CCP4 interest in diffraction integration software

Harvard Medical School

David Waterman

CCP4

July 2012

Acknowledgements

Funding: CCP4 (DGW), EU FP7 BioStruct-X & Diamond Light Source for the rest

Colleagues: Graeme Winter, Gwyndaf Evans and Alun Ashton at Diamond

Thinking: Andrew Leslie, Phil Evans, Gleb Bourenkov, Garib Murshudov

Materials: Andrew Leslie for slides on Mosflm

Collaboration: Nick Sauter's group (NIH grant R01GM095887 for synchrotron data processing)

Overview

- Data reduction with CCP4 now
- The Mosflm legacy
- Beyond Mosflm $\rightarrow$ DIALS
- Error estimation: personal interest in a new framework for integration software

Data reduction with CCP4 now

Mosflm/iMosflm

Data reduction with CCP4 now

Mosflm/iMosflm

- Mosflm is a *very* successful package
- Widely available (through CCP4)
- Typical user experience is *via* the excellent GUI, iMosflm
- This encourages inspection of the images, to identify
  - poor spot shapes
  - anisotropic diffraction
  - multiple lattices
  - very high mosaicity (increase threshold in indexing)
  - incorrect direct beam position (no. 1 cause of indexing failure)
  - shadows
  - ice spots or rings

The Mosflm legacy

- The software is 3 decades old, derived from the Cambridge MOSCO system (Nyborg and Wonacott), developed further at Imperial College and then at the LMB
- Written in FORTRAN 5 for a Data General Nova 3/12 computer with 32 kWords of memory and a 10 Mbyte disk drive. Handswitches provided control of program operation at run time.

- The limited memory dictated that processing involved running a series of separate programs, further divided into overlays, with communication provided by external files

The Mosflm legacy

The drivers for development were technological advances, and to solve specific problems

- Arrival of VAX computers (virtual memory!)
- Image plates
- More “automated” processing for higher throughput
- CCD detectors
- Improve user-friendliness (GUIs) and software “intelligence”
- Fine ϕ slicing
- Pilatus detectors

The Mosflm legacy

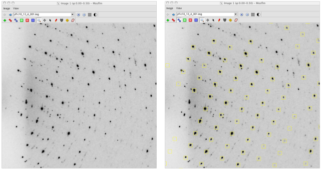
Problems

- The implementation has been overly influenced by hardware limitations
- Development has always been an adaptation of the original specification without changing the basic framework
- Particular issues include:
  - The instrumental correction factor for intensity error estimates
  - A non-general coordinate system (the "Cambridge frame")
  - A confusing change of coordinate system in refinement after autoindexing, for use of REFIX code [1]
  - Crystal orientation parameters "mopping up" unmodelled obliquity of the angle between beam and rotation axis
  - Unphysical parameters such as detector TILT and TWIST
  - Three separate parameter refinement routines
  - Strictly 2D integration

[1] though this is invisible to the user

The Mosflm legacy

- Development of Mosflm continues. Where it can adapt, it will
- Current developments:
  - better handling of Pilatus data (big improvement recently)
  - multiple lattices
  - parallel processing
  - generalised goniometry



Beyond Mosflm → DIALS

New sofware is desired to provide:

- A physically realistic model of the experiment
- 3D parameterised profile fitting
- Rapid processing fully utilising modern computer hardware, keeping pace with high data acquisition rates
- Challenging cases, e.g. deconvolution of overlaps, handling of highly mosaic crystals
- A modular, extensible architecture, suitable for implementation within pipelines as well as interactively

Error estimation

personal interest in a new framework for integration software

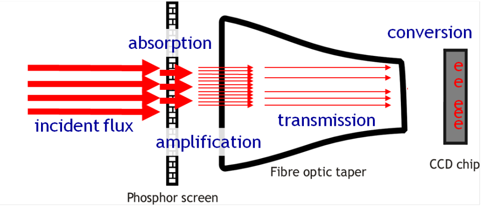
One particular deficiency of current integration programs is poor modelling of the error of integrated intensities

- Non-Poisson statistics of CCD detectors
- Correlations between pixels
- Problems with learned profiles

Error estimation

Non-Poisson detector response

- Real detectors have a DQE < 1, which implies Poisson statistics underestimate the true errors
- It is possible to derive a better estimate for a simulated CCD detector based on the physical processes involved in detection



$\overline{m}_0$ $\overline{m}_1$ $\overline{m}_2$ $\overline{m}_3$ $\overline{m}_4$

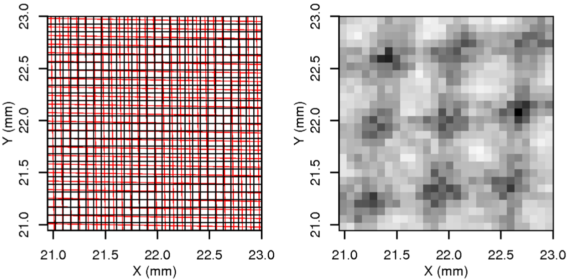
$$\overline{m} = \overline{m}_0 \overline{m}_1 \overline{m}_2 \overline{m}_3 \overline{m}_4$$

$$\sigma_m^2 = \overline{m}_0 \overline{m}_1 \left(\overline{m}_2 \overline{m}_3 \right)^2 \left(1 + \frac{\sigma_{m_2}^2}{\overline{m}_2^2} + \frac{1 - \overline{m}_3}{\overline{m}_2 \overline{m}_3} \right)$$

Error estimation

Non uniform correlation patterns

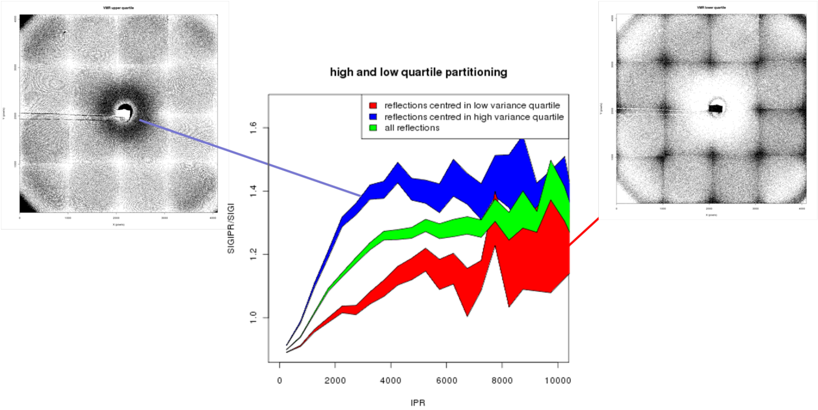
The distortion correction of CCD detectors introduces a Moiré pattern visible on flat field scattering images



Error estimation

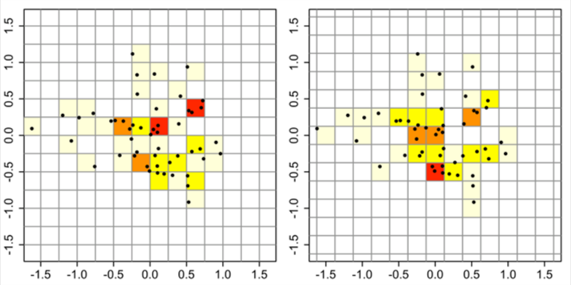
Non uniform correlation patterns

This leaves a detectable signature in profile fitting error estimates



Problems with learned profiles

- Pixellating detectors do not sample a signal, they average in bins (C. Nave)
- The appearance of histograms depends on the anchor points



Problems with learned profiles

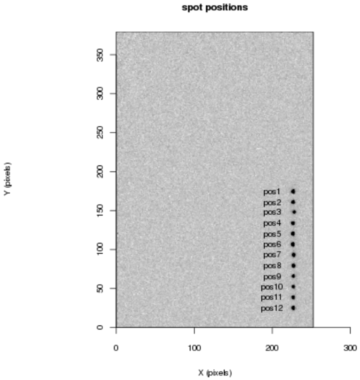
- Rather than learned profiles we'd prefer to use a parameterised model to construct profiles
- This addresses the problem of profile anchor points
- It also provides a way to account for errors properly even in the presence of correlations

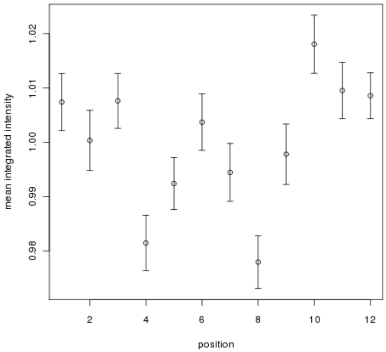
The end

Spatial noise

with James Holton

Position-dependent detector response causes a significant systematic error in the measured intensity





Spatial noise

with James Holton

The spatial period of variation appears to be on the scale of a pixel. Could the cause be phosphor inhomogeneities?