# scientific reports

OPEN

# Data-driven analysis of neutron diffraction line profiles: application to plastically deformed Ta

Aaron E. Tallman[1], Reeju Pokharel[1], Darshan Bamney[2], Douglas E. Spearot[2,3], Bjorn Clausen[1], Ricardo A. Lebensohn[4], Donald Brown[1] & Laurent Capolungo[1]✉

Non-destructive evaluation of plastically deformed metals, particularly diffraction line profile analysis (DLPA), is valuable both to estimate dislocation densities and arrangements and to validate microstructure-aware constitutive models. To date, the interpretation of whole line diffraction profiles relies on the use of semi-analytical models such as the extended convolutional multiple whole profile (eCMWP) method. This study introduces and validates two data-driven DLPA models to extract dislocation densities from experimentally gathered whole line diffraction profiles. Using two distinct virtual diffraction models accounting for both strain and instrument induced broadening, a database of virtual diffraction whole line profiles of Ta single crystals is generated using discrete dislocation dynamics. The databases are mined to create Gaussian process regression-based surrogate models, allowing dislocation densities to be extracted from experimental profiles. The method is validated against 11 experimentally gathered whole line diffraction profiles from plastically deformed Ta polycrystals. The newly proposed model predicts dislocation densities consistent with estimates from eCMWP. Advantageously, this data driven LPA model can distinguish broadening originating from the instrument and from the dislocation content even at low dislocation densities. Finally, the data-driven model is used to explore the effect of heterogeneous dislocation densities in microstructures containing grains, which may lead to more accurate data-driven predictions of dislocation density in plastically deformed polycrystals.

The quantification of dislocation content, and of its evolution, during plastic deformation of metals is critical to furthering our understanding of plasticity[1–4] as well as to the development and validation of microstructure-aware constitutive models[5–7]. Suitably, non-destructive evaluation (NDE) can yield estimates of dislocation densities in bulk[8]. As opposed to microscopy-based methods (e.g., transmission electron microscopy), NDE does not provide a direct observation of defects. Instead NDE measurements are interpreted to identify defect signatures and estimate defect content.

NDE can be performed by either ultrasound spectroscopy[9,10] or diffraction-based methods[11–13]. With the latter, one utilizes diffraction line profiles or alternatively spatially-resolved lattice strains to estimate disloca-tion content. The most commonly used method, diffraction line profile analysis (DLPA) has been used to esti-mate dislocation density in a great number of studies[14–24]. Additionally, much recent work has been done using spatially-resolved lattice strains to characterize microstructure[25–28]. Similarly to the Electron Back Scattering Diffraction (EBSD) method, the crystal orientation and elastic strain fields obtained via High-Energy Diffrac-tion Microscopy (HEDM) (e.g.[25,26]) can be used to estimate the density of geometrically-necessary dislocations (GNDs) density in the medium. Recent techniques use high-resolution EBSD with DLPA models to estimate total dislocation content within a locally resolved volume[27,28]. In all cases, DLPA models analytically relate the shape and breadth of multiple or individual diffraction peaks to the total dislocation density in the bulk[29–34].

Common to all DLPA methods are several assumptions regarding the geometry of dislocation configurations. Straight, parallel dislocations are assumed by a number of DLPA models[29,31,35–38]. These idealized networks can differ vastly from complex dislocation structures, as seen in 2-D with transmission electron microscopy[39–41] or in 3-D with electron tomography[42–44]. Over the past few decades, models accounting for the presence of GNDs and

[1]Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA. [2]Department of Materials Science and Engineering, University of Florida, Gainesville, FL, USA. [3]Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, USA. [4]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA. ✉email: laurent@lanl.gov

for correlations in the positions of dislocations have been proposed[27,28,35,45,46]. These methods use high-resolution diffraction profiles, and relate the breadth of the peak, asymmetry in the broadening, and the asymptotic description of the tail shape to evaluate dislocation density and many other dislocation arrangement-specific measurements. Modern DLPA methods can further estimate the dislocation content on each slip system. This relies on the use of average contrast factors, which describe the effects of the relative orientation of the dislocation line and of its Burgers vector with respect to the diffraction vector on the magnitude of broadening[47].

In practice, broadening is also induced by the instrument itself. Thus, using an instrument with high resolution, removing background scattering, and minimizing experimental noise are all critical to the estimation of the defect content[46]. For example, the extended convolutional multiple whole profile (eCMWP) method[11,21] accommodates instrumental broadening alongside theoretical broadening from defects, to allow dislocation density estimation in the presence of significant instrumental broadening. In the eCMWP approach, an optimization is needed to curve-fit a theoretical profile to the observed profile. This optimization has a sensitivity floor: if strain broadening is small relative to instrument broadening, the optimization cannot be performed. The current work proposes an alternative method to estimate dislocation density from whole diffraction line profiles in the presence of instrumental broadening.

As stated, DLPA provides dislocation content estimates, however quantifying the accuracy of these estimates is non-trivial. Simulations have been proposed as an alternative means of validation for DLPA[28,48]. For example, Balogh et al. simulated single crystal microstructures using discrete dislocation dynamics (DDD)[48]. They employed a simple virtual diffraction approach based on the Stokes-Wilson approximation, calculating broadened peaks directly from strain fields obtained from DDD. They then estimated dislocation density from virtual diffraction profiles with eCMWP to compare with the simulated microstructures. Although this study showed good agreement generally between the dislocation densities extracted from DLPA and generated from dislocation dynamics, when dislocation density varied spatially within a microstructure, significant discrepancies were noted. Nevertheless, both the modest number of cases considered (5) and the approximations made in generating the virtual profiles limited the general applicability of their approach. Since then, several continuum-based methods have been proposed to more accurately generate virtual profiles[49–52].

Recently, improvements to the speed and accuracy of virtual diffraction algorithms[6,48–53] and to the efficiency of mesoscale simulations[54] have enabled the rapid generation of hundreds of diffraction profile-microstructure pairs[50]. For example, Bamney et al.[50] developed two strain-based virtual diffraction algorithms, one using the Stokes-Wilson approximation[34], the other using differential strains to include the effects of spatially correlated strains on broadening. In the same work, a data-driven DLPA model, on which the present work is based, was proposed[50]. This data-driven model uses statistical inference to link virtual diffraction profiles to dislocation density, allowing the core DLPA assumptions of idealized microstructures to be relaxed.

The present study builds upon these developments to derive an extended data-driven DLPA model which can be applied to experimentally gathered peaks with an extended sensitivity floor. The proposed approach further provides a quantification of the uncertainty associated with dislocation density estimates. First, synthetic diffraction profile data are generated from simulations of dislocation networks in pure Ta single crystals with DDD by leveraging the two virtual diffraction algorithms proposed in Bamney et al.[50]. Using gaussian process regression (GPR) in an ensemble method, a data-driven DLPA model is built for each diffraction algorithm. The newly proposed methods are employed to estimate dislocation content from experimentally observed neutron time-of-flight (TOF) profiles of Ta. A baseline of predictions is established for the experimental profiles using the eCMWP software[11]. Both of the ensembles provide plausible predictions of dislocation density as a function of applied strain in the presence of experimental noise and instrumental broadening. The method provides uncertainty intervals alongside the estimates. Data-driven estimates are given for experimental profiles with minimal strain broadening for which eCMWP could not give estimates. The extension of the method to include effects of intergranular heterogeneity in plastically deformed polycrystalline Ta is discussed.

## Results

As established in the methodology section, the data-driven estimates of dislocation density originate from a GPR model-ensemble ("Data-Driven Model Form" section), fit to a database of diffraction profiles (Sect. Diffraction profile preprocessing), made by combining synthetically generated strain broadened profiles with experimentally gathered instrumental broadening. The method relies on two main steps: determining a compact description of profile shape, then relating that description to dislocation density. Notably, machine learning techniques (e.g., GPR) are used to produce dislocation density estimates with a data-driven model ("Data-Driven Methods" section), as opposed to an analytical model. The data-driven model is trained on 177 synthetic profiles, incrementally validated against 16 synthetic profiles, and fully validated against 11 experimentally gathered diffraction profiles from plastically deformed Ta samples (gathered as described in "Experiments" section). Two sets of results are obtained, respective to the two algorithms that are used to calculate the strain broadening via virtual diffraction ("Virtual diffraction" section), one without strain correlations (referred to as the Stokes-Wilson-based model) and one with strain correlations (called the $\epsilon^L$-based model). The entire computational workflow is presented in Fig. 1. Accordingly, the key outcomes of this work are presented in four sections. The synthetic database of microstructures and associated virtual profiles is described in "Database of synthetic profiles" section. The data-driven approximations of the experimental profiles are described statistically in "Approximation of experimental profiles". The dislocation density predictions of each of the two virtual diffraction algorithms are compared in "Application of the two virtual diffraction algorithms against experimental data" section. The data-driven predictions are compared with the predictions of CMWP in Sect. Application of the two virtual diffraction algorithms against experimental data
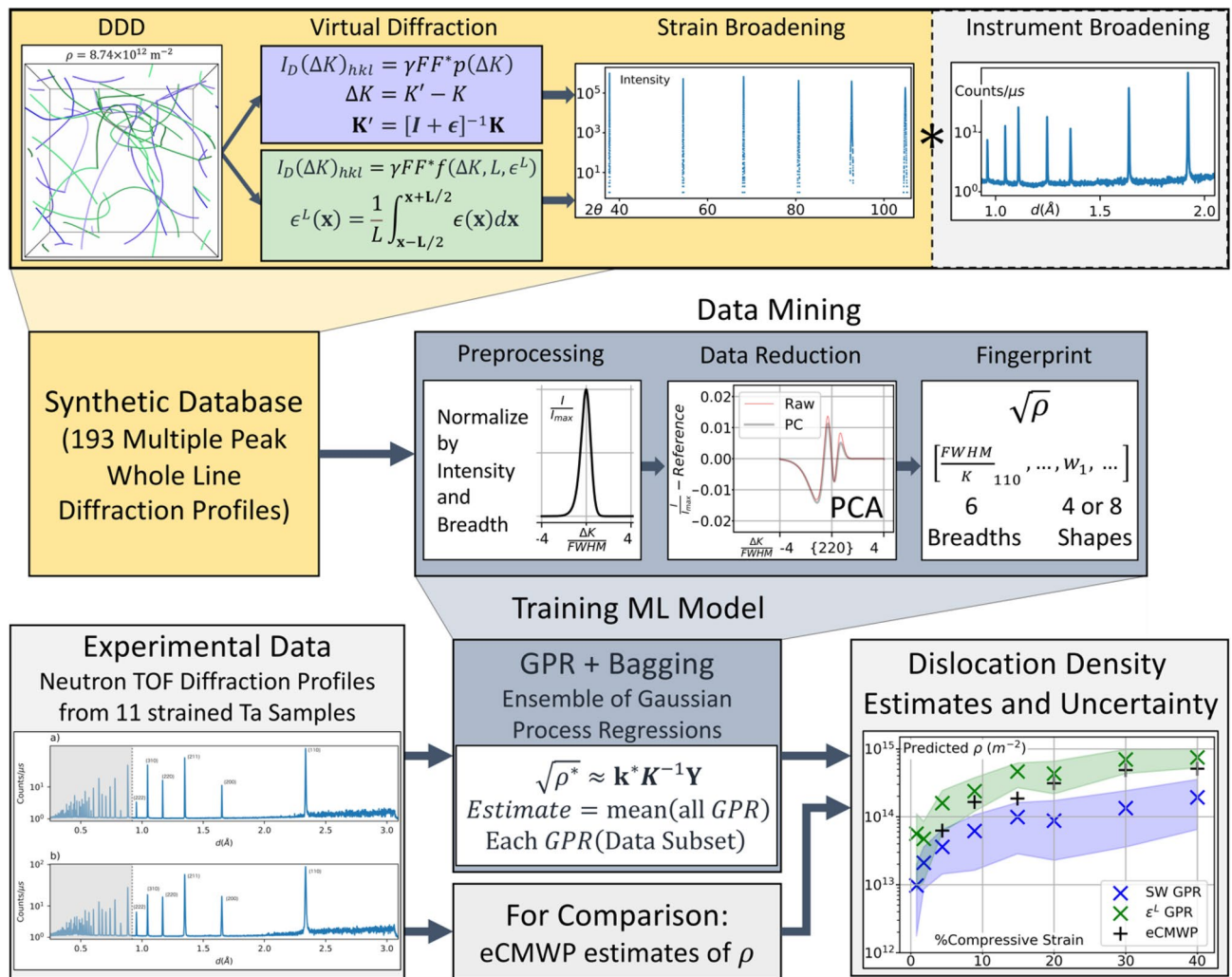
**Figure 1.** The general outline for the approach, beginning with the synthetic database generation, with discrete dislocation dynamics and virtual diffraction algorithms to calculate strain broadening, combined with instrument broadening obtained from calibration experiments. The data mining procedure is shown, including preprocessing steps, the data reduction using principal component analysis (PCA) and resulting in a fingerprint to be used in a machine learning (ML) model. Lastly, the use of the ML model is shown as a comparison with an existing method (eCMWP) in the prediction of dislocation density from experimental neutron diffraction profiles of deformed Ta samples.

**Database of synthetic profiles.** A set of 193 synthetic relaxed microstructures with dislocation densities ranging from $7.23 \times 10^{12}$ to $1.19 \times 10^{15}$ m$^{-2}$ was generated (see "Data-Driven Methods" section for details of the approach used to generate these profiles). A histogram of the dislocation densities in the dataset is shown in Fig. 2. Two microstructures taken from the database are shown in Fig. 3. The dislocation density ($8.74 \times 10^{12}$ and $1.10 \times 10^{15}$ m$^{-2}$) and dimension length ($5000 \cdot a_0$ and $1500 \cdot a_0$) are provided. Notably, both relaxed dislocation networks shown in Fig. 3 contain dislocations on multiple slip systems, junctions, and curved dislocation lines. The simulated microstructures were not observed to contain dislocation cell structures. A direct comparison between the microstructure generated from DDD and those observed by means of TEM remains challenging. This is due in part to the relaxation of the dislocation structures induced by the sample preparation. In recent work[8], such dislocation relaxation process was simulated by means of DDD. The study showed that the dislocation content can reduce significantly during preparation of a thin film, and more importantly–that the dislocation lines are very likely to reorient to minimize their image forces. Nonetheless, the dislocation configurations generated via DDD can be compared qualitatively to those documented experimentally–notably to those presented in seminal studies by Spitzig and Mitchell[55,56]. Clearly the synthetic microstructures generated show, as per Fig. 3, that the configurations are characterized by dislocation tangles (see Fig. 3a) of which density increases with overall dislocation content. However, given that cross glide (i.e. cross slip and climb) processes are not activated, the microstructure simulated precludes one from generating cell-like structures with minimum energy configurations[57]. In comparing the dislocation configurations to those reported, in the case of single crystal deformed in the quasi-static regime, it appears that the microstructures generated synthetically mimic
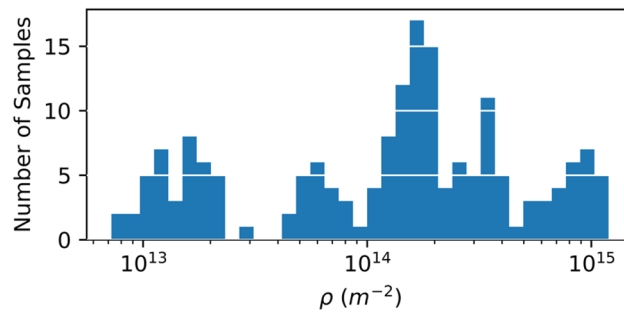
**Figure 2.** Histogram of dislocation densities in the database of Ta single crystal microstructures generated using DDD.
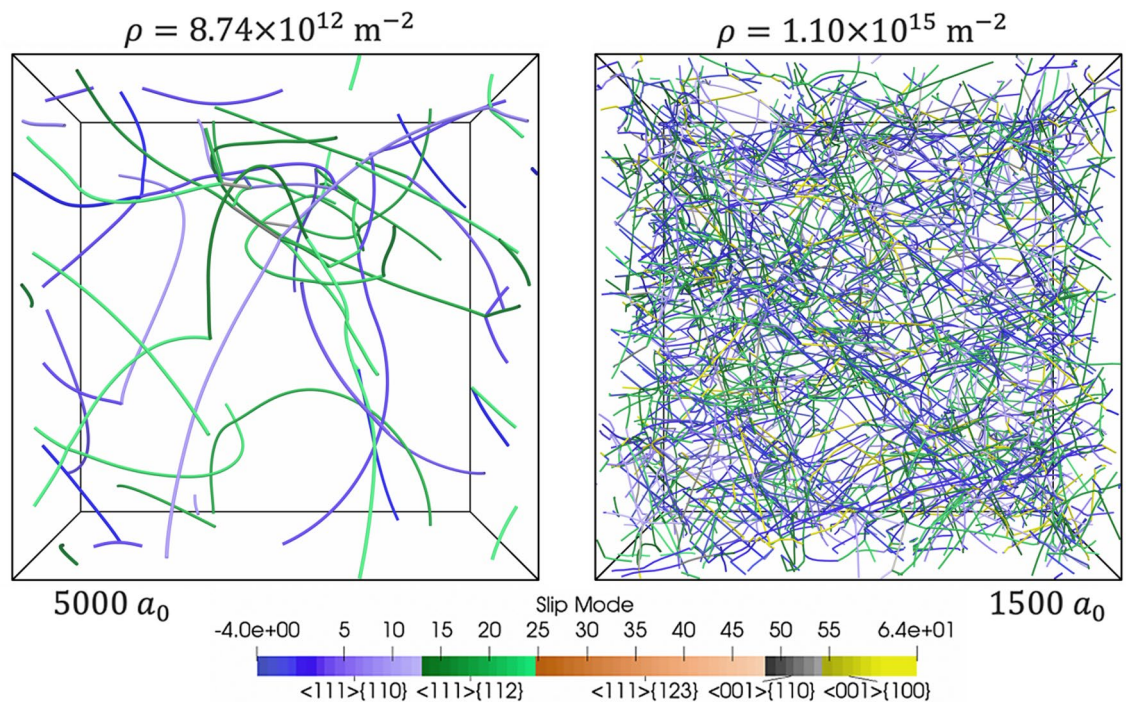


**Figure 3.** Two simulated microstructures of different dislocation densities, with slip systems labelled by slip mode.

those corresponding to stage I and stage II hardening but not of stage III. Thus, the synthetic database is likely to be more representative of microstructures associated with samples deformed up to 20 to 30 percent strain.

Diffraction profiles were calculated from all synthetic microstructures obtained with DDD-FFT, with the intention of capturing experimental diffraction conditions for different configurations. For each dislocation microstructure, the {110}, {200}. {211}, {220} {310}, and {222} peaks were generated numerically using two distinct virtual diffraction algorithms, which include the effect of instrumental broadening. The two algorithms are detailed in "Virtual diffraction" section and are referred to as the S-W-based and $\epsilon^L$-based methods. As the first algorithm is computationally much faster than the more accurate second algorithm, an evaluation of predictions using results of both methods are of interest.

The shape variation of each peak is assessed by normalizing the peak intensity $I$ by the maximum intensity of the peak $I_{max}$ and plotting these as a function of $\frac{\Delta K}{FWHM}$. Here $\Delta K$ denotes the magnitude of the difference between the diffraction vector and the ideal diffraction vector as given by Bragg's condition and FWHM denotes the full width at half maximum. In other words, peaks from the profile were normalized in intensity and peak breadth. Also, the maximum intensity of each peak was shifted to align with $\Delta K = 0$. Then, the peaks that were normalized separately were joined into a single processed profile, an example of which is shown in Fig. 4a. The profiles are also shown in Fig. 4b with a log scale to highlight details in shape of the tails of the peaks. A cutoff of four times the *FWHM* was used to restrict each processed profile to a consistent number of intensity values. Additional processing steps used to remove noise and background from the experimental profiles are detailed in "Diffraction profile preprocessing" section.
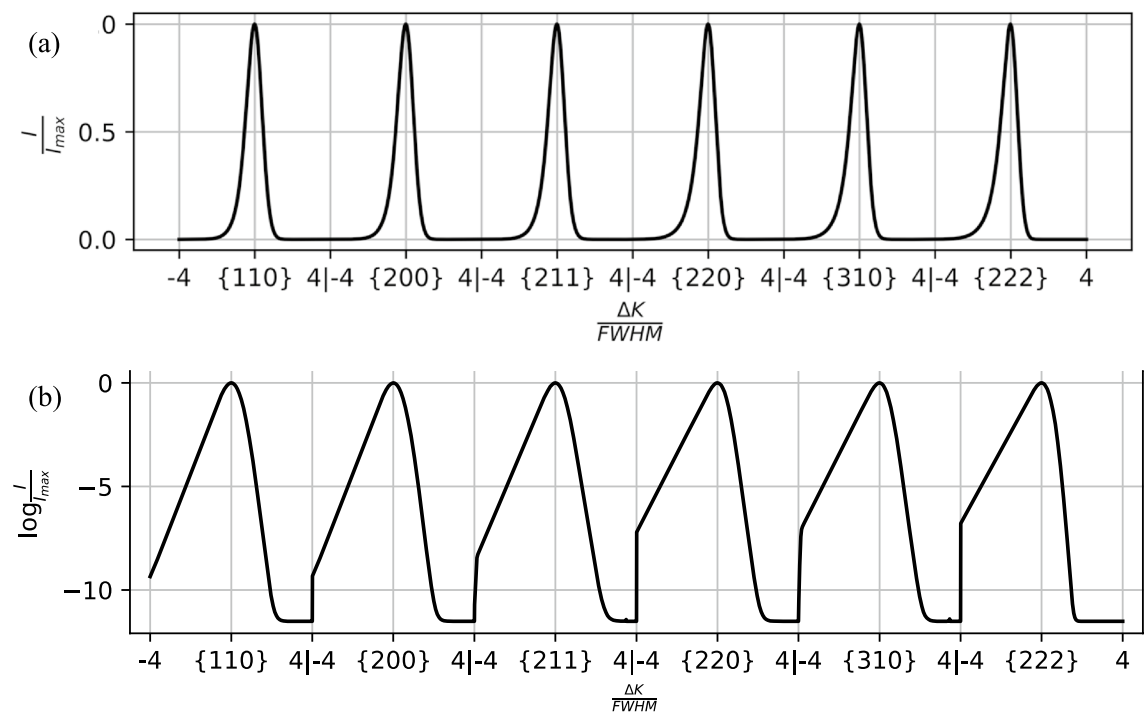
**Figure 4.** (**a**) An example profile (dislocation density of $8.47 \times 10^{12}$ m$^{-2}$) from the $\epsilon^L$-based synthetic database that has been processed to isolate variations in peak shape. (**b**) The profiles shown on a log scale highlight the shape of the tails of the peaks.

| Algorithm | Training set of 177 MSE ($\times 10^{-5}$) | | Synthetic testing set of 16 MSE ($\times 10^{-5}$) | | Experiment set of 11 MSE ($\times 10^{-5}$) | |
|---|---|---|---|---|---|---|
| | Mean | Max | Mean | Max | Mean | Max |
| S-W | 0.0525 | 0.199 | 0.0580 | 0.117 | 21.8 | 31.9 |
| $\epsilon^L$ | 0.0617 | 0.278 | 0.0683 | 0.223 | 17.7 | 25.6 |

**Table 1.** Statistics of the goodness of fit between the PCA representation and the variation of the processed profiles, shown for the two databases. The training set corresponds to the synthetic profiles used to produce the principal components, the synthetic testing set corresponds to synthetic profiles withheld from the PCA, and the experiment set corresponds to the profiles taken from experiments.

**Approximation of experimental profiles.** The data-driven compact representation of diffraction profile shape is evaluated in terms of the accuracy of the approximation, which is important to the subsequent predictions. The compact description of the diffraction profile shape (calculated using principal component analysis, PCA) is used to approximate the experimental profiles. The PCA (described in Sect. Diffraction profile preprocessing) is determined using the synthetic profile database (described in Sect. Database of synthetic profiles), and thus may not be as accurate when used to approximate the experiments. The quality of the GPR ensemble predictions ("Data-Driven Model Form" section ) necessarily depend on the extent to which the synthetic profiles resemble the experimental profiles. To compare the datasets, the mean squared error (MSE) was measured for the PCA of profile shape variations across every profile. The statistics describing each dataset are shown in Table 1. The average value and minimum value of MSE across each set of profiles is shown. The training set is the 177 profiles used in the PCA, the synthetic testing set is the 16 synthetic profiles held in reserve, and the experiment set is the 11 profiles obtained from the Ta samples. The training set statistics show the performance of the PCA in an ideal case, establishing a baseline. The synthetic testing set is included to demonstrate the expected increase in error when statistically equivalent profiles outside the training set are considered. The increase in error for the experimental profiles is substantially larger than that for the synthetic testing set, which points to difference between that data and the synthetic data. The difference that these MSE values indicate is quite pronounced and can be addressed further.

It is not certain whether differences in the experimental and synthetic profiles have bearing on predictions of dislocation density. Differences between the synthetic profiles and the experimental profiles may arise from many sources, such as noise, unremoved background counts, and/or inaccuracy in the single-grain approximation of a polycrystal diffraction profile. These sources may be distinguished by investigating the profiles individually. For example, PCA representations of the experimental profile from one sample are shown in Fig. 5 (sample IP9.3)
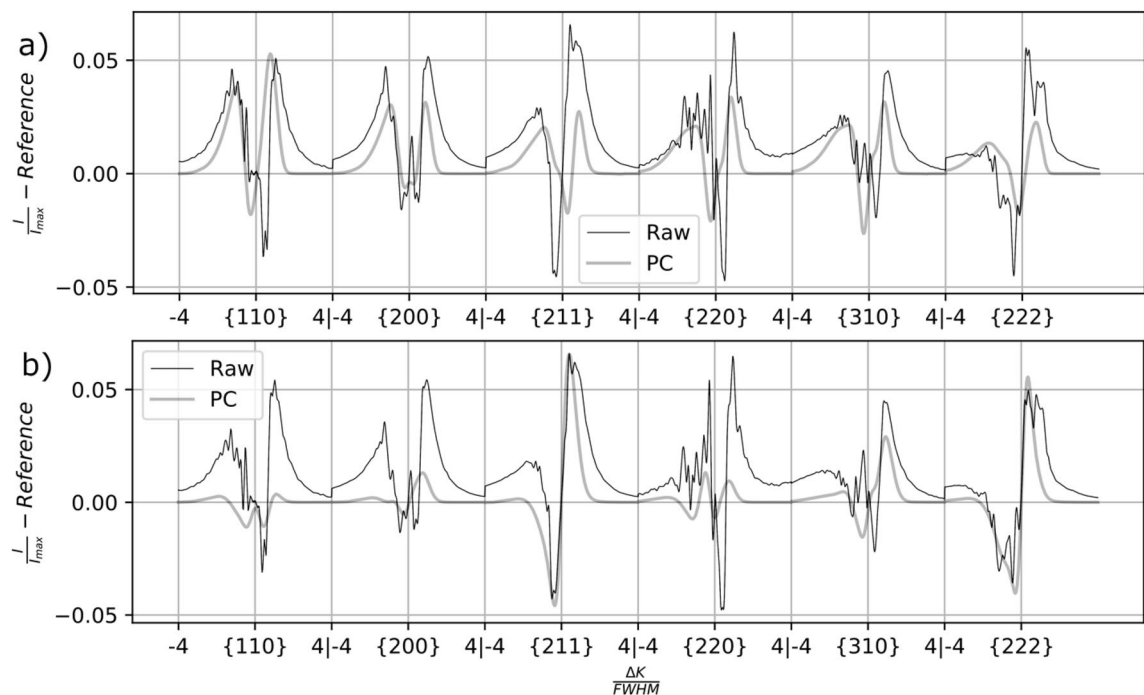
**Figure 5.** Actual (black) and PC (grey) representation of profile shape variation in the profile from sample IP9.3, based on (**a**) the S-W virtual diffraction algorithm and (**b**) the $\epsilon^L$-based virtual diffraction algorithm.

| Ta sample | IP0.8 | IP1.8 | IP4.5 | IP9.3 | IP15 | IP20 | IP30 | IP40 | TT20 | TT30 | TT40 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Final % compressive strain | 0.8 | 1.8 | 4.5 | 9.3 | 15 | 20 | 30 | 40 | 20 | 30 | 40 |
| $FWHM_{110}(d)$ $\text{Å} \times 10^{-3}$ | 4.32 | 4.13 | 4.40 | 5.41 | 5.32 | 5.25 | 5.78 | 6.64 | 5.44 | 5.88 | 6.20 |

**Table 2.** Names used to refer to the 11 Ta samples are shown alongside the compressive strain at which the respective diffraction profiles were taken. Full width at half maximum measurements are included for the {110} peaks.

for both (a) the S-W based dataset and (b) the $\epsilon^L$-based dataset. The actual profile shape variation is shown in black, and the PCA representation is shown in grey. The mismatch observed in this case can be summarized as the contributions of noise and of tails which are much broader than those captured by the PCA. While noise is likely unimportant, the discrepancy in the principal component (PC) representation of the tails of the peaks warrants further attention.

### Application of the two virtual diffraction algorithms against experimental data.

The data-driven DLPA model was trained with each of the synthetic datasets (S-W and $\epsilon^L$-based) and used to make predictions of dislocation density from the processed Ta neutron diffraction profiles. Eleven Ta specimens were compressed at 0.001/s strain rate to maximum true strain levels of 0.8, 1.8, 4.5, 9.3, 15, 20, 30 or 40%. Three specimens were loaded with through thickness (TT) and eight with in plane (IP) orientations ("Experiments" section details the experimental method used). The sample information is shown in Table 2 along with the full width at half maximum (*FWHM*) of the {110} peak.

Both data-driven DLPA ensemble methods were applied to these 11 experimental profiles. Predictions of the dislocation densities using these two methods along with the baseline made using eCMWP are shown in Fig. 6a, b. The data-driven estimates are accompanied by uncertainty intervals, unlike the estimates from eCMWP (Sect. 3.3). In Fig. 6, the S-W data-driven model predictions are shown in blue, and the $\epsilon^L$-based model predictions are shown in green. The ensemble approach consists of 80 GPR predictions for each data-driven model ("Data-driven model form" section). In Fig. 6, the ensembles of predictions are simplified to a mean value and the 5th and 95th percentiles of the ensemble predictions. The percentiles are shown as a shaded region around the plotted mean values for each sample. The ensemble variability was very high for each model: $\left(\frac{max}{min} > 10\right)$ for the S-W model for every sample and $\left(\frac{max}{min} \sim 5\right)$ for the $\epsilon^L$-based model. This suggests the bagging-based estimates (ensemble mean) are more accurate than what could be obtained using all the data to train a single surrogate model, in both cases. Figure 6 shows the dislocation density versus applied strain trends captured by the methods in this study. The trends of dislocation density with strain suggest the new method provides credible estimates of dislocation content.
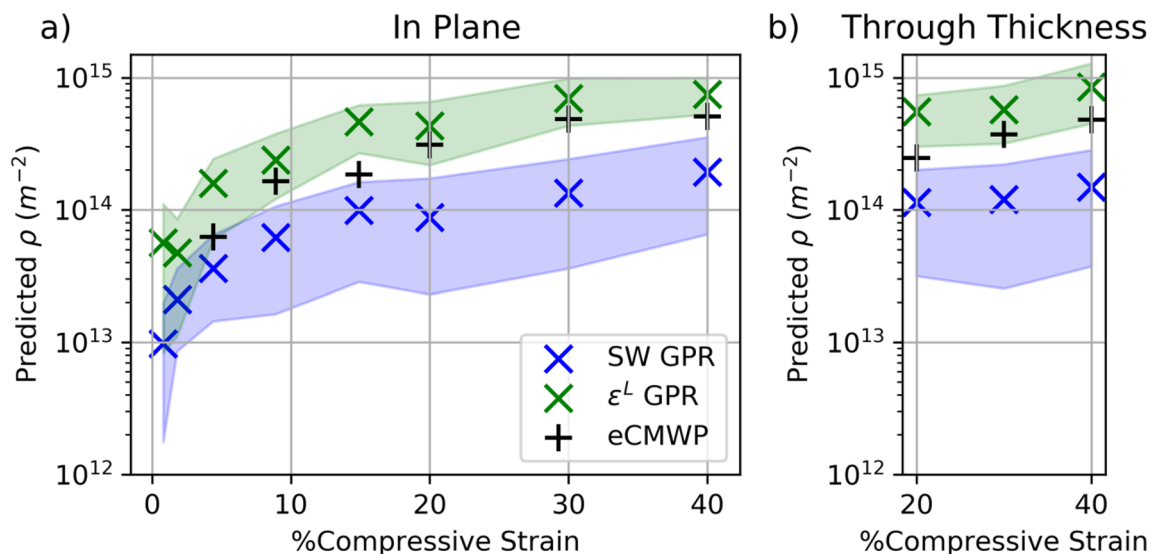
**Figure 6.** Predictions of dislocation density in the Ta samples from each method, as a function of applied true compressive strain. The shaded regions represent the middle 90% of individual GPR predictions in the ensemble. The eCMWP estimates are shown as a baseline for comparison, where the method was able to produce an estimate. (**a**) Shows the predictions for samples taken in the in-plane direction, and (**b**) shows the samples taken in the through-thickness direction.

The baseline prediction of dislocation density made using eCMWP (Sect. 3.3) is also shown in Fig. 6, although for the two samples with the least applied strain, no baseline was available due to the measurement sensitivity limit. In contrast, the data-driven DLPA methods can be used to study dislocation content at the onset of the elasto-plastic transition. The dislocation densities estimated by the S-W surrogate ensemble were lower than the $\epsilon^L$-based estimates (with differences of a factor of roughly 2 to 6). This trend is to be expected for a data-driven model trained using data generated using the S-W approximation, which tends to overestimate broadening. Generally, the $\epsilon^L$-based ensemble predictions were higher in dislocation density and lower in ensemble variability than the S-W-based ensemble predictions. Notably, in the two cases with least applied strain, the $\epsilon^L$-based ensemble variability was substantially higher. This variability indicates a greater degree of uncertainty associated with the estimate, which can explain why the $\epsilon^L$-based estimate of specimen IP0.8 (strained to 0.8 percent) could be higher than that of IP1.8 (strained to 1.8 percent).

## Discussion

The dislocation densities extracted in "Application of the two virtual diffraction algorithms against experimental data" section can be used to parameterize constitutive models. In its simplest form, the flow stress is related to the dislocation density as per Taylor's law:

$$\sigma_y = \alpha M \mu b \sqrt{\rho} \tag{1}$$

where $\sigma_y$ denotes the flow stress, $\rho$ refers to the dislocation content while $\mu$ and $b$ refer to the shear modulus and to the magnitude of the Burgers vector. $\alpha$ is typically a fitting constant while $M$ is the Taylor factor. The former is representative of the strength of interactions between dislocations while the latter is attributed to texture. Figure 7 shows the measured mechanical response of the system loading along the IP and TT directions as well the relationship between the square root of the dislocation density and the experimentally measured flow stress.

As shown in Fig. 7b, c, irrespectively of the method used to estimate the dislocation density (i.e. SW, or eCMWP), one recovers a linear scaling between the square root of the dislocation content and the normalized flows stress. One further notes that the Taylor expression is expected to break down at large dislocation content, as per studies of Queyreau et al.[58,59], which suggest that a logarithmic correction should be appended to Taylor's law at large densities. In the present case though it is found that up to 40% strain leading to estimated dislocation densities greater than $5 \times 10^{14}$ m$^{-2}$ such correction is not necessary. The slopes obtained from a linear fit are for the in-plane loading conditions: GPR-SW: 0.95, GPR-$\epsilon_L$: 0.43, eCMWP: 0.3826. In the case of through thickness loading one obtains: GPR-SW: 1.71, GPR-$\epsilon_L$: 0.47, eCMWP: 0.77. It is thus clear the slopes obtained by the GPR-SW algorithm are far higher than those obtained with GPR-$\epsilon_L$ and eCMWP. Interestingly, it is found that the slopes (i.e. $\alpha M$ in Eq. 1) are found to largely depend on loading direction when using eCMWP. This is not the case when GPR-$\epsilon_L$ is used to extract the dislocation content. Due to the crystal symmetry and materials texture, one does not expect a directional dependence.

From the standpoint of the representativeness of the microstructures used in this study, here single crystal simulations were used to interpret data from polycrystalline samples ("DDD microstructure synthesis" section). In practice though, polycrystalline samples may present variation not captured with the single crystal data. For instance, the single crystal virtual diffraction data could be interpreted as polycrystals of random texture and
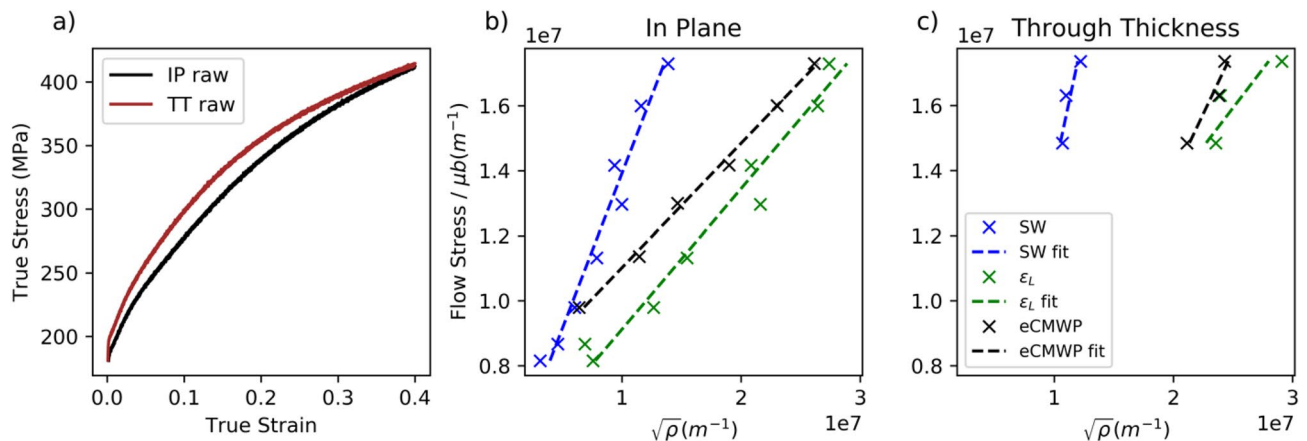
**Figure 7.** (**a**) Measured mechanical response of Ta during in plane and through thickness loading. Normalized Flow stress vs square root of the dislocation densities for samples loaded in plane (**b**) and through thickness (**c**). The dislocation densities extract by GPR-SW, GPR-$\epsilon_L$ and eCMWP are shown in blue, green and black respectively. A linear fit is used to estimate the produce $\alpha M$ in Eq. (1). These fits are shown in dashed lines.

homogeneous dislocation density between grains, in which grains do not interact. Given that the Ta samples in this study were subjected to plastic deformation and that Ta single crystals are plastically anisotropic, heterogeneity in dislocation density is to be expected between the grains in the Ta samples. In the future, polycrystal DDD simulations may be performed to account for intergranular interactions, and the present analysis may be repeated with the new data.

The current model can estimate error due to the presence of heterogeneity in dislocation density between grains. To demonstrate this, a modification of the synthetic data was performed to develop a new dataset against which the data-driven model could be tested. Two profiles were taken from the synthetic database and combined by adding the intensity as a function of $K$, the diffraction vector length, to each profile. The new profile was used to represent a polycrystal with a bimodal distribution of dislocation densities in different grains. Twenty of these profiles were generated using different pairs of original profiles for each. Grain interactions were still neglected in these new "bimodal" profiles. A parameter $H$ was defined to track the heterogeneity in the new profiles as,

$$H = \frac{\rho_{max} - \rho_{min}}{\rho_{mean}} \tag{2}$$

where the values of $\rho_{max}$, $\rho_{min}$, and $\rho_{mean}$ were measured from the two values belonging to the single crystal profiles included in the new profile. The value of $H$ is 0 when the two profiles have the same dislocation density and is 2 when only one profile contains dislocations. A value of $H = 1.64$ reflects an order of magnitude difference between $\rho_{max}$ and $\rho_{min}$. The $\epsilon^L$ GPR ensemble was used to predict dislocation density of the "bimodal" profiles, the results of which are shown in Fig. 8. The predicted values divided by true values of dislocation density are shown as a function of $H$. At high levels of heterogeneity, the surrogate model underpredicts dislocation density.

To investigate this further, the PC representation of a bimodal profile is shown in Fig. 8. The bimodal profile has more emphasis in the tails of the peaks, especially for {310}. This can be seen in the third "hump" that is present for all raw peaks, but is not present or is very small for the PC peaks, except for {222}, where it is with opposite sign. The MSE of the PC representation was $2.72 \times 10^{-5}$, which falls between the values of the validation data and the experiments. A comparison of the profile shape in Fig. 9 with the profile in Fig. 5 shows that heterogeneity may explain some but not all of the discrepancy between the synthetic database and the experiments.

It is expected that the accuracy of the data-driven model will improve as improvements are made to the database. The number of profiles gathered, the volume of the simulated microstructures, and the types of parameters used in the fingerprint in this work should be improved upon. A study of database size effects on predictions could help to establish database size requirement guidelines. Additional parameters could be used to capture the peak shifts omitted in the current model. Critically, any improvements to the database provide an avenue for improvements to the predictions of the data-driven model. In future work, the method could be extended to predict dislocation content in polycrystals with high contrast in dislocation content between grains.

## Methods

**Experiments.** Multiple Ta specimens of 4.2 mm diameter and 8.4 mm height with their axes parallel to the through thickness (TT) and in-plane (IP) directions were electro-discharge machined from a wrought plate. The wrought plate material had an average grain diameter of 30 μm, with nearly equiaxed grains[60]. Eleven Ta specimens (three with TT and eight with IP orientations) were compressed at 0.001/s strain rate to maximum true strain levels of 0.8, 1.8, 4.5, 9.3, 15, 20, 30 or 40%. The diffraction line profile data were recorded on a high-resolution ($FWHM \sim 0.1\%$) backscattering (153 deg) detector bank on the spectrometer for material research at temperature and stress (SMARTS)[61]. The beam width was 10 mm and the beam height was 5 mm, illuminating nearly all of the sample volume, which was about 100 mm³. The diffracted signal was counted for six hours, to
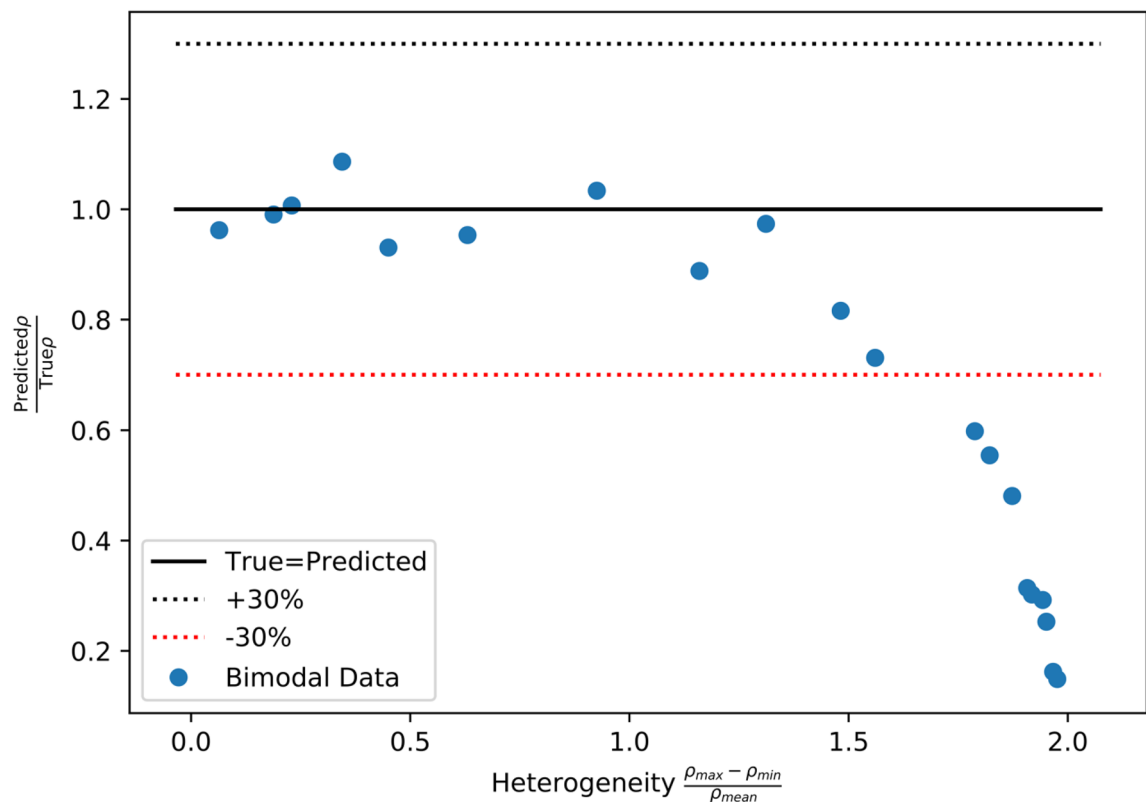
**Figure 8.** Predictions of the $\epsilon^L$-based surrogate model relative to the true values of dislocation density for 20 "bimodal" profiles, as a function of the heterogeneity of the dislocation density between grains.
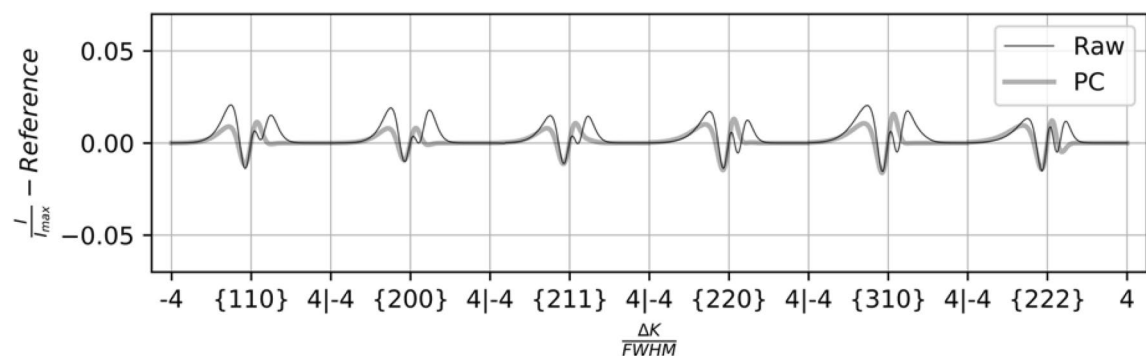


**Figure 9.** Actual and PC representation of the peak shape variations from a reference peak shape of a bimodal profile with a heterogeneity value of 1.976.

ensure high signal to noise ratio. Line profile data were collected for all deformed samples compressed to various levels of final strains. The six peaks with largest interplanar spacing were selected for analysis: {110}, {200}. {211}, {220} {310}, and {222}. Two profiles are shown for reference in Fig. 10. In Fig. 10, the relative intensity of the peaks to the background and noise is shown to be high, and the breadth of peaks was variable between samples. The increase in peak breadth was apparent with increasing macroscopic strain. The peak broadening was asymmetrical for many of the peaks analyzed.

**Data-driven methods.** *DDD microstructure synthesis.* Discrete dislocation dynamics (DDD) simulations represent a dislocated lattice as discrete dislocation lines resolved within a continuous elastic medium. The costliest computations of DDD are the solution of dislocation segment velocities and the solution of the lattice stress and strain within the simulated volume. The second part can be accelerated by using a fast Fourier transform (FFT) and periodic boundary conditions (instead of a finite element solver) to solve for the stress field[54,62]. These algorithmic advances increase the effective time-scale and length-scale of DDD simulations. Here, DDD simulations in the FFT-based framework (DDD-FFT)[54,62] were used to generate microstructures and elastic strain fields associated with relaxed dislocation networks in Ta single crystals. Elastic properties of Ta were taken from experiments reported in[63]. Dislocation motion was assumed to be isotropic and planar. Similar to BCC Fe[64],
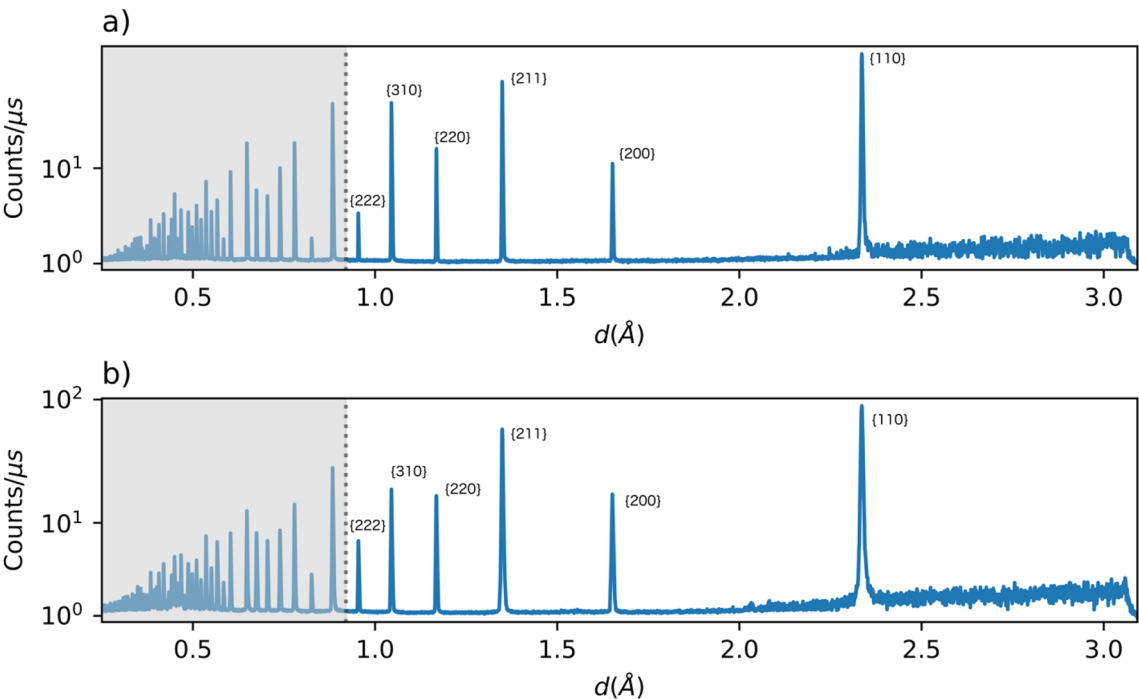
9

**Figure 10.** Neutron diffraction profiles measured for IP1.8 (**a**) and IP30 (**b**) are shown. The first six peaks are labelled with the hkl of the diffraction plane. The portion of the profile not used in the analysis is greyed out.

| Slip mode | Peierls Stress (MPa) | Drag ($10^{-6}$ Pa · s) | Elasticity coefficients | |
|---|---|---|---|---|
| <111>{110} | 10 | 80 | $C_{11}$(GPa) | 260.91 |
| <111>{112} | 10 | 80 | $C_{12}$ | 157.43 |
| <111>{123} | 10 | 80 | $C_{44}$ | 81.82 |
| <001>{110} | 300 | 800 | $a_0$(Å) | 3.37 |
| <001>{100} | 300 | 9000 | | |

**Table 3.** The material parameters used in the DDD simulations of Ta.

slip modes {110}, {112}, and {123} were considered. Further junctions on {100} and {110} were also considered. The parameters used are shown in Table 3. Note that the choice of drag coefficients in Table 3 has little to no influence on the low energy configuration resulting from relaxation. Further, choice was made to set a similar friction stress for all slip mode. This assumption is consistent with recent crystal plasticity based constitutive modeling assumptions in[65]. Note that the values of friction stress utilized in the present study are in the same order of magnitude as those used in[66] where the authors used values ranging from 27 to 37 MPa. The values of the friction stress for junctions was arbitrarily chosen. This choice is known to impact the length of junctions. The impact of this choice on the overall diffraction signature is not expected to be significant but has not been quantified rigorously.

A $64 \times 64 \times 64$ grid of regularly-spaced points where the stress field is computed by means of the FFT-based algorithm was used for all simulations. The size of the simulated volume was adjusted from $1500a_0$ to $5000a_0$ per side to gather a wide range of dislocation densities while providing at least a minimum number of segments. A minimum of 4000 segments in each microstructure was ensured to reduce the effect of the small sample size. Every microstructure generated in this work contained junctions and dislocations on multiple slip systems.

Dislocations were initialized by distributing randomly positioned elliptical shear dislocation loops on each {110} type slip system. The number of dislocations per slip system on average varied per simulation between 0.8 and 4, with random draws being used to resolve fractions into whole number outcomes. For example, an average value of 1.2 was instantiated with a random selection for each slip system, with one dislocation on a slip system at an 80% chance and two dislocations for the remaining 20%. Dislocations were allowed to reach {112} and {123} type slip systems via cross-slip. A two-step process was used to generate equilibrated dislocation microstructures. First, a high stress was applied to induce loop growth for approximately $1 \cdot 10^{-10}$ s, with the following imposed stress tensor,

$$\boldsymbol{\sigma} = \begin{bmatrix} 2 & 1 & -1 \\ 1 & -3 & -2 \\ -1 & -2 & 4 \end{bmatrix} \text{GPa} \tag{3}$$

The shock-level loading led to junction formation as the loops grew and intersected. Second, the system was allowed to relax under internal stresses for $7 \cdot 10^{-10}$ s. This allowed the strain energy of the system to reduce and dislocations to adopt a stable configuration. Notably, the simulated loading has no similarity to the load applied to the experimental samples. The only importance of this stress is that it activates slip on all slip systems. Neither the loading nor the crystal orientation is randomized. Rather, the resulting relaxed dislocation networks have randomness from their initial dislocation loop locations and not from their loading. No attempt was made to generate networks associated with any particular loading history.

*Virtual diffraction.* In the generation of synthetic profiles, broadening contributions from the instrument and from dislocations were included. With this, the diffraction intensity $I(\Delta K)_{hkl}$ is related to $\Delta K$ as follows:

$$I(\Delta K)_{hkl} = I_{instr}(\Delta K)_{hkl} * I_D(\Delta K)_{hkl} \tag{4}$$

for each *hkl* in the six considered peaks (i.e., {110}, {200}. {211}, {220} {310}, and {222}). Size broadening was assumed to be negligible for the single crystal Ta samples in this work.

The virtual diffraction algorithms used here are taken from the work of Bamney et al.[50]. A detailed description of both algorithms is contained in the original work. Here, a brief summary is provided. Both algorithms determine diffracted intensity as a function of diffraction angle using the elastic strain fields calculated using DDD-FFT.

The first algorithm employs the Stokes-Wilson[34] approximation and determines the broadening of a diffraction peak using a probability distribution of apparent strain, i.e., the elastic strain tensor projected along the direction of the diffraction vector. For a hkl-specific diffraction vector **K**, the intensity of the diffracted radiation can be defined as

$$I_D(\Delta K)_{hkl} = \gamma FF^* p(\Delta K = K' - K), \tag{5}$$

where $I_D$ is the intensity of the diffracted beam, $\gamma$ is the Lorentz-polarization factor, $F$ is the structure factor of the unit cell with complex conjugate $F^*$, and $p$ is the probability that the strained lattice will produce a given diffraction vector $\mathbf{K}'$. The diffraction vector for the strained lattice is calculated as

$$\mathbf{K}' = [I + \epsilon]^{-1}\mathbf{K}, \tag{6}$$

with second rank identity tensor $\boldsymbol{I}$, elastic strain tensor $\epsilon$, and diffraction vector associated with the unstrained lattice, **K**. The values of $\gamma$ and $F$ are calculated from the structure of the Ta unit cell and the value of **K**. The algorithm approximates $p$ using the value of elastic strain at the $N_x^{quad} \times N_y^{quad} \times N_z^{quad}$ quadrature points contained in each DDD simulation. Calculations are performed for every plane or *hkl* selected for diffraction. In this work, all planes in a family are assumed to contribute in equal weight to the 1-D profile.

The second virtual diffraction algorithm includes correlations in strain. When lattice strains are correlated across different distances, i.e., at different values of correlation length $L$, the effect of those strains on diffraction peak broadening are diminished. In the previous algorithm these correlations are neglected, which can lead to overestimates of strain broadening. In this method, the diffracted intensity is calculated as,

$$I_D(\Delta K)_{hkl} = \gamma FF^* \sum_{L=-\infty}^{+\infty} \left[A^D(L,\epsilon^L) + iB^D(L,\epsilon^L)\right] \exp[2\pi iL\Delta K], \tag{7}$$

where $\epsilon^L$ is the differential elastic strain resolved along the diffraction vector, and the terms $A^D$ and $B^D$ are the symmetric and asymmetric components of strain broadening, and are defined as,

$$A^D(L,\epsilon^L) = \int_{-\infty}^{+\infty} p(\epsilon^L) \cos\left(2\pi LK\epsilon^L\right)d\epsilon^L \tag{8}$$

$$B^D(L,\epsilon^L) = \int_{-\infty}^{+\infty} p(\epsilon^L) \sin\left(2\pi LK\epsilon^L\right)d\epsilon^L. \tag{9}$$

The value of $p(\epsilon^L)$, the distribution of differential strain for a specific correlation length $L$, as projected onto **K**, was calculated for each correlation length $L$, which can be gathered from the simulated microstructure ($L$ is bounded by the separation between quadrature points and the distance between the periodic boundaries). The distribution $p(\epsilon^L)$ was determined using measurements of differential strain at every quadrature point, **x** as,

$$\epsilon^L(\mathbf{x}) = \frac{1}{L}\int_{\mathbf{x}-L/2}^{\mathbf{x}+L/2} \epsilon(\mathbf{x})d\mathbf{x}, \tag{10}$$

where the integration was performed using a trapezoid rule. Given that the line of **L** typically does not coincide with the array of quadrature points, a linear interpolation was used to determine the projected strain values along **L** used in the integration scheme. These calculations were carried out for each *hkl* selected for measurement.

The strain broadened synthetic profiles were convoluted with a theoretical instrumental broadening function, $H(\Delta t)$. The Cu foil profile was used to calibrate the first theoretical instrumental broadening function

| hkl | $d(\text{Å})$ | $\alpha$ | $\beta$ | $\sigma^2$ |
|---|---|---|---|---|
| 111 | 2.08708 | 0.20804 | 0.045725 | 328.93 |
| 200 | 1.80743 | 0.26294 | 0.050833 | 254.44 |
| 220 | 1.27831 | 0.17400 | 0.054888 | 113.39 |
| 311 | 1.09013 | 0.23225 | 0.056578 | 76.323 |
| 222 | 1.04373 | 0.27436 | 0.061723 | 72.076 |
| 400 | 0.904023 | 0.14311 | 0.051810 | 18.493 |
| 331 | 0.829475 | 0.32658 | 0.058277 | 36.100 |
| 420 | 0.808477 | 0.33464 | 0.059525 | 36.752 |
| 422 | 0.738017 | 0.43557 | 0.060113 | 33.542 |
| 511, 333 | 0.695886 | 0.28550 | 0.063623 | 18.795 |

**Table 4.** Instrument broadening parameters as evaluated with a Cu calibration profile.

from the crystallographic software package General Structure and Analysis System (GSAS) for neutron TOF measurements[67], i.e.,

$$H(\Delta t) = \int G(\Delta t - \tau)E(\tau)d\tau \tag{11}$$

$$E(\tau) = \begin{cases} 2Ne^{\alpha\tau}, \tau < 0 \\ 2Ne^{-\beta\tau}, \tau > 0 \end{cases} \tag{12}$$

$$G(\Delta t) = \frac{1}{\sqrt{2\pi\sigma^2} \exp\left[-\frac{(\Delta t - \tau)}{2\sigma^2}\right]} \tag{13}$$

where $N$ is a normalization constant. $N$ is defined,

$$N = \frac{\alpha\beta}{2(\alpha + \beta)} \tag{14}$$

and where $\alpha$, $\beta$, and $\sigma$ are functions of interplanar spacing, $d$. These functions are determined as interpolations from the values obtained from the calibration profile,

$$\alpha = f_\alpha(d), \ \beta = f_\beta(d), \ \sigma = f_\sigma(d) \tag{15}$$

and the observed values used to define these interpolations are shown in Table 4. The values of $\alpha$, $\beta$, and $\sigma$ and the corresponding values of $d$ were obtained using a numerical fitting to the calibration Cu profile.

To convert $H(\Delta t)$ from TOF microseconds to small difference vector $\Delta K$, the following relation was used,

$$T_{ph} = DIFCd + DIFAd^2 + ZERO, \tag{16}$$

where $DIFA$, $DIFC$ and $ZERO$ are known calibration quantities for the instrument. Using this, $H$ can be expressed in terms of constants and $\Delta K$,

$$I_{instr}(\Delta K)_{hkl} = H\left(\frac{DIFC}{K_{hkl} + \Delta K} + \frac{DIFA}{(K_{hkl} + \Delta K)^2} + ZERO - T_{ph}(d_{hkl})\right) \tag{17}$$

*Diffraction profile preprocessing.* PCA was used to obtain a quantitative description of peak shape variations with a minimal set of values. This procedure is given as a schematic in Fig. 11. The PCA was applied to the database of processed synthetic profiles. At each of the 12,000 values of $\Delta K/FWHM$, the $\frac{I}{I_{max}}$ values of each of the set of profiles is expressed as a vector,

$$x_i = \left[\frac{I}{I_{max}}\left(\frac{\Delta K}{FWHM_i}\right) \text{for all profiles in set}\right], \quad \text{for } i \in [1, \ 12,000] \tag{18}$$

Eigenvectors are obtained for the covariance matrix of $x$,

$$cov[x_i, x_j] = E\left[(x_i - E[x_i])(x_j - E[x_j])\right], \tag{19}$$

where $E[x_i]$ is the expected value (i.e., mean) of $x_i$. The eigenvectors with the largest corresponding eigenvalues become the PCs. Once obtained, the PCs can approximate any profile as a weighted sum, i.e.,
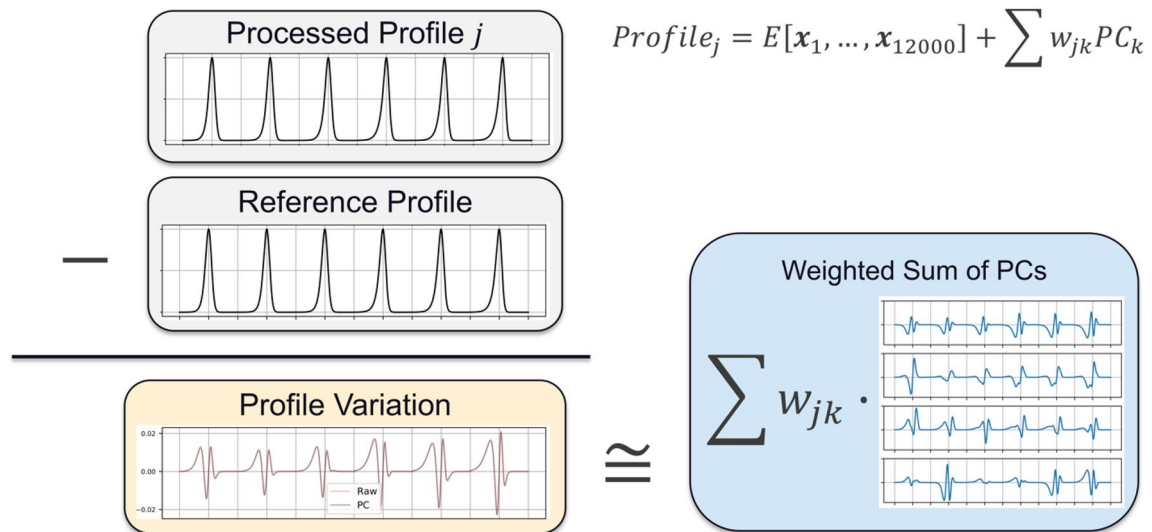
$$Profile_j = E[x_1, \ldots, x_{12000}] + \sum w_{jk} PC_k$$

**Figure 11.** Schematic of how the principal component analysis reproduces a processed profile from principal components.
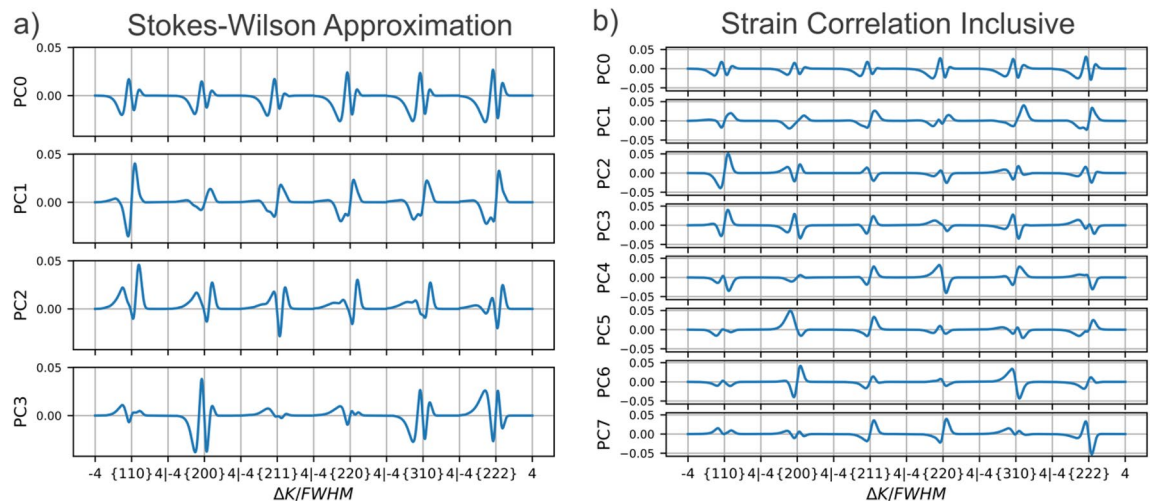


**Figure 12.** The principal components (PCs) which were used to approximate profile shape for the two different profile databases. The components are given in order of decreasing relative contribution to the total variance of the dataset.

$$Profile_j = E[x_1, \ldots, x_{12000}] + \sum w_{jk} PC_k, \tag{20}$$

where $E[x_1, \ldots, x_{12000}]$ is the mean of the set of profiles (i.e. a reference profile). Using the same set of PCs, unique weight vectors, $w_j$, were found for all the profiles in the available set of 177 (16 of the original 193 profiles were reserved for validation). The weight vector, $w_j = [w_0, \ldots, w_k]_j$, associated with each profile is a compressed, approximate, and quantitative signature of the peak shapes. The weights therefore contain the measurable effects of the variation in dislocation density and arrangement present in the database. More than 97.5% of the variation in the $\epsilon^L$-based dataset was captured using 8 PCs, and over 99% of the variation in the S-W dataset was captured using 4 PCs. The PCs are shown in Fig. 12. The PC weights used to approximate each profile were used as the description of peak shape given to the surrogate model.

The compressed data was organized in a "fingerprint" for each profile. The fingerprint contains the $\frac{FWHM}{K}$ of each peak. The series of weights which can reproduce the shape of the profile from the PCs were also included in the fingerprint. The square root of the dislocation density was selected as the dependent variable of the surrogate model. Values of $FWHM$ are expected to vary with $\sqrt{\rho}$[31]. The data used by the surrogate model amounts to a compact vector of quantities,

| Kernel Function Forms |
|---|
| $k_{RBF} = \exp\left(-\frac{1}{2}\left(\frac{d}{l}\right)^2\right), \boldsymbol{\theta} = [l]$ |
| $k_{Matern}(\nu = 1.5) = \left(1 + \frac{\sqrt{3}d}{l}\right)\exp\left(-\frac{\sqrt{3}d}{l}\right), \boldsymbol{\theta} = [l, \nu]$ |
| $k_{Matern}(\nu = 2.5) = \left(1 + \frac{\sqrt{5}d}{l} + \frac{5d^2}{3l^2}\right)\exp\left(-\frac{\sqrt{5}d}{l}\right)$ |
| $k_{RQ} = \left(1 + \frac{d^2}{2\phi l^2}\right)^{-\phi}, \boldsymbol{\theta} = [l, \phi]$ |

**Table 5.** Kernel function forms which are tested to determine the best performing alternative.

$$Output : \sqrt{\rho_i}$$

$$Input : \left[\frac{FWHM}{K}_{110}, \frac{FWHM}{K}_{200}, \frac{FWHM}{K}_{211}, \frac{FWHM}{K}_{220}, \frac{FWHM}{K}_{310}, \frac{FWHM}{K}_{222}, w_0, w_1, w_2, w_3(,\ldots)\right]_i$$

(21)

where the subscript $i$ indicates the $i$-th profile in the dataset. The PCA weights $w_0$, etc., continue to $w_7$ for the fingerprints associated with the $\epsilon^L$-based dataset.

*Data-driven model form.* The data-driven model uses Gaussian process regression (GPR) surrogate models in an ensemble. A GPR uses the covariance between fingerprints, $\mathbf{x}_i$, in the database, $\boldsymbol{X}$, to make predictions for new cases based on the observed dislocation densities and covariance matrix $\boldsymbol{K}$. These predictions can be made after the GPR has been trained, i.e., numerically fitted to a portion of data designated for training the model. The quantity of interest of a new case, $y^* = \sqrt{\rho^*}$, is then predicted using the mean value prediction $\hat{y}^*$ (where $y^*$ may be drawn from a Gaussian distribution with mean = $\hat{y}^*$), i.e.,

$$\hat{y}^* = \mathbf{k}^*(K + \alpha I)^{-1}\mathbf{y},$$

(22)

where $\mathbf{y}$ is the vector of observed values of $y$ in the training data, and $\mathbf{k}^*$ is the vector of covariances between the new fingerprint $\mathbf{x}^*$, and the fingerprint matrix $\boldsymbol{X}$. $\boldsymbol{I}$ is the identity matrix and $\alpha$ is a tunable non-negative quantity which accommodates noise in $\mathbf{y}$. The covariance between fingerprints is defined using a kernel function. This function provides a means of tuning the covariance between two fingerprints, $\mathbf{x}_i$ and $\mathbf{x}_j$, as a function of the distance $d$ (Euclidean or otherwise) between them. The kernel functions used in this work are given as,

$$k\big(d(\mathbf{x}_i, \mathbf{x}_j), \boldsymbol{\theta}\big) \in \big\{k_{Matern}, k_{RBF}, k_{RQ}\big\}$$

(23)

where $\boldsymbol{\theta}$ is a vector of tunable values called hyperparameters and the various kernel functions are defined in Table 5. The chosen kernel function form was the best-fitting to the data, or specifically the highest in terms likelihood of the data, given the optimized hyperparameters and kernel function. The algorithm of the training of GPR models can be found in[68,69].

The GPR was applied in an ensemble approach that uses many random subsamples of 142 of the 177 training profiles to develop 80 GPR models from which an average is taken. The method is known as bagging (from bootstrap-aggregating)[70,71]. This method is recommended in cases where the predictions of a statistical model (here GPR) are subject to large variations when different training data are used. To ascertain if this was the case here, the variation in the bootstrapped GPR predictions was inspected. Where this variation is large, the use of bagging is known to increase the accuracy of the overall predictions. The variation is presented alongside the ensemble predictions in "Application of the two virtual diffraction algorithms against experimental data" section. The variability of the ensemble predictions is also used as an indicator of the uncertainty in the estimates. It is noted that each GPR also produces an estimate of variance with each prediction, however these were not used in this analysis.

**LPA comparison methods.** A baseline set of dislocation density predictions were obtained using existing LPA methods. Semi-quantitative analysis of the microstructural features responsible for the increase in peak breadth was performed using a whole diffraction pattern modeling method implemented in the extended CMWP (eCMWP) software[11]. The formulation of strain broadening used in eCMWP is outlined in "Virtual diffraction" section.

Instrumental broadening was taken from the diffraction of a strain and dislocation free NIST standard Si powder sample. Typically, TOF neutron diffraction data show sharp leading and broad trailing tail-shapes; therefore, a Pearson VII function was used to fit the two sides of the profile to obtain separate breadth and shape parameters for each side of each peak. Linear fits for both shape and breadth versus $K$ were used to calculate the instrument profile at an arbitrary $K$ position. Blanket background subtractions were performed on all diffraction profiles as a pre-processing step. Finally, the calculated whole diffraction pattern was fitted to the measured pattern to refine the microstructural parameters.

A Levenberg–Marquardt algorithm is used to estimate fitting parameters that are used to calculate the arrangement and density of dislocations. Table 6 shows the material parameters used for Ta to determine initial guesses for values needed prior to optimization, specifically the dislocation character parameter ($q$) and average

| Slip Modes | $C_{11}$ (GPa) | $C_{12}$ | $C_{44}$ | $a_0$ (nm) | $b$ (nm) | Zener const. ($A_z$) | $\frac{C_{12}}{C_{44}}$ |
|---|---|---|---|---|---|---|---|
| <111>{110} <br> <111>{112} | 262.77 | 160.88 | 81.44 | 0.33026 | 0.28601 | 1.5986 | 1.9754 |

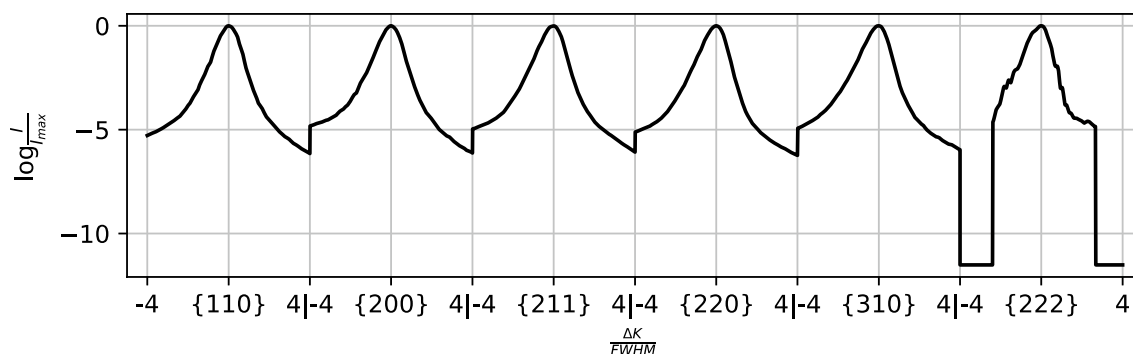**Table 6.** Parameters used in the calculation of the contrast factor and dislocation character as required in eCMWP.



**Figure 13.** A processed profile taken from the set of experimentally obtained neutron diffraction profiles, here shown on a log scale. The peaks shown were processed to remove background intensity, then normalized with respect to maximum intensity and *FWHM*.

contrast factor from Ungar et al.[72]. These initial guesses were $q_{av} = 0.5$, assuming 100% edge type dislocations, and $C_{h00-av} = 0.2$[11]. Estimated values were obtained for area weighted coherently diffracting domain size, dislocation density, dislocation arrangement, and dislocation character. The effects of texture were not included in this analysis. Note, the experimental dislocation densities in as-received Ta and deformation < 5% strain were below the measurement sensitivity and are not reported here.

**Experiment data preprocessing.** To accommodate the data-driven method, background counts and noise were removed from the experimental profiles. Rietveld refinement is commonly used, however the assumption of symmetry in strain broadening is problematic for the whole profile data-driven analysis (that makes no such assumption). A heuristic-based method was developed to avoid biasing the data. The background was subtracted by fitting a Gaussian process to data points manually selected from the background.

Each peak to be used in the surrogate model was extracted from the remaining noisy profile. A combination of an analytical function and the experiment data was used, i.e.,

$$I_{comb} = aI_{ex} + (1 - a)I_{fit} \tag{24}$$

$$a(\Delta K) = \min\left(1, \frac{10 I_{ex}(\Delta K)}{I_{ex,max}}\right) \tag{25}$$

$$I_{fit}(\Delta K) = D\left(\frac{DIFC}{K_{hkl} + \Delta K} + \frac{DIFA}{(K_{hkl} + \Delta K)^2} + ZERO - T_{ph}(d_{hkl})\right) = D(\Delta t) \tag{26}$$

where $I_{ex}$ is the background subtracted peak from the experimental data, $I_{fit}$ is an analytical function fit to the experimental peak, and $I_{comb}$ is a combination of the two that is taken to be used with the surrogate model. The fitted function is defined using $D(\Delta t)$: a convolution of a gaussian, a split exponential, and a Lorentzian function, i.e.,

$$D(\Delta t) = L(\Delta t) * H(\Delta t) \tag{27}$$

$$L(\Delta t) = \frac{l^2}{l^2 + \Delta t^2} \tag{28}$$

where $H$ is defined in Eq. (11) and $l$ is a fitting parameter. The function was fit to each peak using a least squares minimization. In effect, the tails were approximated using an analytical fitting function, while the experimental peak shape was preserved above a noise floor. An example processed experimental profile is shown in Fig. 13.

**eCMWP broadening model.** In eCMWP, a diffraction peak is decomposed into Fourier coefficients to describe broadening in terms of multiple contributing factors: strain broadening, size broadening, and instrument broadening,

$$A_L(\mathbf{K}) = A_L^D(\mathbf{K}) A_L^S(\mathbf{K}) A_L^I(\mathbf{K}) \tag{29}$$

where $A_L^D$ are the strain broadening Fourier coefficients, $A_L^S$ denote size broadening, and $A_L^I$ denote the instrument broadening, all in terms of the Fourier length $L$. The definition of the instrument broadening can vary with experiment, and the form used in this work is described in "Virtual diffraction" section. Size broadening has a negligible influence on this dataset. The strain broadening is chiefly due to dislocations. These effects are calculated as,

$$A_L^D(\boldsymbol{K}) = \exp\left(-2\pi^2 L^2 K^2 \left[\left(\frac{b}{2\pi}\right)^2 \pi \rho C f\left(\frac{L}{R_e^*}\right)\right]\right), \tag{30}$$

with Burgers vector $b$, dislocation density $\rho$, contrast factor $C$, and Wilkens function $f$[31]. $R_e^*$ is a correlation length parameter, which describes the arrangement of dislocations. The strain field of a dislocation is anisotropic, and the strain that is apparent along each diffraction vector $\mathbf{K}$ will vary. The contrast factor $C_{hkl}$ quantifies the apparent strain of each dislocation type at each diffraction vector. Within eCMWP for this work, dislocations are assumed to be equally distributed among slip systems. An average contrast factor is calculated for each $hkl$[38,47,73]. For the cubic material, the average contrast factor $\overline{C}_{hkl}$ is calculated for each peak as,

$$\overline{C}_{hkl} = C_{h00}\left(1 - q_{av}H^2\right) \tag{31}$$

$$H^2 = \frac{h^2k^2 + h^2l^2 + k^2l^2}{\left(h^2 + k^2 + l^2\right)^2} \tag{32}$$

where $q_{av}$ is the dislocation character parameter (determined by the proportion of screw and edge dislocations), and $C_{h00}$ is determined from dislocation crystallography and the elastic properties of the material.

## Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request.

## References
1. Seymour, T. *et al.* Evolution of dislocation structure in neutron irradiated Zircaloy-2 studied by synchrotron X-ray diffraction peak profile analysis. *Acta Mater.* **126**, 102–113 (2017).
2. Wang, H. *et al.* Stress and strain relaxation in magnesium AZ31 rolled plate: In-situ neutron measurement and elastic viscoplastic polycrystal modeling. *Int. J. Plast.* **79**, 275–292 (2016).
3. Zhong, Z. Y. *et al.* Dislocation density evolution of AA 7020–T6 investigated by in-situ synchrotron diffraction under tensile load. *Mater. Charact.* **108**, 124–131 (2015).
4. Brown, D. W., Clausen, B., Sisneros, T. A., Balogh, L. & Beyerlein, I. J. In situ neutron diffraction measurements during annealing of deformed beryllium with differing initial textures. *Metall. Mater. Trans. A* **44**, 5665–5675 (2013).
5. Agnew, S. R., Brown, D. W. & Tomé, C. N. Validating a polycrystal model for the elastoplastic response of magnesium alloy AZ31 using in situ neutron diffraction. *Acta Mater.* **54**, 4841–4852 (2006).
6. Jacques, A. From modeling of plasticity in single-crystal superalloys to high-resolution X-rays three-crystal diffractometer peaks simulation. *Metall. Mater. Trans. A* **47**, 5783–5797 (2016).
7. Miller, M. P. & Dawson, P. R. Understanding local deformation in metallic polycrystals using high energy X-rays and finite elements. *Curr. Opin. Solid State Mater. Sci.* **18**, 286–299 (2014).
8. Kohnert, A. A., Tummala, H., Lebensohn, R. A., Tomé, C. N. & Capolungo, L. On the use of transmission electron microscopy to quantify dislocation densities in bulk metals. *Scr. Mater.* **178**, 161–165 (2020).
9. Zadler, B. J., Le Rousseau, J. H. L., Scales, J. A. & Smith, M. L. Resonant Ultrasound Spectroscopy: theory and application. *Geophys. J. Int.* **156**, 154–169 (2004).
10. Mujica, N., Teresa Cerda, M., Espinoza, R., Lisoni, J. & Lund, F. Ultrasound as a probe of dislocation density in aluminum. *Acta Mater.* **60**, 5828–5837 (2012).
11. Balogh, L., Ribárik, G. & Ungár, T. Stacking faults and twin boundaries in fcc crystals determined by X-ray diffraction profile analysis. *J. Appl. Phys.* **100**, 023512 (2006).
12. Ungár, T. Dislocation densities, arrangements and character from X-ray diffraction experiments. *Mater. Sci. Eng. A* **309–310**, 14–22 (2001).
13. Genzel, C., Denks, I. A. & Klaus, M. Residual stress analysis by X-ray diffraction methods. in *Modern Diffraction Methods* 127–154 (John Wiley & Sons, Ltd, 2013). https://doi.org/10.1002/9783527649884.ch5.
14. Kaganer, V. M., Brandt, O., Trampert, A. & Ploog, K. H. X-ray diffraction peak profiles from threading dislocations in GaN epitaxial films. *Phys. Rev. B* **72**, 045423 (2005).
15. Kapoor, K., Lahiri, D., Rao, S. V. R., Sanyal, T. & Kashyap, B. P. X-ray diffraction line profile analysis for defect study in Zr-2.5% Nb material. *Bull. Mater. Sci.* **27**, 59–67 (2004).
16. Kapoor, K., Lahiri, D., Batra, I. S., Rao, S. V. R. & Sanyal, T. X-ray diffraction line profile analysis for defect study in Cu-1 wt.% Cr-0.1 wt.% Zr alloy. *Mater. Charact.* **54**, 131–140 (2005).
17. Révész, Á., Ungár, T., Borbély, A. & Lendvai, J. Dislocations and grain size in ball-milled iron powder. *Nanostructured Mater.* **7**, 779–788 (1996).
18. Sarkar, A., Mukherjee, P. & Barat, P. X-ray diffraction studies on asymmetrically broadened peaks of heavily deformed zirconium-based alloys. *Mater. Sci. Eng. A* **485**, 176–181 (2008).
19. Sarkar, A., Bhowmik, A. & Suwas, S. Microstructural characterization of ultrafine-grain interstitial-free steel by X-ray diffraction line profile analysis. *Appl. Phys. A* **94**, 943–948 (2009).

20. Mukherjee, P. *et al.* Deformation characteristics of rolled zirconium alloys: A study by X-ray diffraction line profile analysis. *Acta Mater.* **52**, 5687–5696 (2004).
21. Ribárik, G., Gubicza, J. & Ungár, T. Correlation between strength and microstructure of ball-milled Al–Mg alloys determined by X-ray diffraction. *Mater. Sci. Eng. A* **387–389**, 343–347 (2004).
22. Máthis, K. *et al.* The evolution of non-basal dislocations as a function of deformation temperature in pure magnesium determined by X-ray diffraction. *Acta Mater.* **52**, 2889–2894 (2004).
23. Schafler, E., Steiner, G., Korznikova, E., Kerber, M. & Zehetbauer, M. J. Lattice defect investigation of ECAP-Cu by means of X-ray line profile analysis, calorimetry and electrical resistometry. *Mater. Sci. Eng. A* **410–411**, 169–173 (2005).
24. Vives, S., Gaffet, E. & Meunier, C. X-ray diffraction line profile analysis of iron ball milled powders. *Mater. Sci. Eng. A* **366**, 229–238 (2004).
25. Pokharel, R. *et al.* Polycrystal plasticity: Comparison between grain-scale observations of deformation and simulations. *Annu. Rev. Condens. Matter Phys.* **5**, 317–346 (2014).
26. Pokharel, R. *et al.* In-situ observation of bulk 3D grain evolution during plastic deformation in polycrystalline Cu. *Int. J. Plast.* **67**, 217–234 (2015).
27. Kalácska, S., Groma, I., Borbély, A. & Ispánovity, P. D. Comparison of the dislocation density obtained by HR-EBSD and X-ray profile analysis. *Appl. Phys. Lett.* **110**, 091912 (2017).
28. Wilkinson, A. J. *et al.* Measurement of probability distributions for internal stresses in dislocated crystals. *Appl. Phys. Lett* https://doi.org/10.1063/1.4901219 (2014).
29. Krivoglaz, M. A. *X-ray and Neutron Diffraction in Nonideal Crystals* (Springer, 1996).
30. Warren, B. E. & Averbach, B. L. The effect of cold-work distortion on X-ray patterns. *J. Appl. Phys.* **21**, 595–599 (1950).
31. Wilkens, M. The determination of density and distribution of dislocations in deformed single crystals from broadened X-ray diffraction profiles. *Phys. Status Solidi A* **2**, 359–370 (1970).
32. Wilson, A. J. C. The diffraction of X-rays by distorted-crystal aggregates. IV. Diffraction by a crystal with an axial screw dislocation. *Acta Crystallogr.* **5**, 318–322 (1952).
33. Warren, B. E. X-ray studies of deformed metals. *Prog. Met. Phys.* **8**, 147–202 (1959).
34. Stokes, A. R. & Wilson, A. J. C. The diffraction of X rays by distorted crystal aggregates-I. *Proc. Phys. Soc.* **56**, 174–181 (1944).
35. Groma, I. X-ray line broadening due to an inhomogeneous dislocation distribution. *Phys. Rev. B* **57**, 7535–7542 (1998).
36. Gubicza, J., Balogh, L., Hellmig, R. J., Estrin, Y. & Ungár, T. Dislocation structure and crystallite size in severely deformed copper by X-ray peak profile analysis. *Mater. Sci. Eng. A* **400–401**, 334–338 (2005).
37. Ribárik, G. & Ungár, T. Characterization of the microstructure in random and textured polycrystals and single crystals by diffraction line profile analysis. *Mater. Sci. Eng. A* **528**, 112–121 (2010).
38. Ungár, T. & Tichy, G. The effect of dislocation contrast on X-ray line profiles in untextured polycrystals. *Phys. Status Solidi A* **171**, 425–434 (1999).
39. Greer, J. R. & Nix, W. D. Nanoscale gold pillars strengthened through dislocation starvation. *Phys. Rev. B* **73**, 245410 (2006).
40. Hirsch, P. B., Horne, R. W. & Whelan, M. J. Direct observations of the arrangement and motion of dislocations in aluminium. *Philos. Mag. J. Theor. Exp. Appl. Phys.* **1**, 677–684 (1956).
41. Kiener, D. & Minor, A. M. Source truncation and exhaustion: Insights from quantitative in situ TEM tensile testing. *Nano Lett.* **11**, 3816–3820 (2011).
42. Liu, G. S. *et al.* Electron tomography of dislocation structures. *Mater. Charact.* **87**, 1–11 (2014).
43. Chen, C.-C. *et al.* Three-dimensional imaging of dislocations in a nanoparticle at atomic resolution. *Nature* **496**, 74–77 (2013).
44. Scott, M. C. *et al.* Electron tomography at 24-ångström resolution. *Nature* **483**, 444–447 (2012).
45. Groma, I., Ungár, T. & Wilkens, M. Asymmetric X-ray line broadening of plastically deformed crystals. I. Theory. *J. Appl. Crystallogr.* **21**, 47–54 (1988).
46. Ungár, T., Groma, I. & Wilkens, M. Asymmetric X-ray line broadening of plastically deformed crystals. II. Evaluation procedure and application to [001]-Cu crystals. *J. Appl. Crystallogr.* **22**, 26–34 (1989).
47. Ungár, T. & Borbély, A. The effect of dislocation contrast on x-ray line broadening: A new approach to line profile analysis. *Appl. Phys. Lett.* **69**, 3173–3175 (1996).
48. Balogh, L., Capolungo, L. & Tomé, C. N. On the measure of dislocation densities from diffraction line profiles: A comparison with discrete dislocation methods. *Acta Mater.* **60**, 1467–1477 (2012).
49. Upadhyay, M. V., Capolungo, L. & Balogh, L. On the computation of diffraction peaks from discrete defects in continuous media: Comparison of displacement and strain-based methods. *J. Appl. Crystallogr.* **47**, 861–878 (2014).
50. Bamney, D., Tallman, A., Capolungo, L. & Spearot, D. E. Virtual diffraction analysis of dislocations and dislocation networks in discrete dislocation dynamics simulations. *Comput. Mater. Sci.* **174**, 109473 (2020).
51. Eloh, K. S., Jacques, A., Ribarik, G. & Berbenni, S. The effect of crystal defects on 3D high-resolution diffraction peaks: A FFT-based method. *Mater. Basel Switz.* **11**, 1669 (2018).
52. Bertin, N. & Cai, W. Computation of virtual X-ray diffraction patterns from discrete dislocation structures. *Comput. Mater. Sci.* **146**, 268–277 (2018).
53. Coleman, S. P., Spearot, D. E. & Capolungo, L. Virtual diffraction analysis of Ni [010] symmetric tilt grain boundaries. *Model. Simul. Mater. Sci. Eng.* **21**, 055020 (2013).
54. Bertin, N., Upadhyay, M. V., Pradalier, C. & Capolungo, L. A FFT-based formulation for efficient mechanical fields computation in isotropic and anisotropic periodic discrete dislocation dynamics. *Model. Simul. Mater. Sci. Eng.* **23**, 065009 (2015).
55. Mitchell, T. E. & Spitzig, W. A. Three-stage hardening in tantalum single crystals. *Acta Metall.* **13**, 1169–1179 (1965).
56. Spitzig, W. A. & Mitchell, T. E. Dislocation arrangements in tantalum single crystals deformed in tension at 373°K. *Acta Metall.* **14**, 1311–1323 (1966).
57. Kuhlmann-Wilsdorf, D. Theory of plastic deformation: - Properties of low energy dislocation structures. *Mater. Sci. Eng. A* **113**, 1–41 (1989).
58. Queyreau, S., Monnet, G. & Devincre, B. Orowan strengthening and forest hardening superposition examined by dislocation dynamics simulations. *Acta Mater.* **58**, 5586–5595 (2010).
59. Queyreau, S., Monnet, G. & Devincre, B. Slip systems interactions in α-iron determined by dislocation dynamics simulations. *Int. J. Plast.* **25**, 361–377 (2009).
60. Jones, D. R., Fensin, S. J., Martinez, D. T., Trujillo, C. P. & Gray, G. T. Effect of peak stress and tensile strain-rate on spall in tantalum. *J. Appl. Phys.* **124**, 085901 (2018).
61. Bourke, M. A. M., Dunand, D. C. & Ustundag, E. SMARTS—A spectrometer for strain measurement in engineering materials. *Appl. Phys. A* **74**, s1707–s1709 (2002).
62. Moulinec, H. & Suquet, P. A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Comput. Methods Appl. Mech. Eng.* **157**, 69–94 (1998).
63. Featherston, F. H. & Neighbours, J. R. Elastic constants of tantalum, tungsten, and molybdenum. *Phys. Rev.* **130**, 1324–1333 (1963).
64. Shi, X. J., Dupuy, L., Devincre, B., Terentyev, D. & Vincent, L. Interaction of ⟨100⟩ dislocation loops with dislocations studied by dislocation dynamics in α-iron. *J. Nucl. Mater.* **460**, 37–43 (2015).
65. Lim, H. *et al.* Investigating active slip planes in tantalum under compressive load: Crystal plasticity and slip trace analyses of single crystals. *Acta Mater.* **185**, 1–12 (2020).

66. Lim, H. *et al.* Grain-scale experimental validation of crystal plasticity finite element simulations of tantalum oligocrystals. *Int. J. Plast.* **60**, 1–18 (2014).
67. Von Dreele, R. B., Jorgensen, J. D. & Windsor, C. G. Rietveld refinement with spallation neutron powder diffraction data. *J. Appl. Crystallogr.* **15**, 581–589 (1982).
68. Rasmussen, C. E. & Williams, C. K. *Gaussian Process for Machine Learning* (MIT Press, 2006).
69. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
70. Breiman, L. Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996).
71. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (CRC Press, 1994).
72. Ungár, T., Dragomir, I., Révész, Á. & Borbély, A. The contrast factors of dislocations in cubic crystals: The dislocation model of strain anisotropy in practice. *J. Appl. Crystallogr.* **32**, 992–1002 (1999).
73. Borbély, A., Dragomir-Cernatescu, J., Ribárik, G. & Ungár, T. Computer program ANIZC for the calculation of diffraction contrast factors of dislocations in elastically anisotropic cubic, hexagonal and trigonal crystals. *J. Appl. Crystallogr.* **36**, 160–162 (2003).

## Acknowledgements

## Author contributions

L.C., D.B., A.E.T., R.P. and B.C conceived of and designed this project. A.E.T. and D.B. developed the new code used in this work. B.C. collected the neutron diffraction profiles. A.E.T. and R.P. performed the data analysis. L.C., D.B., and D.E.S. acquired funding and managed the project. A.E.T. and R.P. provided the initial draft of the paper with contributions from the other authors. All of the authors discussed the contents, edited, and approved the publication of this paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to L.C.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.