Supplementary Information

Performance of Deep Convolutional Neural Network to Classify Crystal Structures Using Selected Area Electron Beam Diffraction Patterns Containing Lattice Defect Information

Jae Min Jeong,[†a] Moonsoo Ra[†b], Jinha Jeong[*b], and Woong Lee[*a,c]

a Dept. of Materials Convergence and System Engineering, Changwon National University, 20 Changwondaehak-ro, Changwon-si, Gyeongsangnam-do 51140, Republic of Korea

b LightVision Inc., 20 Seongsuil-ro 12-gil, Seongdong-gu, Seoul 04793, Republic of Korea.

c School of Materials Science and Engineering, Changwon National University, 20 Changwondaehak-ro, Changwon-si, Gyeongsangnam-do 51140, Republic of Korea.

† These authors equally contributed as the first authors.

∗ Corresponding Authors. E-mails: trizmaster@gmail.com (J. Jeong), woonglee@changwon.ac.kr (W. Lee)

S1. Preparation of CIFs for Materials Having Defects Overview

The selected area electron beam diffraction (SAD) simulation code developed by the authors' research tram requires crystal information files (CIFs), written in accordance with the syntax structures provided by the Materials Project [S1], as inputs. Inside the code, the unit cell is repeated along three major crystallographic axes to create a virtual crystal. Typical number of repetitions is 20 × 20 × 40 along the length, width, and thickness directions, respectively. This procedure generates a virtual specimen with a thickness of 16.2 nm and 11.5 nm for Al and Fe, respectively. Therefore, the first step of this study is to prepare CIFs containing defect-related information such as changes in the unit cell geometry (shapes and lattice parameters), disordered positions of the lattice atoms (errors in atomic positions) due to the presence of defects, position of point defects when present, etc. Such necessary information was obtained in the following manner for each defect type considered, taking the lattice geometry and atomic positions in the original CIFs obtained from Materials Project for defect-free pristine materials as references. Once the necessary calculations were carried out, these original CIFs were modified to include the changes in the lattice geometry and atomic positions to prepare the CIFs for materials having defects to generate corresponding SADPs using TEM electron diffraction simulations.

S1. Preparation of CIFs for Strained Materials

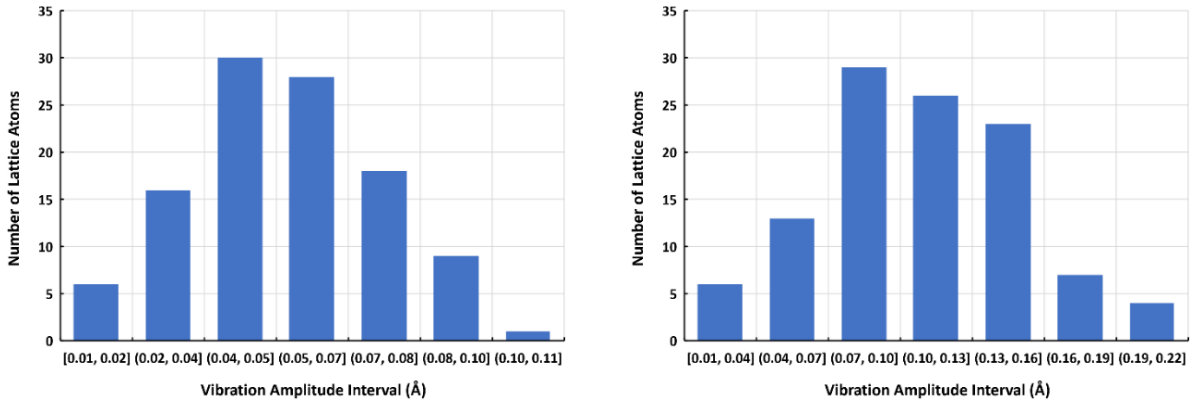
To generate SADPs for materials under mechanical strains, six lattice parameters, namely *a*, *b*, *c*, α, β, and γ in the original CIFs should be modified to store the strain state defined by six strain components ε_{xx} (= ε_1), ε_{yy} (= ε_2), ε_{zz} (= ε_3), ε_{yz} (= ε_3), ε_{zx} (= ε_4), and ε_{xy} (= ε_6) properly. The general relation between applied stress and accompanying strain is determined by the Hook's law which can be stated as following constitutive equation [S2]:

$$[\varepsilon_i] = [S_{ij}][\sigma_i] \text{ or } [\sigma_i] = [S_{ij}]^{-1}[\varepsilon_i] \ (i = 1 \cdots 6, j = 1 \cdots 6) \qquad \text{(S1)}$$

where $[\varepsilon_i]$ is the elastic strain tensor, $[S_{ij}]$ is the compliance tensor determined by the appropriate symmetry of the material of interest, and $[\sigma_i]$ is the stress tensor applied to the material. For materials belonging to the cubic system, normal strain along one direction due to a uniaxial stress accompanies strains in other two directions determined by Eqn. (S1), which is called Poisson effect, whereas shear strain components are mutually independent. Hence, in estimating the three normal strain components developing under uniaxial tension along the *a*-direction, Eqn. (S1) was solved numerically by varying σ_1 (= σ_{xx}) until ε_1 (= ε_{xx}) reached a preset value. In this process, the other two components (Poisson strains ε_{yy} and ε_{zz} along the *b*- and *c*-directions) were automatically calculated by Eqn. (S1). Once the normal strain components along the *a*-, *b*-, and *c*-directions were determined, the lattice parameters *a*, *b*, and *c* in the original CIF were adjusted accordingly to include the strain information. Concerning the shear strain component, the angle between *a*- and *b*-axes, viz the lattice parameter γ, were modified in the CIF to account for shear strains ε_{xy} (= ε_6). Since shear strain components are independent of each other if the material has cubic symmetry, it was not necessary to solve Eqn. (S1) for shear strains. The modified CIFs for appropriate strain states were then used as inputs to the diffraction simulation code to generate SADPs revealing strained lattice information. Since the positions of atoms are provided as reduced coordinate (point coordinate), modifying the lattice parameters automatically changes the cartesian coordinate of the lattice atoms when the CIF is read by the TEM SAD simulation code.

S2. Preparation of CIFs for Disordered Lattice by Thermal Vibrations

Consideration of thermal vibrations of the lattice atoms in generating SADPs requires the information about the instantaneous positions of the atoms which have been randomly displaced from their ordered position (the positions at 0 K). In this study, the disordered positions of lattice atoms at the temperatures of 100 K and 300 K were predicted by ab initio molecular dynamics (AIMD) simulations for the model system of Al (space group 225). Since the displacements of atoms due to thermal vibrations occur in random directions while the distribution of the vibration amplitudes is expected to follow the Maxwell-Boltzmann distribution, it is desirable to carry out the calculations for the largest possible supercells to account for the disordered lattice structure fully. However, limitations in computational cost allowed the AIMD simulations only on a 3 × 3 × 3 supercell which accommodates 108 Al atoms over 27 unit cells—face centered cubic structure means four atoms in every unit cell. An open source first principles simulation package ABINIT [S3] was used to predict the instantaneous positions of the atoms after lapse of 1 ps. Together with the disordered atomic positions, thermal expansion of the lattice was predicted simultaneously by the same simulation to maintain constant pressure condition while maintaining the cubic lattice structure. For the model system of Al considered in this study, the AIMD calculations predicted that the vibration amplitudes roughly follow the Maxwell-Boltzmann distribution as shown in the histogram in Fig. S1. The root mean square displacement along each crystallographic axis ($(\overline{u_i^2})^{1/2}$, $i = x, y, z$) were predicted to be in the range between 0.032 to 0.035 at 100 K and between 0.063 to 0.074 Å at 300 K, with the standard deviations between 0.018 to 0.021 Å at 100 K and between 0.037 to 0.043 Å at 300 K. These predictions appear to be reasonable compared with the experimental observations and theoretical predictions on other materials [S4, S5]. Based on the AIMD simulation results, the input CIFs were prepared by including the thermal expansion of the lattice as well as the displaced positions of all the 108 constituent atoms in the 3 × 3 × 3 in reduced coordinate. These CIF were used as input to the TEM SAD simulation code to generate SADPs for Al at temperatures of 100 K and 300 K respectively.



(a) (b)

Fig. S1 Histograms showing the distributions of the vibration amplitudes of lattice atoms in Al super lattice at (a) T = 100 K and (b) T = 300 K.

S3. Preparation of CIFs for Distorted Lattice by Point Defects

The presence of point defects displaces lattice atoms in their proximity to achieve new equilibria. Simply removing an atom from a lattice point, substituting a lattice atom with an impurity atom, placing an impurity atom at an interstitial site while retaining the lattice geometry intact do not properly reproduce the lattice structures with point defects. To find the equilibrium positions of the lattice atoms under the presence of the point defects, geometric optimization was carried out by first principles calculations using ABINIT. The model geometries consisted of a 2 × 2 × 2 supercell and a 3 × 3 × 3 supercell for Al, and a 3 × 3 × 3 supercell and a 4 × 4 × 4 supercell for Fe. Vacancy was modeled by removing one lattice atom from arbitrary lattice points in these supercells. Concerning the substitutional impurities, two model systems of Al-Si and Fe-Au systems were considered. For the Al-Si system, one lattice site in a 2 × 2 × 2 supercell and a 3 × 3 × 3 supercell was substituted with Si giving the impurity concentrations of 3.12 at% and 0.93 at%, respectively. For the Fe-Au system, one lattice site in a 3 × 3 × 3 supercell and a 4 × 4 × 4 supercell was substituted with Si giving the impurity concentrations of 1.85 at% and 0.78 at%, respectively. In the case of interstitial impurities, the Al-C binary system was considered. In the Al supercells, one C atom was placed in an arbitrary tetrahedral site. One C atom in a 3 × 3 × 3

supercell of Al corresponds to 0.92 at% of impurity concentration. The calculations were carried out to allow full optimization of the supercells including the changes in the dimension of the superlattice while seeking the equilibrium positions of the lattice atoms in the presence of the point defects using the Broyden-Fletcher-Goldfarb-Shanno minimization (BFGS) [S6] within the framework of generalized gradient approximation (GGA). At the same time, constraint condition was also applied to maintain the overall cubic symmetry of the lattice shape itself. Once the geometry optimizations were completed, CIFs were prepared by including the disordered positions of all the atoms in the corresponding supercells in reduced coordinate. These CIF were subsequently used as input to the TEM SAD simulation code to generate SADPs for materials having point defects as considered.

S4. Preparation of CIFs for Lattice Containing Edge Dislocation and Twin Boundary

An edge dislocation is formed when an extra atomic half plane exists midway through the crystal which results in an imperfection in lattice match between one half section with the extra plane and the other half section without the extra plane. Therefore, there exists imperfection in the lattice structure lacking appropriate symmetry around the dislocation. Dislocations do not exist arbitrarily and their presence in a crystal structure must comply with the slip systems of the corresponding material. In this study, Al system was considered to model the superlattices containing edge dislocations. In these cases, the slip system is <110>{111}. Instead of introducing extra half planes manually, a lattice manipulation package ATOMSK [S7], which has a capability of generating superlattices with edge dislocations, was used to create an input CIF for a supercell having an edge dislocation on a slip plane. The supercell with dislocations as created by ATOMSK is shown in Fig. S2. The supercell consists of 13824 atoms with a dislocation density of 2.61×10^{16} m^{-2}. This dislocation density is comparable to those observed in metals undergone plastic work such as rolling [S8]. In generating SADPs by simulation, a virtual crystal was created by repeating this supercell twice along *x*-, *y*-, and *z*-directions, respectively, to account for arrays of

dislocations as in real materials. This gives four and two equally spaced dislocations along the [110] and the [$\bar{1}$11] directions, respectively.

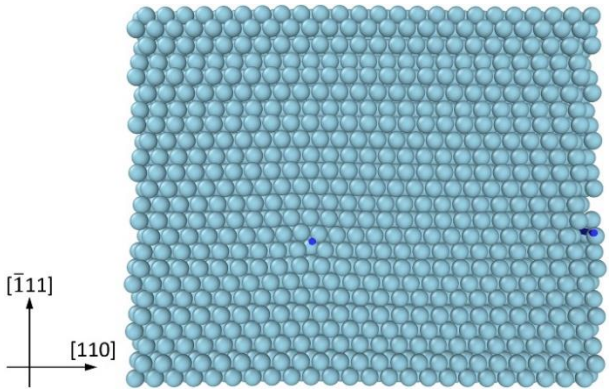
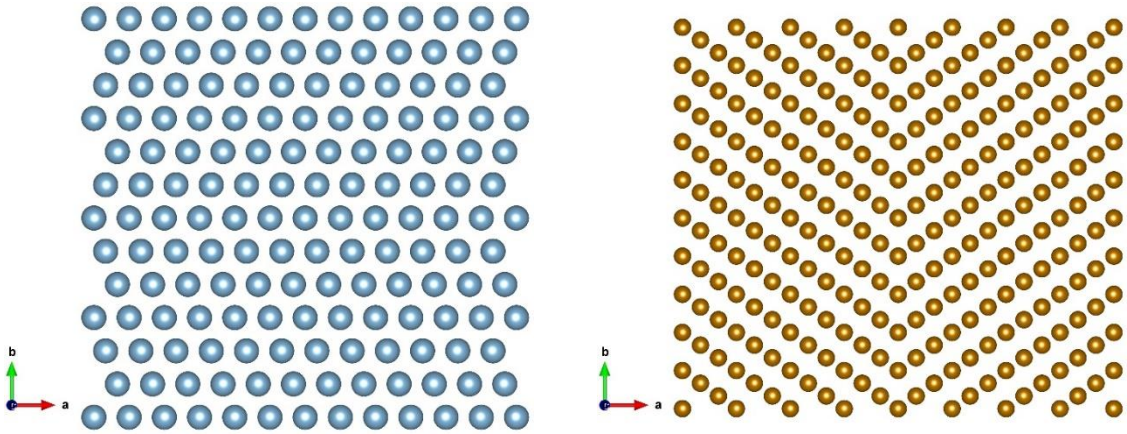


Fig. S2 A supercell of Al having an edge dislocation with the Burgers vector of [110]/2 which is marked by a blue dislocation line extending along the direction normal to the figure.



(a) (b)

Fig. S3 Visualized CIFs for twin systems of (a) Al and (b) Fe.

When sequences of stacking atomic plane change during the formation of lattice structure or the lattice is mechanically deformed along specific crystallographic directions on a specific plane, resulting lattice structure may appear mirror symmetric with respect to that specific plane called twin plane or twin boundary. Twin boundaries are defined according to the twin systems depending on crystal structures. In the Al and Fe model systems, the twin systems are <112>{111} and <110>{112}, respectively. To introduce twin boundaries in the model Al and Fe superlattices, twin construction capability of ATOMSK was used to create CIFs for Al and Fe having a twin

boundary. The resulting supercells are shown in Fig. S3.

References

S1 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Materials*, 2013, 1, 011002.

S2 T. Gnäupel-Herold, P. C. Brand, H. J. Prask, *Journal of Applied Crystallography*, 1998, 31, 929-935.

S3 X. Gonze, B. Amadon, G. Antonius, F. Arnardi, L. Baguet, J.-M. Beuken, J. Bieder, F. Bottin, J. Bouchet, E. Bousquet, N. Brouwer, F. Bruneval, G. Brunin, T. Cavignac, J.-B. Charraud, W. Chen, M. Côté, S. Cottenier, J. Denier, G. Geneste, J. W. Zwanziger, *Computer Physics Communications*, 2020, 248, 107042.

S4 A. Dygo, P. J. M. Smulders, D. O. Boerma, *Nuclear Instruments and Methods in Physics Research B*, 1992, 64, 701-705.

S5 A. Fomin, *Multiple Scattering Effects on the Dynamics and Radiation of Fast Charged Particles in Crystals. Transients in the Nuclear Burning Wave Reactor*, Ph.D. Thesis, Université Paris-Saclay, 2017.

S6 H. B. Schlegel, *Journal of Computational Chemistry*, 1982, 3, 214-218.

S7 P. Hirel, *Computer Physics Communications*, 2015, 197, 212-219.

S8 J. D. Verhoeven, *Fundamentals of Physical Metallurgy*, John Wiley and Sons, New York, 1975.