Data in Brief 55 (2024) 110689

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Dataset of simulated vibrational density of states and X-ray diffraction profiles of mechanically deformed and disordered atomic structures in Gold, Iron, Magnesium, and Silicon

Daniel Vizoso, Rémi Dingreville*

Center for Integrated Nanotechnologies, Sandia National Laboratories, 1515 Eubank Blvd SE, Albuquerque, NM 87185, USA

ARTICLE INFO

Article history:
Received 19 February 2024
Revised 19 May 2024
Accepted 20 June 2024
Available online 3 July 2024

Dataset link: Simulated Vibrational Density of States and X-ray Diffraction Profiles of Mechanically Deformed and Disordered Atomic Structures in Gold, Iron, Magnesium, and Silicon (Original data)

Keywords:
Spectroscopy
Diffractograms
Radiation damage
Molecular dynamics

ABSTRACT

This dataset is comprised of a library of atomistic structure files and corresponding X-ray diffraction (XRD) profiles and vibrational density of states (VDoS) profiles for bulk single crystal silicon (Si), gold (Au), magnesium (Mg), and iron (Fe) with and without disorder introduced into the atomic structure and with and without mechanical loading. Included with the atomistic structure files are descriptor files that measure the stress state, phase fractions, and dislocation content of the microstructures. All data was generated via molecular dynamics or molecular statics simulations using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) code. This dataset can inform the understanding of how local or global changes to a materials microstructure can alter their spectroscopic and diffraction behavior across a variety of initial structure types (cubic diamond, face-centered cubic (FCC), hexagonal close-packed (HCP), and body-centered cubic (BCC) for Si, Au, Mg, and Fe, respec-

* Corresponding author.
E-mail address: rdingre@sandia.gov (R. Dingreville).
Social media: @DingrevilleRemi (R. Dingreville)

https://doi.org/10.1016/j.dib.2024.110689

tively) and overlapping changes to the microstructure (i.e., both disorder insertion and mechanical loading).

Specifications Table

Subject	Materials Science: Material Characterization
Specific subject area	Quantitative description of materials for high-throughput materials analysis.
Type of data	Tables: Raw and Analyzed dump.* or *.dump files (LAMMPS dump files) *.dat (velocity autocorrelation function data) *.out (XRD compute output) *.txt (Processed vibrational density of states files or characterization data of the LAMMPS dump files) Supporting Materials (example simulation scripts, README files, interatomic potential files)
Data collection	Data collected from LAMMPS molecular dynamics simulations of bulk single crystal silicon (Si), gold (Au), magnesium (Mg), and iron (Fe) with and without disorder insertion and with and without mechanical loading.
Data source location	Data generated at the Center for Integrated Nanotechnologies at Sandia National Laboratories, New Mexico, United States of America
Data accessibility	Repository name: Materials Data Facility (MDF) Data identification number: doi: 10.18126/tacz-v14v Direct URL to data: https://acdc.alcf.anl.gov/mdf/detail/sandia_vdos_xrd_si_fe_mg_si_v1.1/ Instructions for accessing these data: By following the provided URL, the user can transfer the dataset from the MDF Globus storage location to another Globus collection. To download the data, the user must install Globus Connect Personal, which can then be used to download the data to a personal machine.
Related research article	A portion of this dataset was used for the following publication [1]: Daniel Vizoso, Ghatu Subhash, Krishna Rajan, Rémi Dingreville *Chemistry of Materials* **2023** *35* (3), 1186–1200

1. Value of the Data

- With more than 11,000 unique atomic structure files along with their corresponding spectroscopy and diffraction characterizations, this dataset provides researchers access to a substantial range of material states in different material systems without requiring to perform additional costly atomistic simulations or associated generation of spectroscopy and diffraction data.
- Atomistic structure files may be used by researchers to perform their own characterization using a variety of publicly available software packages if they desire access to material properties that were not included in the dataset (for example, point defect densities, stacking fault area, etc.).
- With access to both spectroscopy and diffraction data, researchers, material scientists, and data scientists can compare the atomic structure files or the provided material properties to either modality or perform multimodal analysis that combines the information encoded in the spectroscopy and diffraction data to create better correlations to material properties or atomic structures.
- This dataset can be used to develop or test empirical relations between spectroscopy and diffraction data to global or local microstructural properties, or to aid in the development of machine-learning methods for performing a variety of tasks related to connecting atomic structures, material properties, and popular materials characterization techniques.

2. Background

Spectroscopy and diffraction measurements are frequently used non-destructive characterization techniques to extract a wide variety of material properties. Diffraction techniques, such as X-ray (XRD) or neutron diffraction, provide information on the spatial arrangement of atoms within crystalline materials. When X-rays encounter a crystalline sample, they diffract in specific directions due to the periodic arrangement of atoms in the crystal lattice. This diffraction pattern contains information about the spacing and orientation of the crystal planes within the material. By analysing the angles and intensities of the diffracted X-rays, researchers can deduce the crystal structure, including parameters such as lattice constants, unit cell dimensions, and atomic positions. On the other hand, vibrational spectroscopy, such as infrared or Raman spectroscopy, probe the atomic structure vibrations and interactions within materials, offering information regarding the nature of chemical composition and bonds. Different chemical bonds exhibit characteristic vibrational frequencies, resulting in a shift in wavelength. The corresponding vibrational density of state (VDoS) spectrum provides information about atomic vibrations, crystal symmetry, and chemical bonding, complementing the insights obtained from diffractograms. As such, in the dataset provided in this Data in Brief, we generated a database of atomic structures with and without disorder and with and without mechanical loading. For each of these atomic structures we generated the corresponding XRD and VDoS spectra.

Traditional analysis of spectroscopy and diffraction spectra is typically performed manually via peak analysis and is prone to biases introduced by the practitioner or by the specific analysis techniques used that can lead to inconsistent results or inaccurate predictions. This dataset was created for the purpose of testing various machine-learning (ML) methods for extracting material properties from spectroscopic and diffraction data [1,2], with sub-objectives of testing how well ML techniques can be used to automate the analysis of these types of datasets without requiring human intervention as well as to test whether additional material properties could be extracted from these datasets via ML techniques that are typically considered inaccessible via traditional analysis methods. A portion of the Si data provided in this dataset was used for a previous publication [1], and this represents a substantially expansion of that work, providing researchers the ability to test their approaches on multiple characterization modalities and across multiple material systems.

3. Data Description

The dataset is available with open access [3]: https://acdc.alcf.anl.gov/mdf/detail/sandia_vdos_xrd_si_fe_mg_si_v1.1/. Instructions on how to download the data can be found on Globus: https://docs.globus.org/guides/tutorials/manage-files/transfer-files/. This dataset is comprised of 11,813 atomistic structure files (see Fig. 1(a)) and their corresponding XRD and VDoS profiles (see Fig. 1(b) and (c) respectively). An example of these modalities is presented in Fig. 1 for a bulk, single-crystal, iron atomic structure with an amount of disorder corresponding to a displacement damage level of 0.15 displacements per atom (dpa) (see Methods section for description of displacement damage) and a hydrostatic tensile strain of 0.05 (see Methods section for description of tensile strain). As such, this dataset provides a comprehensive library of atomic structures in various state of mechanical deformation and/or disorder along with their corresponding XRD and VDoS spectra.

The structure of this dataset is broken into the following main directories: 'Python Scripts', 'Si', 'Au', 'Mg', and 'Fe'. The 'Si', 'Au', 'Mg', and 'Fe', as well as the head directory containing these subdirectories, each contain a 'README_*.txt' file that provides descriptions of the contents of each directory and its subdirectories. Organization of the dataset is primarily handled through nested directories. Tabular descriptions of each of the main directories are provided in Tables 1-5.

Table 1
Description of python scripts directory: python scripts/contents.

File Name	Description
VDoS_Compute.py	Python script for computing the vibrational density of states from a data file containing velocity autocorrelation function data.

Table 2
Description of silicon (Si) subdirectory: Si/ Contents.

File/Directory Name	File/Directory Description
README_Si.txt	README file that describes organization of data.
LAMMPS_Inputs/	Directory containing LAMMPS input scripts for recreating the data provided in this dataset, as well as all necessary supplementary files.
Postprocessing_Scripts/	Directory containing an OVITO script (phase_fraction.py) for measuring the phase fraction of atomistic structures.
Molecular_Dynamics_Files/	Directory containing atomistic structure files and corresponding phase fraction, dislocation content, and stress measurements.
VDoS_Files/	Directory containing raw velocity autocorrelation function (VACF) measurements and vibrational density of state (VDoS) files derived from these VACF measurements for the atomistic structures provided in Molecular_Dynamics_Files/.
XRD_Files/	Directory containing raw X-ray diffraction (XRD) measurements for the atomistic structures provided in Molecular_Dynamics_Files/.
LAMMPS_Inputs/ Contents:	
Si_1.meam.spline	Interatomic potential file used for LAMMPS simulations
in.Si_*	LAMMPS input scripts, purpose of script is defined by name of file that replaces the asterisk (*).
Molecular_Dynamics/_Files, VDoS_Files/, and XRD_Files/ share the following subdirectory architecture:	
FP_Structures/	Data related to atomistic simulations involving Frenkel pair insertion.
Mechanical_Loading/	Data related to atomistic simulations of the mechanical loading of initially pristine or disordered single crystal structures.
Mechanical_Loading/ Contents:	
Pristine/, 0p10/,0p20/, 0p30/…	Subdirectories that store information for mechanical loading simulations with different initial microstructures, which are described by the name of the subdirectory.
Contents of subdirectories in Mechanical_Loading/:	
(Hydrostatic or Uniaxial)_(Compression or Tension)/	Each of these subdirectories stores information related to mechanical loading simulations involving either hydrostatic or uniaxial compression or tension.
Molecular_Dynamics_Files/ Data File Naming Conventions:	
.dump or dump.	LAMMPS dump file of an atomistic structure. The asterisk (*) is replaced by text indicating the type of simulation that was performed or the material state described in that file (i.e., number of Frenkel pairs, strain, etc.).
*.dat	Text file with stress information, broken into the following components: x, y, z, xy, xz, yz. The asterisk is replaced by text describing the type of simulation.
*_phase_fraction.txt	Text file with phase fraction information. The asterisk (*) is replaced by text describing the specific simulation that the data file is related to.
VDoS_Files/ Data File Naming Conventions:	
*.dat	Text file with VACF output information from LAMMPS simulations. The asterisk (*) is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.
pdos_*.out	Text file with calculated VDoS data. The asterisk (*) is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.
XRD_Files/ Data File Naming Conventions:	
*_xrd.out	Text file with XRD data from LAMMPS. The asterisk (*) is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.

Table 3
Description of gold (Au) subdirectory: Au/ Contents.

File/Directory Name	File/Directory Description
README_Au.txt	README file that describes organization of data.
LAMMPS_Inputs/	Directory containing LAMMPS input scripts for recreating the data provided in this dataset, as well as all necessary supplementary files.
Postprocessing_Scripts/	Directory containing an OVITO script (phase_get.py) for measuring phase fraction and dislocation content of atomistic structures.
Molecular_Dynamics_Files/	Directory containing atomistic structure files and corresponding phase fraction, dislocation content, and stress measurements.
VDoS_Files/	Directory containing raw velocity autocorraltion function (VACF) measurements and vibrational density of state (VDoS) files derived from these VACF measurements for the atomistic structures provided in Molecular_Dynamics_Files/.
XRD_Files/	Directory containing raw X-ray diffraction (XRD) measurements for the atomistic structures provided in Molecular_Dynamics_Files/.
LAMMPS_Inputs/ Contents:	
Au_stopping.txt	Stopping power data used for some LAMMPS simulations.
Au.lammps.eam	Interatomic potential file used for LAMMPS simulations.
in.Au_*	LAMMPS input scripts, purpose of script is defined by name of file that replaces the asterisk (*).
Molecular_Dynamics/_Files, VDoS_Files/, and XRD_Files/ share the following subdirectory architecture:	
Single_Cascades/	Data related to atomistic simulations of single collision cascades at different energies, which are stored in separate subdirectories.
ROAC_Structrues/	Data related to atomistic simulations of disorder insertion.
Mechanical_Loading/	Data related to atomistic simulations of the mechanical loading of initially pristine or disordered single crystal structures.
Single_Cascades/ Contents:	
*_keV_PKA/	Subdirectories that store information for single cascade simulations that were performed at a single event energy, where the energy replaces the asterisk in the directory name.
Mechanical_Loading/ Contents:	
pristine/, 0p05/,0p10/, 0p15/...	Subdirectories that store information for mechanical loading simulations with different initial microstructures, which are described by the name of the subdirectory.
Contents of subdirectories in Mechanical_Loading/:	
(Hydrostatic or Uniaxial)_(Compression or Tension)/	Each of these subdirectories stores information related to mechanical loading simulations involving either hydrostatic or uniaxial compression or tension.
Molecular_Dynamics_Files/ Data File Naming Conventions:	
.dump or dump.	LAMMPS dump file of an atomistic structure. The asterisk is replaced by text indicating the type of simulation that was performed or the material state described in that file (i.e., dpa, strain, etc.).
*_stress.dat	Text file with stress information, broken into the following components: x, y, z, xy, xz, yz. The asterisk is replaced by text describing the type of simulation.
*_disl_+.txt	Text file with dislocation content information. The asterisk (*) is replaced by text describing the specific simulation that the data file is related to, and the plus symbol (+) is replaced by either fcc, bcc, or hcp, relating to the input crystal type that was used when computing the dislocation content in OVITO.
*_phase_info.txt	Text file with phase fraction information. The asterisk (*) is replaced by text describing the specific simulation that the data file is related to.
VDoS_Files/ Data File Naming Conventions:	
*.dat	Text file with VACF output information from LAMMPS simulations. The asterisk (*) is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.
*_vdos.txt	Text file with calculated VDoS data. The asterisk (*) is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.
XRD_Files/ Data File Naming Conventions:	
*_xrd.out	Text file with XRD data from LAMMPS. The asterisk is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.

Table 4
Description of magnesium (Mg) subdirectory: Mg/ Contents.

File/Directory Name	File/Directory Description
README_Mg.txt	README file that describes organization of data.
LAMMPS_Inputs/	Directory containing LAMMPS input scripts for recreating the data provided in this dataset, as well as all necessary supplementary files.
Postprocessing_Scripts/	Directory containing an OVITO script (phase_get.py) for measuring phase fraction and dislocation content of atomistic structures.
Molecular_Dynamics_Files/	Directory containing atomistic structure files and corresponding phase fraction, dislocation content, and stress measurements.
VDoS_Files/	Directory containing raw velocity autocorrelation function (VACF) measurements and vibrational density of state (VDoS) files derived from these VACF measurements for the atomistic structures provided in Molecular_Dynamics_Files/.
XRD_Files/	Directory containing raw XRD measurements for the atomistic structures provided in Molecular_Dynamics_Files/.
LAMMPS_Inputs/ Contents:	
Mg_stopping.txt	Stopping power data used for some LAMMPS simulations.
Mg.lammps.eam	Interatomic potential file used for LAMMPS simulations.
in.Mg_*	LAMMPS input scripts, purpose of script is defined by name of file that replaces the asterisk (*).
Molecular_Dynamics_Files/, VDoS_Files/, and XRD_Files/ share the following subdirectory architecture:	
Single_Cascades/	Data related to atomistic simulations of single collision cascades at different energies, which are stored in separate subdirectories.
ROAC_Structrues/	Data related to atomistic simulations of disorder insertion.
Mechanical_Loading/	Data related to atomistic simulations of the mechanical loading of initially pristine or disordered single crystal structures.
Single_Cascades/ Contents:	
*_keV_PKA/	Subdirectories that store information for single cascade simulations that were performed at a single event energy, where the energy replaces the asterisk (*) in the directory name.
Mechanical_Loading/ Contents:	
pristine/, 0p05/,0p10/, 0p15/...	Subdirectories that store information for mechanical loading simulations with different initial microstructures, which are described by the name of the subdirectory.
Contents of subdirectories in Mechanical_Loading/:	
Compression_*/ or Tension_*/	Each of these subdirectories stores information related to mechanical loading simulations involving either uniaxial tension or compression, with the asterisk (*) being replaced by the direction of loading (either X, Y, or Z).
Molecular_Dynamics_Files/ Data File Naming Conventions:	
.dump or dump.	LAMMPS dump file of an atomistic structure. The asterisk (*) is replaced by text indicating the type of simulation that was performed or the material state described in that file (i.e., dpa, strain, etc.).
*_stress.dat	Text file with stress information, broken into the following components: x, y, z, xy, xz, yz. The asterisk is replaced by text describing the type of simulation.
*_disl_+.txt	Text file with dislocation content information. The asterisk (*) is replaced by text describing the specific simulation that the data file is related to, and the plus symbol (+) is replaced by either fcc, bcc, or hcp, relating to the input crystal type that was used when computing the dislocation content in OVITO.
*_phase_info.txt	Text file with phase fraction information. The asterisk (*) is replaced by text describing the specific simulation that the data file is related to.
VDoS_Files/ Data File Naming Conventions:	
*.dat	Text file with VACF output information from LAMMPS simulations. The asterisk (*) is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.
*_vdos.txt	Text file with calculated VDoS data. The asterisk (*) is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.
XRD_Files/ Data File Naming Conventions:	
*_xrd.out	Text file with XRD data from LAMMPS. The asterisk (*) is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.

Table 5
Description of iron (Fe) subdirectory: Fe/ Contents.

File/Directory Name	File/Directory Description
README_Fe.txt	README file that describes organization of data.
LAMMPS_Inputs/	Directory containing LAMMPS input scripts for recreating the data provided in this dataset, as well as all necessary supplementary files.
Postprocessing_Scripts/	Directory containing an OVITO script (phase_get.py) for measuring phase fraction and dislocation content of atomistic structures.
Molecular_Dynamics_Files/	Directory containing atomistic structure files and corresponding phase fraction, dislocation content, and stress measurements.
VDoS_Files/	Directory containing raw velocity autocorrelation function (VACF) measurements and vibrational density of state (VDoS) files derived from these VACF measurements for the atomistic structures provided in Molecular_Dynamics_Files/.
XRD_Files/	Directory containing raw XRD measurements for the atomistic structures provided in Molecular_Dynamics_Files/.
LAMMPS_Inputs/ Contents:	
Fe_stopping.txt	Stopping power data used for some LAMMPS simulations.
M07_eam.fs	Interatomic potential file used for LAMMPS simulations.
in.Fe_*	LAMMPS input scripts, purpose of script is defined by name of file that replaces the asterisk (*).
Molecular_Dynamics/_Files, VDoS_Files/, and XRD_Files/ share the following subdirectory architecture:	
Single_Cascades/	Data related to atomistic simulations of single collision cascades at different energies, which are stored in separate subdirectories.
ROAC_Structrues/	Data related to atomistic simulations of disorder insertion.
Mechanical_Loading/	Data related to atomistic simulations of the mechanical loading of initially pristine or disordered single crystal structures.
Single_Cascades/ Contents:	
*_keV_PKA/	Subdirectories that store information for single cascade simulations that were performed at a single event energy, where the energy replaces the asterisk (*) in the directory name.
Mechanical_Loading/ Contents:	
pristine/, 0p05/,0p10/, 0p15/...	Subdirectories that store information for mechanical loading simulations with different initial microstructures, which are described by the name of the subdirectory.
Contents of subdirectories in Mechanical_Loading/:	
(Hydrostatic or Uniaxial)_(Compression or Tension)/	Each of these subdirectories stores information related to mechanical loading simulations involving either hydrostatic or uniaxial compression or tension.
Molecular_Dynamics_Files/ Data File Naming Conventions:	
.dump or dump.	LAMMPS dump file of an atomistic structure. The asterisk (*) is replaced by text indicating the type of simulation that was performed or the material state described in that file (i.e., dpa, strain, etc.).
*_stress.dat	Text file with stress information, broken into the following components: x, y, z, xy, xz, yz. The asterisk (*) is replaced by text describing the type of simulation.
*_disl_+.txt	Text file with dislocation content information. The asterisk (*) is replaced by text describing the specific simulation that the data file is related to, and the plus symbol (+) is replaced by either fcc, bcc, or hcp, relating to the input crystal type that was used when computing the dislocation content in OVITO.
*_phase_info.txt	Text file with phase fraction information. The asterisk (*) is replaced by text describing the specific simulation that the data file is related to.
VDoS_Files/ Data File Naming Conventions:	
*.dat	Text file with VACF output information from LAMMPS simulations. The asterisk (*) is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.
*_vdos.txt	Text file with calculated VDoS data. The asterisk is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.
XRD_Files/ Data File Naming Conventions:	
*_xrd.out	Text file with XRD data from LAMMPS. The asterisk is replaced by text describing the simulation, which relates each file to a corresponding LAMMPS dump file in the matching Molecular_Dynamics_Files/ subdirectory.

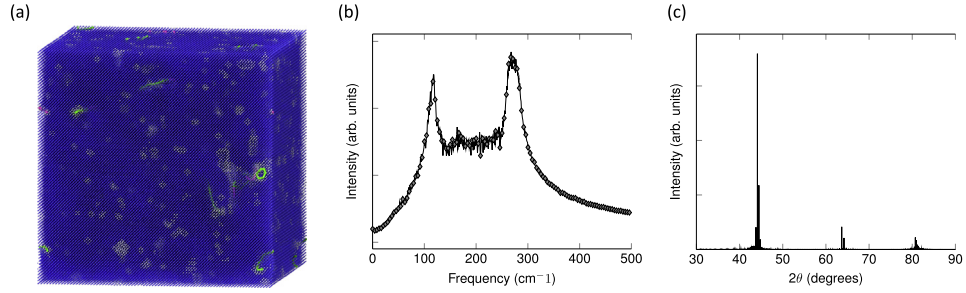

Fig. 1. (a) Render of a bulk Fe microstructure at 0.15 dpa and 0.05 hydrostatic tensile strain. Atoms are coloured according to their phase: blue is BCC, white is disordered. Dislocation lines are coloured according to their type: green are ½⟨111⟩ and magenta are ⟨100⟩. Corresponding (b) VDoS profile, truncated at 500 cm^{-1} and (c) XRD profile for the microstructure rendered in (a).

4. Experimental Design, Materials and Methods

This dataset contains and is derived from a set of monocrystalline, fully periodic (i.e., bulk) silicon (Si), gold (Au), magnesium (Mg), and iron (Fe) atomic structures that were created and simulated using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) [4]. The interatomic potentials used to describe the interatomic interactions in these four materials as well as sample LAMMPS input scripts for all the simulations performed to create the data provided in this dataset are provided as part of the dataset. XRD profiles were created using the *compute xrd* modifier in LAMMPS [5], while VDoS profiles were created via the velocity autocorrelation function (VACF) method [1], with the VACF computed using the *compute vacf* modifier in LAMMPS. The vibrational density of states (VDoS) was calculated from the collected VACF data via the FFT method in a Python script, which is also included as part of the dataset. Measurements of characteristic materials descriptors of the simulated atomic structures such as phase fractions and dislocation densities were performed using the Open Visualization Tool (OVITO) [5], with the scripts used to generate the provided outputs present in the dataset itself.

Molecular dynamics simulations of Si were performed with a modified embedded atom method (MEAM) interatomic potential developed by Lenosky et al. [6], with the cubic diamond structure having an initial lattice constant of 5.431 Å. Simulations of the compression or tension of Si without the introduction of disorder were performed with a supercell consisting of $12\times12\times12$ unit cells containing 13,824 atoms. Simulations of Si involving the introduction of disorder (see description of the method later on in this manuscript) were performed with a supercell consisting of $19\times19\times19$ unit cells containing 54,872 atoms. Simulations of Au were performed with an embedded atom method (EAM) potential developed by Sheng et al. [7], with an initial lattice constant of 4.078 Å and an initial crystallographic structure of face-centered cubic (FCC). Most Au simulations were performed with a supercell consisting of $63\times63\times63$ units cell containing a total of 1000,188 atoms, while the size of the simulation box for the single cascade simulations were scaled based on the energy of the Primary Knock-on Atom (PKA) . Simulations of Mg were performed with an EAM potential developed by Pei et al. [8], with an initial lattice constant of 3.2094 Å and an initial crystallographic structure of hexagonal close-packed (HCP). Most Mg simulations were performed with a supercell consisting of $90\times52\times53$ unit cells containing a total of 992,160 atoms, while the size of the simulation box for the single cascade simulations were scaled based on the energy of the PKA. Simulations of Fe were performed with an EAM potential that included a modification styled after Finnis-Sinclair potentials [9], with an initial lattice constant of 2.85528 Å and an initial crystallographic structure of body-centered cubic (BCC). Most Fe simulations were performed with a supercell consisting of $80\times80\times80$ unit cells containing a total of 1024,000 atoms, while the size of the simulation box for the single cascade simulations were scaled based on the energy of the PKA.

All the cubic structures (Si, Au, and Fe) were oriented such that the ⟨100⟩ family of crystallographic directions were along the *x*, *y*, and *z* axes of the simulation box. For Mg, the [0001] direction was oriented along the *z* axis of the simulation box, the $[10\bar{1}0]$ direction was oriented along the *x* axis, and the $[01\bar{1}0]$ direction was oriented along the *y* axis.

Prior to the introduction of disorder or the application of any mechanical load, each of the atomic structures was equilibrated at 300 K and zero pressure under an isothermal-isobaric (NPT) ensemble for a minimum of 200 ps. All simulations of compression or tension were performed under a canonical (NVT) ensemble at a constant temperature of 300 K, with the compression or tension of the simulation cell being handled by the *fix deform* command in LAMMPS using atomic displacements. All mechanical simulations were strain-controlled with constant engineering strain rates of 10^{-4} ps^{-1} (10^{8} s^{-1}).

For Si, Fe, and Au simulations, simulations of uniaxial tension or compression were performed along the ⟨001⟩ direction while maintaining the dimensions of the simulation cell in the ⟨100⟩ and ⟨010⟩ directions. For Mg simulations, uniaxial tension or compression simulations were performed along the [0001], $[10\bar{1}0]$, or $[01\bar{1}0]$ directions while keeping the other two dimensions constant. Hydrostatic simulations of tension or compression in the cubic structures (i.e., Si, Fe, and Au) were performed by reducing or expanding the dimensions of the simulation cells along each of the ⟨100⟩ directions at strain rates computed with the following relation: $1+\underset{tot}{\varepsilon} = (1+\varepsilon_{xx})(1+\varepsilon_{yy})(1+\varepsilon_{zz})$, where $\underset{tot}{\varepsilon}$ is the previously defined strain rate of 10^{-4} ps^{-1}, and $\underset{xx}{\varepsilon}$, $\underset{yy}{\varepsilon}$, and $\underset{zz}{\varepsilon}$ are the strain rates along each of the simulation cell axes. All simulations of uniaxial or hydrostatic tension or compression were performed to a final engineering strain of 30 %.

Disorder was introduced to the atomic structures via two different approaches. For the Si structures, disorder was introduced in the form of Frenkel pairs (vacancy and self-interstitial atom pairs), as described in [1,10–12]. In this approach, disorder was introduced in a step-wise fashion where during each disorder insertion step, 50 randomly selected atoms were displaced from their original position by a randomly sampled distance between 20 and 60 Å. The positions of the displaced atoms were adjusted until no atoms were closer than 2 Å from any neighbor atoms. After 50 atoms were selected and displaced, the system was allowed to evolve under an NPT ensemble maintained at zero pressure for 0.5 ps with a timestep of 0.1 fs, followed by a 2.5 ps evolution with a timestep of 0.5 fs. These disorder insertion steps were repeated until the total number of displaced atoms equaled the number of atoms within the structure.

For Au, Mg, and Fe structures, disorder was introduced using the reduced-order atomistic cascade (ROAC) method [10]. In the ROAC method, multiple damage events consisting of thermally excited core regions and ballistic shells are inserted into the microstructure and evolved in time, resulting in thermally induced mixing in the core regions and the creation of interstitial-type defects in the shell regions. The number of atoms involved in each region for each damage event is determined with an extension of the Norgett-Robinson-Torrens (NRT) dpa model [13,14], as well as an event energy that is sampled from a primary recoil spectrum computed for each material [10,15]. For all materials, an incident ion energy of 1 MeV was used to compute the primary recoil spectra, with the sampling of the spectrum being truncated to exclude the top 5 % of energies for the Au and Fe material systems while the top 3 % of energies were excluded for the Mg material system. For disordering simulation of Au, 800 damage events were inserted during each damage step, with the dynamics of the system between damage insertion steps progressing as 0.5 ps with a timestep size of 0.1 fs, then 70 ps with a timestep size of 1 fs. For disordering simulations of Fe and Mg, 1500 and 500 damage events were inserted per damage step respectively, with the dynamics of the systems between damage insertion stpes progressing as 70 ps with a timestep of 1 fs. Table 6 contains the material constants for the modified NRT-dpa defect production model used as part of the ROAC method. LAMMPS input scripts for performing defect insertion via both methods are provided in the dataset.

The process of attaining the constants provided in Table 6 is described in references [10,15], and involves performing a series of single cascade (or PKA) simulations at different event ener-

Table 6
Material constants for modified NRT-dpa model using in the ROAC model (see [10]).

Material	b_{dpa}	c_{dpa}	b_{rpa}	c_{rpa}	E_d (eV)
Au	−1.1	0.10	1800	1.05	35.0
Fe	−0.85	0.11	2500	0.95	40.0
Mg	−1.0	0.315	434.835	1.34866	16.5

gies. Defected structures produced by these single cascade simulations are also included in the dataset, alongside their corresponding VDoS and XRD spectra.

Simulations that involved both disorder and mechanical loading were performed in the following manner. First, disorder insertion simulations were performed to a desired damage level using the disorder insertion method appropriate for the material as defined above. Then, compressive or tensile loading simulations were performed with the disordered structures directly, with no stress relaxation prior to the beginning of the mechanical loading simulations.

Stress information was computed directly within the molecular dynamics simulations and written to a text files by LAMMPS directly. Stress data is stored in the dataset as directional components in the following order: x, y, z, xy, xz, yz. Phase fraction information was computed from the atomistic structure files created during the molecular dynamics simulations via the 'common neighbor analysis' modifier [16] within OVITO [17] using default parameters. Dislocation content information was computed from the atomistic structure files via the 'dislocation analysis' modifier [18] within OVITO. For all of the Au, Mg, and Fe structures, the 'dislocation analysis' modifier was run with input crystal types of FCC, HCP, and BCC.

VDoS data is obtained via the following process. First, a molecular dynamics simulation is performed where the VACF is measured using the *compute vacf* command in LAMMPS. This simulation is performed under an NVT ensemble maintained at 300 K for a simulated time of 400 ps, where the velocity autocorrelation value is recorded every 0.01 ps. The VDoS is then computed by taking the Fourier transform of the VACF data to convert it into the frequency domain. The result of the Fourier transform contains both real and imaginary components; the data provided in the 'VDoS_Files' subdirectories in the dataset only contain the real components.

XRD data is obtained directly from a LAMMPS simulation, where the atomic structure file is read in and the 'compute xrd' command is called in LAMMPS. Simulations for determining the XRD profiles of Si microstructures were performed with an X-ray wavelength of 1.541838 Å, a 2θ range from 20° to 85°, and a reciprocal lattice resolution of 0.006 $Å^{-1}$. Simulations for determining the XRD profiles of Au, Fe, and Mg microstructures were performed with an X-ray wavelength 0f 1.54 Å, a 2θ range from 30° to 90°, and a reciprocal lattice spacing of 0.005 $Å^{-1}$.

Limitations

Not Applicable.

Ethics Statement

The authors confirm that this dataset has been collected, curated, and reported in compliance with the ethical publishing requirements for this journal and this publisher. This work did not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

Simulated Vibrational Density of States and X-ray Diffraction Profiles of Mechanically Deformed and Disordered Atomic Structures in Gold, Iron, Magnesium, and Silicon (Original data) (Materials Data Facility)

CRediT Author Statement

Daniel Vizoso: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft, Writing – review & editing; **Rémi Dingreville:** Conceptualization, Writing – original draft, Writing – review & editing, Resources, Project administration, Funding acquisition.

Acknowledgements

Computational and modeling resources are supported by the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. Department of Energy. This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC, under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title, and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (https://www.energy.gov/sites/prod/files/2014/08/f18/DOE_Public_Access%20Plan_FINAL.pdf).

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.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.110689.

References

[1] D. Vizoso, G. Subhash, K. Rajan, R. Dingreville, Connecting vibrational spectroscopy to atomic structure via supervised manifold learning: beyond peak analysis, Chem. Mater. 35 (2023) 1186–1200, doi:10.1021/acs.chemmater.2c03207.
[2] C. Kunka, et al., Decoding defect statistics from diffractograms via machine learning, npj Comput. Mater. 7 (1) (2021) 67, doi:10.1038/s41524-021-00539-z.
[3] D. Vizoso, R. Dingreville, Simulated vibrational density of states and x-ray diffraction profiles of mechanically deformed and disordered atomic structures in gold, iron, magnesium, and silicon, Mater. Data Facil. (2023), doi:10.18126/tacz-v14v.
[4] A.P. Thompson, et al., LAMMPS – a flexible simulation tool for particle-based materials modelling at the atomic, meso, and continuum scales, Comp. Phys. Comm. 271 (2022) 10817, doi:10.1016/j.cpc.2021.108171.
[5] S.P. Coleman, D.E. Spearot, L. Capolungo, Virtual diffraction analysis of Ni [010]symmetric tilt grain boundaries, Model. Simul. Mater. Sci. Eng. 21 (2013) 055020, doi:10.1088/0965-0393/21/5/055020.
[6] T.J. Lenosky, et al., Highly optimized empirical potential model of silicon, Model. Simul. Mater. Eng. 8 (2000) 825, doi:10.1088/0965-0393/8/6/305.
[7] H.W. Sheng, M.J. Kramer, A. Cadien, T. Fujita, M.W. Chen, Highly optimized embedded-atom-method potentials for fourteen fcc metals, Phys. Rev. B 83 (2011) 134118, doi:10.1103/PhysRevB.83.134118.
[8] Z. Pei, H. Sheng, X. Zhang, R. Li, B. Svendsen, Tunable twin stability and an accurate magnesium interatomic potential for dislocation-twin interactions, Mater. Des. 153 (2018) 232–241, doi:10.1016/j.matdes.2018.04.085.
[9] L. Malerba, et al., Comparison of empirical interatomic potentials for iron applied to radiation damage studies, J. Nucl. Mater. 406 (2010) 19–38, doi:10.1016/j.jnucmat.2010.05.017.
[10] E.Y. Chen, C. Deo, R. Dingreville, Reduced-order atomistic cascade method for simulating radiation damage in metals, J. Phys. Condens. Matter 32 (2020) 045402, doi:10.1088/1361-648X/ab4b7c.
[11] A. Chartier, C. Meis, J.-P. Crocombette, W.J. Weber, L.R. Corrales, Molecular dynamic simulation of disorder induced amorphization in pyrochlore, Phys. Rev. Lett. 94 (2005) 025505, doi:10.1103/PhysRevLett.94.025505.

[12] A. Chartier, C. Onofri, L. Van Brutzel, C. Sabathier, O. Dorosh, J. Jagielski, Early stages of irradiation induced dislocations in urania, Appl. Phys. Lett. 109 (2016) 181902, doi:10.1063/1.4967191.
[13] M.T. Robinson, I.M. Torrens, Computer simulation of atomic-displacement cascades in solids in the binary-collision approximation, Phys. Rev. B 9 (1974) 5008, doi:10.1103/PhysRevB.9.5008.
[14] K. Nordlund, et al., Improving atomic displacement and replacement calculations with physically realistic damage models, Nat. Commun. 9 (2018) 1084, doi:10.1038/s41467-018-03415-5.
[15] D. Vizoso, et al., Size-dependent radiation damage mechanisms in nanowires and nanoporous structures, Acta Mater 215 (2021) 117018, doi:10.1016/j.actamat.2021.117018.
[16] J.D. Honeycutt, H.C. Anderson, Molecular dynamics study of melting and freezing of small Lennard-Jones clusters, J. Phys. Chem. 91 (1987) 4950–4963, doi:10.1021/j100303a014.
[17] A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool, Model. Simul. Mater. Eng. 18 (2010) 015012, doi:10.1088/0965-0393/18/1/015012.
[18] A. Stukowski, V.V. Bulatov, A. Arsenlis, Automated identification and indexing of dislocations in crystal interfaces, Model. Simul. Mater. Sci. Eng. 20 (2012) 085007, doi:10.1088/0965-0393/20/8/085007.