modulus was the largest among all.

The Grüneisen parameters obtained using NNP reproduced the DFT results well at $T > 200$ K. Even though the prediction at $T < 200$ K was not accurate, a qualitative agreement was achieved: the V_N^0 and V_N^{1+} (V_N^{2+} and V_N^{3+}) exhibited larger (smaller) Grüneisen parameters than that of the pristine case. Because calculations require a highly accurate prediction performance, we might need DFT calculations or ML potentials with significantly smaller RMSE values as criteria. Possibly, charged defect corrections for atomic forces may require to be considered. We leave this aspect for future studies. It is worth noting that the Grüneisen parameters of V_N systems approached the pristine value when the defect density decreased as expected. We found that a $6 \times 6 \times 4$ model may be sufficient to examine the dilute-limit properties (see Fig. S9 in the Supplemental Material for the model size dependence of the Grüneisen parameters [29]).

C. Defect formation energies

Finally, we calculated the vacancy formation energies using the proposed NNP. In formation energy calculations of charged defects in periodic systems, corrections to eliminate spurious electrostatic interaction energy among the periodic images are necessary. Various techniques have been proposed in this regard [36–38]. In this paper, we did not apply such corrections owing to the inaccessibility of electrostatic potentials by NNP. However, this problem can be resolved by accounting for corrections as a preprocess for the preparation of training data.

Figure 5(a) depicts the calculated vacancy formation energies of V_N^q as functions of E_F. Note that we considered the Ga-rich condition and used the reference valence-band maximum and chemical potential values obtained using DFT for NNP calculations. Near the valence-band side, we obtained the transition from V_N^{3+} to V_N^{1+} (3+/1+). In addition, the transition from V_N^{1+} to V_N^0 (1+/0) could be seen at the conduction-band side. These tendencies agreed with those reported in previous studies [40]. The transition levels obtained using NNP reproduced the DFT results well for $V_N^{0,1+,2+}$.
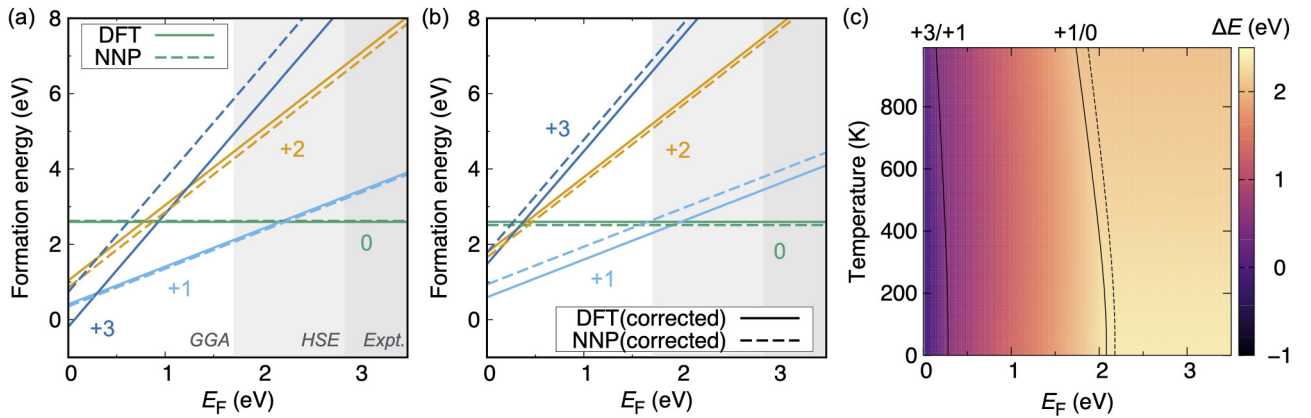
FIG. 5. (a) Calculated defect formation energies as functions of E_F. The background colors indicate the band-gap energies of GGA (1.71 eV), Heyd-Scuseria-Ernzerhof (2.85 eV), and experiment (ca. 3.50 eV) [39]. (b) Calculated defect formation energies with the charge corrections. (c) Calculated temperature dependence on transition levels of V_N^q. Solid (dotted) line depicts the DFT (NNP) results.

However, an error of 1.03 eV remained in the V_N^{3+} case. One of the reasons for this could be insufficient sampling near near the optimized structure in the training data. The current training dataset generated by adding thermal fluctuations (MD trajectories) could not cover the particular structure-energy relationship. Thus, a structural sampling method that can resolve this problem would be necessary for defective systems. In addition, we found that the electronic states of the highly charged V_N^q are sensitive to structural changes especially near the most stable structure, which may also increase the error. Another reason could be the absence of correction energies for the spurious electrostatic interactions of the charged defects in periodic systems. The correction energies in the present systems exhibited noticeable variation depending on the supercell size, atomic configuration, and charge state. The distribution was spread over a few eV in the V_N^{3+} case, which presumably degraded the prediction accuracy of NNP. In fact, the prediction error in V_N^{3+} was suppressed when considering the energy as the sum of the total energy and correction energy for each structure during NNP training as shown in Fig. 5(b). For more details on the above discussion, see Section S6 in the Supplemental Material [29].

Figure 5(b) depicts the temperature-dependent transition levels of V_N^q including the phonon contributions. The color map depicts the Gibbs free energies, and the transition levels are indicated as lines (solid for DFT; dotted for NNP). In both $3+/1+$ and $1+/0$, the transition levels decreased when the temperature increased because the phonon contributions (phonon energy and entropy terms) were larger for lower valence states. The V_N^0 case had the largest contributions, resulting in the largest shift in the transition level. The proposed NNP successfully reproduced these shifts, although the $3+/1+$ transition was predicted inside the valence band.

IV. CONCLUSIONS

To summarize, we developed a scheme of the NNP to analyze the point defect behavior in multiple charge states. Our results indicate that minimal modifications to NNP—we added only a system charge node in the input layer of the conventional NNP—can significantly improve the training results of such systems. We demonstrated the performance of the proposed NNP using wurzite GaN including a nitrogen vacancy with 0, 1+, 2+, and 3+ charge states, as a prototype material. We constructed the proposed NNP using various GaN structures and the corresponding total energies and atomic forces obtained using the DFT calculations. The resultant NNP accurately reproduced the phonon band structures and thermodynamic properties of defective systems. Preparation of training data using hybrid functional calculations should further improve the prediction performance. The proposed scheme is expected to pave the way for further advancement in potential applications of ML and growth in materials science research.

[1] F. Oba and Y. Kumagai, Appl. Phys. Express **11**, 060101 (2018).

[2] B. Monemar and O. Lagerstedt, J. Appl. Phys. **50**, 6480 (1979).

[3] P. Bogusl/awski, E. L. Briggs, and J. Bernholc, Phys. Rev. B **51**, 17255(R) (1995).

[4] K. H. Ploog and O. Brandt, J. Vac. Sci. Technol. A **16**, 1609 (1998).

[5] C. Mion, J. F. Muth, E. A. Preble, and D. Hanser, Appl. Phys. Lett. **89**, 092123 (2006).

[6] M. Fujikane, A. Inoue, T. Yokogawa, S. Nagao, and R. Nowak, Phys. Status Solidi C **7**, 1798 (2010).

[7] P. G. Moses, M. Miao, Q. Yan, and C. G. Van de Walle, J. Chem. Phys. **134**, 084703 (2011).

[8] R. Gillen and J. Robertson, J. Phys.: Condens. Matter **25**, 405501 (2013).

[9] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).

[10] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).

[11] A. V. Shapeev, Multisc. Model. Simul. **14**, 1153 (2016).

[12] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, J. Comput. Phys. **285**, 316 (2015).

[13] W. Li, Y. Ando, E. Minamitani, and S. Watanabe, J. Chem. Phys. **147**, 214106 (2017).

[14] E. Minamitani, M. Ogura, and S. Watanabe, Appl. Phys. Express **12**, 095001 (2019).

[15] S. Watanabe, W. Li, W. Jeong, D. Lee, K. Shimizu, E. Minamitani, Y. Ando, and S. Han, J. Phys. Energy **3**, 012003 (2021).

[16] K. Shimizu, E. F. Arguelles, W. Li, Y. Ando, E. Minamitani, and S. Watanabe, Phys. Rev. B **103**, 094112 (2021).

[17] N. Artrith, T. Morawietz, and J. Behler, Phys. Rev. B **83**, 153101 (2011).

[18] T. W. Ko, J. A. Finkler, S. Goedecker, and J. Behler, Nat. Commun. **12**, 398 (2021).

[19] C. Zhang, J. Jin, W. Na, Q.-J. Zhang, and M. Yu, IEEE Trans. Microw. Theory Techn. **66**, 3781 (2018).

[20] H. Kim, J. Ohta, K. Ueno, A. Kobayashi, M. Morita, Y. Tokumoto, and H. Fujioka, Sci. Rep. **7**, 2112 (2017).

[21] S. N. Mohammad, A. A. Salvador, and H. Morkoc, Proc. IEEE **83**, 1306 (1995).

[22] F. H. Stillinger and T. A. Weber, Phys. Rev. B **31**, 5262 (1985).

[23] K. Momma and F. Izumi, J. Appl. Cryst. **44**, 1272 (2011).

[24] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).

[25] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

[26] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[27] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[28] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[29] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.106.054108 for computational conditions of NNP and additional calculations and analyses which support the results of the main text.

[30] A. Togo and I. Tanaka, Scr. Mater. **108**, 1 (2015).

[31] P. B. Allen, T. Berlijn, D. A. Casavant, and J. M. Soler, Phys. Rev. B **87**, 239904(E) (2013).

[32] https://gitlab.abinit.org/xuhe/unfolding.git.

[33] T. Ruf, J. Serrano, M. Cardona, P. Pavone, M. Pabst, M. Krisch, M. D'Astuto, T. Suski, I. Grzegory, and M. Leszczynski, Phys. Rev. Lett. **86**, 906 (2001).

[34] I. C. Diallo and D. O. Demchenko, Phys. Rev. Applied **6**, 064002 (2016).

[35] H. Iwanaga, A. Kunishige, and S. Takeuchi, J. Mater. Sci. **35**, 2451 (2000).

[36] C. Freysoldt, J. Neugebauer, and C. G. Van de Walle, Phys. Rev. Lett. **102**, 016402 (2009).

[37] Y. Kumagai and F. Oba, Phys. Rev. B **89**, 195205 (2014).

[38] M. C. da Silva, M. Lorke, B. Aradi, M. F. Tabriz, T. Frauenheim, A. Rubio, D. Rocca, and P. Deák, Phys. Rev. Lett. **126**, 076401 (2021).

[39] B. Monemar, Phys. Rev. B **10**, 676 (1974).

[40] C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004).

[41] Editage, www.editage.com.