**Diffuse Multiple Scattering**

Standard X-ray scattering from a single crystal produces diffraction spots, characteristic of the crystal structure which is scattering the X-ray. This process is displayed in Figure 1. From this pattern information about the structure of the crystal and the lattice parameters of the crystal can be determined. One of the most important pieces of information that can be obtained is the structure type of the crystal and at a finer level of detail the space group. The possible Bravais lattices are shown in Figure 2. One serious drawback of regular X-ray diffraction in samples where more than one crystal structure is present is that the method only gives information for the area directly under the beam and therefore generally only refers to one phase per measurement.

A picture containing object, clock, table, room

Description automatically generated

Figure 1X-ray scattering process producing a diffraction pattern

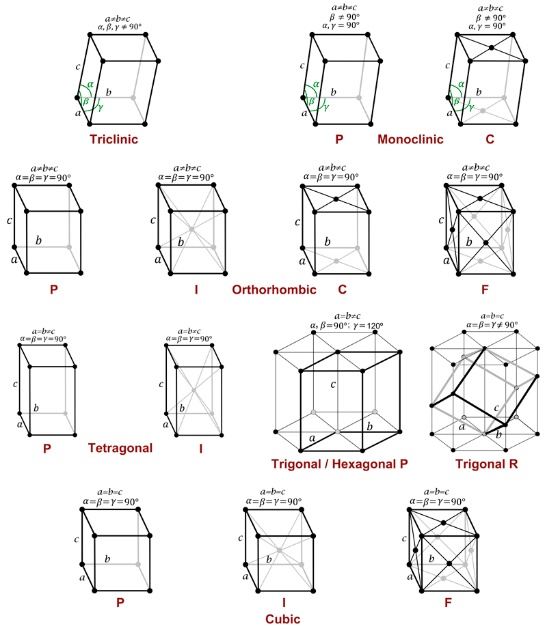
Diffuse multiple scattering takes place when scattered waves interact with one another. The waves can interact constructively or destructively. Depending on the source of the second wave, if the first scattered wave interacts with it, the result can produce lines of increased intensity in the diffraction patterns. These lines are far fainter than diffraction peaks and have only become visible with the advent of high intensity light sources and very sensitive detectors. However these lines can contain information about much larger regions of the sample than single diffraction peaks and so can be used to investigate sample that have more than one crystal structure *in a single measurement.* The lines are also exquisitely sensitive to the lattice parameters of the crystal and can be used to infer information about strain in the crystal in much greater detail than possible with regular diffraction. An example of the lines produced in diffuse multiple scattering experiments are shown in Figure 3.

Figure 2 The 14 possible Bravais lattices for crystal structures

DMS patterns are usually solved by comparing the experimental pattern to simulated patterns and minimising the difference between the two, by altering the simulation parameters. However, to have a starting point for the minimisation it is necessary to have a lot of information about the crystal structure(s) in the sample. One must also manually work out information about orientations by finding points of exact intersection of three lines (triple intersections). All of this means that the analysis of DMS patterns although extremely powerful is only possible on a very small scale and by a handful of highly expert practitioners.

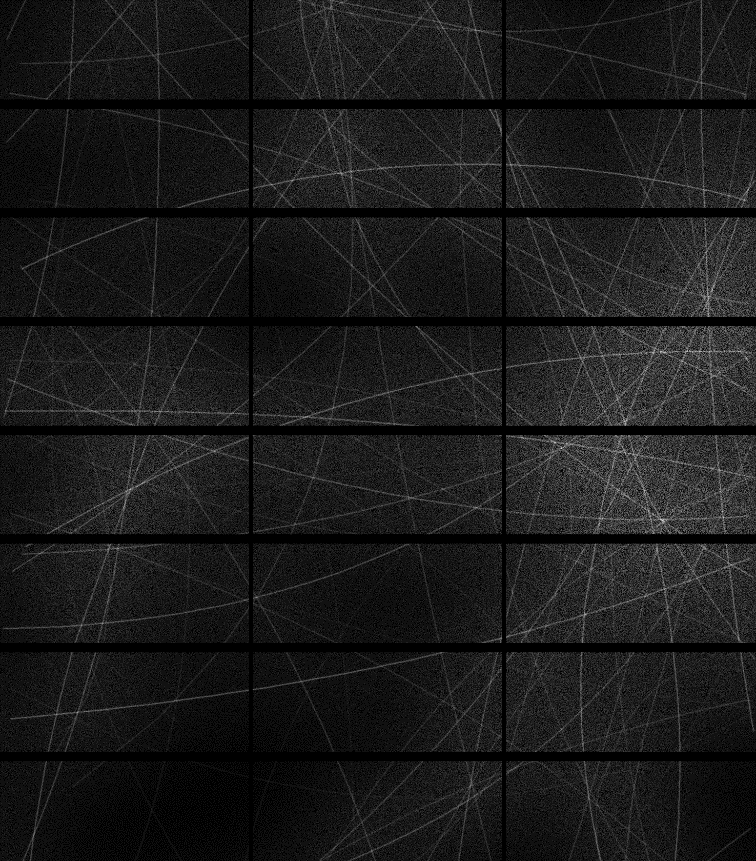


Figure 3 A Diffuse multiple scattering pattern

**Benchmark description**

DiffuseScatter (Material Science, Condensed Matter):Diffuse Multiple Scattering (DMS) is a phenomenon that has been observed in X-ray patterns for many years, but has only become accessible as a useful tool for analysis with the advent of modern X-ray sources and sensitive detectors in the past decade. The method is very promising, allowing for investigation of multi-phase materials from a single measurement – something not possible with standard X-ray experiments. However, analysis currently relies on extremely laborious searching of patterns to identify important motifs (triple intersections) that allow for inference of information. This task can only be performed by expert beam scientists and severely limits the application of this promising technique.

Here we have two benchmarks (i) a simple binary classification to distinguish between two possible phases. (ii) A multi-phase classification problem where we classify between six different phases or (possibly) combinations of two of the phases.

**Benchmark (i)** the data is for the older generation detectors and has dimensions of 487x195, with three channels. The data relates to two possible crystal structures monoclinic and tetragonal structures. There are 4030 examples of each lattice in the training data. The benchmark model is a CNN with two convolutional blocks, block 1 has 4x(4x4) filters, followed by (2x2) maxpooling and batch normalisation, block 2 has 8x(2x2) filters followed by (2x2) maxpooling and batch normalisation. The outputs connect to a dense layer with 8 nodes, which then connects to a binary classification layer. Between the dense layer and the classification layer a dropout rate of 0.3 was applied.

The dataset for this benchmark, is of size 8.6GB and consists of 8,060 DMS diffraction patterns, with each pattern of  487x195 pixels with  three channels. The data was generated using the [DMS simulator code](https://zenodo.org/record/12866).

**Benchmark (ii)** the data is for newer detectors from which we have taken a region of interest of 385x485 pixels. The data relates to six possible crystal structure types (Rhombohedral, Orthorhombic, MonoclinicA, MonoclinicB, MonoclinicC, Tetragonal, Cubic) with one or two different phase types possible per sample, it is also possible to have two Rhombohedral phases per sample. The labels are encoded in a one-hot-vector of length seven, one element for each phase and an additional element for a second rhombohedral phase. The benchmark model is a *ResNet* implementation, identified by neural architecture search using *AutoKeras*.

The dataset for this benchmark, is of size 110GB and consists of 78693 examples. Each sample is 385x485 pixels with one channel. The data was generated using the DMS [simulator code](https://zenodo.org/record/12866).

|  |  |
| --- | --- |
| **Benchmark Descriptor** | DiffuseScatter - Binary Case |
| **Parent Descriptor** | DiffuseScatter |
| **Main Domain** | Material Sciences |
| **Sub Domain** | Condensed Matter Physics |
| **Task** | Image classification |
| **Data Type** | Images [487 x 195 x 3] |
| **Data Size** | 8.6 GB |