Fast GPU-based absolute intensity determination for energy-dispersive X-ray Laue diffraction

Fast GPU-based absolute intensity determination for energy-dispersive X-ray Laue diffraction

F. Alghabi,[a,1] **S. Send,**[b] **U. Schipper,**[a] **A. Abboud,**[b] **U. Pietsch**[b] **and A. Kolb**[a]

[a]*Institute for Vision and Graphics, University of Siegen,*
 Hölderlinstr. 3, 57076 Siegen, Germany

[b]*Solid State Physics, University of Siegen,*
 Walter-Flex Str. 3, 57072 Siegen, Germany

 E-mail: farhoosh.alghabi@uni-siegen.de

ABSTRACT: This paper presents a novel method for fast determination of absolute intensities in the sites of Laue spots generated by a tetragonal hen egg-white lysozyme crystal after exposure to white synchrotron radiation during an energy-dispersive X-ray Laue diffraction experiment. The Laue spots are taken by means of an energy-dispersive X-ray 2D pnCCD detector. Current pnCCD detectors have a spatial resolution of 384×384 pixels of size $75 \times 75 \, \mu m^2$ each and operate at a maximum of 400 Hz. Future devices are going to have higher spatial resolution and frame rates.

The proposed method runs on a computer equipped with multiple Graphics Processing Units (GPUs) which provide fast and parallel processing capabilities. Accordingly, our GPU-based algorithm exploits these capabilities to further analyse the Laue spots of the sample. The main contribution of the paper is therefore an alternative algorithm for determining absolute intensities of Laue spots which are themselves computed from a sequence of pnCCD frames. Moreover, a new method for integrating spectral peak intensities and improved background correction, a different way of calculating mean count rate of the background signal and also a new method for n-dimensional Poisson fitting are presented.

We present a comparison of the quality of results from the GPU-based algorithm with the quality of results from a prior (base) algorithm running on CPU. This comparison shows that our algorithm is able to produce results with at least the same quality as the base algorithm. Furthermore, the GPU-based algorithm is able to speed up one of the most time-consuming parts of the base algorithm, which is n-dimensional Poisson fitting, by a factor of more than 3. Also, the entire procedure of extracting Laue spots' positions, energies and absolute intensities from a raw dataset of pnCCD frames is accelerated by a factor of more than 3.

KEYWORDS: Data processing methods; Pattern recognition, cluster finding, calibration and fitting methods; Analysis and statistical methods

---

[1]Corresponding author.

Contents

1 Introduction

The availability of highly brilliant synchrotron radiation allows for precise analyses of crystalline materials by means of X-ray diffraction techniques. However, experimental access to structural parameters is mainly determined by the spectroscopic performance of the used detector system. Nowadays, pixilated area detectors (e.g. CCDs) widely serve to record the diffraction signal of the sample exposed to monochromatic X-rays. Those detectors provide two-dimensional position resolution depending on the pixel size and high dynamic range but insufficient energy resolution. Conventional X-ray structure analysis in macromolecular crystallography or determination of the full strain tensor in material science is then achieved by rotating the crystal around well-defined axes and recording series of diffraction images. Especially in the case of polycrystalline materials experimental difficulties arise as a result of sample rotation since the number of illuminated grains cannot be controlled.

New possibilities for structure analysis are offered by fully depleted pn-junction charge coupled devices (pnCCDs) enabling simultaneous position- and energy-resolved X-ray detection close to the Fano limit of silicon [1, 2]. In this way, energy-dispersive diffraction techniques can be realized based on the use of a pnCCD in combination with white synchrotron radiation to irradiate the sample. In a previous application it was shown that by means of energy-dispersive Laue diffraction both mono- and polycrystalline materials can be quantitatively analyzed in a single-shot exposure of the arbitrarily oriented crystal [3]. The possibility of unit-cell calculation without additional information and the need for sample rotation is a clear advantage of the Laue method compared to monochromatic applications. Moreover, absolute intensity determination allows to

extract kinematical structure-factor moduli of individual position- and energy-resolved Bragg peaks of the crystal from the pnCCD-generated data sets with adequate accuracy [4]. Recently, energy-dispersive Laue diffraction with a pnCCD was considered in the ultra-hard X-ray regime between 40 keV and 140 keV providing a sufficiently large Laue spot density to investigate inorganic crystals in one single measurement [5].

The routine use of pnCCDs for energy-dispersive Laue diffraction with white synchrotron radiation suffers from the generation of large raw data sets. Presently, the University of Siegen is equipped with a pnCCD module comprising 384×384 pixels in the image area with a total area of 28.8×28.8 mm^2. When operated at the largest possible frame rate of about 400 Hz, the amount of generated data is in the order of 112 MB/s. Typical recorded raw data include image sequences of 100000 frames corresponding to a total data volume of up to 27.5 GB. The gain of structural information from a raw data set of this size requires highly efficient data processing procedures in order to prevent long processing times.

During the past few years GPUs have become widespread in processing of large data sets. Areas like physically-based simulation, signal and image processing and even data mining [6] have profited from fast processing capabilities of GPUs for more than a decade. Areas such as medical physics [7, 8], high-energy physics [9–11] and lattice quantum chromodynamics [12, 13] have also witnessed increasing use of GPUs for acceleration of scientific computing. Recently, a GPU data processing scheme was successfully applied to accelerate the extraction of the Laue spots' positions and energies from a pnCCD data set for energy-dispersive Laue diffraction of hen egg-white lysozyme, resulting in a 7 times faster processing time [14–16].

This paper presents a GPU-based algorithm for calculating the absolute intensities (mean photon count rates) of individual Laue spots from a set of pnCCD images generated during an energy-dispersive X-ray Laue diffraction experiment. The GPU-based algorithm is developed using the base algorithm in [17], which runs on CPU, as the main reference and involves some algorithmic innovations. Also some implementation-specific novelties are realized through the use of CUDA which is an interface for programming NVIDIA GPUs [18]. Major algorithmic innovations include a different way of integrating peak intensities and improved background correction, a new method for computing the mean count rate of the background signal (λ_{bg}) and also a new approach for n-dimensional Poisson fitting. Also we quantitatively compared the results from the CPU and GPU-based algorithms using fitting error. The comparison shows that the GPU-based algorithm can produce results of par quality while effectively speeding up one of the most time-consuming parts of the base algorithm by a factor of more than 3. Also, the whole procedure for extraction of Laue spots' positions, energies and absolute intensities from a raw dataset of pnCCD frames is accelerated by a factor of more than 3 when CPU-based algorithms are replaced by GPU-based ones.

The remainder of the paper is organized as follows. Section 2 briefly describes the base algorithm for calculating absolute intensities and highlights the major areas where it is different from the GPU-based algorithm. In section 3 the developed GPU algorithm is explained. Section 4 discusses the results of the GPU algorithm and its comparison with the CPU algorithm and finally section 5 concludes the paper.
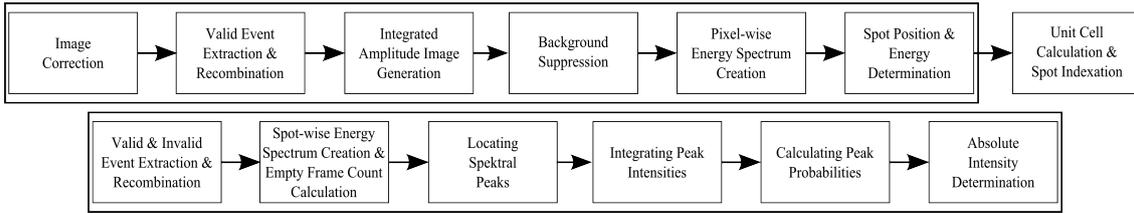
Figure 1. Block diagram of major steps for calculating absolute Laue spot intensities from raw data set of pnCCD images: the steps in the row above are already described and implemented in a prior work (except the last step). Steps in the row below are discussed in this paper.

2 Base algorithm

From a conceptual point of view absolute intensity determination is the most sophisticated step in processing data recorded by the pnCCD in an energy-dispersive Laue diffraction experiment. Due to the fact that the Bragg peaks of a crystal give rise to high local count rates in the plane of detection different photons are accumulated with a high pile-up probability. In the case of a pile-up event the simultaneous position and energy information about individual photons is lost while their energies add up. In general, many photons may be spatially overlapped to create pile-up signals with large integrated energies from which the numbers of involved photons cannot be deduced. However, spectral analyses can be performed in a statistical way using the Poisson distribution of detected photon numbers including background corrections [17]. This data treatment was realized by means of a CPU-based algorithm applied to the measured energy spectrum diffracted by a crystalline material in a well-defined direction when exposed to white X-rays: The energy spectrum consists of sharp peaks at integer multiples of a minimum energy, also called base energy in this paper, associated with the lowest diffraction order in case of which the Miller indices h, k and l are coprime. Obviously, statistical analyses of pile-up events are required for absolute intensity determination whereas the positions and energies of individual Laue spots can be extracted from events generated by single photons.

The base algorithm for calculating absolute intensities comprises the processing steps outlined in the following. These steps form the basis for GPU-based algorithm whose major steps are shown in the bottom row of figure 1. We try to make this relationship clear by mentioning the corresponding step in the GPU-based algorithm in section 3.

1. An energy window is defined. Due to the experimental conditions in this paper it is set to be [8.5 keV, 30 keV]: the lower bound is due to the absorption by filters while the upper bound is due to the limited usable spectral range of bending magnet radiation provided at the storage ring BESSY II (see section 3).

2. The highest peak's order is determined. This is done by computing the greatest common divisor of Miller indices ($gcd(h, k, l)$) (see section 3.3).

3. The base energy is computed. This value is the lowest possible diffracted energy associated with coprime values of Miller indices. Diffracted photons of this energy are either present

($\geq 8.5\,\text{keV}$) or not ($< 8.5\,\text{keV}$). Every observed peak is an integer multiple of the base energy (see section 3.3).

4. Spot spectrum and background spectrum are extracted. Note that the latter is computed from the surroundings of the defined spot area of 5×5 pixels (see section 4.5.2 in [17]). Note that the choice of 5×5 pixels is empirical. The chosen region of interest has to account for beam size and parallax effects, i.e. sample-detector distance. Previous analyses have shown that in case of 3×3 pixels, about 20% of the signal is lost.

5. Gaussian functions are fitted to the highest peak and all observed peaks with multiples of the base energy in the energy spectrum.

6. The fit dimensionality is computed as the number of peaks in the energy window (see section 3.6).

7. In this step the background is subtracted. This is done by cutting the peaks from the spot spectrum and fitting the background spectrum to the spot spectrum over the whole range [8.5 keV, 30 keV]. The result is a fit factor between the background and the spot spectrum which was also applied to peaks at energies higher than 30 keV.

8. Background-corrected peaks (including correction with factor $1 + \lambda_{\text{bg}}$) are integrated to absolute intensity values and then normalized to experimental probabilities $P_{k,\text{exp}}$, $k = 1, 2, \ldots, K$ corresponding to the first, second,. . . and K^{th} observed peaks in the energy spectrum where K is the number of spectral peaks (see section 4.5.3 in [17] and section 3.5).

9. The zero peak probability is calculated from empty frame count (including single background events). This value expresses the probability of having no diffracted photons (see section 3.2 and section 3.5).

10. All possible pile-up combinations to the measured multiples of the base energy are determined (see section 3.6).

11. Finally Poisson fitting is done (see section 3.6).

Note that a Poisson fit is only applied if more than two peaks are found in the spectrum. If we get only one peak of low intensity and, additionally, the empty frame count, a Poisson fit is not necessary (even the empty frame count is not needed then). In such a case we take the absolute intensity directly from the background-corrected peak.

The CPU-based algorithm assumes a square shape of 5×5 pixels for each spot and then extracts information for background spectrum and mean count rate of background from the surroundings of each spot. This approach is not easily adaptable for the GPU-based algorithm due to the non-regular shapes of spots extracted by the GPU algorithm in [14]. Instead we propose a different method for calculation of the mean count rate of the background directly from the spot spectrum. This method is also used for fitting the background in the spot spectrum and consequently for peak integration thus leading to a different method than the CPU algorithm for the same tasks (see section 3.4 and section 3.5). In order to take profit of the parallel processing capability of GPUs, the GPU-based

algorithm exploits a different method for n-dimensional Poisson fitting which is known as genetic algorithm. Genetic algorithms are suitable for parallelization and through this property we gain speed-up advantage over the CPU-based counterpart (see section 3.6).

3 GPU-based absolute intensity determination

The GPU-based algorithm developed for the calculation of absolute intensities consists of 6 major steps (see bottom row in figure 1). These 6 processing steps together with the 7 steps shown in the top row of figure 1 constitute a complete processing chain for calculation of absolute intensities from a raw data set of pnCCD frames: while the steps in the top row are responsible for determining the unit cell and indexation of spots, the 6 steps in the bottom row use their processing results to further compute the absolute intensities. Note that the first 6 steps in the top row are already implemented in the prior GPU-based algorithm in [14].

Before going into details, we give an overview of the algorithm using figure 2. This figure shows the histogram of events, the *energy spectrum*, with various energies occurring in a spot. An event can be defined as a continuous pattern of non-zero pixels. Each event is generated by one or more photons hitting the pnCCD image plane. On a coarse-level these photons can be classified as either diffracted by the crystal or air-scattered. The purpose of the algorithm is to find useful information about diffracted photons. Specifically, the algorithm tries to find the energy and number of diffracted photons, also called absolute intensities, that give rise to the energy spectrum. This is achieved through a detailed analysis of the energy spectrum as described in the following 6 subsections corresponding to the 6 steps of the GPU-based algorithm. Note that all the processing done is performed on a per-spot basis except for the first step, i.e. from step two until the end the spots are processed independently.

3.1 Valid & invalid event extraction & recombination

In this step we extract single events, valid and invalid split events as well as pile-up events from corrected pnCCD images. Single events are those comprised of only one pixel while split events spread over two, three and four pixels resulting in double, triple and quadruple events. Four valid event patterns exist for every type of split event according to the various splitting directions in the pixel array of the pnCCD [19]. In the case of triple events, the measured signal is called "valid" if the highest signal amplitude is located in the corner of the event. For quadruple events, the amplitudes need to be distributed over a square with a size of 2×2 pixels in such a way that the highest and lowest signals are recorded along one diagonal of the event (see figure 3). Pile-up events are those resulting from the spatial and temporal overlap of charge clouds generated by two or more photons. In most cases, pile-up events spread over more than four adjacent pixels or create triple and quadruple events with "forbidden" amplitude distributions. In both situations, the recorded event pattern is denoted as "invalid".

The split events and pile-up events are then recombined into individual photon hits by calculating the center-of-mass coordinates for each event and then rounding them to the nearest integer pixel positions (rounded recombined events). For each recombined event the energy is calculated by accumulating the signal amplitudes of constituent pixels. Due to the experimental conditions, only events with energies above a threshold of 8.5 keV are considered for further analyses. This step
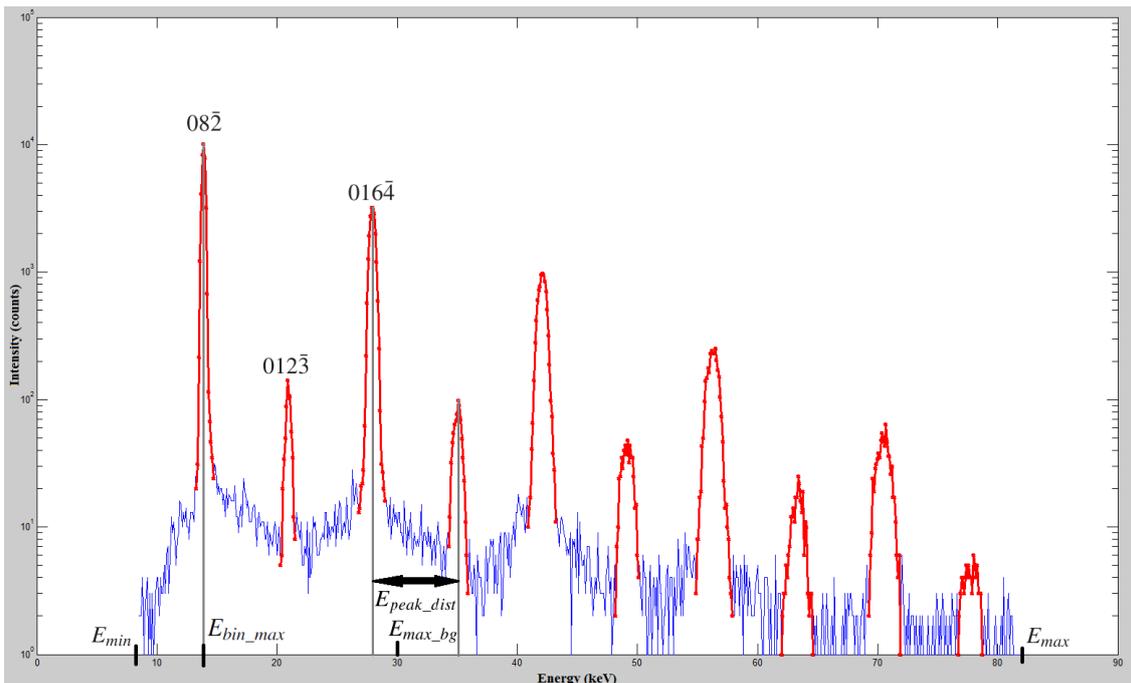
Figure 2. Energy spectrum of a GPU-computed spot: the energy bins are of size 100 eV each. The frequency values (intensities) on the vertical axis are shown in logarithmic scale. The logarithmic scale is chosen so that peaks (red curves) with lower intensities are also recognizable.
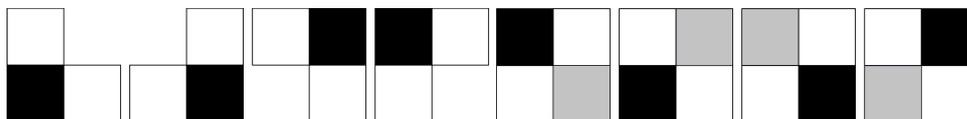
Figure 3. Valid patterns for triple and quadruple events: the black and gray pixels show the highest and lowest signal amplitudes in a pattern, respectively.

is similar to the event extraction & recombination done in section 3.1 of [14] and the only difference is that, here we also consider invalid events. Also see [20] for a discussion on valid patterns for split events and event recombination.

3.2 Spot-wise energy spectrum creation & empty frame count calculation

In this step, as the name suggests, we form an energy spectrum separately for each spot using all extracted events in the last step. To this end for each spot we use all events occurring in any of the spot's pixels to form an accumulated energy spectrum for that spot. This is done by storing in each energy bin of the spectrum the total number of all events whose energies are in the energy range represented by the bin.

Figure 2 shows an instance of an energy spectrum measured at the position of the $08\bar{2}$ reflection of hen egg-white lysozyme and computed by the GPU algorithm. The spectrum consists of several peaks shown by red curves on top of a continuous distribution of background events and pile-up events (blue curve).

Another quantity we calculate in this step is the number of empty frames per spot. If during data acquisition no event occurred in a pnCCD frame at the considered spot position then the frame is said to be empty for that spot. For each spot we count the number of empty frames over the whole set of corrected pnCCD images and report this number as the empty frame count for that spot. We calculate this number for each spot separately and denote it by N_0.

3.3 Locating spectral peaks

The purpose of this step is to find the positions of the spectral peaks. The processing done in this step is guided by the Miller indices of the considered reflection which was extracted after unit-cell calculation. The main idea of the procedure for this step is outlined below:

1. The energy bin with the maximum intensity value, i.e. frequency, is determined. We use the notation $E_{\text{bin\_max}}$ to indicate the energy value represented by this energy bin. Also we use the notations E_{min} and E_{max} to indicate the minimum and maximum energy values represented by the spectrum bins, respectively (see figure 2). Note that these two values are the same for all spots.

2. The greatest common divisor of the given Miller indices ($gcd(h, k, l)$) for the current spot is computed.

3. The distance between two consecutive spectral peaks $E_{\text{peak\_dist}}$ is computed using the following formula:

$$E_{\text{peak\_dist}} = \frac{E_{\text{bin\_max}}}{gcd(h, k, l)}. \tag{3.1}$$

4. The locations of spectral peaks are said to be those energy bins which contain energy values in the following set:

$$\{E_{\text{bin\_max}} + n \cdot E_{\text{peak\_dist}} | n \in \mathbb{Z} \wedge n_{\text{min}} \leq n \leq n_{\text{max}}\}. \tag{3.2}$$

where n_{min} is the smallest integer value such that $E_{\text{bin\_max}} + n_{\text{min}} \cdot E_{\text{peak\_dist}} \geq E_{\text{min}}$ and n_{max} is the largest integer value such that $E_{\text{bin\_max}} + n_{\text{max}} \cdot E_{\text{peak\_dist}} \leq E_{\text{max}}$.

To further clarify the procedure we show the principles for a sample spot spectrum. Consider the spot spectrum shown in figure 2 as an instance. The largest intensity of the spectrum is located in the bin with a representative energy of 13.9 keV ($E_{\text{bin\_max}} = 13.9$ keV) and the spectrum extends from the bin with a representative energy of 8.5 keV ($E_{\text{min}} = 8.5$ keV) to the bin with a representative energy of 81.9 keV ($E_{\text{max}} = 81.9$ keV). Also, during unit cell calculation, the Miller indices assigned to the dominant peak of the spot with this spectrum are $(h, k, l) = (0, 8, \bar{2})$ and thus $gcd(h, k, l) = 2$. Based on this information, we determine $E_{\text{peak\_dist}}$ as 6.95 keV (= 13.9/2). E_{min}, E_{max}, $E_{\text{bin\_max}}$ and $E_{\text{peak\_dist}}$ are also shown in figure 2. Consequently, 10 energy bins containing energy values in the set {$E_1 = 13.9$ keV ($hkl = 08\bar{2}$), $E_2 = 20.85$ keV ($hkl = 012\bar{3}$), $E_3 = 27.8$ keV ($hkl = 016\bar{4}$), 34.75 keV, 41.7 keV, 48.65 keV, 55.6 keV, 62.55 keV, 69.5 keV, 76.45 keV} are determined as peak positions.

Note that the quantity $E_{\text{peak\_dist}}$ corresponds to the energy of the lowest possible diffraction order at the spot's position. Depending on its energy, the lowest possible diffraction order is either

present or not in the spot spectrum. In the example of figure 2 the lowest possible diffraction order is represented by the $04\bar{1}$ reflection with the energy $E_{\text{peak\_dist}} = 6.95\,\text{keV}$. This energy is below $8.5\,\text{keV}$ and is filtered out during the pnCCD image acquisition. Therefore, it is not present in the spot spectrum.

3.4 Integrating peak intensities

In order to compute the absolute intensity for each Bragg peak we need to calculate the peak intensities in the corresponding spectrum. A naive definition would describe the peak intensity as the area below the curve which represents the peak, thereby suggesting the result of integrating the area under the peak as the desired value. A closer look at the spectral peaks, however, reveals that the problem is not that simple. For instance, consider the first peak at $13.9\,\text{keV}$ in the spot spectrum depicted in figure 2. As can be obviously seen, the peak is located on top of another curve (background) with values well exceeding 10 in the vicinity of the peak. Thus, following the naive procedure would simply result in an erroneous value for the peak intensity since it incorporates the intensity of the background into the final result. In order to resolve this problem we need to apply a suitable background correction. The general approach would be to exploit the energy bins in the vicinity of spectral peaks in order to approximate the background intensities within the peak areas. These approximated values are then used in conjunction with original intensities in the peak areas to calculate the background-corrected integrated peak intensities. In the following, the main steps of our specific method for peak intensity integration with background correction are explained:

1. The peak range is defined as all of the energy bins from $B_{\text{center}} - R_{\text{peak}}$ to $B_{\text{center}} + R_{\text{peak}}$ where B_{center} denotes the peak position as determined in section 3.3 and R_{peak} is a non-negative integer number showing the number of used bins on either side belonging to the peak. Thus, we have a total of $2R_{\text{peak}} + 1$ energy bins in the peak.

2. The intensities of all of the $2R_{\text{peak}} + 1$ energy bins are accumulated. The result is denoted by N_{acc}.

3. The background is approximated by a polynomial of degree 2. Intensities at energy bins $[B_{\text{center}} - R_{\text{peak}} - \delta, B_{\text{center}} - R_{\text{peak}} - 1]$ and $[B_{\text{center}} + R_{\text{peak}} + 1, B_{\text{center}} + R_{\text{peak}} + \delta]$, $\delta \geq 2, \delta \in \mathbb{N}$ are used for fitting the polynomial. The polynomial is then evaluated at all energy bins from $B_{\text{center}} - R_{\text{peak}}$ to $B_{\text{center}} + R_{\text{peak}}$, i.e. for all of the energy bins in the peak. The evaluated values are then accumulated and the result is denoted by N_{bg}.

4. Finally the intensity is reported as $N_k = N_{\text{acc}} - N_{\text{bg}}$ for k^{th} peak ($k \geq 1$).

Note that the above procedure serves to compute the intensity of a single peak and it is applied independently to each peak in the spectrum. Also R_{peak} is not the same for all peaks and it increases when we go to higher energy peaks. This is to account for the physical effect that the peaks are broadened with increasing energy. This effect is also observable in figure 2.

3.5 Calculating peak probabilities

The empty frame count and peak intensity values computed in previous steps undergo another step of processing to be prepared for use in the calculation of absolute intensities. In order to correct for

pile-up signals of diffracted photons and background photons we calculate the corrected values N_k' from N_k and to this end we need the mean count rate of the background signal λ_{bg} (see section 4.5.3 in [17]). The procedure for computing λ_{bg} is straightforward: the intensities of all energy bins in the spectrum which do not belong to any peak are accumulated. For those energy bins which belong to a peak the intensities are evaluated from the corresponding fitted polynomial for the background and accumulated. The sum of these numbers gives the background intensity. Finally, the mean background count rate is determined as the result of division of the background intensity by the total number of pnCCD frames N. In order to calculate the peak probabilities we need to know the total number of frames in which no photon diffracted by the crystal was measured (N_0'). Also we need to know the number of frames which contain k photons diffracted by the crystal for $k > 0$ (N_k'). The values $N_k', k \geq 0$ are computed using the following formula (see section 4.5.3 in [17]):

$$N_k' = \begin{cases} N_k + N\lambda_{\text{bg}}, & \text{if } k = 0 \\ N_k(1 + \lambda_{\text{bg}}), & \text{otherwise} \end{cases} \tag{3.3}$$

Having computed N_k', we are able to calculate experimental peak probabilities using the formula:

$$P_{k,\text{exp}} = \frac{N_k'}{\sum\limits_{i=0}^{K} N_i'}, k = 0, 1, \dots, K. \tag{3.4}$$

where K is the number of spectral peaks. Note that the empty frame count (see section 3.2) also leads to an experimental peak probability value. This peak is referred to as 'zero' peak and its probability is denoted by $P_{0,\text{exp}}$. Physically speaking, this peak shows the number of frames in which no photon diffracted by the crystal was measured at the considered spot position.

3.6 Absolute intensity determination

This is the last step towards determination of the absolute intensity. To make the things simpler we explain the processing in this step using the sample spectrum in figure 2. Before starting note that, as already mentioned in section 3.3, the first diffraction order at energy of 6.95 keV, though physically allowed, can't be measured.

The explanation starts with how the spectral peaks are formed. The peak at the energy 13.9 keV is formed by photons with an energy of 13.9 keV. Similarly, the peak at the energy 20.85 keV is formed by photons with an energy of 20.85 keV. The situation for the spectral peak at energy 27.8 keV, however, is different. Although photons with an energy of 27.8 keV can give rise to this peak there is also the possibility of a pile-up event generated by two photons with an energy of 13.9 keV each which contribute to the peak intensity. Thus the peak at the energy 27.8 keV is formed in two different ways.

As can be seen, there can be many different ways of forming spectral peaks especially at higher energies. This, in turn, would increase the complexity of the analysis used to determine the absolute intensities. There is, however, an inherent restriction which effectively reduces this complexity. According to this restriction photons with energies higher than a specific threshold can not be measured due to the low incident photon flux and the reduced sensitivity of the pnCCD at

these energies (also see the sensitivity of the detector as a function of energy in figure 2.15 of [17]). Here, we use the notation $E_{\mathrm{max\_bg}}$ to denote this energy threshold.

In case of the spectrum shown in figure 2, $E_{\mathrm{max\_bg}}$ is set to be equal to 30 keV. Considering the restriction just explained, a similar line of reasoning can well be applied to peaks at higher energies, i.e. a peak at a specific energy can be formed either by photons of that energy if contained in the primary beam spectrum or by any linear combination (with integer coefficients) of different photon energies of peaks below 30 keV. For instance, the peak at energy 34.75 keV can be formed in only one way: by two photons with energies of 13.9 keV and 20.85 keV. As a further example, the peak at energy 41.7 keV can be formed in three different ways: by three photons with an energy of 13.9 keV each, by two photons with an energy of 20.85 keV each or by two photons with energies of 13.9 keV and 27.8 keV. These statements can be symbolically written as follows:

$$
\begin{cases}
13.90\,\mathrm{keV} & \leftarrow 1 * 13.90\,\mathrm{keV} \\
20.85\,\mathrm{keV} & \leftarrow 1 * 20.85\,\mathrm{keV} \\
27.80\,\mathrm{keV} & \leftarrow 2 * 13.90\,\mathrm{keV} \mid 1 * 27.80\,\mathrm{keV} \\
34.75\,\mathrm{keV} & \leftarrow 1 * 13.90\,\mathrm{keV} + 1 * 20.85\,\mathrm{keV} \\
41.70\,\mathrm{keV} & \leftarrow 3 * 13.90\,\mathrm{keV} \mid 2 * 20.85\,\mathrm{keV} \mid 1 * 13.90\,\mathrm{keV} + 1 * 27.80\,\mathrm{keV}
\end{cases}
$$

where energy values on the left and right sides of the arrows represent energies of peaks and energies of photons, respectively. The dimensionless factors on the right side correspond to photon numbers and vertical bars are used to separate different ways of forming spectral peaks. This set of symbolic equations can be translated into a set of algebraic equations involving probability terms as follows:

$$
\begin{cases}
P_{0\,\mathrm{keV}} & = P_0(\lambda_1)P_0(\lambda_2)P_0(\lambda_3) \\
P_{13.90\,\mathrm{keV}} & = P_1(\lambda_1)P_0(\lambda_2)P_0(\lambda_3) \\
P_{20.85\,\mathrm{keV}} & = P_0(\lambda_1)P_1(\lambda_2)P_0(\lambda_3) \\
P_{27.80\,\mathrm{keV}} & = P_2(\lambda_1)P_0(\lambda_2)P_0(\lambda_3) + P_0(\lambda_1)P_0(\lambda_2)P_1(\lambda_3) \\
P_{34.75\,\mathrm{keV}} & = P_1(\lambda_1)P_1(\lambda_2)P_0(\lambda_3) \\
P_{41.70\,\mathrm{keV}} & = P_3(\lambda_1)P_0(\lambda_2)P_0(\lambda_3) + P_0(\lambda_1)P_2(\lambda_2)P_0(\lambda_3) + P_1(\lambda_1)P_0(\lambda_2)P_1(\lambda_3)
\end{cases}
\tag{3.5}
$$

where $P_{x\,\mathrm{keV}}$ represents the probability of the occurrence of an event with an accumulated energy of x keV, which are the same as the experimental peak probability values computed by eq. (3.4) and $P_m(\lambda_i)$ is the probability of simultaneous occurrence of m photons of the same energy with a mean count rate of λ_i. There are three λ_i values ($i \in \{1, 2, 3\}$) corresponding to three possible photon energies which generate the spectral peaks ($\lambda_1 \leftrightarrow 13.90$ keV, $\lambda_2 \leftrightarrow 20.85$ keV and $\lambda_3 \leftrightarrow 27.80$ keV). Note that there is one equation at the beginning of this set which does not exist in the set of symbolic equations. This equation expresses the probability $P_{0\,\mathrm{keV}}$ of the simultaneous occurrence of no photons with the mean count rates λ_1, λ_2 and λ_3. Actually, as expressed in the equation, this peak is formed when neither of these three photons occur in a pnCCD frame at the considered spot position, thereby giving the name 'zero' to this peak.

The $P_m(\lambda_i)$ terms in eq. (3.5) are given by the Poisson distribution of detected photon numbers:

$$
P_m(\lambda_i) = \frac{\lambda_i^m e^{-\lambda_i}}{m!}.
\tag{3.6}
$$

Expanding $P_m(\lambda_i)$ terms in eq. (3.5) using eq. (3.6) results in a set of equations in which the only unknowns are the mean photon count rates, i.e. λ_i, $i = 1, 2, 3$. In order to find the values of these unknowns we need to solve this set of equations. Our solution to this non-linear problem is the application of a class of algorithms known as genetic algorithms [21]. Genetic algorithms try to find solutions to problems such as optimization by mimicking the process of natural selection.

To solve the non-linear equation system shown in eq. (3.5) using genetic algorithms we interpret the right side of each equation as a function whose value is defined in terms of a number of unknown parameters λ_i. Then the goal is defined as fitting these functions collectively to the values $P_{k,\exp}$ on the left side of the equations by finding proper values for the unknown parameters. This newly-interpreted problem, which we refer to as n-dimensional Poisson fitting, can then be solved through the definition of an objective function. The fitting error can well serve this purpose and be used as an objective function whose value should be optimized (minimized). We have chosen the least squares error as the fitting error whose value is computed as follows:

$$f(\lambda_1, \lambda_2, \lambda_3) = \sum_{k=0}^{K} (P_{k,\exp} - P_{k,\mathrm{fit}})^2. \tag{3.7}$$

where K is the total number of spectral peaks as determined in section 3.3 and $P_{k,\mathrm{fit}}$ is the fitted probability for the kth peak.

4 Results

In order to compare the CPU and GPU-based algorithms we need a ground truth. Unfortunately, the problem of absolute intensity determination for energy-dispersive X-ray Laue diffraction suffers from the lack of such a ground truth. Instead, we use an alternative quality measure: as already mentioned, the fitting error provides the fitting procedure (genetic algorithm) with useful information about the quality of the solutions. In addition to this, the fitting error can also well serve to judge the quality of the overall procedure (i.e. the CPU and GPU-based algorithms). Specifically, the fitting error gives a measure of how well the following two objectives are realized:

1. Generation of experimental peak probabilities (data points) which can well be approximated by an n-dimensional Poisson distribution (model).

2. Computing λ_i value(s) for the n-dimensional Poisson distribution.

Apparently, the first objective corresponds to all the processing done prior to the fitting whereas the second objective corresponds to the fitting procedure. Consequently, the fitting error can also be used as a means of verification for the whole procedure.

Following the discussion above, we have applied the procedure described in section 3 to a number of spots and analysed the results using the fitting error. The spots we have used in this experiment are those computed by the GPU algorithm in [14] from a raw data set of 100000 pnCCD frames ($N = 100000$) generated during an energy-dispersive X-ray Laue diffraction experiment. The experiment is carried out at room temperature with no cooling and the pnCCD image plane is perpendicular to the incident beam. The crystal sample is approximately of size $0.5 \times 0.5 \times 0.2\,\mathrm{mm}^3$

Table 1. Rate of success for computing the absolute intensities for 284 spots using the GPU algorithm and 319 spots using the CPU algorithm with various thresholds for the average fitting error.

Average Fitting Error	GPU Method	CPU Method
$\leq 10^{-2}$	284/284 (100%)	319/319 (100%)
$\leq 10^{-3}$	282/284 (99.30%)	319/319 (100%)
$\leq 10^{-4}$	275/284 (96.83%)	318/319 (99.69%)
$\leq 10^{-5}$	274/284 (96.48%)	318/319 (99.69%)
$\leq 10^{-6}$	268/284 (94.37%)	316/319 (99.06%)
$\leq 10^{-7}$	253/284 (89.08%)	304/319 (95.30%)
$\leq 10^{-8}$	230/284 (80.99%)	241/319 (75.55%)
$\leq 10^{-9}$	169/284 (59.51%)	158/319 (49.53%)

and randomly oriented. The cycle time is 11.5 ms and the exposure time is 1150 seconds (= 100000 * 11.5 ms). The used X-ray beam flux is about $5 * 10^8$ photons per second and the total amount of radiation to which the crystal is exposed is about $6 * 10^{11}$ photons. Since the procedure relies on the knowledge of Miller indices (see section 3.3) we have used only a subset of those spots which are correctly indexed with adequate accuracy during unit cell calculation and spot indexation (see figure 1). The subset we have chosen comprises 284 spots which were indexed by a tolerance of 20% (see table 2 in [14]). Also, the parameters E_{min}, E_{max}, E_{max_bg} and δ in section 3 are set to 8.5 keV, 81.9 keV, 30 keV and 2 for all the spots, respectively. Table 1 provides detailed statistics of the fitting errors calculated for each of the spots. Since the number of spectral peaks in each spot and thus the number of data points in each fitting is different, the table reports the average fitting error in order to provide a comparable quantity. The same statistics are also shown for a set of 319 reflections where the base algorithm running on CPU has been applied to the same set of pnCCD images. As can be seen the quality of results computed by the GPU algorithm is well comparable to the quality of results computed by CPU algorithm in terms of the average fitting error. For the specific spot whose spectrum is shown in figure 2, the GPU-computed results are $I_1 = N \cdot \lambda_1 = 79623$ ($\lambda_1 = 0.79623$), $I_2 = N \cdot \lambda_2 = 501$ ($\lambda_2 = 0.00501$) and $I_3 = N \cdot \lambda_3 = 27$ ($\lambda_3 = 0.00027$) and the average fitting error is $1.0 * 10^{-4}$. Also figure 4 shows the fitted probabilities when overlaid on the measured peak probabilities for this spot. As can be seen, I_3 is very small and effectively negligible. Hence, one would not use this value for structure refinement based on experimental structure-factor moduli due to its large statistical uncertainty.

The performance evaluation shows that the GPU algorithm spends most of the time for two tasks while the time spent for the rest of processing is comparably negligible. The first task is going through the whole set of pnCCD images in order to extract the individual events and count the number of empty frames for each spot. This task takes less than 4 minutes for 100000 pnCCD frames while running on 4 Tesla C2050 GPUs. The other time-consuming task is the n-dimensional Poisson fitting. This task takes approximately 45 seconds when we compute the absolute intensities for the above set of 284 spots using one Tesla C2050 GPU. The CPU algorithm, on the other hand, spends approximately 0.5 seconds for n-dimensional Poisson fitting of each spot, i.e. almost 142 seconds for the same set of 284 spots.
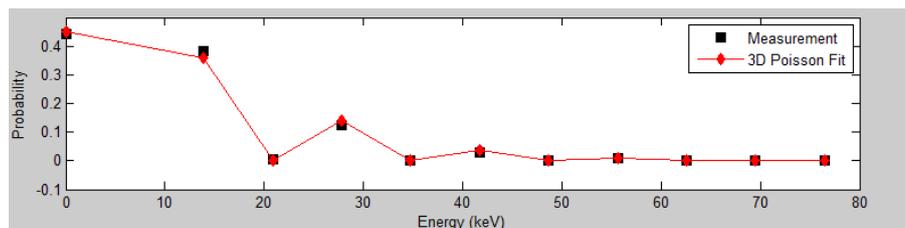
Figure 4. Result of the 3D Poisson fit for a GPU-computed Laue spot with the dominating Bragg peak reflection $08\bar{2}$: the fitted probabilities are computed using $\lambda_1 = 0.79623$, $\lambda_2 = 0.00501$ and $\lambda_3 = 0.00027$ (red curve) and are compared with the measured probabilities (black squares) for 11 peaks (including the 'zero' peak).

5 Conclusion

In this paper we presented a GPU-based algorithm for calculating the absolute intensities of individual spatially overlapping Bragg peaks generated by a hen egg-white lysozyme crystal in an energy-dispersive X-ray Laue diffraction experiment. The presented algorithm can be considered as a further development of the prior work in [14]. The GPU-algorithm was able to compute the absolute intensities with a quality comparable to that of the CPU algorithm while effectively speeding up the n-dimensional Poisson fitting by a factor of more than 3 (using one Tesla C2050 GPU). In total, the GPU algorithms in [14] and this paper are able to speed-up major parts of the processing done by the CPU-based algorithms for determining the Laue spots' positions, energies and absolute intensities from a raw dataset of pnCCD frames by a factor of more than 3. Note that a potential possibility for increasing this factor is to extract both valid and invalid events only at one place instead of two (as shown in figure 1). This is possible following the same processing for both valid and invalid events (see section 3.1). This can well be considered as an idea for future work to further improve the GPU-based algorithm.

Acknowledgments

References

[1] L. Strüder et al., *The European photon imaging camera on XMM-Newton: the pn-CCD camera*, *Astron. Astrophys.* **365** (2001) L18.

[2] N. Meidinger et al., *PNCCD for photon detection from near infrared to X-rays*, *Nucl. Instrum. Meth.* **A 565** (2006) 251.

[3] S. Send et al., *Analysis of polycrystallinity in hen egg-white lysozyme using a pnCCD*, *J. Appl. Cryst.* **45** (2012) 517.

[4] S. Send et al., *Energy-dispersive Laue diffraction by means of a frame-store pnCCD*, *J. Appl. Cryst.* **42** (2009) 1139.

[5] S. Send et al., *Application of a pnCCD for energy-dispersive Laue diffraction with ultra-hard X-rays*, *J. Appl. Cryst.* **48** (2015), submitted.

[6] D. Luebke et al., *GPGPU: general-purpose computation on graphics hardware*, *Proceedings of the 2006 ACM/IEEE conference on Supercomputing*, Tampa U.S.A. (2006), pg. 208.

[7] X. Jia et al., *GPU-based high-performance computing for radiation therapy*, *Phys. Med. Biol.* **59** (2014) R151.

[8] G. Pratx, L. Xing, *GPU computing in medical physics: A review*, *Med. Phys.* **38** (2011) 2685.

[9] J. Kanzaki, *Application of graphics processing unit (GPU) to software in elementary particle / high energy physics field*, *Proc. Comput. Sci.* **4** (2011) 869.

[10] G. Lamanna et al., *GPUs for fast triggering and pattern matching at the CERN experiment NA62*, *Nucl. Instrum. Meth.* **A 628** (2011) 457.

[11] G. Collazuol et al., *Fast online triggering in high-energy physics experiments using GPUs*, *Nucl. Instrum. Meth.* **A 662** (2012) 49.

[12] M.A. Clark et al., *Solving Lattice QCD systems of equations using mixed precision solvers on GPUs*, *Comput. Phys. Commun.* **181** (2010) 1517.

[13] R. Babich et al., *Scaling lattice QCD beyond 100 GPUs*, *Proceedings of International Conference for High Performance Computing, Networking, Storage and Analysis*, Seattle U.S.A. (2011), article No. 70.

[14] F. Alghabi et al., *Fast GPU-based spot extraction for energy-dispersive X-ray Laue diffraction*, 2014 *JINST* **9** T11003.

[15] F. Alghabi et al., *A scalable software framework for stateful stream data processing on multiple GPUs and applications*, in *Proceedings of Symposium on GPU Programming and Applications, LNCS*, Singapore (2013), Springer, Berlin Germany (2015).

[16] F. Alghabi et al., *Real-Time processing of pnCCD images using GPUs*, *14th International workshop on radiation imaging detectors*, Figueira da Foz Portugal (2012), page. 178.

[17] S. Send, *Utilization of a frame store pnCCD for energy-dispersive Laue diffraction with white synchrotron radiation*, Ph.D. Thesis, University of Siegen, Siegen Germany (2013).

[18] NVIDIA Corporation, *NVIDIA CUDA C programming Guide*, *Version 4.0* (2011).

[19] A. Abboud et al., *Sub-pixel resolution of a pnCCD for X-ray white beam applications*, 2013 *JINST* **8** P05005.

[20] R. Andritschke et al., *Data Analysis for Characterizing pnCCDs*, *IEEE Nucl. Sci. Symp. Conf. Rec.* **2008** (2008) 2166.

[21] A.E. Eiben et al., *Introduction to Evolutionary Computing*, Springer, Heidelberg Germany (2008).