Data collection and data reduction

Harry Powell

ccp4-dev meeting 2nd March 2004

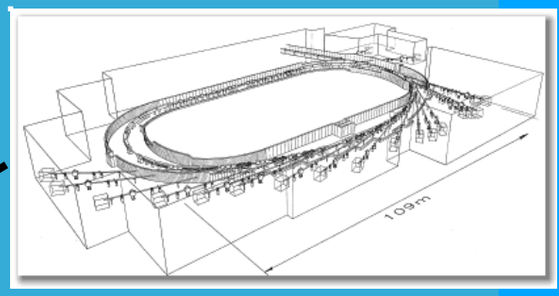
Which projects are involved?

- *DNA* - automation of whole process from crystal to $|F_{obs}|$ and beyond...
- *BEST* - analytical strategy program
- *Mosflm* - improvements to data integration
- *Mosflm* - *ccp4i* task
- *Scala* - improvements to scaling

DNA - what's it supposed to do?

- Automate data collection and processing at synchrotron beamlines
- Use & develop existing software to automate data processing
- Write new software to coordinate DC and DP

http://www.dna.ac.uk



DNA - who's involved (*where*)?

BEST & Mosflm

- *BEST* is a strategy program which uses more information than the current strategy option, and provides more detail for the optimum data collection

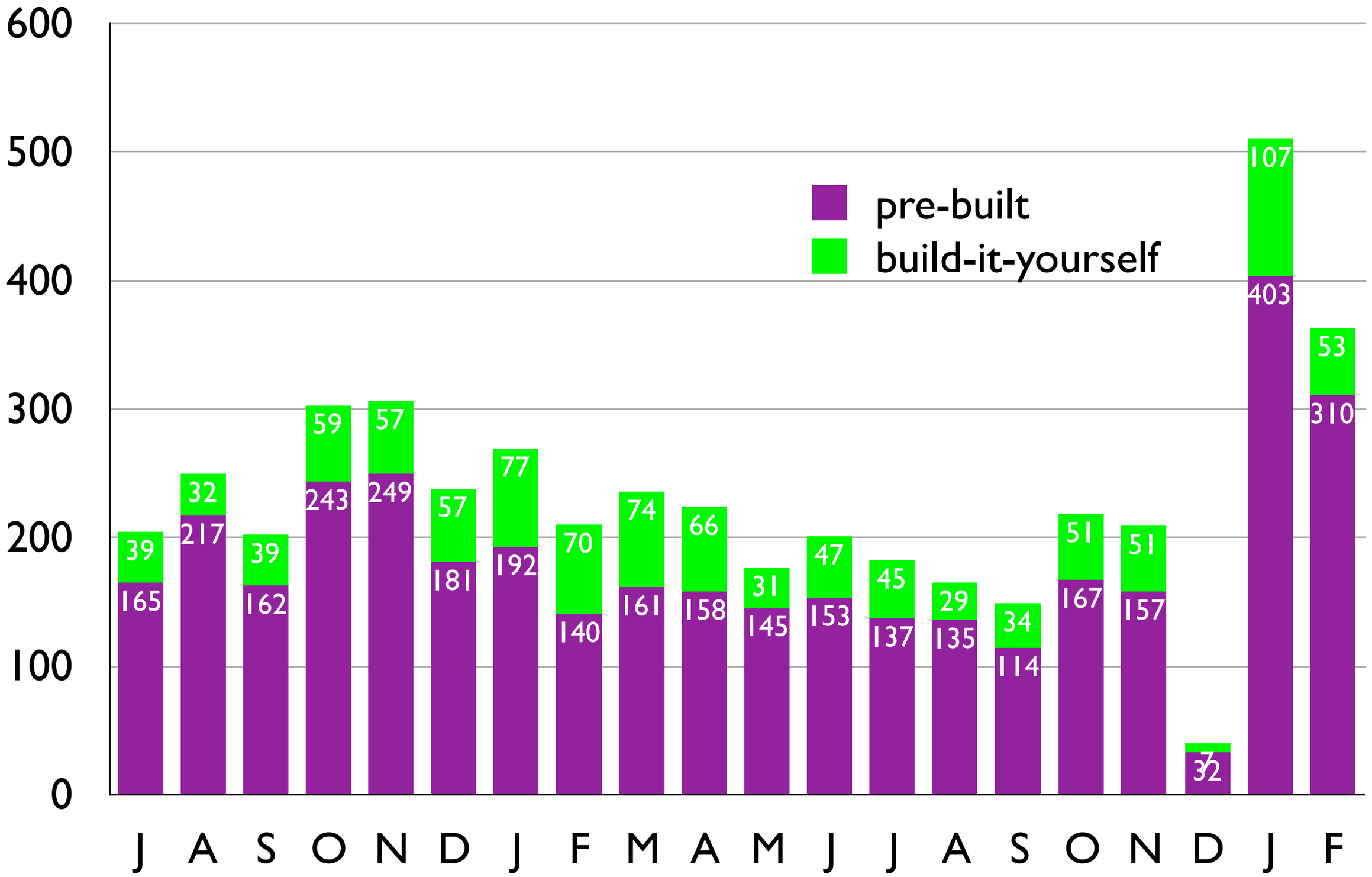
- currently being;
  - modified to use output from *Mosflm*
  - incorporated into *DNA* using output from *Mosflm*

Sasha Popov and Gleb Bourenkov (*EMBL, Hamburg*)

Mosflm - who does what?

- Andrew Leslie - coordinates whole thing
- Harry Powell (Andrew Leslie) - bug fixing, maintenance, development, user support, *etc, etc*.
- Geoff Battye (Harry Powell) - new GUI

complete downloads of Mosflm v6.2.2 & v6.2.3 by month



Mosflm - recent past & short term future

- major new developments
  - new GUI
  - spot positions in postrefinement
- minor new features
  - ice-ring exclusion in autoindexing
  - performance improvement (buffered *i/o* for SPOTOD)
  - *o/p* modified for *BEST* & *DNA*
- improvements
  - mosaicity estimation
  - autoindexing
- maintenance
  - bug fixing, additional detectors (automatic recognition)...
  - user support, better diagnostics

New *Mosflm* GUI

- Designing a new GUI for *Mosflm* with the aim of providing an attractive and straightforward interface to data integration in a step-by-step approach.
- communicates with *Mosflm* via *TCP/IP* sockets using a custom XML

Geoff Battye (*MRC-LMB, Cambridge*)

New GUI - timescale

- alpha test version (locally in LMB) - end June
- beta release (developers) - end September
- beta release (public) - before Christmas

MOSFLM version 6.2.3 Image Display

Edits allowed

Processing params

a :	0.00
b :	0.00
c :	0.00
alpha :	0.00
beta :	0.00
gamma :	0.00
PsiX :	0.00
PsiY :	0.00
PsiZ :	0.00
Mosaic :	0.300
Divh :	0.000
Divv :	0.000
Lambda :	1.542
Distance:	194.20
Beam X :	149.78
Y :	151.45
CCOMEGA :	0.000
ROFF :	0.00
TOFF :	0.00
YSCAL :	1.0000
Pick area: X:	11
Y:	11
Int threshold:	20
Vector scale	1
Two theta	0.00
Resolution	0.00
SPOT SEARCH	
Threshold	5.00
Rmin	15.00
Rmax	135.00
X offset	0.00
Y offset	0.00
Min X size	0.50
Max X size	2.00
Min Y size	0.50
Max Y size	2.00
Min no of pix	6
X splitting	0.30
Y splitting	0.30
AUTOINDEXING	
Min I/sig(I):	20

Prompts [On]
Update display:
After refinement [No]
After integration [No]

Timeout mode [Off]

Select item

Main menu

- Read image
- Find spots
- Edit spots
- Clear spots
- Select images
- Autoindex
- Estimate mosaicity
- Predict
- Clear prediction
- Adjust
- Refine cell
- Integrate
- Strategy
- Keyword input
- Find hkl
- Pick
- Measure cell
- **Circles**
- Beam / backstop
- Save/Exit

Output

Pixel X,Y	2001	0	
XC,YC mm	300.0	0.0	
Resolution		0.00	
Indices	0	0	0
F Phi	0.00	width	0.00
Intensity			0
Sigma			0
Spacing A			0.000
Average			0.0
Rms			0.0
Number			0
Zoomfactor			0

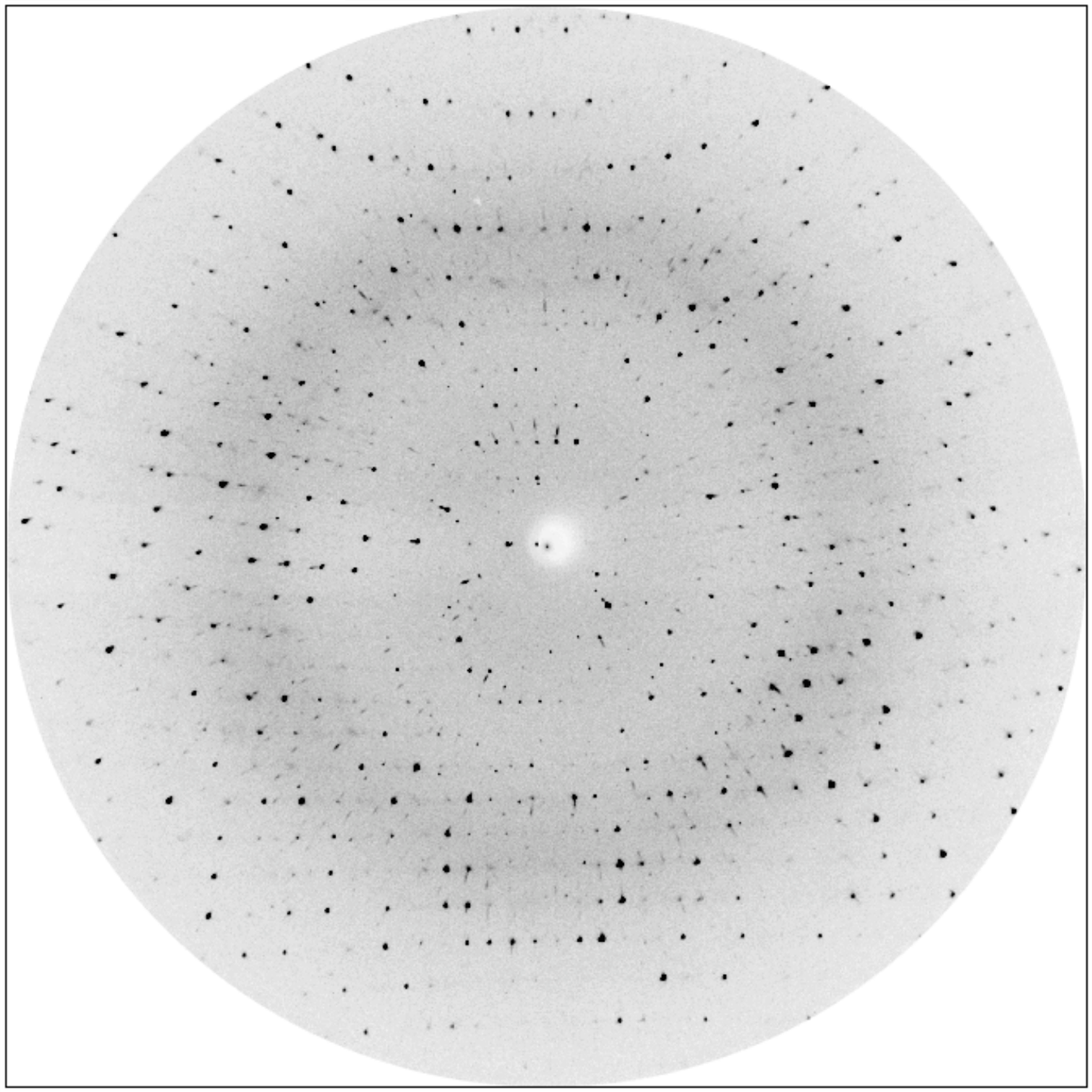
Circle resolution A
0.0 0.0 0.0 0.0

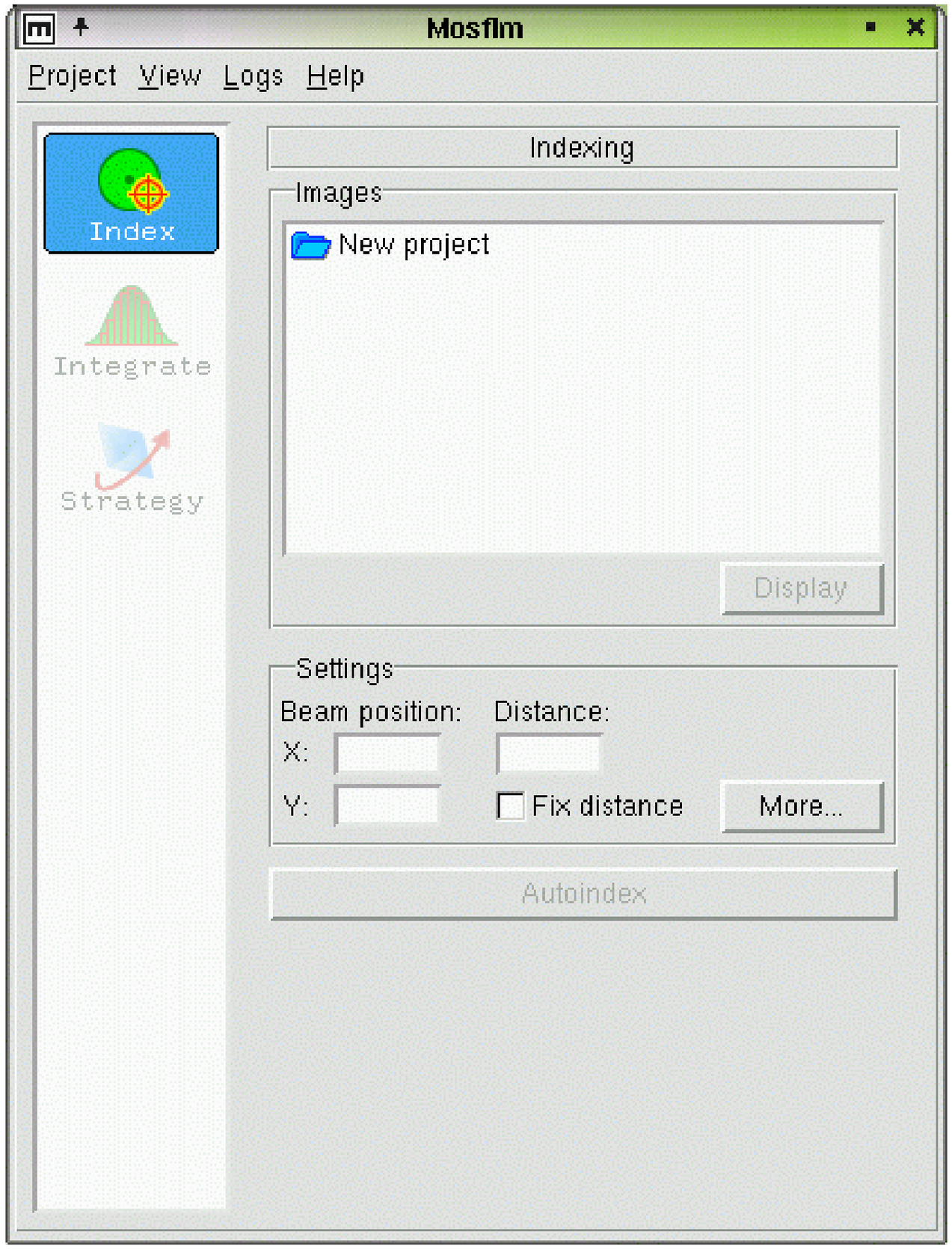
Phi	0.00	0.20

Missets ThetaX,Y,Z
0.00 0.00 0.00

lys_fine_002.pck

Min 1 Max 253 Cursor position
Overlay [on] Contrast
Colour [Black on whi] Mag [x4] PS Zoom




Mosflm
Project View Logs Help
Index
Integrate
Strategy
Indexing
Images
New project
Display
Settings
Beam position:
Distance:
X:
Y:
Fix distance
More...
Autoindex

Mosflm ccp4i task

- designed only for integrating images - no functionality is intended for indexing or postrefinement as these are best performed under supervision

Pete Briggs (*CCP4, Daresbury*)

Run MOSFLM in batch mode

Help

Run MOSFLM in batch mode to integrate diffraction images

Job title

Load parameters from command file | Browse | View

Load parameters from command file

Specify where the images are located

Use current working directory /gr0/harry/lys_fine

Image file name template:

Matrix file trypsin | Browse | View

MTZ out trypsin | Browse | View

Dataset and Harvesting Parameters

Create harvest file in project harvesting directory

Crystal | belonging to Project trypsin

Dataset name

Images to Integrate

Edit list | Add processing run

Crystal Parameters

Space group

Cell a | b | c | alpha | beta | gamma

Mosaicity

Wavelength of radiation used in the experiment | Angstrom

Detector Parameters

Set detector type to Mar Research IP and over-ride default determined by MOSFLM

Run | Save or Restore | Close

Scala - progress

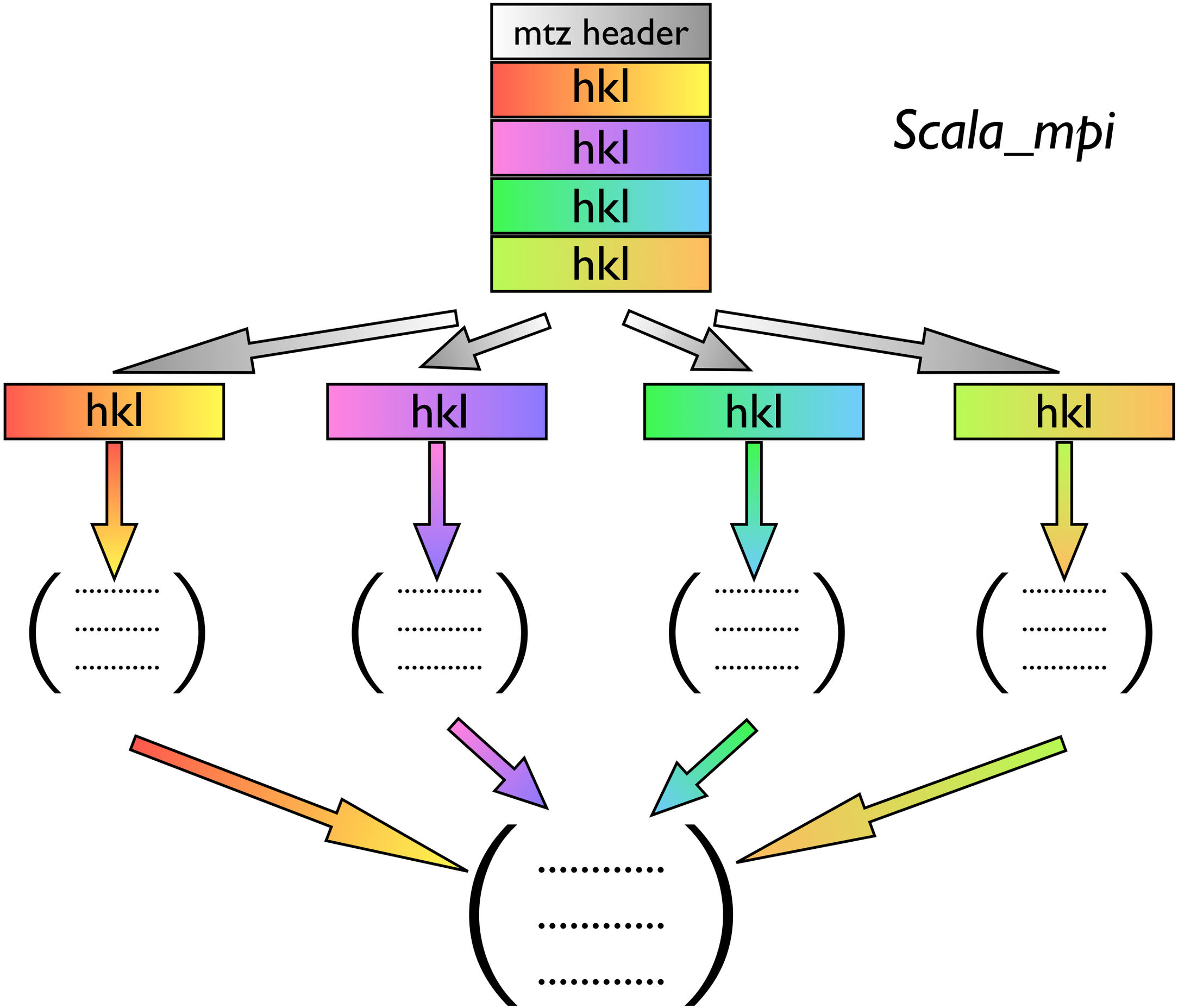
- rewrite using OOP methods
- unmerged reflection class which can be used for scoring all possible point groups (initially in a standalone application, later to be incorporated in Scala)

Phil Evans (*MRC-LMB, Cambridge*)

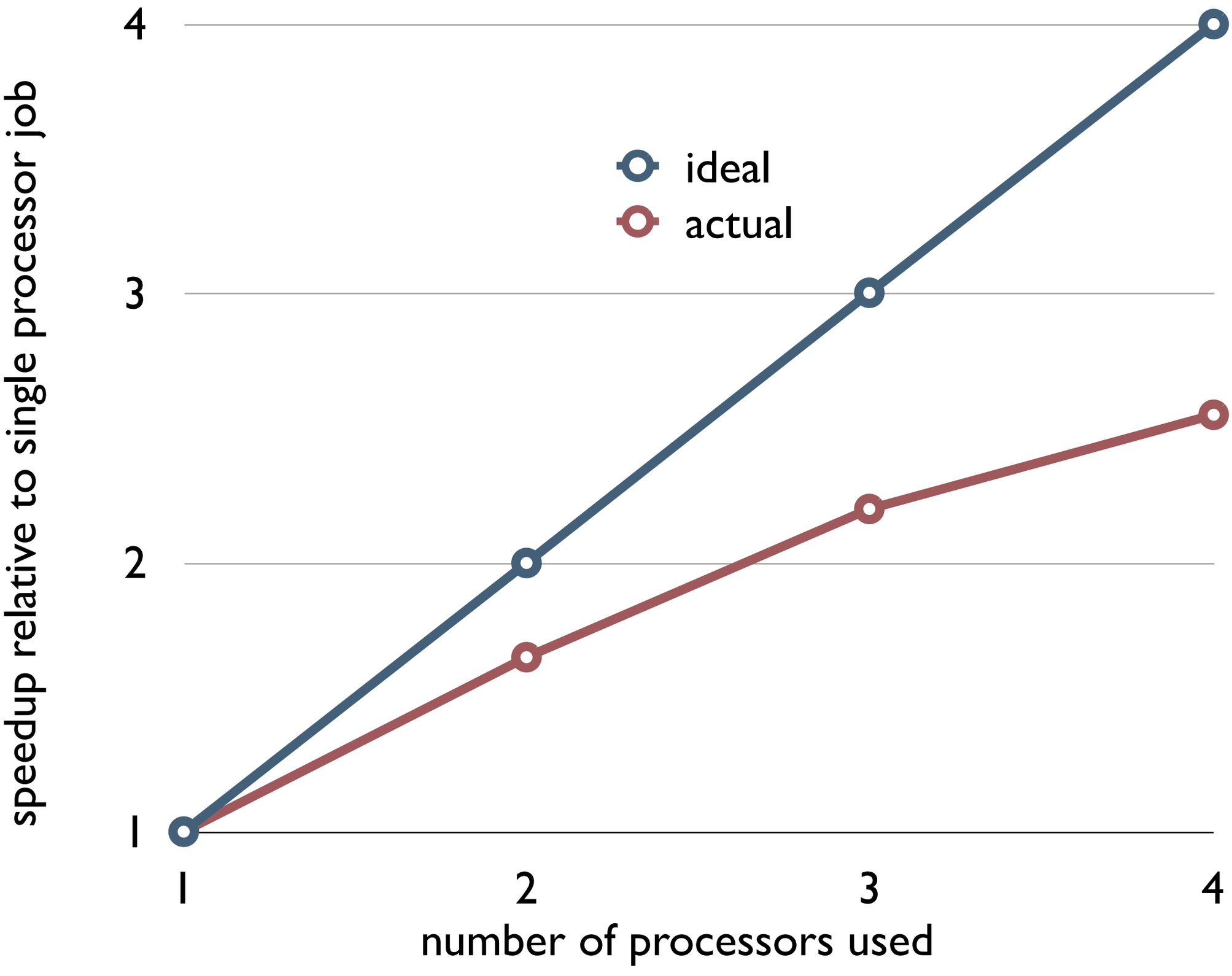
Scala_MPI - parallel version

- Improved performance by parallelizing the rate determining steps.
- Main bottleneck is the construction of a parameter matrix from the reflection data which is used to scale and merge the data; it is constructed by looping over all the observations in the MTZ file.
- Each available processor takes its own section of the MTZ file and compiles its own parameter matrix. These are then collected and summed into a single matrix.
- Code parallelised using MPI for use on distributed memory clusters.

Ronan Keegan (*eHTPX, Daresbury*)


mtz header
hkl
hkl
hkl
hkl
hkl
hkl
hkl
hkl
Scala_mpi

Scala_mpi: speedup of parallel code on up to 4 processors



that's all folks!