

# Prediction of positron-annihilation parameters for vacancy-type defects in ternary alloy semiconductors by data-scientific approach

Shoji Ishibashi<sup>1,2</sup>, Hiori Kino<sup>2</sup>, Akira Uedono<sup>3</sup>, Takashi Miyake<sup>1,2</sup> and Kiyoyuki Terakura<sup>2</sup>

<sup>1</sup> Research Center for Computational Design of Advanced Functional Materials (CD-FMat), AIST, Tsukuba, Ibaraki 305-8568, Japan

<sup>2</sup> Center for Materials research by Information Integration (CMI<sup>2</sup>), NIMS, Tsukuba, Ibaraki 305-0047, Japan

<sup>3</sup> Division of Applied Physics, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan

E-mail: [shoji.ishibashi@aist.go.jp](mailto:shoji.ishibashi@aist.go.jp)

**Abstract.** We calculated positron annihilation parameters for mono- and di-vacancies in ternary semiconductors  $\text{Al}_{0.5}\text{Ga}_{0.5}\text{N}$  and  $\text{In}_{0.5}\text{Ga}_{0.5}\text{N}$ . It has been found that the obtained annihilation parameters are well correlated with structural parameters. By constructing multiple linear regression models using selected (about 1/4 of the total) datasets as training sets in order to reduce computational cost, we could predict annihilation parameters for the rest.

## 1. Introduction

Group-III nitrides are important materials for optoelectronic and power devices. Since vacancy-type defects often play a critical role in device performances, we have been investigating such defects by the positron annihilation technique [1, 2, 3]. As is well known, the positron annihilation technique is a powerful tool to study vacancy-type defects in various solids [4, 5]. Alloying is one of the key technologies for semiconductor devices by controlling the band gap and/or the lattice parameters. In alloy semiconductors, atomic sites are not equivalent to one another because of randomness. In our previous studies, we calculated positron annihilation  $S$  and  $W$  parameters as well as positron lifetimes  $\tau$  for 64 types of cation mono-vacancies and those for 104 types of cation-nitrogen di-vacancies among 256 possibilities in  $\text{In}_{0.5}\text{Ga}_{0.5}\text{N}$  [2, 3]. In the present study, we have completed calculations for the rest of the di-vacancies and calculated positron annihilation parameters for mono- and di-vacancies in  $\text{Al}_{0.5}\text{Ga}_{0.5}\text{N}$  also.

## 2. Computational details

For calculations of positron states and annihilation parameters, we used our computational code QMAS (Quantum MAterials Simulator), which is based on the plane-wave basis and the projector augmented-wave (PAW) method [6]. Ternary semiconductors were modeled as random alloys by the special quasirandom structure (SQS) approach [7]. Their lattice parameters and high-frequency dielectric constants were evaluated by averaging experimental values for binary systems. Further details are described in Ref. [2]. For data-scientific analyses, we used the **lm()**

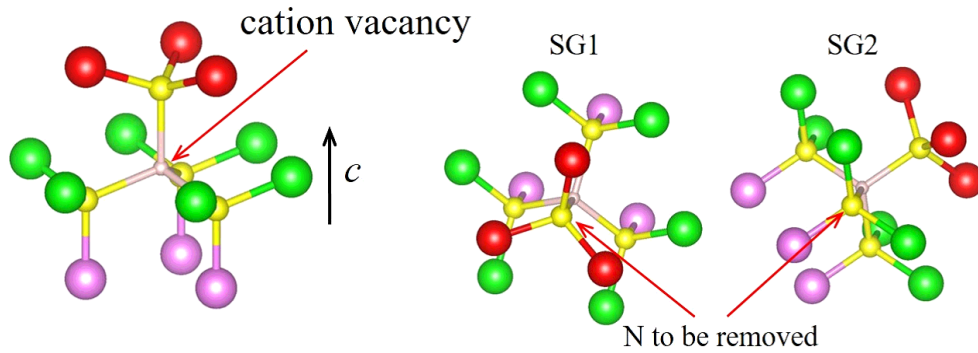


function of R [8] to perform multiple linear regression. With the help of the `step()` function, we selected descriptors in regression models.

### 3. Results and discussion

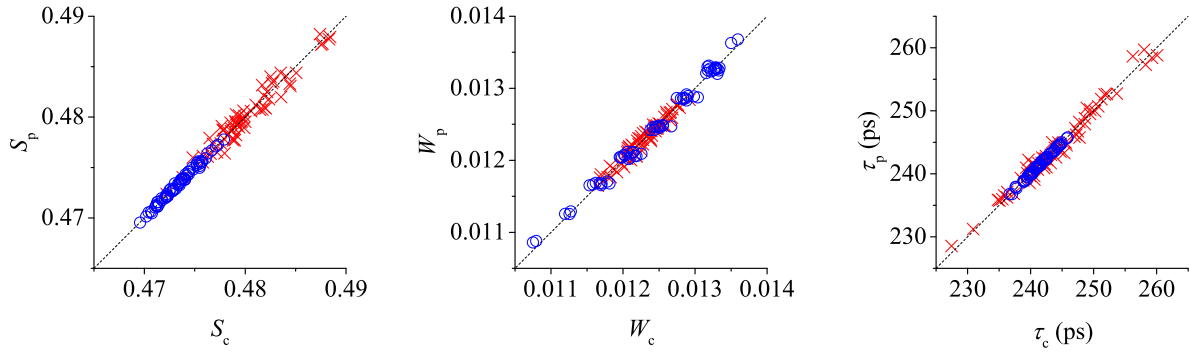
#### 3.1. Multiple linear regression with optimized structural parameters

In Fig. 1, an atomic configuration around a cation mono-vacancy is shown together with two types of configurations for di-vacancies. For cation mono-vacancies, we used the following 3 descriptors:  $x_{\text{Ga}}$ , Ga concentration among neighboring 12 cations,  $d_{\text{N}}$ , average distance of 4 nitrogen atoms from the center of tetrahedron which 4 nitrogen atoms form,  $d_{\text{C}}$ , average distance of 12 cations from the center of the tetrahedron. In Fig. 2, predicted positron annihilation parameters are plotted as a function of originally-calculated values, which were obtained using QMAS with atomic positions computationally-optimized. Root-mean-square-error (RMSE) values in percentage for  $S$ ,  $W$  and  $\tau$  are 0.037 %, 0.61 % and 0.061 % for  $\text{Al}_{0.5}\text{Ga}_{0.5}\text{N}$ , 0.16 %, 0.56 %, 0.42 % for  $\text{In}_{0.5}\text{Ga}_{0.5}\text{N}$ , respectively. Multiple linear regression works very well for these cases. Here, we do not present regression coefficients because their values do not have essential meanings when there is significant correlation among descriptors.



**Figure 1.** Atomic configuration around cation mono-vacancy (left) and two types of configurations for di-vacancy (middle and right). Yellow small balls represent nitrogen atoms while larger balls represent cations. In case of an ideal wurtzite structure, balls with the same color are crystallographically equivalent to each other.  $d_{\text{N}}$  and  $d_{\text{C}}$  in the text are defined as average distances from the center of the nitrogen tetrahedron of 4 nitrogen atoms and 12 cations, respectively. SG1 and SG2 correspond to two subgroups of di-vacancies described in the text.

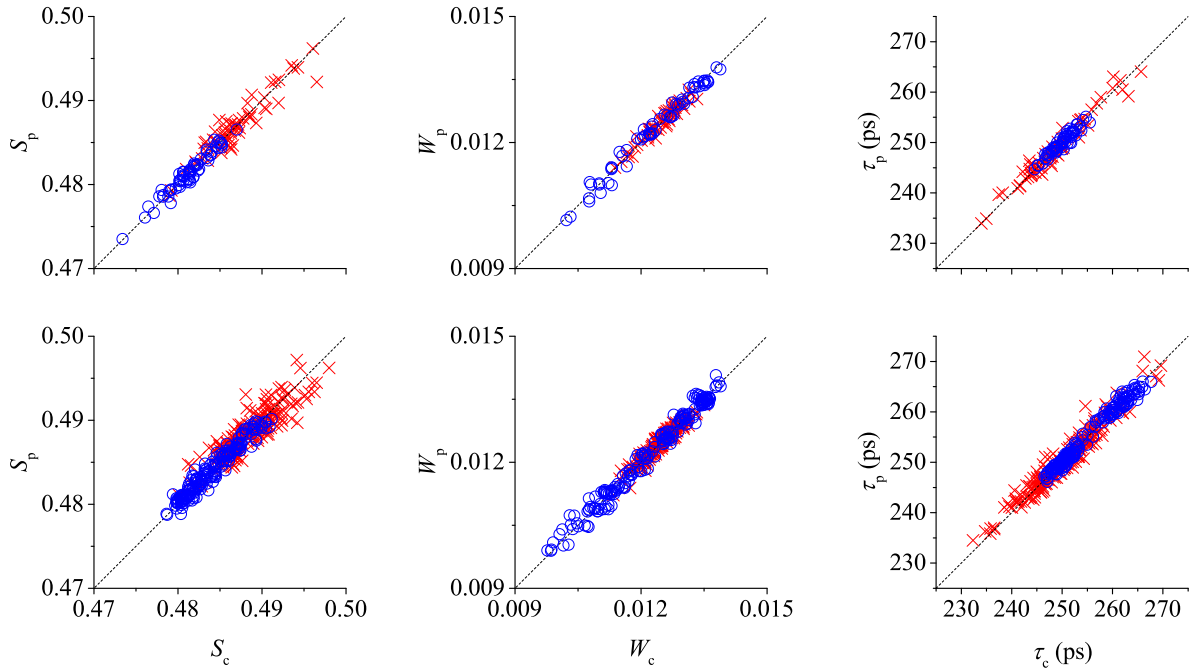
Among 4 nitrogen atoms around a cation mono-vacancy, one nitrogen shows a different geometry from those for the rest because of the wurtzite structure. We therefore divided 256 patterns of di-vacancies into two subgroups consisting of 64 and 192 patterns (denoted as SG1 and SG2), respectively. Surrounding 12 cations should be classified into two categories depending on the position of the vacant nitrogen site  $V_{\text{N}}$  (a part of di-vacancy). For the di-vacancy case, the following 5 descriptors were chosen:  $x1_{\text{Ga}}$ , Ga concentration among 3 cations adjacent to  $V_{\text{N}}$ ,  $x2_{\text{Ga}}$ , that among the other 9 cations,  $d_{\text{N}}$ , average distance of 3 nitrogen atoms from the center of tetrahedron formed by 3 nitrogen atoms and the center of 3 cations,  $d1_{\text{C}}$ , average distance of 3 cations from the center of the tetrahedron,  $d2_{\text{C}}$ , that of the other 9 cations. In Fig. 3, predicted positron annihilation parameters are plotted as a function of originally-calculated values. RMSE values for the SG1 are 0.12 %, 1.04 % and 0.28 % for  $\text{Al}_{0.5}\text{Ga}_{0.5}\text{N}$ , 0.24 %, 0.88 %, 0.54 % for  $\text{In}_{0.5}\text{Ga}_{0.5}\text{N}$ , and those for the SG2 are 0.15 %, 1.20 % and 0.33 % for  $\text{Al}_{0.5}\text{Ga}_{0.5}\text{N}$ , 0.30 %, 0.85



**Figure 2.** Positron annihilation parameters predicted by linear regression models ( $S_p, W_p, \tau_p$ ) for cation mono-vacancies in  $\text{Al}_{0.5}\text{Ga}_{0.5}\text{N}$  (open circles) and  $\text{In}_{0.5}\text{Ga}_{0.5}\text{N}$  (crosses) plotted as a function of originally-calculated values ( $S_c, W_c, \tau_c$ ).

%, 0.66 % for  $\text{In}_{0.5}\text{Ga}_{0.5}\text{N}$ . Although these are slightly worse than those for the mono-vacancy case, multiple linear regression is still working well.

Apart from the regression issue, differences in the calculated parameters between two materials should be mentioned. Whether for mono- or di-vacancy,  $S_c$  and  $\tau_c$  show wider distributions while  $W_c$  shows a narrower one in the case of  $\text{In}_{0.5}\text{Ga}_{0.5}\text{N}$  as shown in Figs. 2 and 3. It is thought to be closely related to the difference in ionic radii of cations and to the presence or absence of  $d$  electrons. This problem will be discussed elsewhere.



**Figure 3.** Positron annihilation parameters predicted by linear regression models ( $S_p, W_p, \tau_p$ ) for di-vacancies in  $\text{Al}_{0.5}\text{Ga}_{0.5}\text{N}$  (open circles) and  $\text{In}_{0.5}\text{Ga}_{0.5}\text{N}$  (crosses) plotted as a function of originally-calculated values ( $S_c, W_c, \tau_c$ ). The top and bottom panels are for the SG1 and SG2.

### 3.2. Multiple linear regression with pre-optimized structural parameters

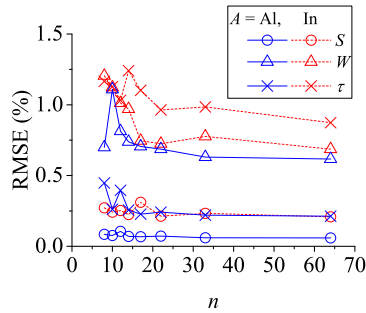
To obtain descriptors  $d_N$  and  $d_C$ , structural optimization is required and its computational cost is significant. We tried to predict positron annihilation parameters using pre-optimized structures, which are the SQS-modeled vacancy-free structures with internal atomic positions fully-relaxed. As a first attempt, we replaced  $d_N$  and  $d_C$  with values evaluated on the vacancy-free structures but, as a matter of course, it was not successful. The degree of structural relaxation accompanied with the vacancy creation is crucial for the prediction of positron annihilation parameters. It is not determined solely by the distances from the vacancy site. Considering that the structural optimization is performed using force values acting on the constituent atoms, we added two more descriptors  $f_N$  and  $f_C$ , which are force components of 4 nitrogen atoms and 12 cations toward the vacancy just after removal of the cation. One-shot calculations of the force components require much less computational time than full optimization. Thus, the following 5 descriptors were chosen:  $x_{Ga}$  (exactly the same as the above),  $d_N^b$ , average distance of 4 nitrogen atoms from the cation to be removed to create a mono-vacancy on the vacancy-free structure,  $d_C^b$ , average distance of 12 cations from the cation,  $f_N$ , average force component of 4 nitrogen atoms toward the vacancy just after removal of the cation,  $f_C$ , average force component of 12 cations toward the vacancy.

To construct regression models, we must know several correct answers for target parameters ( $S, W, \tau$ ). Full structural optimization is required to obtain them. We call the selected datasets “training set”. To assess the minimal number of datasets for constructing regression models, we evaluated RMSE values as a function of the number of datasets in the training set  $n$ . As training sets, in order of  $d_N^b$ , the 1st,  $k + 1$ st,  $2k + 1$ st,  $3k + 1$ st, ... ( $k = 2, 3, 4, 5, 6, 7, 9$ ) and 64th datasets among the 64 datasets are chosen. Figure 4 represents training-set size dependence of RMSE values of positron annihilation parameters for mono-vacancies. The results at  $n = 64$ , where all the datasets were used to construct the regression models, are not much different from those obtained using fully-optimized structures. Down to  $n = 17$ , RMSE values are similar to those at  $n = 64$ . Even with the pre-optimized structures, the regression models were successfully constructed using about 1/4 of the total datasets.

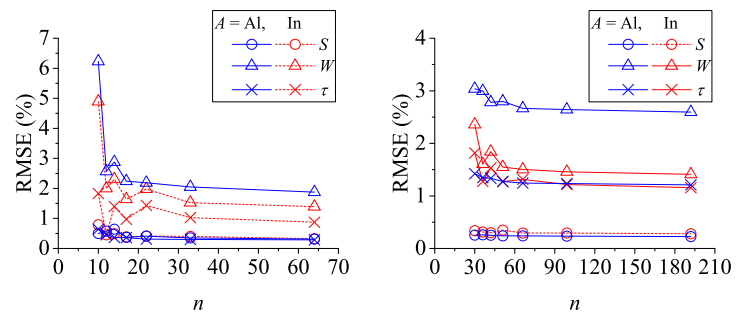
For the di-vacancy case, similar replacements of descriptors to those in the mono-vacancy case were made. The resultant descriptors are  $x1_{Ga}$  (exactly the same as the above),  $x2_{Ga}$  (exactly the same as the above),  $d_N^b$ , average distance of 3 nitrogen atoms from the cation to be removed on the vacancy-free structure,  $d1_C^b$ , average distance of 3 cations from the cation,  $d2_C^b$ , that of the other 9 cations,  $f_N$ , average force component of 3 nitrogen atoms toward the cation vacancy just after removal of the cation-nitrogen pair  $f1_C$ , average force component of 3 cations,  $f2_C$ , that of the other 9 cations. In Fig. 5, training-set size dependence of RMSE values is shown. Even at  $n = 64$  for the SG1 and at  $n = 192$  for the SG2, RMSE values are significantly worse than those for predictions from the optimized structures especially for  $W$ . It shows room for improvement in choosing descriptors. Despite this situation, at most 3 % errors are often acceptable depending on purposes. Again, utilization of about 1/4 of the total datasets brings similar results to those obtained using all the datasets.

## 4. Summary and concluding remarks

Close relationships between positron annihilation parameters and structural parameters are confirmed for mono- and di-vacancies in  $Al_{0.5}Ga_{0.5}N$  and  $In_{0.5}Ga_{0.5}N$ . It is shown that positron annihilation parameters are predictable by constructing multiple linear regression models with descriptors from the pre-optimized structures and that about 1/4 of the total datasets are sufficient to prepare the training set though there is room for improvement in choosing descriptors in the di-vacancy case. In future, the present study will be extended utilizing nonlinear regression techniques such as the Gaussian process.



**Figure 4.** Training-set size dependence of RMSE values of positron annihilation parameters for mono-vacancies.



**Figure 5.** Training-set size dependence of RMSE values of positron annihilation parameters for di-vacancies. The left panel is for the SG1 and the right panel is for the SG2.

### Acknowledgments

This work was supported in part by “Materials research by Information Integration” Initiative (MI<sup>2</sup>I) project of the Support Program for Starting Up Innovation Hub from JST.

### References

- [1] Uedono A, Ishibashi S, Oshima N and Suzuki R 2013 *Jpn. J. Appl. Phys.* **52** 08JJ02
- [2] Ishibashi S and Uedono A 2014 *J. Phys.: Conf. Ser.* **505** 012010
- [3] Ishibashi S and Uedono A 2016 *J. Phys.: Conf. Ser.* **674** 012020
- [4] Krause-Rehberg R and Leipner H S 1999 *Positron Annihilation in Semiconductors* (Berlin: Springer)
- [5] Tuomisto F and Makkonen I 2013 *Rev. Mod. Phys.* **85** 1583
- [6] Blöchl P E 1994 *Phys. Rev. B* **50** 17953
- [7] Zunger A, Wei S-H, Ferreira L G and Bernard J E 1990 *Phys. Rev. Lett.* **65** 353
- [8] R Core Team 2015 *R: A language and environment for statistical computing* (Vienna: R Foundation for Statistical Computing)