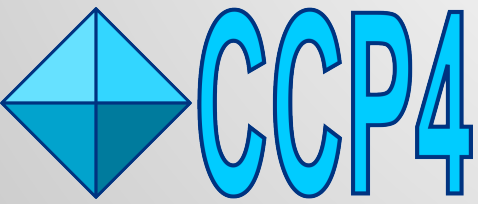
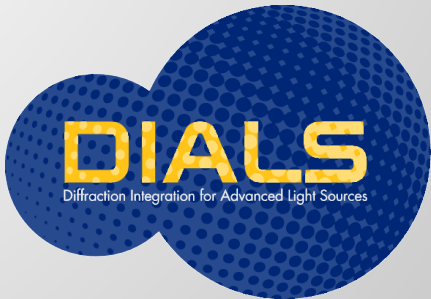
Model refinement

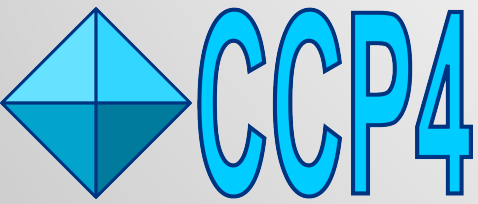
David Waterman

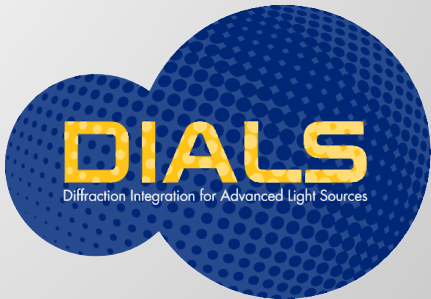
DIALS-6 / LBNL / May 27 2015



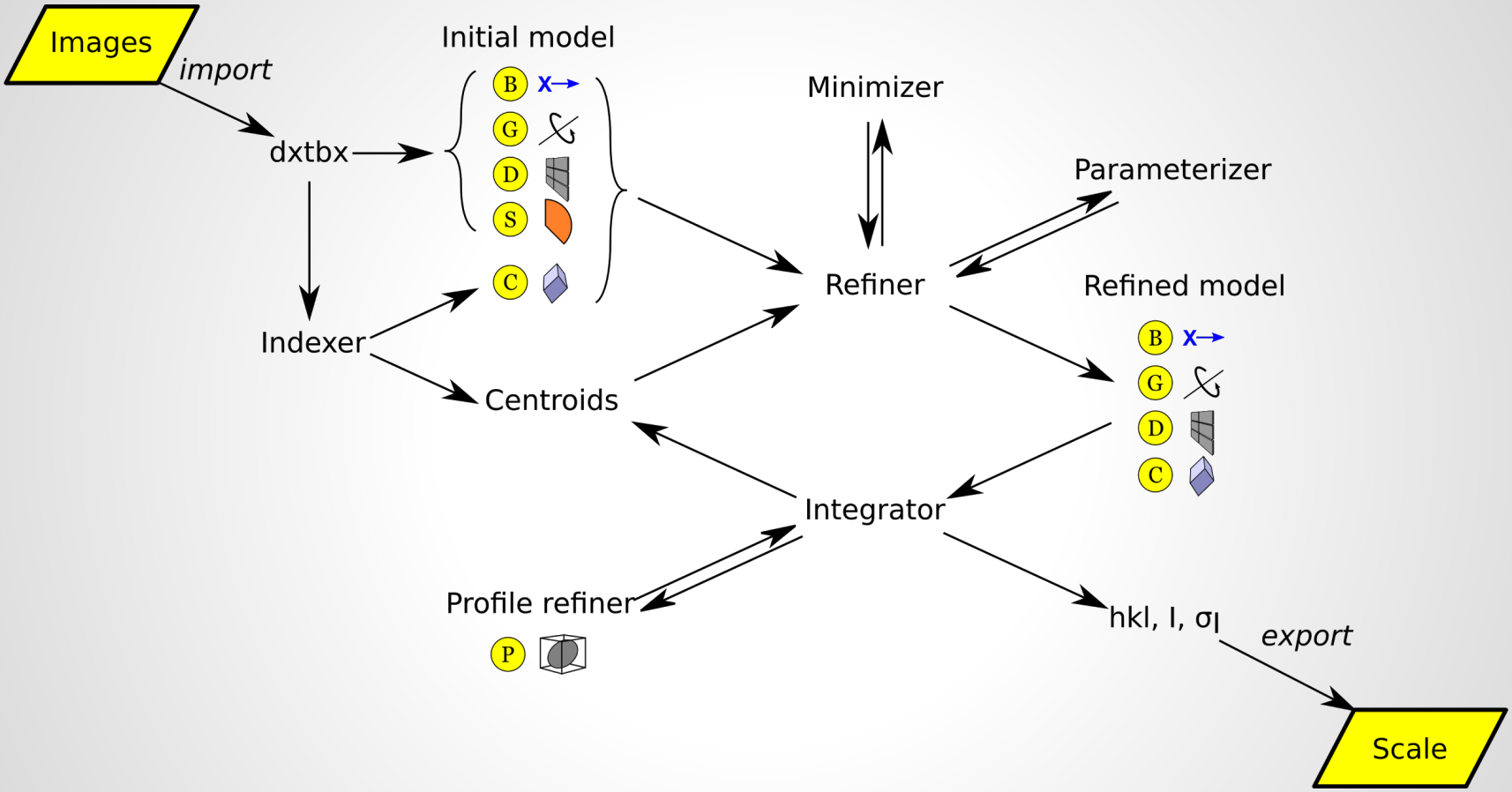


1. Centroid refinement
2. Parameterisation
3. Modularity
4. Scan-varying refinement
5. Joint multiple experiments
6. Multi-tile metrology
7. Tools for large jobs





Centroid refinement

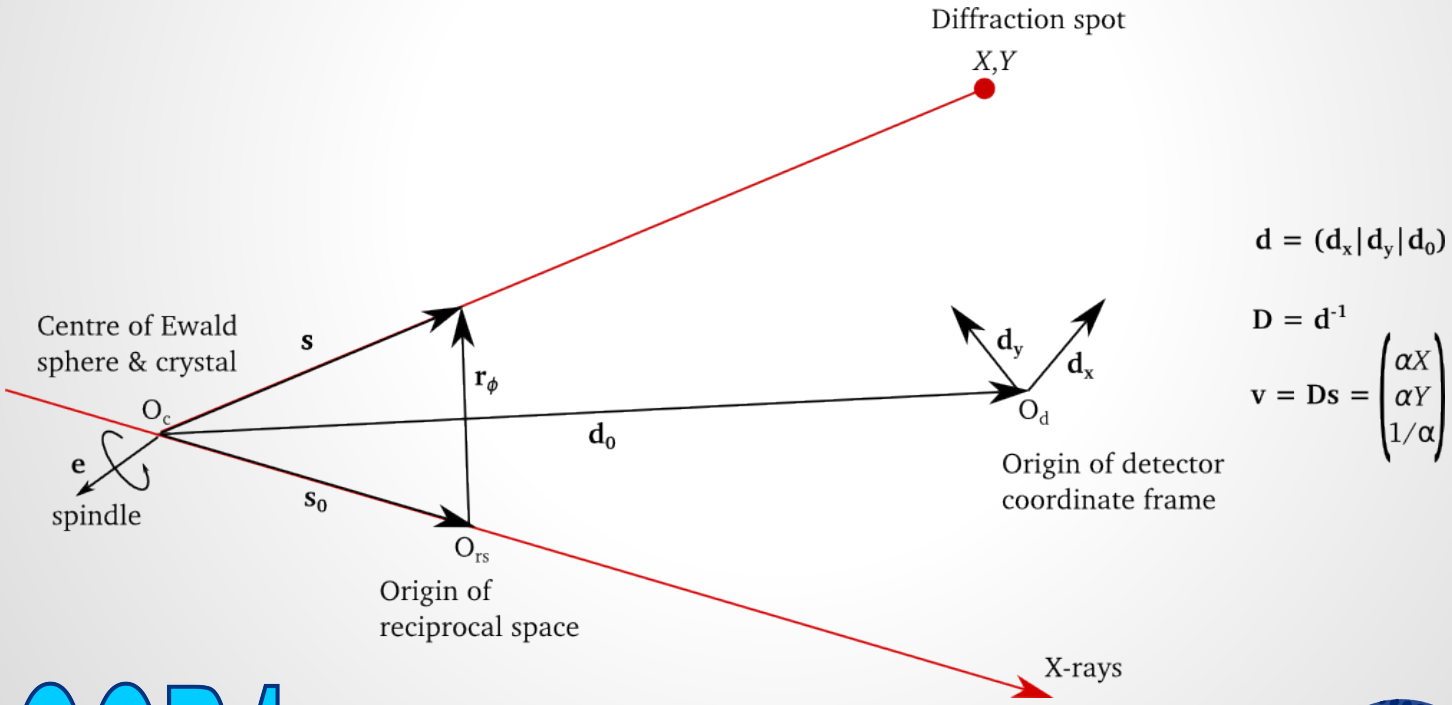


CCP4

DIALS
Diffraction Integration for Advanced Light Sources

Centroid refinement

- Refine parameters that affect *central impacts**
- Parameters that affect *general impacts* (mosaicity, $\Delta\lambda$, …) are determined by profile modelling



$$\mathbf{d} = (\mathbf{d}_x | \mathbf{d}_y | \mathbf{d}_0)$$

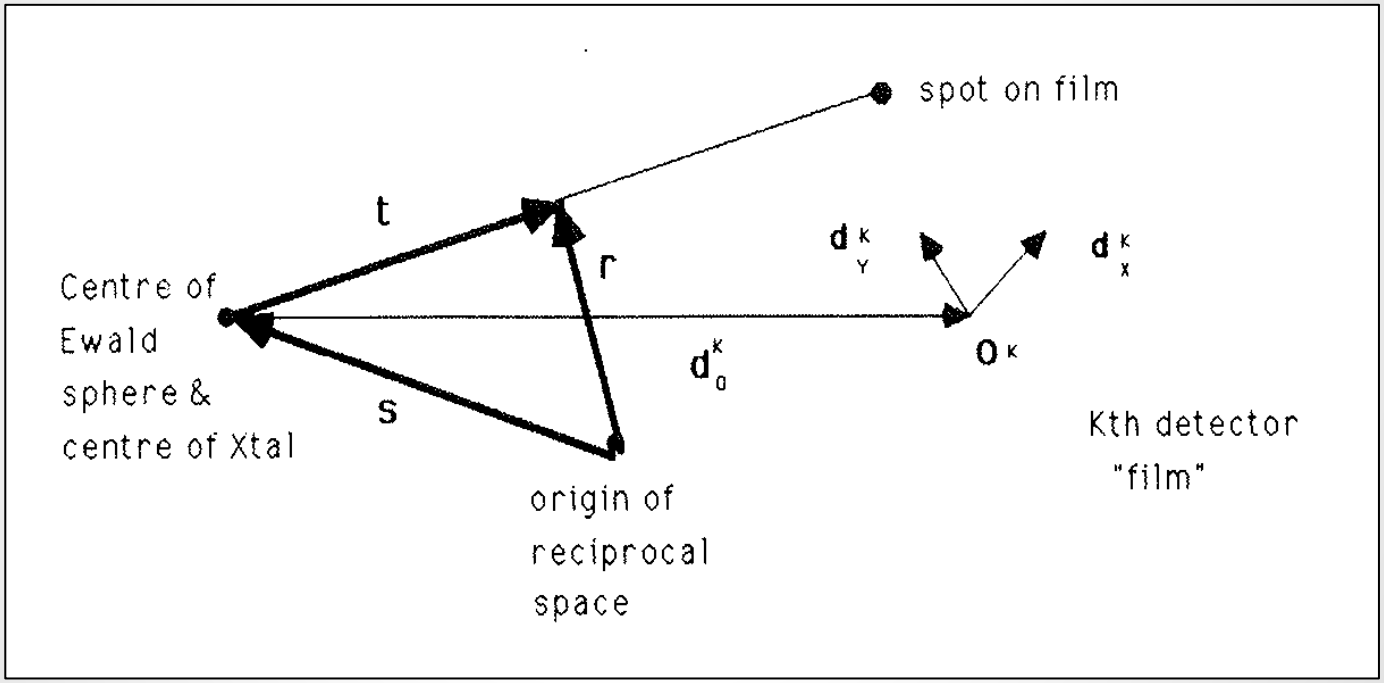
$$\mathbf{D} = \mathbf{d}^{-1}$$

$$\mathbf{v} = \mathbf{D}\mathbf{s} = \begin{pmatrix} \alpha X \\ \alpha Y \\ 1/\alpha \end{pmatrix}$$

*cf. *EVAL* package
J. Appl. Cryst. **36**, (2003) 220-229

Centroid refinement

- Resembles generalised vector description of MADNES (not by accident)



Bricogne, G (1987) *Proc. CCP4 Daresbury Study Weekend*

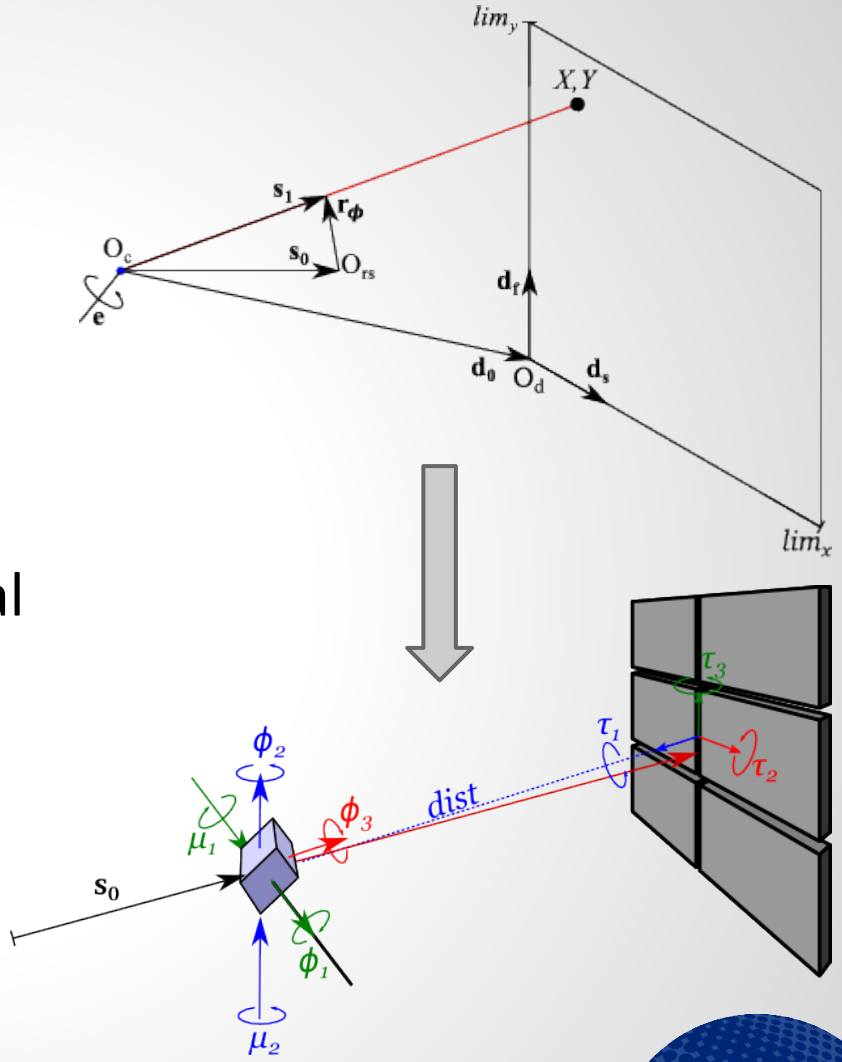
Various authors (1986) *Proc. EEC Cooperative Workshop on Position Sensitive Detector Software I-III*

Parameterisation

- Prediction equation parameterisation:
- Calculate

  (X, Y, ϕ) & $\left(\frac{\partial X}{\partial p}, \frac{\partial Y}{\partial p}, \frac{\partial \phi}{\partial p}\right)$

  for an abstract parameter *p*
- Implement **p** suitable for typical rotation method experiment



CCP4

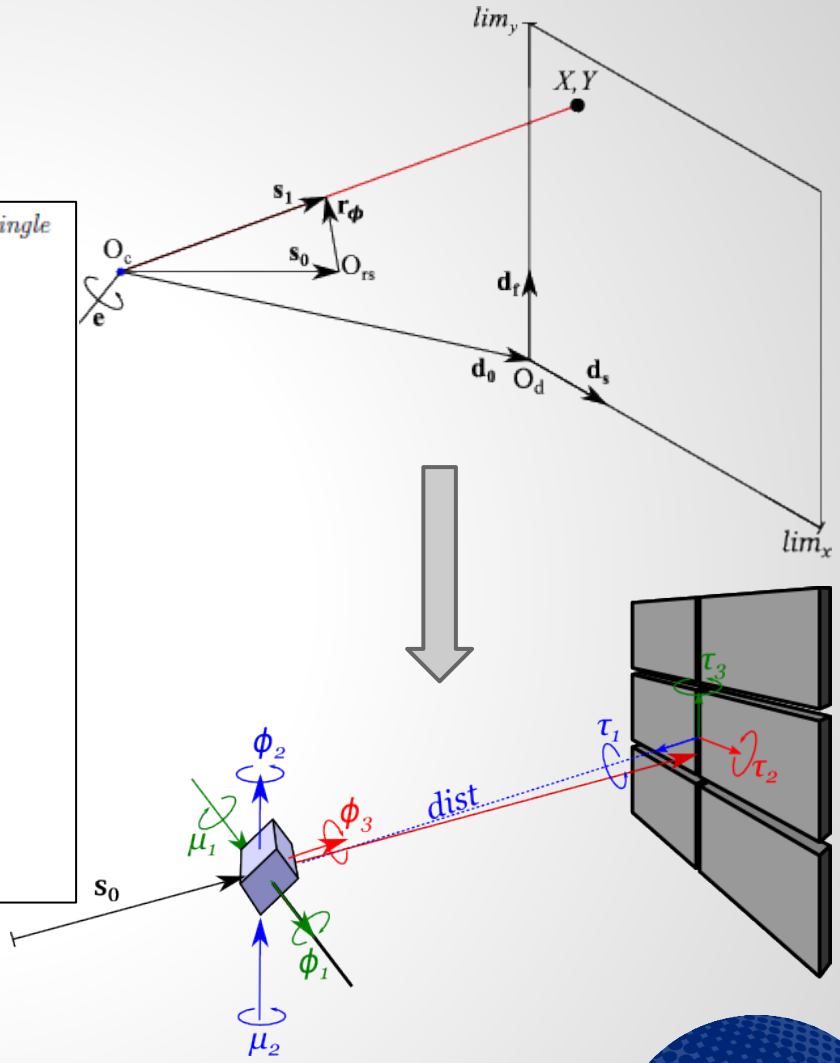
DIALS
Diffraction Integration for Advanced Light Sources

Parameterisation

There are 18 parameters in the *P* 1 case:

Table 1. *Default parameterisation in* dials.refine *for scan-static refinement using a single panel detector.*

Parameterisation	Model state	Parameters	Action
Beam	$\mathbf{s_0}$	μ_1	rotation about initial $\hat{\mu_2} \times \hat{\mathbf{s_0}}$
		μ_2	rotation about initial $\hat{\mathbf{s_0}} \times \hat{\mathbf{e}}$
		ν	set length of s_0 (wavenumber)
Crystal orientation	$\mathbf{U}$	ϕ_1	rotation about laboratory X
		ϕ_2	rotation about laboratory Y
		ϕ_3	rotation about laboratory Z
Crystal unit cell	$\mathbf{B}$	g^*_{11}	set metrical matrix elements
		g^*_{22}	
		g^*_{33}	
		g^*_{12}	
		g^*_{13}	
		g^*_{23}	
Detector	$\mathbf{d}$	p_0	set distance along initial $\hat{\mathbf{d_f}} \times \hat{\mathbf{d_s}}$
		t_1	translation along initial $\hat{\mathbf{d_f}}$
		t_2	translation along initial $\hat{\mathbf{d_s}}$
		τ_1	rotation about initial $\hat{\mathbf{d_f}} \times \hat{\mathbf{d_s}}$
		τ_2	rotation about initial $\hat{\mathbf{d_f}}$
		τ_3	rotation about initial $\hat{\mathbf{d_s}}$



Usually ν and μ_1 are fixed

CCP4

DIALS
Diffraction Integration for Advanced Light Sources

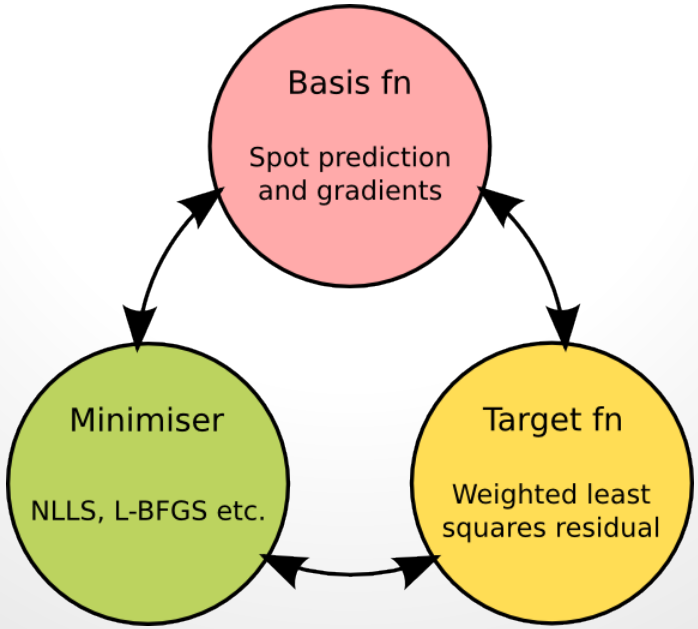
Modularity

We approach this as a traditional non-linear least squares problem (no reason not to - it works well for other programs)

$$L = \frac{1}{2}\sum_{i=1}^{n} w_{i,X}(X - X_{\text{obs}})^2 + w_{i,Y}(Y - Y_{\text{obs}})^2 + w_{i,\phi}(\phi - \phi_{\text{obs}})^2$$

$$\frac{\partial L}{\partial p} = \sum_{i=1}^{n} w_{i,X}\frac{\partial X}{\partial p}(X - X_{\text{obs}}) + w_{i,Y}\frac{\partial Y}{\partial p}(Y - Y_{\text{obs}}) + w_{i,\phi}\frac{\partial \phi}{\partial p}(\phi - \phi_{\text{obs}})$$

However, taking heed of Phil Evans' advice we keep things flexible



CCP4

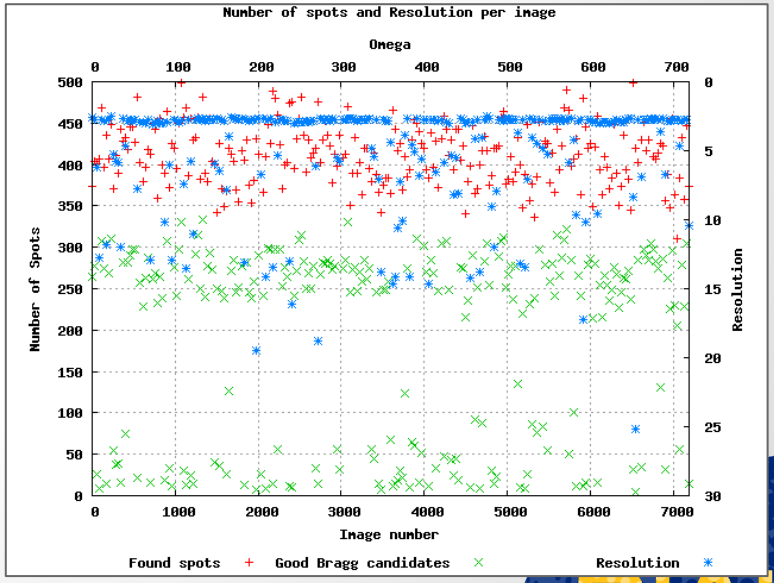
DIALS
Diffraction Integration for Advanced Light Sources

Scan-varying refinement

- We do global, not local, refinement
- How to model changes to the crystal model over time?

Example

- 720° of tetragonal thaumatin data collected at 0.1°/image, 40Hz, 3% transmission at DLS I03

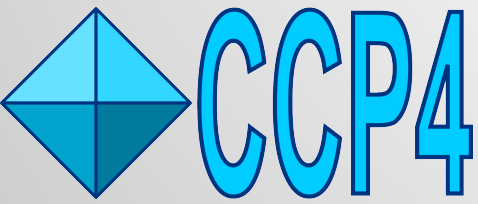


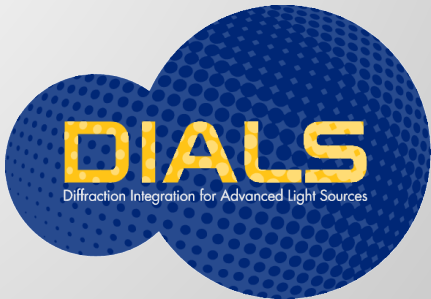
CCP4

DIALS
Diffraction Integration for Advanced Light Sources

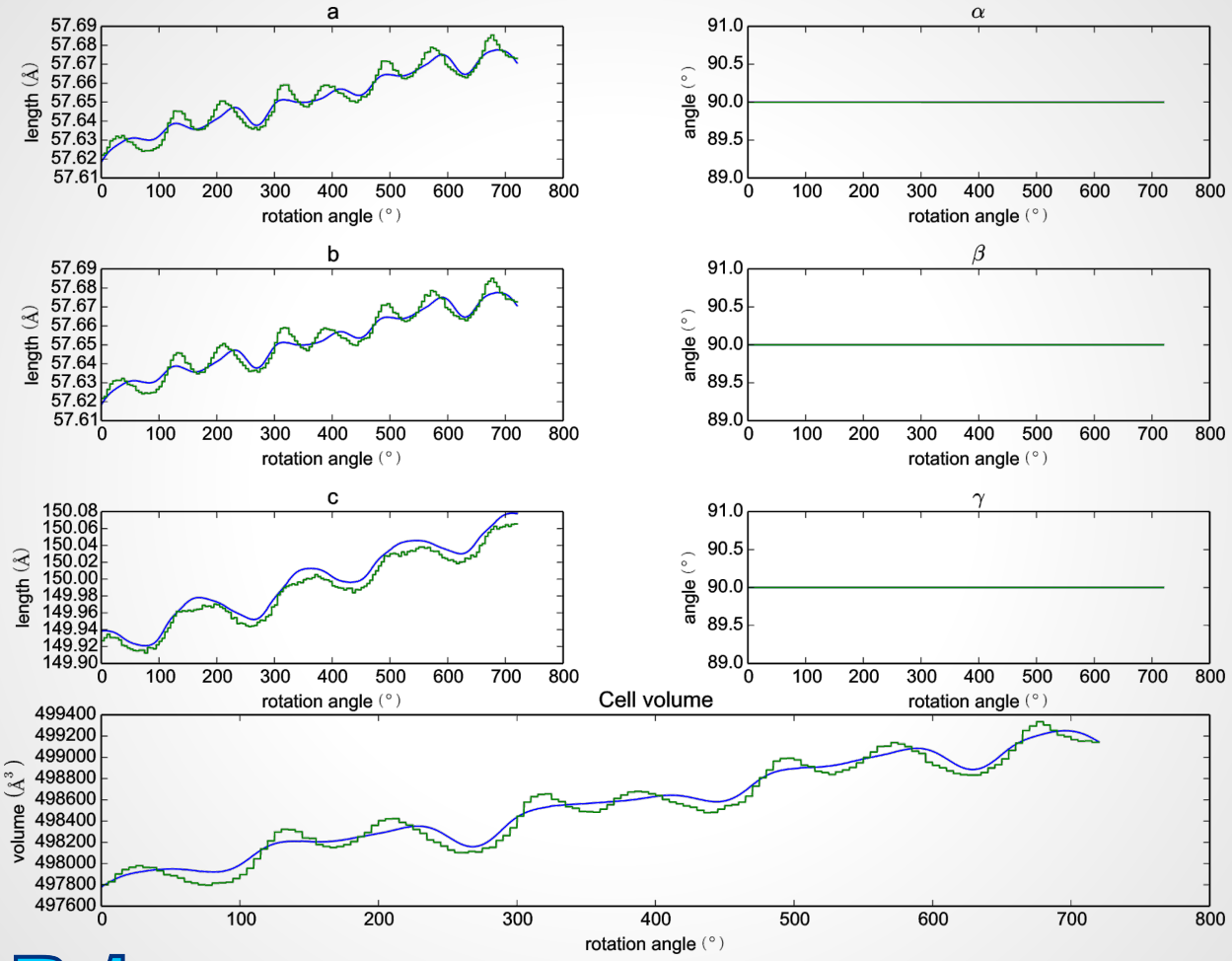
Scan-varying refinement

- Scan divided into equal-sized intervals
- Crystal parameterisation split over sample points
- Gaussian smoother, inspired by AIMLESS
- 117 parameters:
  - 6 detector
  - 1 beam
  - 3 crystal orientation × 22 "samples"
  - 2 unit cell parameters × 22 "samples"





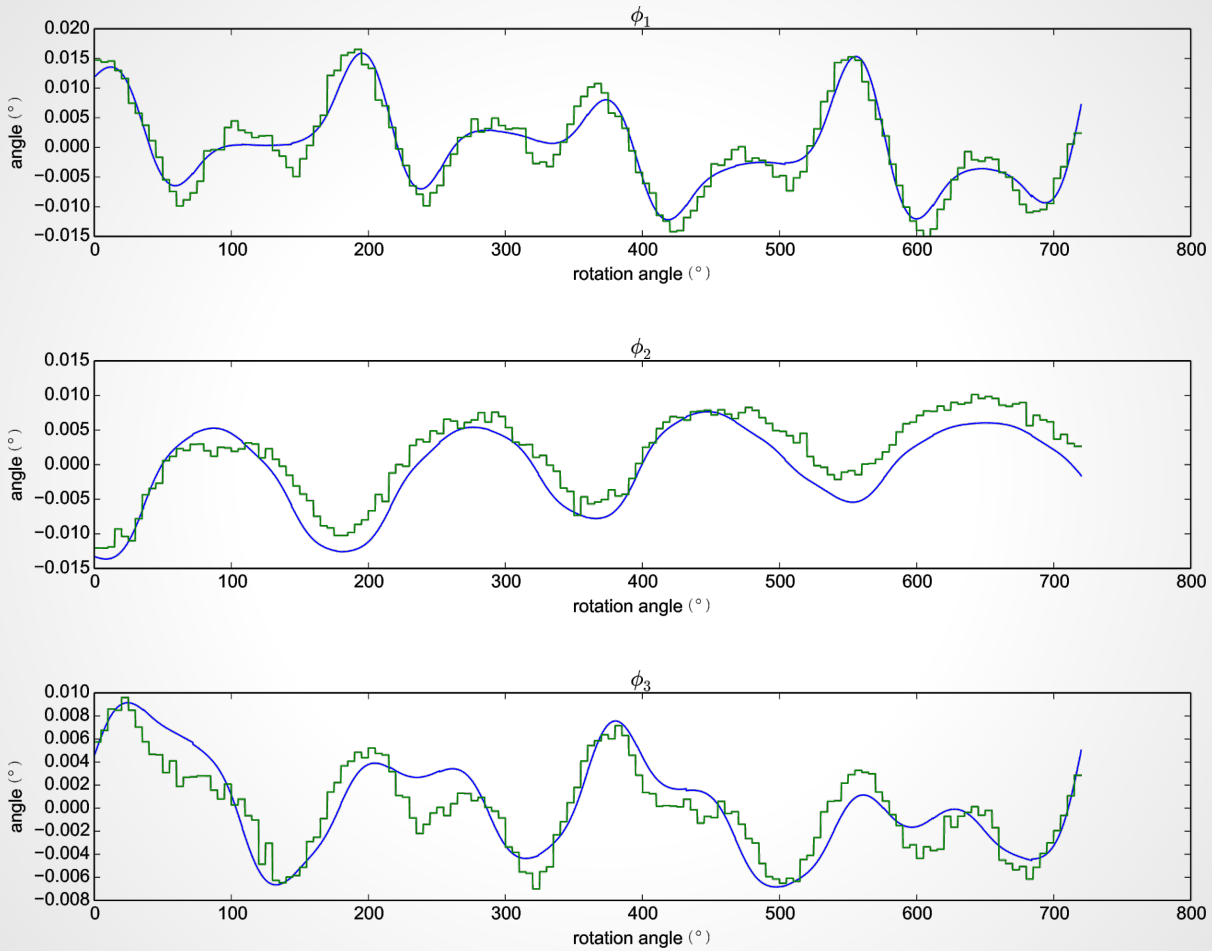
Scan-varying refinement



CCP4

DIALS
Diffraction Integration for Advanced Light Sources

Scan-varying refinement

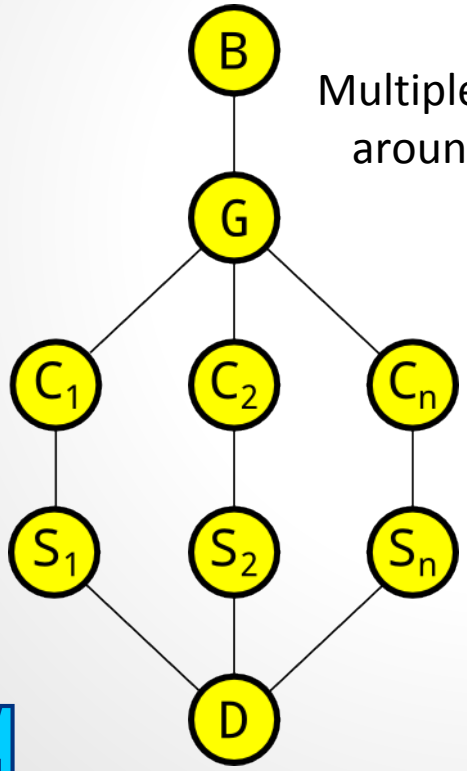


CCP4

DIALS
Diffraction Integration for Advanced Light Sources

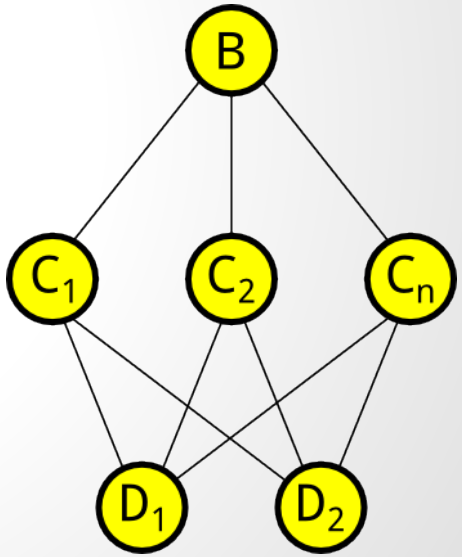
Multiple Experiments

- Global refinement across datasets that share some models
- Typical use cases involve multiple crystals



Multiple rotation scans around the same axis

XFEL shots with two detectors

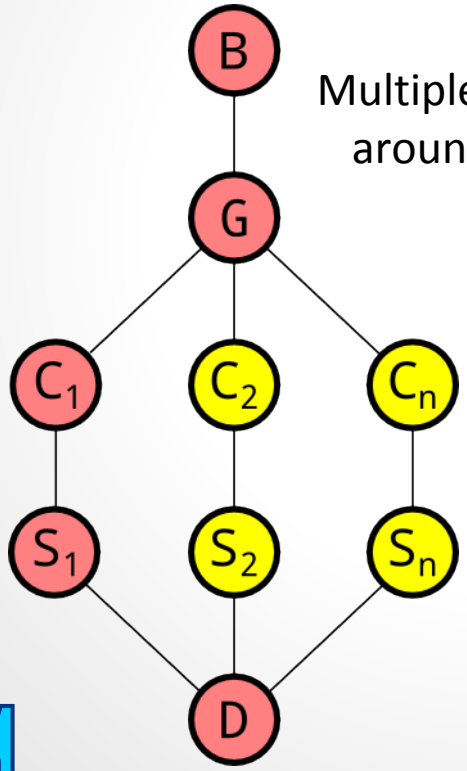


CCP4

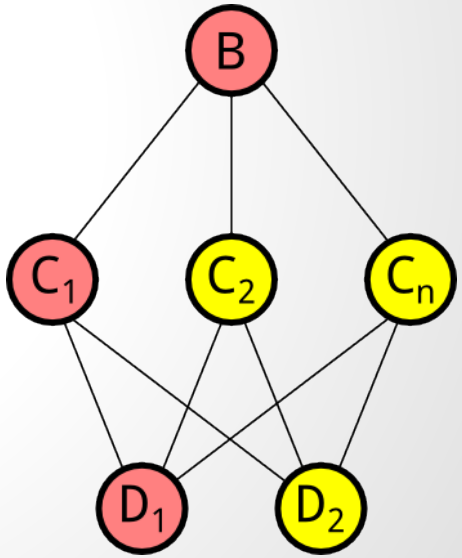
DIALS
Diffraction Integration for Advanced Light Sources

Multiple Experiments

- Global refinement across datasets that share some models
- Typical use cases involve multiple crystals



Multiple rotation scans around the same axis



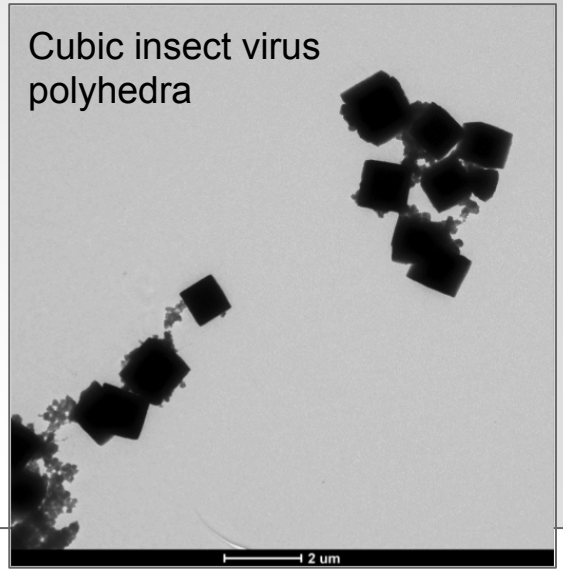
XFEL shots with two detectors

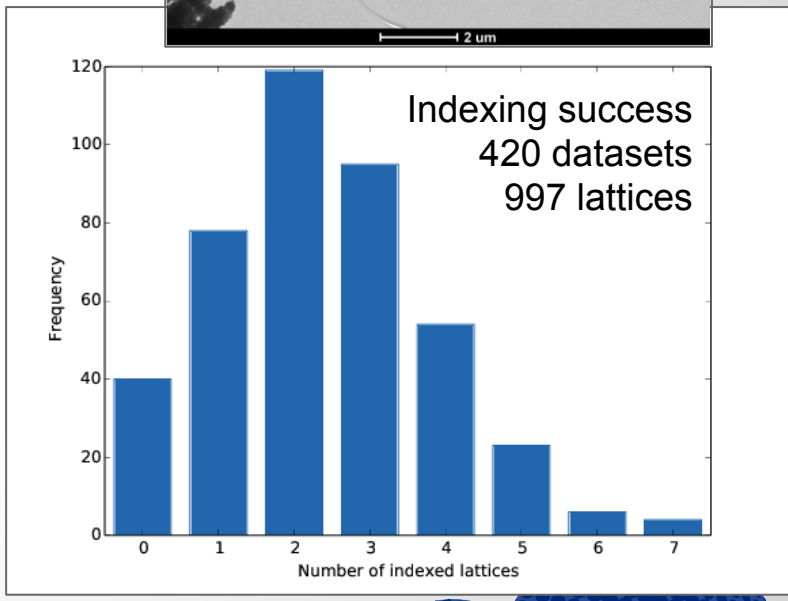
CCP4

DIALS
Diffraction Integration for Advanced Light Sources

Multiple Experiments

- *DIALS* contains an indexing algorithm that is very successful at identifying multiple lattices*
- This even works when lattices diffract equally well, and only a narrow wedge of data is available
- As additional lattices are found, joint refinement reduces correlations between crystal and detector parameters





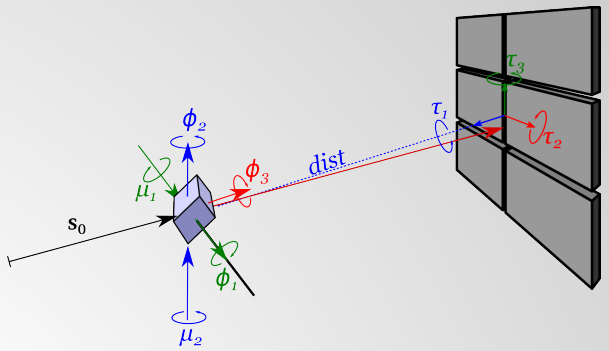
CCP4

*Gildea et al. (2014) *Acta Crystallogr. D Biol. Crystallogr.* **70**, 2652-66

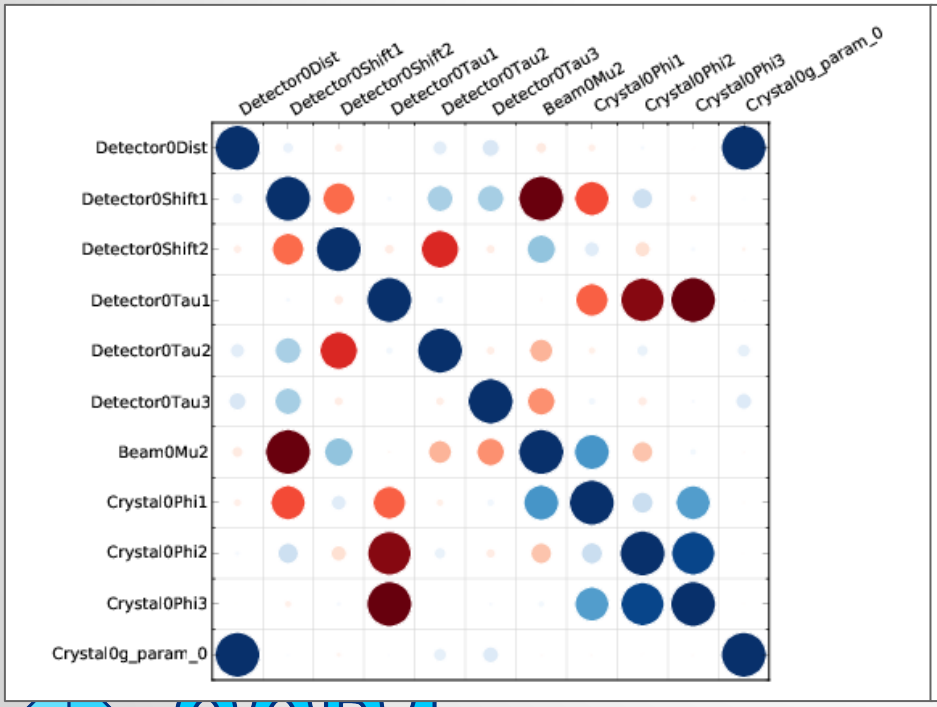
DIALS
Diffraction Integration for Advanced Light Sources

Multiple Experiments

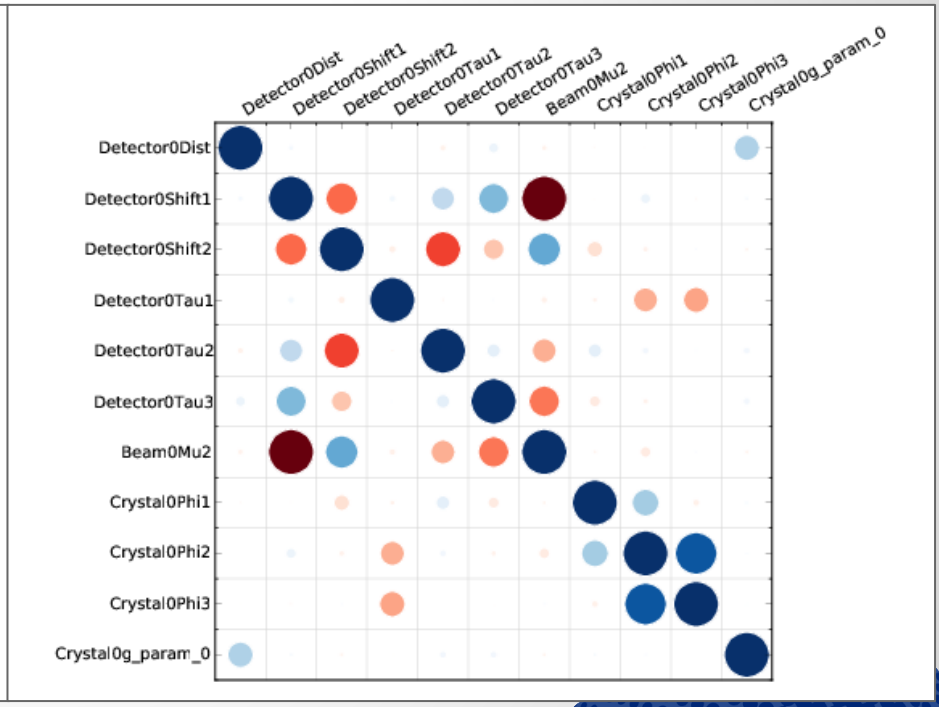
Cubic polyhedrin crystals, 1° scans



One lattice



5 sweeps (16 lattices)

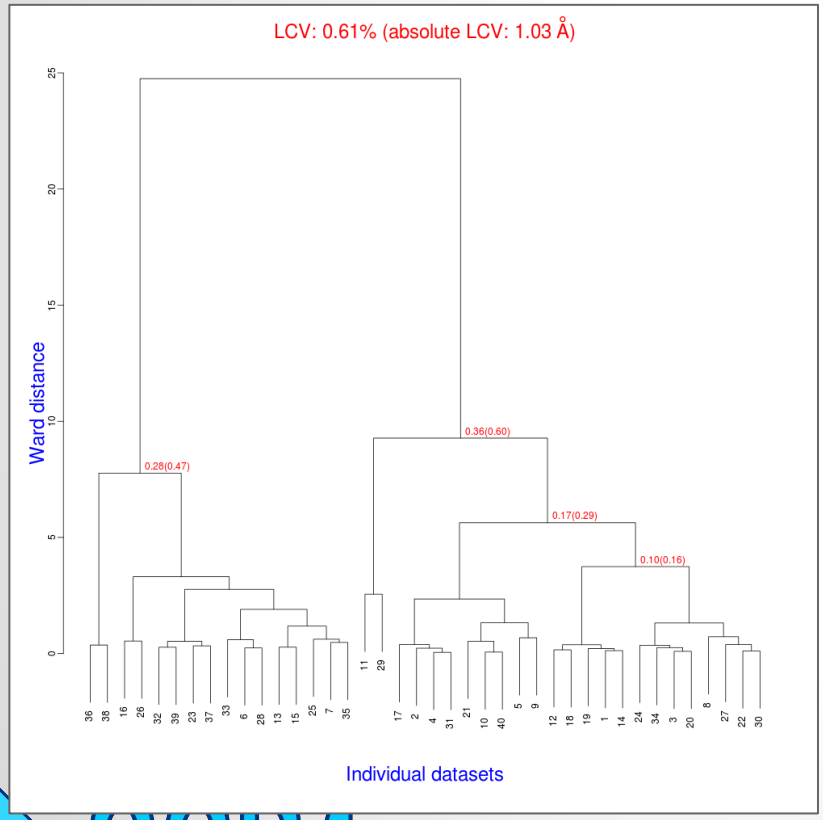


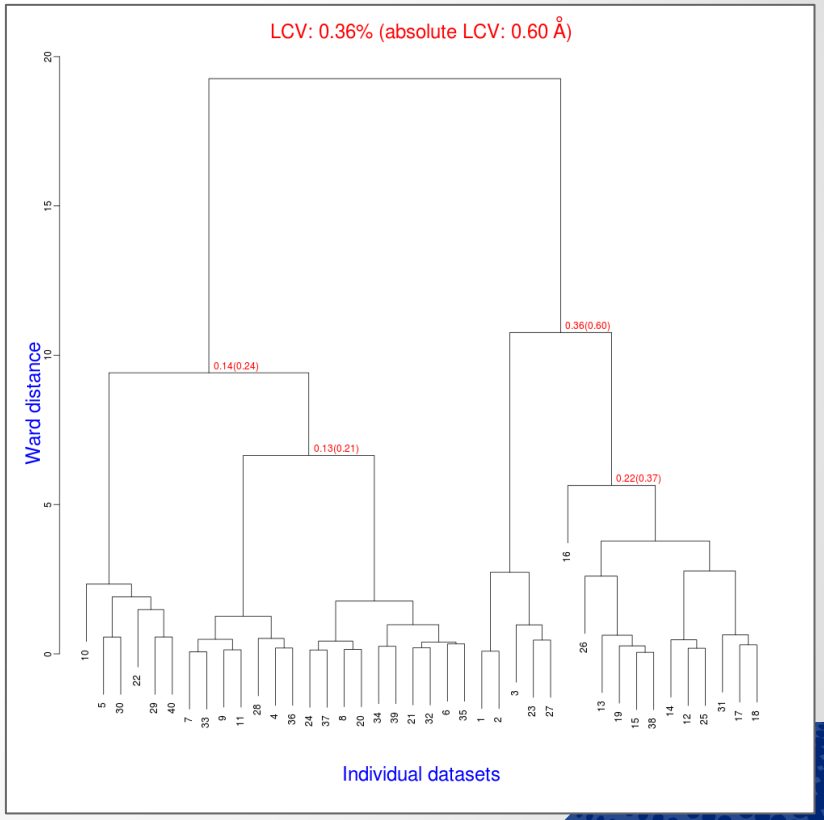
CCP4

DIALS
Diffraction Integration for Advanced Light Sources

Multiple Experiments

- We are investigating the use of joint refinement as a preparatory step for *BLEND*





TehA data. See forthcoming *Acta Cryst. D***71** (June 2015) for original analysis

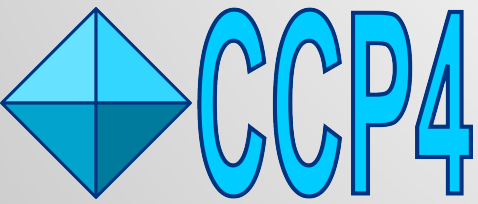
DIALS
Diffraction Integration for Advanced Light Sources

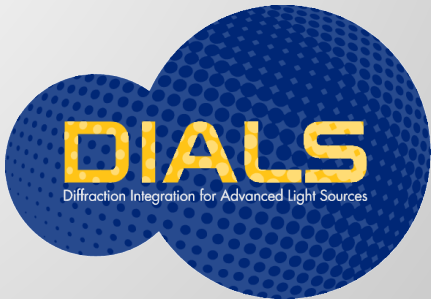
Multi-tile detector metrology

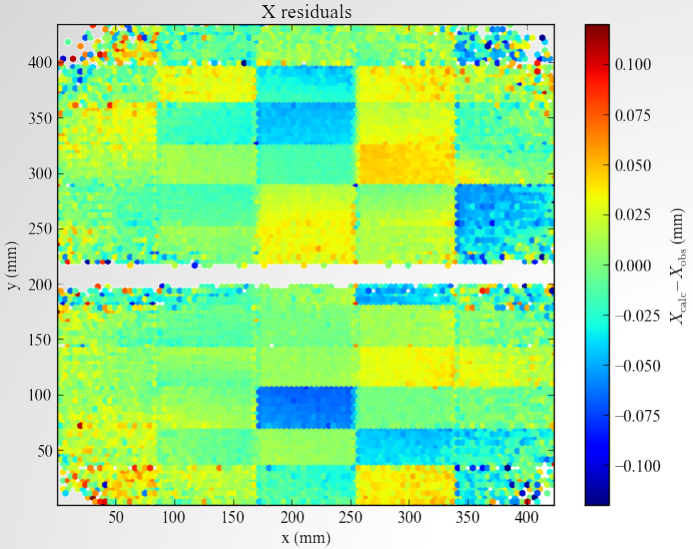
- DECTRIS Pilatus PADs are extremely well-made
- The spatial calibration of panels is measured at the factory
- Correction tables are supplied with the detector
- However, *we don't typically use this information for MX*
- Nevertheless its effect is detectable in processed data

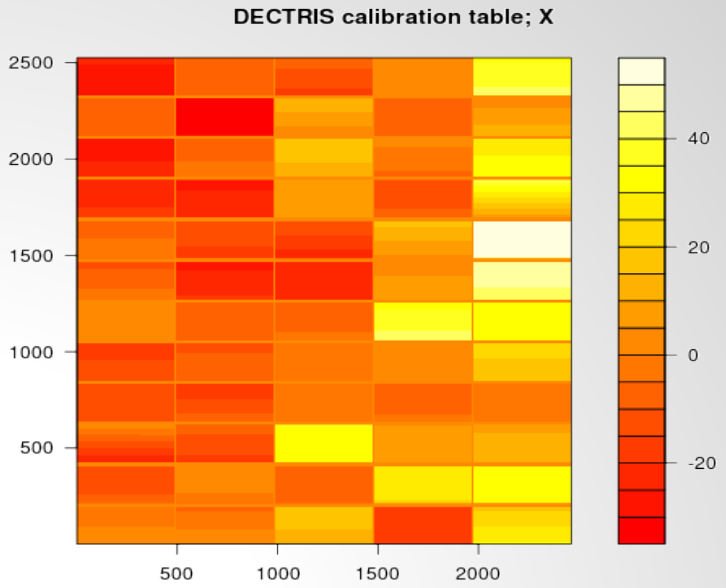
Example

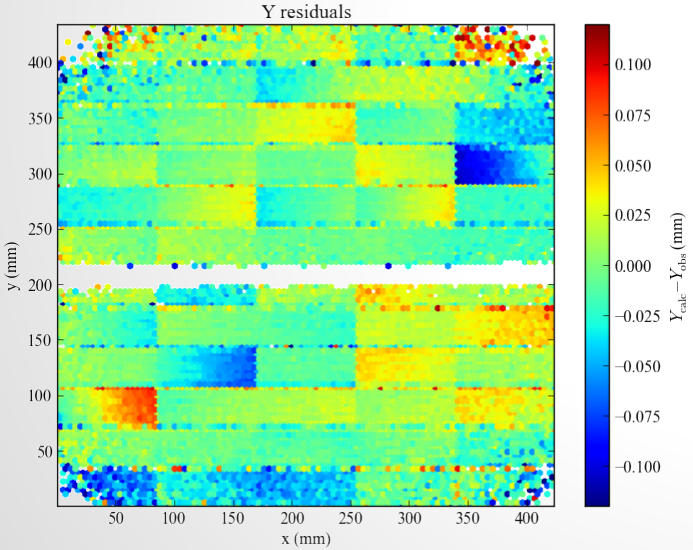
- 720° of thaumatin data collected at 0.1°/image, 40Hz, 3% transmission at DLS I03
- diffraction to 1.8 Å in the corners

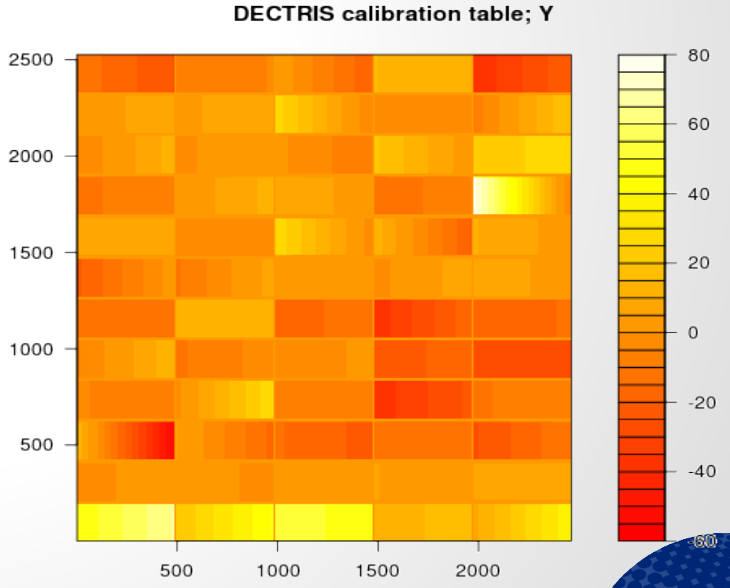






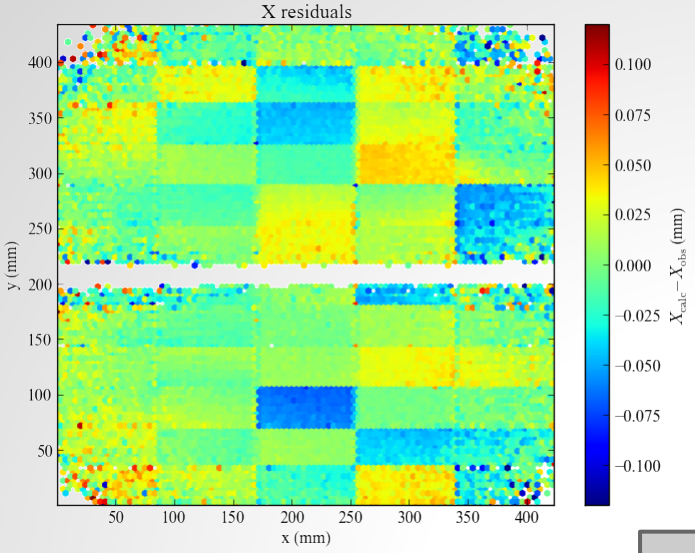


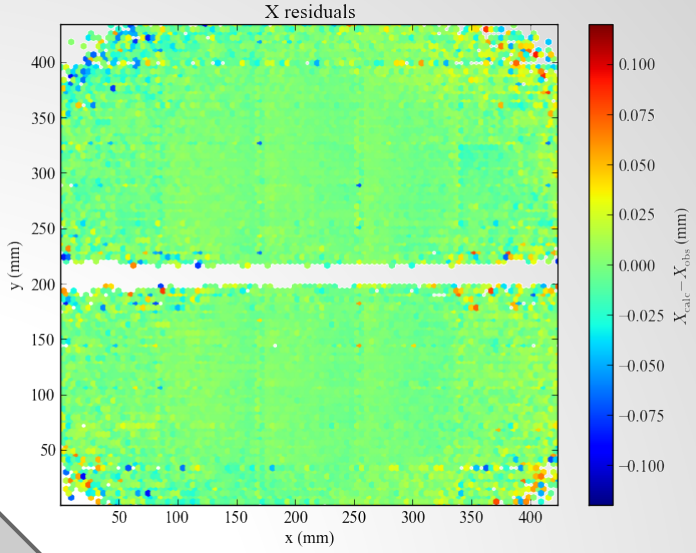




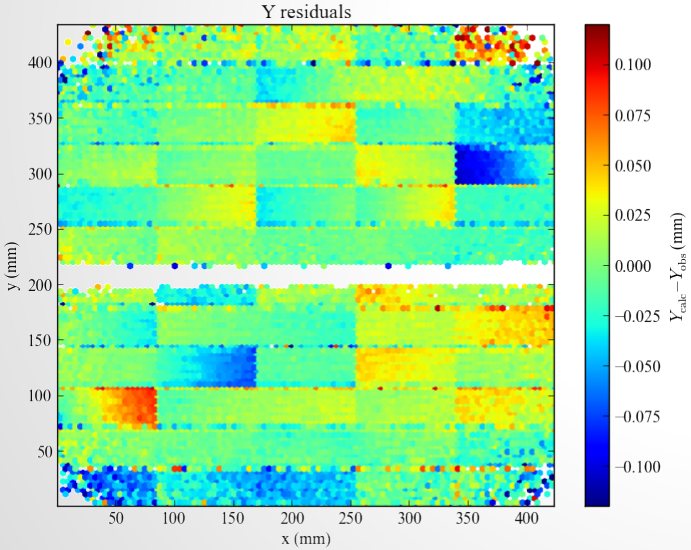
CCP4

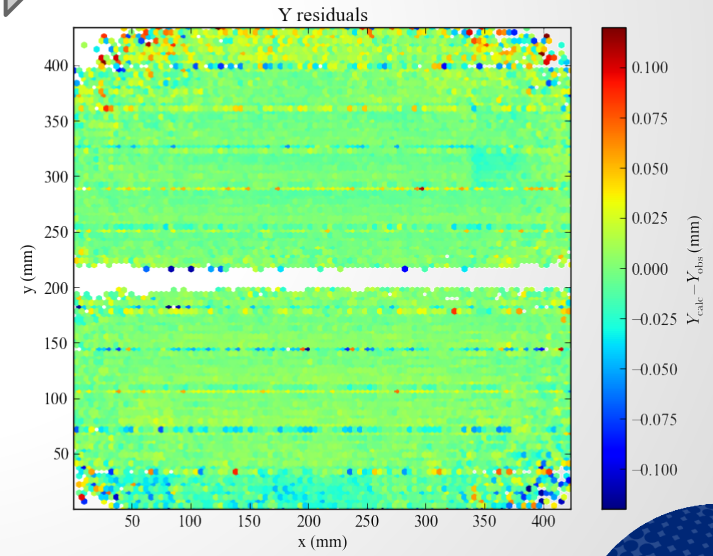
DIALS
Diffraction Integration for Advanced Light Sources





refinement



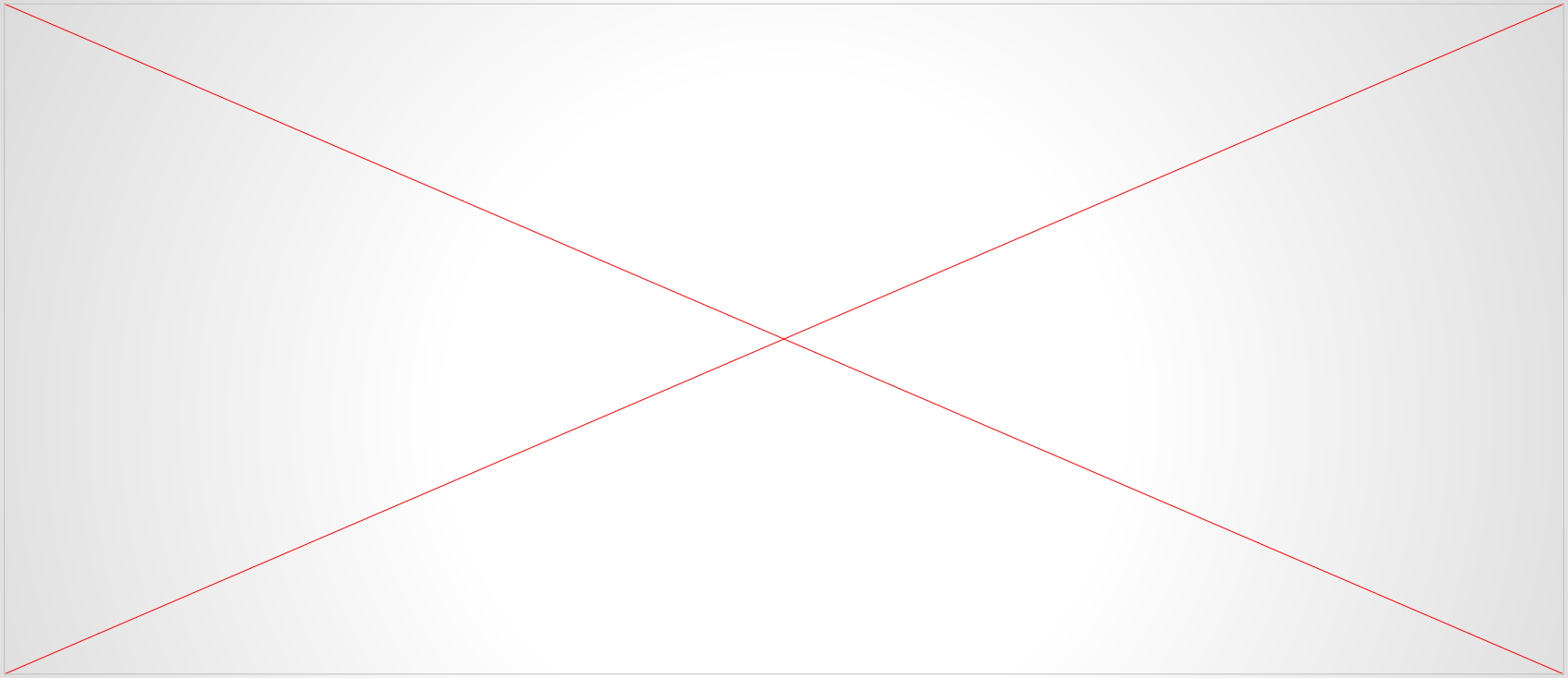


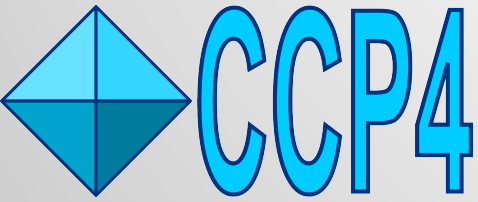
CCP4

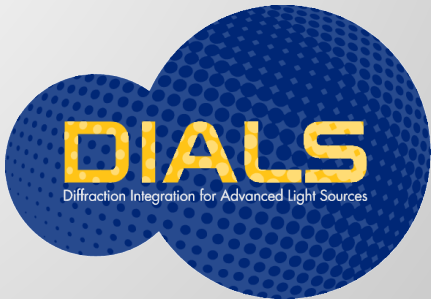
DIALS
Diffraction Integration for Advanced Light Sources

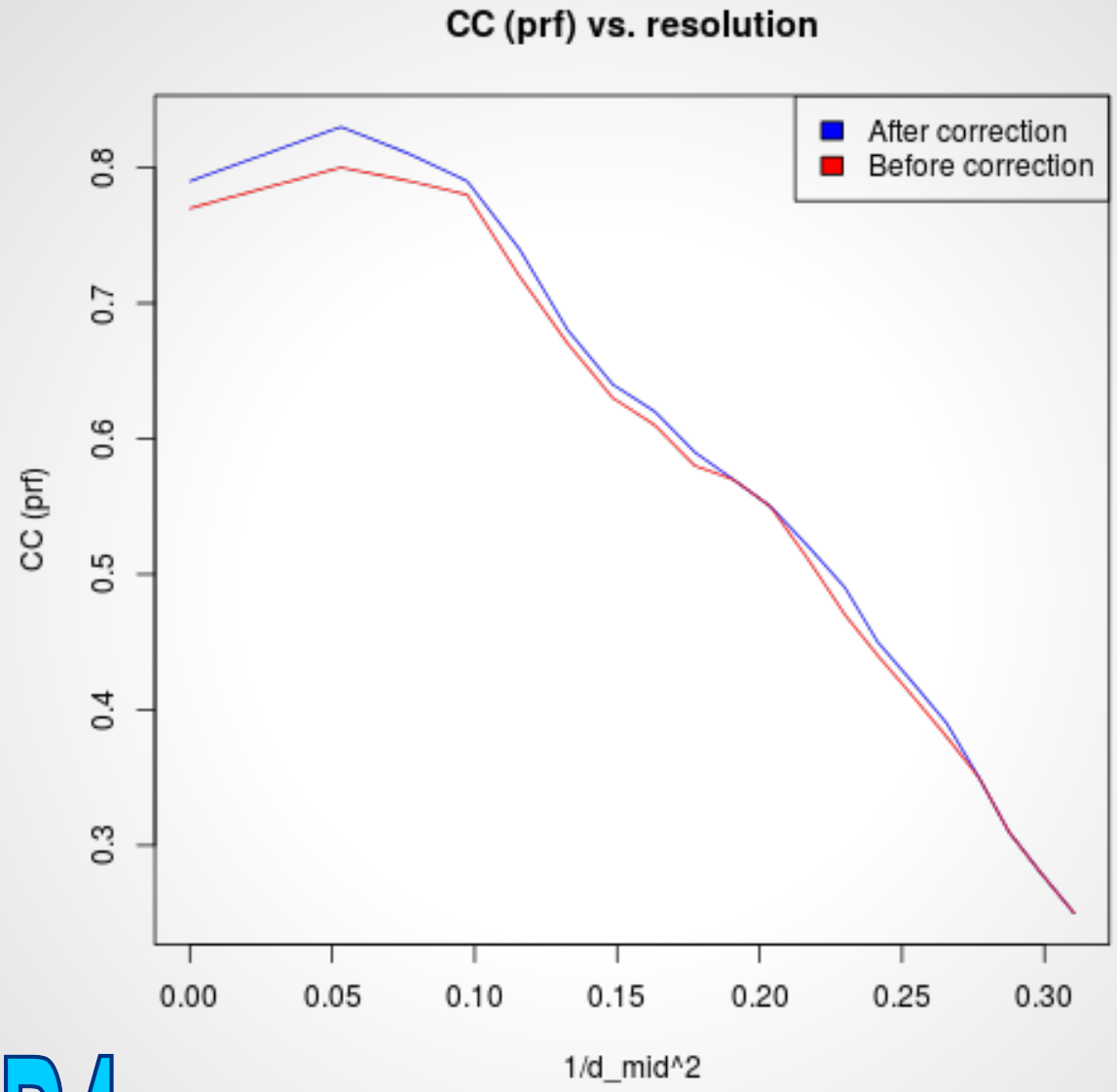
Before correction

After correction









CCP4

DIALS
Diffraction Integration for Advanced Light Sources

Before correction

	Overall	InnerShell	OuterShell
Low resolution limit	150.00	150.00	1.83
High resolution limit	1.79	8.97	1.79
Rmerge (within I+/I-)	0.057	0.036	0.147
Rmerge (all I+ and I-)	0.058	0.037	0.169
Rmeas (within I+/I-)	0.059	0.037	0.190
Rmeas (all I+ & I-)	0.059	0.038	0.195
Rpim (within I+/I-)	0.013	0.007	0.118
Rpim (all I+ & I-)	0.009	0.006	0.094
Rmerge in top intensity bin	0.036	-	-
Total number of observations	631003	9164	594
Total number unique	18352	271	147
Mean((I)/sd(I))	48.7	69.4	7.5
Mn(I) half-set correlation CC(1/2)	1.000	1.000	0.948
Completeness	74.4	99.9	10.7
Multiplicity	34.4	33.8	4.0
Anomalous completeness	73.7	100.0	8.6
Anomalous multiplicity	18.3	25.0	2.3
DelAnom correlation between half-sets	0.054	0.563	-0.182
Mid-Slope of Anom Normal Probability	1.047	-	-

After correction

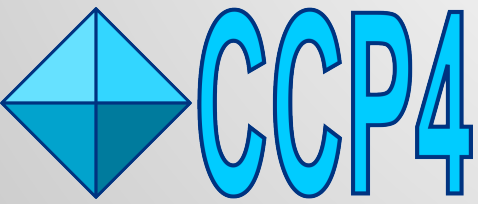
	Overall	InnerShell	OuterShell
Low resolution limit	150.00	150.00	1.83
High resolution limit	1.79	8.97	1.79
Rmerge (within I+/I-)	0.057	0.037	0.154
Rmerge (all I+ and I-)	0.058	0.037	0.173
Rmeas (within I+/I-)	0.059	0.037	0.198
Rmeas (all I+ & I-)	0.059	0.038	0.198
Rpim (within I+/I-)	0.013	0.007	0.121
Rpim (all I+ & I-)	0.009	0.006	0.094
Rmerge in top intensity bin	0.036	-	-
Total number of observations	631669	9225	601
Total number unique	18355	271	147
Mean((I)/sd(I))	48.7	69.6	7.5
Mn(I) half-set correlation CC(1/2)	1.000	1.000	0.959
Completeness	74.5	99.9	10.7
Multiplicity	34.4	34.0	4.1
Anomalous completeness	73.7	100.0	8.6
Anomalous multiplicity	18.3	25.3	2.4
DelAnom correlation between half-sets	0.059	0.394	-0.237
Mid-Slope of Anom Normal Probability	1.045	-	-

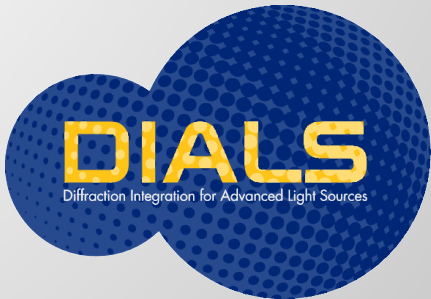
CCP4

DIALS
Diffraction Integration for Advanced Light Sources

Tools for large refinement jobs

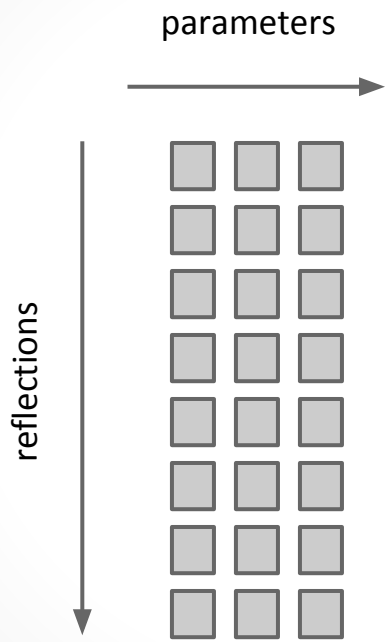
- Normal matrix methods scale as $O(mn^2)$
- The defaults of *dials.refine* are not well-suited to large jobs
- Two approaches
  1. Use a quasi-Newton method such as L-BFGS
  2. Partition the problem: hybrid Refiner
- Note that gradient calculations are expensive, but calculations per reflection are independent
- *dials.refine* optionally allows the use of a memory-efficient L-BFGS engine with gradient calculations in parallel blocks





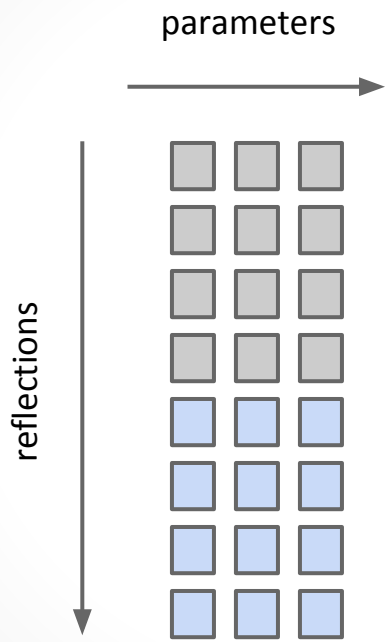
Tools for large refinement jobs

Calculation of the whole Jacobian



Tools for large refinement jobs

Calculation of the Jacobian in parallel blocks

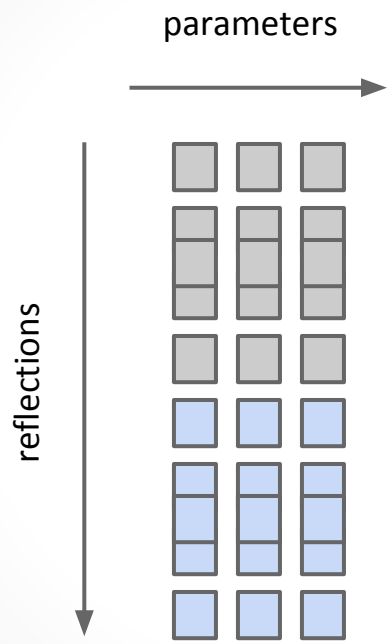


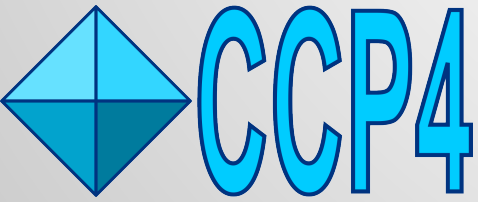
CCP4

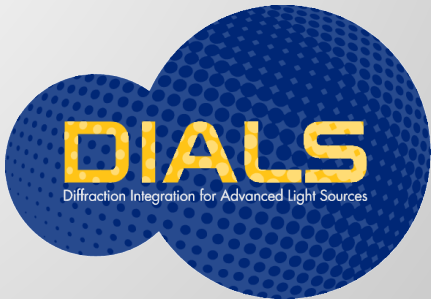
DIALS
Diffraction Integration for Advanced Light Sources

Tools for large refinement jobs

Calculation of individual dL/dp in parallel (for L-BFGS)





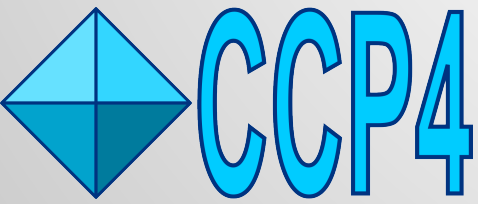


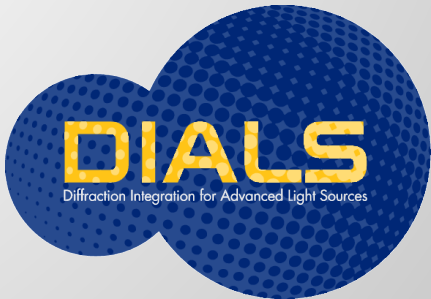
Tools for large refinement jobs

- Only really for the largest jobs
- Each step has a serial component, which is a bottleneck
- The `parallel_map` overhead may be significant
- But we can now tackle large problems with parallel L-BFGS
- Example results for P6M 60 panel metrology:

Engine	nproc	steps	memory	time
LevMar	1	16	2.9 GB	9m59s
LevMar	7	16	~3.4 GB	10m15s
LBFGScurvs	1	119	1.3 GB	22m13s
LBFGScurvs	7	119	?	10m16s

- Need larger jobs still before this starts to shine, such as multi-crystal joint refinement, e.g. XFEL stills





Acknowledgements

See Graeme's slides!

However, I would like to emphasise special thanks from me to Andrew Leslie and Phil Evans from the MRC LMB, Cambridge.