# Processing scanning Laue microdiffraction patterns with machine learning algorithms: A case-study with fatigued polycrystalline copper

## Introduction

Laue diffraction, that may occur when a polychromatic X-ray beam illuminated a crystal, was first discovered in 1912, and has revealed both the electromagnetic nature of X-ray and the periodic ordering of atoms in crystal [1]. Thanks to the polychromaticity of the employed X-ray, multiple diffraction peaks can be recorded in a single exposure without any rotation, thereby excluding the ambiguity of the illuminated volume [2]. With the development of polychromatic beam focusing optics, notably Kirkpatrick–Baez mirrors, micron-sized high-brilliance polychromatic X-ray beam can be produced at synchrotron radiation sources and directed to probe inside materials with submicrometric spatial resolution, i.e. Laue microdiffraction. In analogy to EBSD (electron backscatter diffraction) technique, Laue microdiffraction technique serves by raster scanning the sample to generate the lattice orientation and distortion maps from the one-by-one analysis of the diffraction pattern emanating from each scanned spot [3-5]. The two techniques are comparable [6] and complementary to each other [7]. It is generally accepted that EBSD has an edge on finer spatial resolution of nanoscale, whilst Laue microdiffraction can have a much better accuracy on the lattice orientation and distortion with an attainable order of 10-9 [8].

A salient feature of Laue microdiffraction is its sensitivity to the local misorientation inside the illuminated volume [9, 10], more specifically, the fragmentation of Laue spot may indicate the presence of geometrically necessary boundaries and the elongation of Laue spot the presence of geometrically necessary dislocations. Although a critical aspect of the spot shape analysis lies on the assumption that the dislocations were dominantly edge-type in the illuminated volume, a recent study with focused ion beam transmission electron microscopy confirmed that this analysis stood still if the dislocations had predominately screw-type [11]. With the aid of a wire profiler (typically Pt), the shape of spot can be spatially resolved to yield subsurface, 3D mapping of lattice orientation and distortion non-destructively [12-14], namely the differential-aperture X-ray microscopy (DAXM) technique [15].

Despite the wealth of information behind Laue microdiffraction pattern, the interpretation of Laue microdiffraction pattern is not straightforward since the wavelength or indexation pertaining to each diffraction peak is unknown. Standard treatment involves modulating the orientation and calibration parameters to minimize the discrepancy between the simulated and experimental diffraction pattern, and has been implemented in software such as *XMAS* [16] and *LaueTools* (<https://gitlab.esrf.fr/micha/lauetools>). The standard treatment is in essence a trial-and-error process that usually suffers from inefficiency, especially for the raster scanned diffraction patterns which has to be treated one by one. Therefore, any additional information concerning the scanned microstructure would facilitate the process, for example, Örs et al. [7] used the orientation obtained by EBSD to overcome the difficulty in indexing the Laue microdiffraction patterns; Kou et al. [17] suggested indexing one Laue microdiffraction pattern per grain as the reference with which the rest patterns of the grain could be analyzed without indexation.

Concurrently, the emerging image processing techniques have demonstrated their potential in the interpretation of Laue microdiffraction patterns: Petit et al. [18] introduced the digital image correlation (DIC) technique to have a better measurement of relative lattice distortion with reference to an assumed stress-free position, i.e. Laue-DIC method, Zhang et al. [19] later extended the framework of Laue-DIC to get rid of the dependence on the stress-free reference; Zhang et al. [20] used DIC to correct the misalignment of the investigated volume in the experiments of DXAM;

In a word, the indexation of diffraction peaks is the key to full interpretation of diffraction pattern. However, in certain circumstances, full interpretation of diffraction patterns is unnecessary or fast parallel computing capabilities are unavailable, thereby necessitating the development of indexation-free approach towards on-the-fly analysis of raster scanned diffraction patterns. Zhou et al. [5] proposed using the distribution of average recorded intensities and average filtered intensities of the raster scanned diffraction patterns to visualize the characteristics microstructural features. The recently emerging convolutional neural networks (CNN) has been used to extract features from diffraction patterns for further clustering and labeling the raster scanned diffraction patterns [21].

In the present work, we demonstrated the application of machine learning algorithms to the raster scanned diffraction patterns of the fatigued polycrystalline copper. Substantial dislocation structures will grow in copper after the cyclic loading [22], deteriorating the identifiability of the diffraction pattern. Although template matching schemes have been shown applicable to the indexation and misorientation analysis of smeared diffraction patterns, huge amount of calculation was still inevitable [23] and the reliability of outcome would be degraded in line with the formation of dislocation structures. On the other hand, machine learning algorithms, which were developed to handle big data, were possible to circumvent the difficulty of indexation and segment raster scanned diffraction patterns according to their features, thereby mapping phases, grains, or grain substructures. Since the microstructure of the scanned area was not known *a priori*, unsupervised labeling algorithms had to be adopted. The objective of this paper is to: (i) outline the computational pipeline from the raw data to the clustering of diffraction patterns; (ii) compare the performance of several algorithms; and (iii) discuss the influence of the diffraction patterns on the results of clustering.

## Experiment

The diffraction patterns were collected from raster scanning of fatigued polycrystalline copper. The sample was designed in accordance with the ASTM/E606 standard. The tensile strength at room temperature was given in Fig 1 along with the EBSD (electron backscattered diffraction) mapping in the inset. The sample, cyclically loaded in stress-control mode with the stress varying sinusoidally within the range 0 ~ 140 MPa, has undergone a maximum strain of ~ 10% in the initial cycle from Figure 1 and cyclic creep in the subsequent cycles. The sample was fatigued up to 109 cycles with a frequency of 10 Hz.



Figure 1 The tensile curve of the sample along with the orientation mapping in the inset.

The Laue raster scanning over the sample was performed in beamline 4B of Pohang Light Source. The raster scanning was over a grid with the horizontal direction parallel to the fatigue loading wherein . The step size was 2 μm in both the horizontal and vertical directions of the grid. The obtained Laue microdiffraction pattern was extremely blurred with almost no discernable diffraction peaks (Figure 2a).

## Data reduction

The original Laue microdiffraction pattern has 1024×1024 pixels. Although one can directly use the full gray levels of images to cluster the grid points from a theoretical point of view, it is either impractical or unnecessary to handle such huge amount of data, thus calling for a data reduction process to reduce the diffraction patterns into a manageable number of latent features.

To begin with, each image needed to be normalized to eliminate systematic errors. Normalization was accomplished by subtracting the mean gray level from the gray level and dividing by the standard deviation of the gray levels (Figure 2a). Then the normalized images were compressed to 512×512 pixels by 2×2 averaging binning to compress the data and smooth the noise (Figure 2b). The images in spatial domain could be equivalently expressed in frequency domain (Figure 2c) by discrete sine transformation (DST), which transformed the image into the weighted sum of sinusoids with discrete frequencies. It was obvious from Figure 2c that the components of high frequencies were negligible compared to those of low frequencies.



Figure 2 (a) The normalized image of the diffraction pattern; (b) the shrunk image of the Figure 2a after 8×8 averaging binning; (c) the DST of Figure 2b.

Latent features can be extracted from either spatial domain or frequency domain. Song et al. applied CNN to extract latent features [21]. However, at present, the authors did not have sufficient patterns to train the CNN, therefore unsupervised learning algorithms that did not require training dataset were employed herein to extract latent features of each pattern. If the latent features were properly extracted, it was possible to classify the diffraction patterns to the grains that they belonged to. Here we used the hierarchical agglomerative clustering (HAC) algorithm backed by Scikit-Learn [24]. When treating the scanning diffraction images with HAC algorithm, each pixel corresponded to a vector comprised of the values at the pixel in all images whether they be in spatial or frequency domain; then a metric (Euclidean distance, maximum distance, etc.) was used to quantify the dissimilarity between the pairs of pixels; then pixels with high similarities were merged to form a feature according to a linkage criterion. The connectivity of pixels could be exploited to facilitate the merging process such that only the pairs of adjacent pixels were under consideration. In this work, a total of latent features were to be extracted from the shrunk images wherein the Euclidean distance was used as the metric of dissimilarity, and the linkage criterion employed was “ward” aiming at minimizing the sum of squared differences within all clusters.

Figure 2d and e display the results of feature clustering of Figure 2b and c respectively, and Figure 2f plots the histograms of the numbers of pixels contained by each feature in both spatial and frequency domain. The feature clustering was rather uniform in the spatial domain whilst the features concentrated on the lower frequencies in the frequency domain. A highly skewed distribution of pixel numbers in frequencies domain was identified, suggesting that the majority of high frequency components could be merged into one feature, whereas the pixel numbers distributed uniformly among features.



Figure 3

Figure 3a and b give two examples of the diffraction images, as well as their restored images from the latent features extracted from spatial and frequency domain. Conspicuously, images restored from latent features extracted from spatial domain are more similar to the original images than those from frequency domain. A more quantitative comparison employs the normalized cross-correlation (NCC) coefficient, which quantifies the resemblance between the original image and the restored image.

where and are the gray levels in the original and restored images. NCC is strictly within the range and larger NCC indicates better representability of the latent features. For two identical images, the NCC equals one. Figure 3g gives the distributions of NCC for latent features extracted from both spatial and frequency domain. It is apparent that the latent features extracted from spatial domain are more representative of the original images that those extracted from frequency domain. Therefore, latent features from spatial domain will be used for clustering hereinafter. In this manner, the originally diffraction patterns of 1024×1024 pixels were reduced to a data matrix (, ).



Figure 4

However, each latent feature is only representative of its corresponding pixels of the diffraction pattern. For example, Figure 5a displays the mappings of the first five latent features over the scanned area, where only highlighted are the parts of the scanned area in which the latent feature is active. A principal component analysis (PCA) would be useful to transform the latent features of the diffraction patterns into highly distinctive features among diffraction patterns. The PCA is in essence the singular value decomposition (SVD) of the data matrix :

is a diagonal matrix consisting of the square roots of the eigenvalues of the covariance matrix , with their eigenvectors stored in the columns of orthogonal matrix and termed as principal components hereinafter. The diagonal elements of are in proportion to the variances of their corresponding principal components.

We applied the PCA to the data matrix with Scikit-Learn [24], and displayed the variances of principal components in descending order in Figure 4c, in which the variances of the first 64 principal components were shown in the inset. Figure 4b displays the mappings of the first five principal components over the scanned grid, wherein the grain structures can be more clearly identified than using the latent features. The data matrix could be effectively represented by retaining the most significant principal components () while discarding the rest. The choice of is empirical, and its influence will be discussed in the later section.



Figure 5

The authors noticed that in both Figure 5a and b, there existed a few conspicuous spots that differed significantly from their neighbors. The diffraction pattern of one of these spots was shown in Figure 5d in which a strong background noise was observed. The authors speculated that the distinctively strong background noise was due to the artifacts in the scanning process rather than due to the microstructural variation of the sample. Figure 5e and f presented two diffraction patterns from two adjacent regions in the mapping of the 5th principle component. Homology between the two patterns could be identified as the spots in the two patterns seemed displaced, suggesting a relative rotation between the two underlying lattices; moreover, the spots in one pattern (Figure 5e) were elongated indicating the storage of geometrically necessary dislocations in a single slip system, whilst the spots in another pattern (Figure 5g) exhibited relative roundness. To further clarify the relationship between the two diffraction patterns, Figure 5f gave the diffraction pattern in between the aforementioned two patterns, wherein the bifurcated spots suggested that multiple slip systems had been triggered. It was clear that the two adjacent regions highlighted in the inset of Figure 5b-v belonged to the same grain before the fatigue loading, however, during the fatigue loading, elements other than the Schmid factors influenced the activation of slips systems, i.e., the constrain imposed by the neighboring grains, leading to the development of subgrain structures that differed in the activation of slip systems.



Figure 6

## Grid clustering

After mapping each diffraction pattern into features, the scanning grid was ready to be segmented by merging the samples with “similar” features. Three clustering algorithms, namely K-Means, affinity propagation (AP), and the aforementioned HAC, were investigated.

AP algorithm clusters samples by sending messages between pairs of samples iteratively and the number of clusters can be determined automatically. Figure 7a presents the clustering results of AP algorithms with different dimension of features, i.e. in which 78, 150, 210, and 268 clusters were identified respectively. The iteration count for AP algorithm to converge also increased with and larger might result in difficulty of convergence within an acceptable time. In contrary to AP algorithm, both K-Means and HAC algorithms require the numbers of clusters to be entered *a priori*. For comparison with AP algorithm, the number of clusters specified for both K-Means and HAC algorithms were set to be consistent with the automatically determined cluster number in AP algorithm for each , and their results are shown in Figure 7b and c respectively.



Figure 7

The CPU times required for all algorithms were shown in Figure 8. HAC algorithm took much less CPU times than the other algorithms since the connectivity of pixels were exploited that only the adjacent pixels were under consideration for merging. Another advantage of HAC algorithm brought by the connectivity was that the clustered regions were continuous. Therefore, HAC algorithm would be employed in the following discussion.



Figure 8



Figure 9

[1] M. Eckert, Disputed discovery: the beginnings of X-ray diffraction in crystals in 1912 and its repercussionsThis Laue centennial article has also been published in Zeitschrift fur Kristallographie [Eckert (2012). Z. Kristallogr. 227, 27-35], Acta Crystallogr. Sect. A 68(1) (2012) 30-39.

[2] J.-S. Chung, G.E. Ice, Automated indexing for texture and strain measurement with broad-bandpass x-ray microbeams, J. Appl. Phys. 86(9) (1999) 5249-5255.

[3] R. Spolenak, W.L. Brown, N. Tamura, A.A. MacDowell, R.S. Celestre, H.A. Padmore, B. Valek, J.C. Bravman, T. Marieb, H. Fujimoto, B.W. Batterman, J.R. Patel, Local Plasticity of Al Thin Films as Revealed by X-Ray Microdiffraction, Phys. Rev. Lett. 90(9) (2003) 096102.

[4] N. Tamura, A.A. MacDowell, R. Spolenak, B.C. Valek, J.C. Bravman, W.L. Brown, R.S. Celestre, H.A. Padmore, B.W. Batterman, J.R. Patel, Scanning X-ray microdiffraction with submicrometer white beam for strain/stress and orientation mapping in thin films, J. Synchrotron Rad. 10(2) (2003) 137-143.

[5] G. Zhou, J. Kou, Y. Li, W. Zhu, K. Chen, N. Tamura, Quantitative Scanning Laue Diffraction Microscopy: Application to the Study of 3D Printed Nickel-Based Superalloys, Quantum Beam Sci. 2(2) (2018) 13.

[6] E. Plancher, J. Petit, C. Maurice, V. Favier, L. Saintoyant, D. Loisnard, N. Rupin, J.B. Marijon, O. Ulrich, M. Bornert, J.S. Micha, O. Robach, O. Castelnau, On the Accuracy of Elastic Strain Field Measurements by Laue Microdiffraction and High-Resolution EBSD: a Cross-Validation Experiment, Exp. Mech. 56(3) (2016) 483-492.

[7] T. Örs, J.-S. Micha, N. Gey, V. Michel, O. Castelnau, R. Guinebretiere, EBSD-assisted Laue microdiffraction for microstrain analysis, J. Appl. Crystallogr. 51(1) (2018) 55-67.

[8] C. Zhang, T.R. Bieler, P. Eisenlohr, Exploring the accuracy limits of lattice strain quantification with synthetic diffraction data, Scr. Mater. 154 (2018) 127-130.

[9] R. Barabash, G.E. Ice, B.C. Larson, G.M. Pharr, K.-S. Chung, W. Yang, White microbeam diffraction from distorted crystals, Appl. Phys. Lett. 79(6) (2001) 749-751.

[10] R.I. Barabash, G.E. Ice, F.J. Walker, Quantitative microdiffraction from deformed crystals with unpaired dislocations and dislocation walls, J. Appl. Phys. 93(3) (2003) 1457-1464.

[11] C. Zhang, S. Balachandran, P. Eisenlohr, M.A. Crimp, C. Boehlert, R. Xu, T.R. Bieler, Comparison of dislocation content measured with transmission electron microscopy and micro-Laue diffraction based streak analysis, Scr. Mater. 144 (2018) 74-77.

[12] W. Yang, B.C. Larson, J.Z. Tischler, G.E. Ice, J.D. Budai, W. Liu, Differential-aperture X-ray structural microscopy: a submicron-resolution three-dimensional probe of local microstructure and strain, Micron 35(6) (2004) 431-439.

[13] R.I. Barabash, G.E. Ice, W. Liu, O.M. Barabash, Polychromatic microdiffraction characterization of defect gradients in severely deformed materials, Micron 40(1) (2009) 28-36.

[14] S. Das, F. Hofmann, E. Tarleton, Consistent determination of geometrically necessary dislocation density from simulations and experiments, Int. J. Plast. 109 (2018) 18-42.

[15] B.C. Larson, W. Yang, G.E. Ice, J.D. Budai, J.Z. Tischler, Three-dimensional X-ray structural microscopy with submicrometre resolution, Nature 415(6874) (2002) 887-890.

[16] N. Tamura, XMAS: A Versatile Tool for Analyzing Synchrotron X-ray Microdiffraction Data, Strain and Dislocation Gradients from Diffraction2014, pp. 125-155.

[17] J. Kou, K. Chen, N. Tamura, A peak position comparison method for high-speed quantitative Laue microdiffraction data processing, Scr. Mater. 143 (2018) 49-53.

[18] J. Petit, O. Castelnau, M. Bornert, F.G. Zhang, F. Hofmann, A.M. Korsunsky, D. Faurie, C. Le Bourlot, J.S. Micha, O. Robach, O. Ulrich, Laue-DIC: a new method for improved stress field measurements at the micrometer scale, J. Synchrotron Rad. 22(4) (2015) 980-994.

[19] F.G. Zhang, O. Castelnau, M. Bornert, J. Petit, J.B. Marijon, E. Plancher, Determination of deviatoric elastic strain and lattice orientation by applying digital image correlation to Laue microdiffraction images: the enhanced Laue-DIC method, J. Appl. Crystallogr. 48(6) (2015) 1805-1817.

[20] C. Zhang, Y. Zhang, G. Wu, W. Liu, R. Xu, D. Juul Jensen, A. Godfrey, Alignment of sample position and rotation during in situ synchrotron X-ray micro-diffraction experiments using a Laue cross-correlation approach, J. Appl. Crystallogr. 52(5) (2019) 1119-1127.

[21] Y. Song, N. Tamura, C. Zhang, M. Karami, X. Chen, Data-driven approach for synchrotron X-ray Laue microdiffraction scan analysis, Acta Crystallogr. Sect. A 75(6) (2019) 876-888.

[22] H. Mughrabi, Cyclic Slip Irreversibilities and the Evolution of Fatigue Damage, Metall. Mater. Trans. B 40(4) (2009) 431-453.

[23] V.K. Gupta, S.R. Agnew, Indexation and misorientation analysis of low-quality Laue diffraction patterns, J. Appl. Crystallogr. 42(1) (2009) 116-124.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res. 12(null) (2011) 2825–2830.