Computational Techniques for Efficient Conversion of Image Files from Area Detectors

Taha Sochi[*]

2011

[*]University College London, Department of Physics & Astronomy, Gower Street, London, WC1E 6BT. Email: t.sochi@ucl.ac.uk.

Contents

List of Figures

1 Abstract

Area detectors are used in many scientific and technological applications such as particle and radiation physics. Thanks to the recent technological developments, the radiation sources are becoming increasingly brighter and the detectors become faster and more efficient. The result is a sharp increase in the size of data collected in a typical experiment. This situation imposes a bottleneck on data processing capabilities, and could pose a real challenge to scientific research in certain areas. This article proposes a number of simple techniques to facilitate rapid and efficient extraction of data obtained from these detectors. These techniques are successfully implemented and tested in a computer program to deal with the extraction of X-ray diffraction patterns from EDF image files obtained from CCD detectors.

Keywords: area detector; computational techniques; image processing; data extraction algorithms; CCD; EDF.

2 Introduction

In the recent years, the technology of detectors and data acquisition systems has witnessed a huge revolution. This, accompanied by the wide availability of intense radiation sources such as synchrotron radiation beams, contributed to the huge increase in the volume of data obtained in a typical experiment. It is not unusual these days to collect hundreds of thousands of experimental data files occupying several tera bytes of magnetic storage from a number of correlated measurements within just a few days. This situation necessitates the development of new computational algorithms and strategies to process and analyze such massive data sets. The current article presents a number of simple techniques that were developed and used recently by the author to deal with the processing of huge quantities of EDF (which stands for European Data Format) binary image files obtained from charge-coupled devices (CCD) on synchrotron X-ray beamlines to extract numeric data in the form of 1D diffraction patterns. These techniques are simple and general and hence can be easily implemented and used by the interested scientists as a substitute for commercial and free software that rely on more sophisticated but slower algorithms. In the following we describe these techniques in the context of data extraction from binary image files of EDF format obtained from charge-coupled devices, although they can be equally applied to other data formats obtained by other types of detector.

3 EDF Image Processing

Charge-coupled devices consist of an array of light-sensitive solid-state cells that convert photons into quantified electric charges. These charges measure the intensity of the photon source in terms of energy and number of counts. The 2D spatial distribution of these cells produce a 2D image of the source object. Hence, an

image of an object detected by a CCD device consists of a 2D matrix of the same dimensions as the CCD array where each entry in the matrix indicates the intensity of radiation at the corresponding cell of the CCD array. Charge-coupled devices are used in many scientific and technological applications such as astronomy and X-ray imaging. Because CCDs are efficient area detectors, they can reduce the acquisition time substantially with improved resolution. A typical CCD used for X-ray imaging on a synchrotron beamline consists of an array of more than five million cells (2640×1920). Figure 1 is a simple demonstration of the setting of a charge-coupled device in an X-ray diffraction experiment.

The purpose of EDF image processing is to convert binary image files obtained from CCD detectors to ASCII numeric format. While the binary image of an EDF file consists of a 2D rectangular matrix where the intensity of the diffracted radiation at each cell is given as a function of implicit xy coordinates of the pixel in the matrix, the extracted ASCII numeric data represent a 1D diffraction pattern of total intensity as a function of scattering angle. The radial dependence of the concentric rings in the 2D pattern is correlated to the scattering angle in the 1D pattern by a simple geometric relation. An image of the data contained in a typical
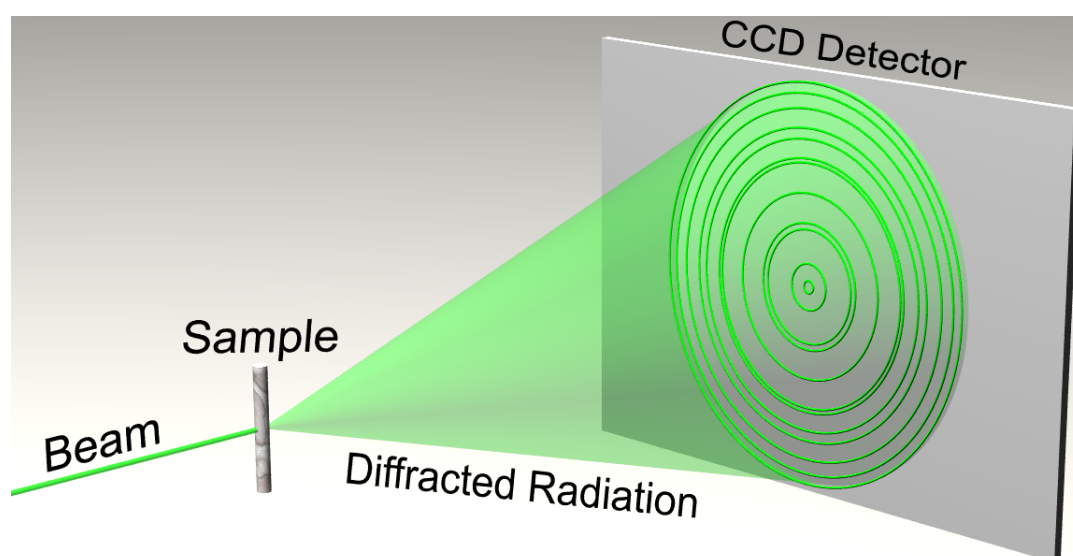

Figure 1: The setting of a charge-coupled device in an X-ray diffraction experiment.

EDF file is displayed in Figure 2 while a sample of the extracted 1D pattern is shown in Figure 3. Each EDF file contains, beside the binary data of the rectangular matrix, an ASCII header which normally consists of 24 lines of text. This header includes, among other things, the endian bit type (little or big), the data type (e.g. unsigned short or long integer), the horizontal and vertical dimensions of the image matrix in pixels, the size of the data file in bytes, and the date and time of data acquisition. In the following sections we outline the main steps of data extraction of these files.

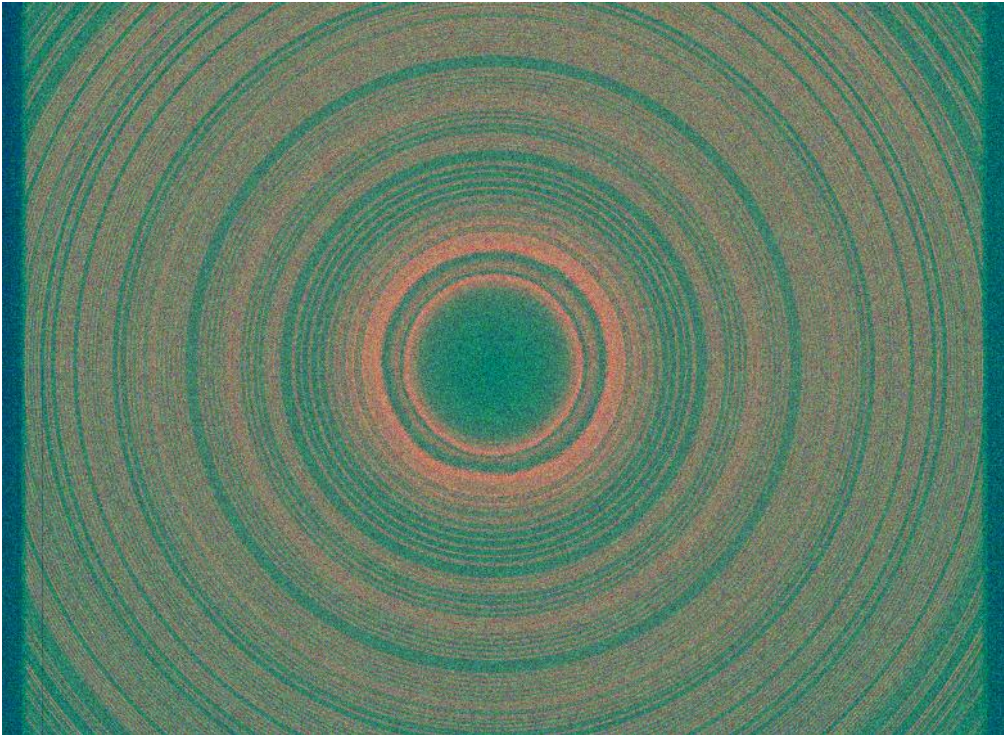Figure 2: Image of a 2D diffraction pattern contained in a typical EDF binary file. The 1D pattern of this image is shown in Figure 3.

3.1   Radial Data Vector

To convert the rings of pixel values in the rectangular data matrix to a 1D pattern, a radial 1D vector is used to store the cumulative intensity of each ring. Each cell in the rectangular data grid is assigned a certain radius according to its distance

from the image center taking into account the tilt in the horizontal and vertical orientations as will be discussed in section 7. As the pixel values are read, they are assigned to the radial cells in the 1D vector immediately. This ensures rapid processing with minimal computing resources since no memory space is required to store the data in an intermediate processing stage. The radial dependence is computed from the pixel implicit coordinates in the rectangular matrix when the routine is run in single mode, while it is obtained in a more efficient way by using lookup table when the routine is executed in a multi batch mode as will be outlined later.

Because the pixel can be too coarse as a unit of length and as a unit of intensity storage due to its finite measurable size resulting in uneven distribution of intensity in the data points of the 1D pattern, the pixels of the CCD grid can be split to take into account these two factors. This splitting takes two forms:

1. By taking the radial unit length as a fraction of a pixel so that more than one ring can fit within one unit pixel of radial distance. The radial assignment of a pixel then depends on its radial distance from the image center as a float quantity rather than as an integer quantity.

2. By splitting the intensity of the pixel according to the area contained within the closest ring so each ring represents a strip with finite width in the radial direction. To do this, each pixel is divided into a grid of small squares (e.g. 10×10). The distance between the center of each small square in this grid and the center of the image is then computed and allocated to the nearest ring. The number of allocated points to a particular ring is then divided by the total number of points in the pixel (i.e. 100 for a 10×10 grid) and the fraction of the total intensity of that pixel is added to the corresponding ring.

3.2   Tilt Correction

Because the CCD device may not be perpendicular to the primary beam line direction, the diffracted ray can be deflected to a higher or lower radial distance resulting in errors in the scattering angle. Therefore, the tilt in the vertical and horizontal orientations requires correction to obtain accurate diffraction patterns. The derivation of this correction is presented in section 7. To ensure rapid data processing and minimal computing resources, this correction is computed only once from a representative image in any single run regardless of the number of files processed in that run.

3.3   Missing-Rings Correction

Because charge-coupled devices have a rectangular shape, the rings whose radius exceeds a certain limit, by having a radius greater than the distance between the ring center and one or more of the rectangle sides, will not be complete (refer to the rings near the corners in Figure 2) and therefore the pattern will have reduced intensity at high scattering angles. To compensate for the loss of intensity for the incomplete diffraction rings, a simple and time saving technique is used. This technique scales the incomplete rings to the continuum value at that radius by multiplying the total intensity of incomplete rings by the ratio of the continuum circumference at that radius to the actual number of pixels in that ring. This number is obtained once in any single operation while serially reading the pixel values. These scale ratios are stored in a 1D double vector to be used at the end of each application of the extraction routine on an individual binary data file when processing multiple files.

Figure 4 displays an example of the actual (discrete) number of pixels of the image rings as a function of radius in pixels alongside the continuum value of $2\pi r$ represented by the straight line. As can be seen, the two curves match very well for

the complete rings. This indicates that the continuum is a very good approximation apart from the rings that are too close to the image corners. The meaning of the two cusps is obvious as the radius of the rings increases and exceeds the rectangle sides in one direction (i.e. vertical or horizontal) and then in the other direction. Figure 3 presents a sample diffraction pattern obtained from an EDF image with and without the application of missing-rings correction.
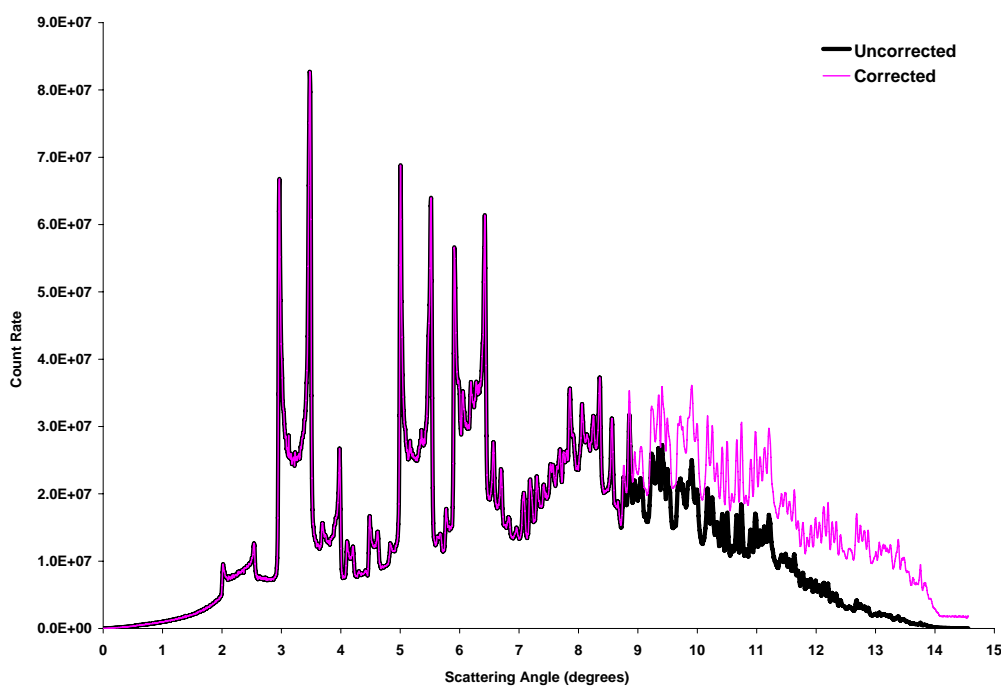
Figure 3: Example of a 1D diffraction pattern extracted from a 2D EDF pattern. The pink curve is obtained with the application of missing-rings correction while the black is obtained without this correction.

3.4 Multi Batch Operation

One of the strategies that we employ in EDF image extraction routine is multi batch processing where the routine is applied in a single run on a data directory that includes multiple data sets each stored within a specific folder. This multi batch processing is applied in an organized fashion and the resulted data are archived systematically using meaningful naming to facilitate managing and post precessing.
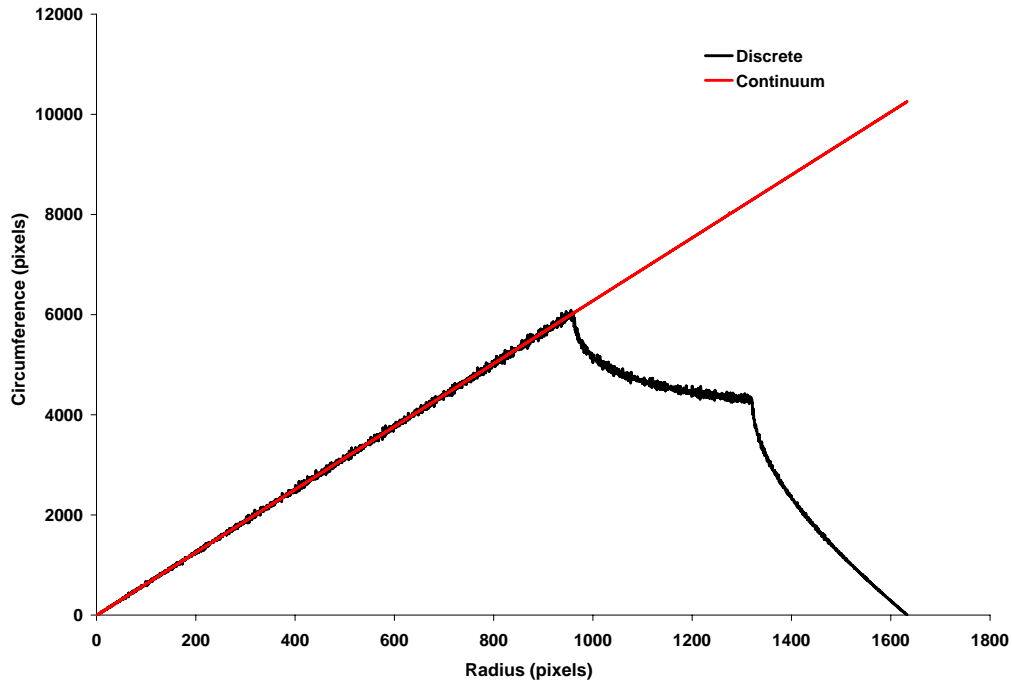
Figure 4: The continuum approximation alongside the discrete value for the number of pixels in a ring.

As a time-saving measure, when EDF data extraction routine is run in a multi batch mode the header is read only once from a representative image file to obtain the required information. Another time saving measure used in multi batch operations is the use of lookup matrix for radial assignment of the individual pixels. These assignments are computed only once at the start of operation and stored in a 2D vector for the use in the subsequent data extraction operation without repeating these lengthy calculations. The radial assignment of pixels include tilt correction in the horizontal and vertical directions. In multi batch mode the scale ratios which are required for missing-rings correction are also computed once at the start of operation and stored in a 1D lookup vector as indicated already. All these measures ensure rapid data processing and considerable save in computational resources.

4   Case Study

The data extraction techniques which we described have been implemented in a rapid-analysis software called EasyDD [2]. EasyDD has been used in a number of studies (e.g. [1, 3, 6, 7]), and is in use by the High Energy X-Ray Imaging Technology (HEXITEC) project [5]. The EDF extraction algorithm requires as an input the CCD image center, the detector tilt and the physical dimensions of the detector system. The user has the opportunity to select the type of the extraction operation by determining the radial size of the extracted pattern, the application of missing-rings correction and the radial split factor.

In a recent study, EasyDD has been used for extracting and processing data obtained from the European Synchrotron Radiation Facility (ESRF). The measurements were carried out at the beamline ID15B which is dedicated to applications that require very high energy X-ray radiation up to several hundreds of keV [4]. The data, which were collected over a few days, consist of about 254 thousand EDF image files in 179 data sets with total size of about 2.45 tera bytes. EasyDD was used to extract the data and convert these images to 1D spectral patterns in ASCII numeric format. This was performed, using multi-batch mode, in about 36 hours of CPU time on an ordinary desktop computer. This time, when compared to an estimated 2 months on a rival commercial software, reveals the efficiency of our EDF extraction techniques and the crucial role that they can play in rapid processing of huge data sets. A sample of these data sets is presented in Figure 5 for one of the diffraction peaks.

Finally, it should be remarked that despite the fact that the data conversion algorithm in its current state is for numeric conversion of EDF files, it can be easily extended to other data types with similar structure. The algorithm can also convert the EDF binary images to 2D visual images in a number of formats (png, jpg and bmp). An example of these images is given in Figure 2. This type

of operation is approximately as fast as numeric conversion to 1D patterns. The algorithm can also perform fidelity check to find and compensate for the missing files in an integrated data set consisting of files with regular naming pattern.
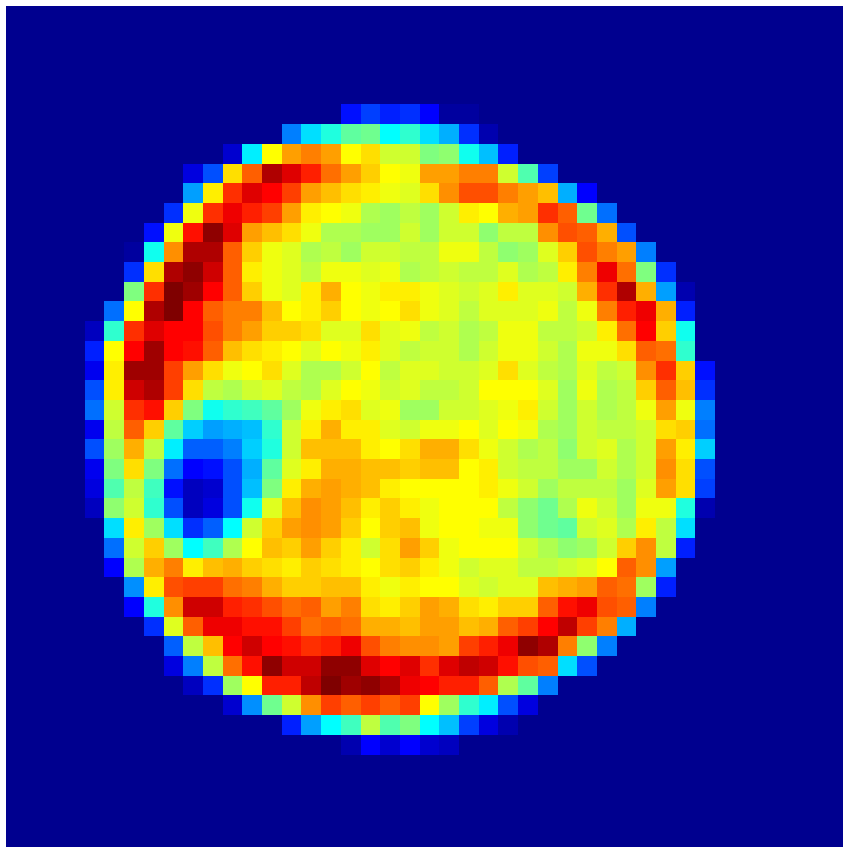

Figure 5: A tomographic image of a diffraction peak obtained from a nickel compound on a cylindrical alumina extrudate sample. The numeric data are obtained from EasyDD using EDF extraction, back projection and curve-fitting routines.

5   Conclusion

The recent developments in the detectors technology and data acquisition techniques coupled with the readily-available bright radiation sources, such as high-energy synchrotrons, have led to the situation where the accumulation of massive data sets became commonplace in many scientific disciplines. This requires the adoption of new data management strategies and the development of highly effi-

cient computational techniques to overcome this bottleneck and make processing and analyzing these massive data possible within reasonable time frames. The development of the image conversion techniques presented in this article and implemented in the rapid analysis software 'EasyDD' comes as a response to this need. These computational techniques have been tested by several researchers working in a number of research institutions and proved to be efficient and reliable. A vivid manifestation of the success of these techniques can be seen for example in [7] where several large and highly-complex chemical and crystallographic data sets, obtained from a number of alumina-based molybdenum and nickel catalysis systems, have been rapidly analyzed by these techniques and the fine details of the systems were monitored during dynamic phase transformation. The high quality of the results is a signature for the quality of the underlying computational techniques that are employed in these studies.

6   References

[1] Cernik R.J., Hansson C.C.T, Martin C.M., Preuss M., Attallah M., Korsunsky A.M., Belnoue J.P., Jun T.S., Barnes P., Jacques S., Sochi T. and Lazzari O. (2011) A synchrotron tomographic energy-dispersive diffraction imaging study of the aerospace alloy Ti 6246. Journal of Applied Crystallography 44(1): 150-157. 12

[2] EasyDD website: www.scienceware.net/id3.html. 12

[3] Espinosa-Alonso L., O'Brien M.G., Jacques S.D., Beale A.M., de Jong K.P., Barnes P. and and Weckhuysen B.M. (2009) Tomographic energy dispersive diffraction imaging to study the genesis of Ni nanoparticles in 3D within γ-Al_2O_3 catalyst bodies. Journal of the American Chemical Society 131(46): 16932-16938. 12

[4] European Synchrotron Radiation Facility (ESRF) website: www.esrf.eu/. 12

[5] High Energy X-Ray Imaging Technology website: www.hexitec.co.uk/index.htm. 12

[6] Lazzari O., Jacques S., Sochi T. and Barnes P. (2009) Reconstructive colour X-ray diffraction imaging - a novel TEDDI imaging method. Analyst, 134(9): 1802-1807. 12

[7] Sochi T. (2010) High Throughput Software for Powder Diffraction and its Application to Heterogeneous Catalysis. PhD thesis, Birkbeck College London. 12, 14

7   Appendix: Tilt Correction Derivation

In this derivation we use a standard Cartesian coordinate system in a plane perpendicular to the original beam direction that passes through the center of the image and hence the origin of the coordinate system. For simplicity we use a unit circle centered at the origin and lying in this plane. In the following derivation we have:

ϕ is the angle of an arbitrary point on the unit circle $(0 \leq \phi < 2\pi)$

θ_x is the tilt in the x direction

θ_y is the tilt in the y direction

θ is the actual tilt at an arbitrary point on the tilted CCD image. It is obvious that θ is a function of ϕ and can have negative as well as positive values.

From a simple geometric argument, the actual tilt as a function of ϕ is given by

$$\theta(\phi) = \arctan[\cos \phi \times \tan \theta_x + \sin \phi \times \tan \theta_y] \tag{1}$$

As can be seen, this formula retrieves the correct (and obvious) tilt in the x, y, $-x$ and $-y$ directions.

Now referring to Figure 6, which is in the plane identified by the original beam and ray directions, and on applying the cosine rule on the triangle ABC we have

$$c = \sqrt{a^2 + b^2 - 2ab \cos \left(\theta + \frac{\pi}{2} \right)} \tag{2}$$

Using the sine rule we obtain the scattering angle

$$\psi = \arcsin \left[b \times \frac{\sin(\theta + \pi/2)}{c} \right] \tag{3}$$

And finally the required quantity, d, can be obtained from

$$d = a \times \tan \psi \tag{4}$$

It is obvious that there are other methods for finding d. Although this derivation is demonstrated for $\theta > 0$, as seen in Figure 6, it is general and valid for $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$.
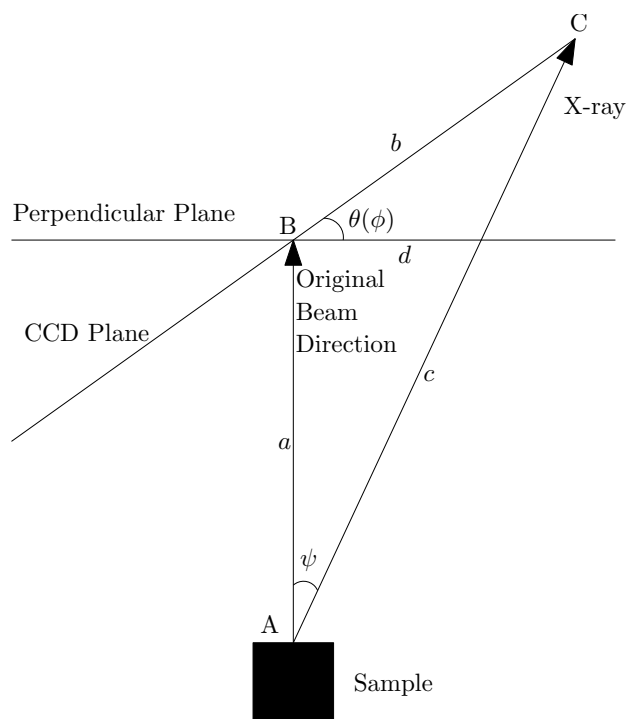

Figure 6: A schematic diagram for demonstrating the derivation of tilt correction.