ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST046

MACHINE LEARNING APPLIED FOR THE CALIBRATION OF THE HARD X-RAY SINGLE-SHOT SPECTROMETER AT THE EUROPEAN XFEL

C. Grech*, M. Guetg, Deutsches-Elektronen Synchrotron, DESY, Hamburg, Germany
G. Geloni, EuXFEL, Schenefeld, Germany

Abstract

Single-crystal monochromators are used in free electron lasers for hard x-ray self-seeding, selecting a very narrow spectral range of the original SASE signal for further amplification. When rotating the crystal around the roll and pitch axes, one can exploit several symmetric and asymmetric reflections as established by Bragg's law. This work describes the implementation of a machine learning classifier to identify the crystal indices corresponding to a given reflection, and eventually calculate the difference between the photon energy as measured by a single-shot spectrometer and the actual one. The image processing techniques to extract the properties of the crystal reflection are described, as well as how this information is used to calibrate two spectrometer parameters.

INTRODUCTION

Hard X-Ray Self-Seeding (HXRSS) is based on the principle of using crystal monochromators to narrow down the spectral range of self-amplified spontaneous emission (SASE) free electron lasers (FELs) whilst increasing the peak spectral brightness [1, 2]. In its simplest configuration, a self-seeded XFEL consists of an input undulator and an output undulator separated by a single crystal monochromator. In order to ease heat loading effects at high repetition rate and low photon energies, at the European XFEL (EuXFEL) we installed two cascaded HXRSS systems at the SASE2 undulator line, as shown in Figure 1.

HXRSS can only be achieved when the incident beam hits the crystal at the Bragg angle corresponding to the seed frequency for a specific reflection. Thus, the seeded XFEL output is reliant on the diffraction process in the crystal, which highly depends on the crystal orientation with respect to the incident beam direction. The crystal orientation can be controlled in pitch and roll by means of dedicated stages, where the pitch angle θ_p can move in the range $30^\circ \leq \theta_p \leq 120^\circ$. The yaw plane controlled by θ_y is kept constant and unchanged throughout operation. Figure 1 shows the crystal rotational convention considered in this work. The pitch rotation axis is orthogonal to the beam incident direction and parallel to the floor, while the roll rotation axis has an axis lying parallel to the beam incident direction when $\theta_p = 0^\circ$.

The HIgh REsolution hard X-ray single-shot (HIREX) spectrometer installed at the SASE2 undulator beamline of the European XFEL is employed to measure the spectrum

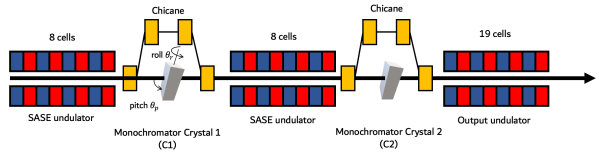

Figure 1: EuXFEL's SASE2 undulator two-crystal hard x-ray self-seeding scheme.

of individual photon pulses. The spectrometer is based on diamond gratings, bent crystals, and a MHz-repetition-rate strip Gotthard detector [3]. An X-ray Gas Monitor (XGM) detector provides a complementary measure of the total x-ray pulse energy, without spectral information. The pySpectrometer photon diagnostic software developed for the European XFEL [4] is used to monitor the measured spectrum, as well as control the calibration of the HIREX spectrometer. Two spectrometer calibration parameters can be controlled through this diagnostic tool: the reference energy and the pixel calibration. The reference energy parameter E_0 provides an absolute energy value, as a reference for the measured spectrum. It does not necessarily correspond to the seeding energy. The second parameter, the pixel calibration, is defined as the change in energy corresponding to a displacement of one pixel, with respect to the pitch angle.

Identifying the actual photon energy to obtain the desired crystal reflection can be arduous. In fact, there often are many neighbouring (in terms of pitch angle and photon energy) reflections, and it can be difficult to identify the actual one based on visual inspection. Machine Learning (ML) methods applied to particle accelerator controls are becoming more common, with the focus being on the efficient use of large amounts of sensor data. In this paper, we propose and demonstrate a machine learning technique, which can be applied at any XFEL facility utilising a monochromator crystal setup to identify crystal reflections, and as a result to identify the operational absolute photon energy. In this study, data obtained at the EuXFEL over two years of HXRSS commissioning and operation is considered, with photon energies ranging from 6 keV to 18 keV. A measurement model is developed based on Bragg's law, and is used to train a classifier that can identify the crystal reflection. Crystal pitch angle and photon energy scans in the form of an image are obtained from the spectrometer, fed to the model and all reflections present in the image are identified. In the end, this information is used to update the two spectrometer calibration parameters. Figure 2 summarizes the complete process from obtaining the image to determining the calibra-

* christian.grech@desy.de

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST046

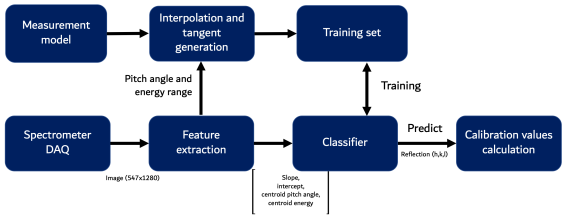

Figure 2: Schematic technique of the absolute energy calibration based on the machine learning to predict the crystal reflection identifier.

tion values. This approach can be potentially used to speed up crystal-based setups at XFEL machines and contribute to the calibration of energy measurement instruments.

MEASUREMENT MODEL

Whilst Bragg law defines the relationship between the crystal rotational position and the photon energy (E_{ph}), it does not consider systematic errors due to the crystal misalignment in the beamline. This work considers five corrective parameters that can be applied to match the theoretical model to the actual (or expected) crystal behaviour. Eq.1 shows the proposed measurement model for the absolute photon energy E_{ph}, where f_{Bragg} represents Bragg law. The pitch angle error ($\delta\theta_p$) is a constant associated with a fixed deviation from the actual pitch angle, and similarly the yaw angle error ($\delta\theta_y$) in the case of the yaw angle. The roll angle error is linearly associated with the pitch angle, represented by two constants (k_{r1}, k_{r2}) as shown in Eq. 2.

$$E_{ph} = f_{Bragg}(\theta_p + \delta\theta_p, \theta_r + \delta\theta_r, \theta_y + \delta\theta_y) + \delta E \quad (1)$$

$$\delta\theta_r = k_{r1} + k_{r2}\theta_p \quad (2)$$

Finally, δE represents the offset between the set spectrometer reference energy and the actual photon energy. This value is not fixed but can vary from one measurement session to another, depending on the spectrometer initial setup.

FEATURE EXTRACTION

Using pySpectrometer, when a monochromator crystal is rotated in one of the axes, the photon energy measured from the single-shot spectrometer can be correlated with the angle of rotation and exported for further processing. The information in the form of an image highlights any reflection present, as shown for example in Figure 3a).

In order to segment the information in the foreground from the background, thresholding is applied using the method proposed by Yen [5]. This creates a binary image as shown in Figure 3b). In addition, morphological closing is performed. Morphological closing is useful for filling small holes from an image while preserving the shape and size of the objects in the image. The closing operation dilates an image and then erodes the dilated image, using the same structuring element for both operations.

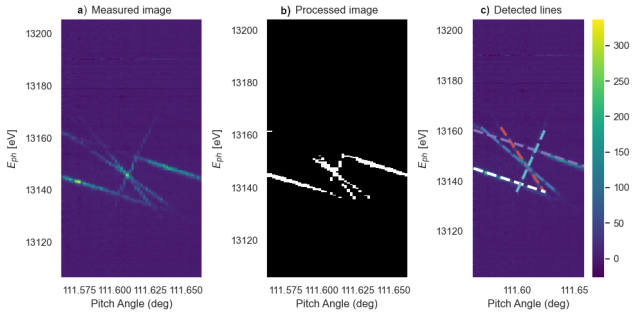

Figure 3: Example of a) a captured image, b) a processed version (binarization, dilation, erosion) and c) the detected lines (Hough Transform).

For the identification of lines within the binary images, the classical Hough transform method is implemented. For each pixel in the foreground, a number of lines are plotted going through it, all at varying angles (θ). The Hough transform accumulates contributions from all pixels in the detected edge. The distance between an origin point and each pixel is calculated, and when multiple points have similar (r, θ) values, these are considered to pertain to the same line. In this way, a number of lines can be detected from the image as shown in Figure 3c). As a result, several line properties, namely the line slope, intercept and centroid co-ordinates can be determined.

In order to be able to match the detected lines with the measurement model curves, tangent lines are generated for each Bragg curve in the measured pitch angle range. This makes the comparison between the measurement model curves and the detected lines possible as both datasets consist of straight lines.

CLASSIFIER TRAINING

A machine learning classifier is used to match detected image lines with the tangent lines derived from the measurement model. The classifier considers the features of a given reflection line, and predicts the corresponding Miller indices $[h, k, l]$. Comparing three multi-class classifiers, the k-nearest neighbors algorithm is found to be the best fit for this scenario. The nearest neighbor algorithm is a supervised machine learning method that creates 'neighborhoods' for objects with similar features. The category of an unclassified point is found by comparing the point features against those of the previously classified k neighbors, which exist as a subset of a larger space of previously classified points.

The classifier is trained with the properties of the tangent line of the measurement model as input (slope, intercept, centroid pitch angle, centroid energy) and the reflection identifier as output. The algorithm calculates the proximity between the measurement model features and the features in the detected lines and picks the closest class. Once the model is trained with the measurement model tangent infor-

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST046

mation, the detected line properties are used as inputs and an identifier is obtained for each line.

ABSOLUTE ENERGY CALIBRATION

Once an identifier is obtained for each detected reflection, the vertical offset between the two lines can be determined, which corresponds to the δE parameter in Eq. 1. As a result the spectrometer reference energy parameter E_0 can be relatively adjusted by this amount to calibrate the absolute energy value. The pixel calibration can also be adjusted by comparing the slope of the measured curve with the model curve and applying the gain between the two values. A PyQt GUI tool integrated with pySpectrometer provides users with the ability to upload an image, perform feature extraction, train the classifier, identify the reflections and recalculate the spectrometer calibration parameters.

RESULTS

Thresholding Accuracy

The identification of straight lines from the captured images by thresholding the image is not a perfect method for noting all visible lines, as some faint lines can be mistaken for noise or background. In addition the accuracy is lower in cases where the contrast between foreground and background is low. By default lines with a slope of zero or infinity are ignored to avoid identifying measurement artifacts as reflections. From over 200 images containing 446 lines (visible and faint lines), the Yen thresholding method is able to extract 303 of these lines, corresponding to an accuracy of 68 %. What makes this method desirable for this scenario is the very low number of false positives (line detection where no reflection is present), which amounts to 3 lines in over 200 images.

Classifier Performance

Three classifiers are compared in this study: the nearest neighbor, random forest and decision tree algorithms. The performance criteria used for the comparison are the precision score, recall score, F1 score, and balanced accuracy score. The precision score for multi-class classification can be defined as the sum of true positives (TP) across all classes, divided by the sum of true positives and false positives (FP) across all classes. This is shown in Eq. 3.

$$Precision = \sum_{c=1}^{C} TP_c / \sum_{c=1}^{C} (TP_c + FP_c) \quad (3)$$

The recall score, on the other hand sums the true positives across all classes, and divides by the sum of true positives and false negatives across all classes, as shown in Eq. 4.

$$Recall = \sum_{c=1}^{C} TP_c / \sum_{c=1}^{C} (TP_c + FN_c) \quad (4)$$

The F1 score is the harmonic mean of both the precision and recall score. The balanced accuracy score in multi-class classification cases is introduced to deal with imbalanced datasets. This is true in this case as some reflections are represented more than others. The balanced accuracy score is the average of recall obtained on each class. Table 1 shows a comparison between these classifiers. Across all performance metrics, the nearest neighbor algorithm was found to perform better than the other two algorithms.

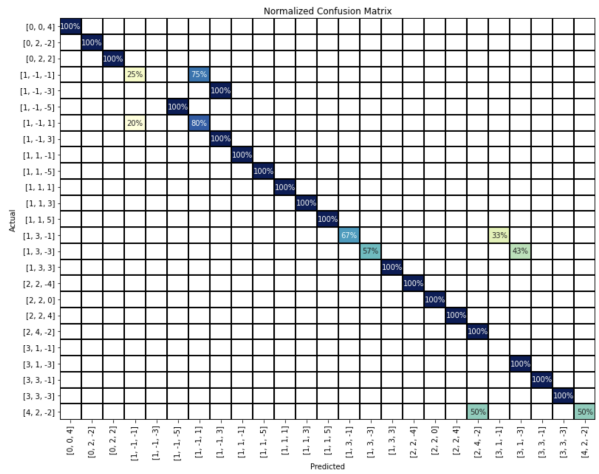

Figure 4: Normalized confusion matrix.

Table 1: ML Classifier Comparison

Classifier	Nearest Neighbor	Random Forest	Decision Tree
Precision score	0.81	0.78	0.74
Recall score	0.83	0.75	0.73
F1 score	0.79	0.74	0.71
Bal. Accuracy	0.87	0.84	0.82

DISCUSSION AND PROSPECTS

In this work, we proposed a method for determining the absolute photon energy based on the identification of crystal reflections. Crystal reflections are first scanned at different pitch angles using the spectrometer, and the image obtained is binarized. With a 68 % extraction rate and a very low false positive rate, lines with a high intensity and good contrast from the background can be detected using the Hough Transform method. A ML classifier then compares a preset measurement model and the extracted lines and identifies the reflections based on different line properties. The confusion matrix in Fig. 4 shows which reflections were successfully identified, with most errors noted in cases where multiple parallel lines occur close to each other (for example classifying [1, -1, -1] as [1, -1, 1]). Improvements in the future can be made in extracting more information from the raw images by trying different image processing techniques.

ACKNOWLEDGEMENTS

The authors would like to thank the colleagues at DESY and European XFEL: N. Kujala, S. Liu, W. Qin, S. Serkez and S. Tomin and for their help in the measurements and fruitful discussions.

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST046

REFERENCES

[1] G. Geloni, "Self-Seeded Free-Electron Lasers", in *Synchrotron Light Sources and Free-Electron Lasers*, E. J. Jaeschke, Ed. Springer International Publishing, 2016, pp. 161–193. doi:10.1007/978-3-319-14394-1_4

[2] G. Geloni, V. Kocharyan, E. Saldin, "A Novel Self-Seeding Scheme for Hard X-ray FELs", *J. Mod. Opt.*, vol. 58, p. 1391, 2011. doi:10.1080/09500340.2011.586473

[3] N. Kujala, *et al*, "Hard x-ray single-shot spectrometer at the European X-ray Free-Electron Laser", *Rev. Sci. Instrum.*, vol. 91, p. 103101, Oct. 2020. doi:10.1063/5.0019935

[4] Tomin, S., Serkez, S., PySpectrometer (pyHIREX), https://github.com/sergey-tomin/pyhirex

[5] J.-C. Yen, F.J. Chang and S. Chang, "A new criterion for auto-matic multilevel thresholding," *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 370-378, 1995. doi:10.1109/83.366472