Agenda

Meeting title:	CCP4 Working Group 2 meeting	
Date:	Wednesday 27th April 2022	**Time:** 09:00 – 12.00
Location:	Zoom	
Circulation:	ccp4wg2@stfc.ac.uk	
Attending:	James Beilstein, Nick Devenish, Paul Emsley (PE), Luis Fuentes-Montero, Mike Hough, David McDonagh, Randy Read, David Waterman (DW), Keith Wilson (KSW), Graeme Winter (GW) ← please add all attendees	

We will break at least 5 mins around the full hour

1. DIALS updates
    a. new team members, releases
    b. conda (Graeme Winter and Nick Devenish)
    c. Neutrons (David McDonagh)
    d. Electrons (David Waterman)
2. Feature presentations
    a. xia2 for stills (James Beilstein)
    b. coot v.1 (Paul Emsley)
3. DUI
    a. DUI WG / DUI 2 / web DUI (Luis Fuentes Montero)
4. Take note of the date for the next meeting (proposal 25.5.2022)
5. AOB

Minutes

1. **DIALS updates (Graeme Winter)**

    a. **Team changes, releases (Graeme Winter, Nick Devenish)**

Team changes: Rohan and Elena were joining since last WG2 meeting, Nick is moving over to Diamond MX analysis team.

Summary of releases:

- major releases every run (3.5, 3.6 etc.)
- point releases during the run maybe every 2 weeks
- currently have two release trees - 3.9.x and 3.8.x LTS which supports Python 3.7 - back port bug fixes to end 2022 for CCP4
- 3.9.x and later on Python 3.8
- derive support schedule following NEP-29 (numpy deprecation policy) - we are slightly trailing this
- decent release notes thanks to news fragments etc. using town crier via PR

Packaging improvements: we are trying to make things simpler for end users etc. This is held back by libtbx build system. CCP4 also did a lot of work to support Windows but the core group struggle to provide 1st party support.

    b. **Conda (Nick Devenish)**

Nick presented some detail on what conda is. Now you can conda install DIALS \o/. But conda is very slow → use mamba. Building conda / mamba package required also using cmake for the builds - but that should help CCP4. Also required some re-arrangement of the dials, dxtbx and xia2 source trees. Developer builds still use cctbx build system. Now we have conda support across usual platforms even M1 mac builds.

Questions about libtbx build systems - still an open discussion. Express enthusiasm for moving to a numpy based system, also now able to use DIALS as a library in a sane install.

    c. **Neutrons (David McDonagh)**

There is growing use of neutron TOF with macromolecules, but this is really struggling for software. We are still using Daresbury Laue suite! New instruments are also being developed … why not mantid? One package required to support data analysis for both. DIALS is designed to be flexible and can be used as a toolkit to start development. There is also an interest in analysis of non-monochromatic X-ray scatter. David describes SXD at ISIS (small molecule instrument, but ideas are relevant for macromolecules).

Every pixel records TOF as well but the data are binned upstream so we get histograms. [DIALS also does a good job of complex experimental geometry e.g. i23]. What is hard doing this work? Non-monochromatic, 3rd data dimension is TOF - does not fit neatly across the existing data models, so have had to extend beam, scan model etc. This involved generalising some of the ideas. With those changes we can use the image viewer, spot finding, geometry viewer etc. Showed some nice examples of how the data are processed, seems to work as well as or better than SXD2001 which is the current benchmark.

Discussion on the TOF distribution, diffuse tails etc. as well as making it easier to move data to and from DIALS, to find spots in DIALS, index in mantid and so on. For indexing the dials algorithms actually work well, even without prior, and is much more robust than the existing software. Have specific refinement for TOF data allows detector to move, unit cell to change, but is not refining against the time of flight which may improve the results in future. Integration is largely following the current model, though not quite the same as

the Kabsch transformation (very rotation scan specific). For NaCl there are few spots; straight summation is better than profile fitting right now. Present work is on scaling improvements, extending integration and refinement, as well as some additional GUI elements. Code on a fork, being slowly merged into DIALS.

d. Neutrons and Electrons (David Waterman)

"From one massive particle to another!" The technique is 3DED or μED and depends which lab you come from. Continuous rotation monochromatic data collection on nanocrystals. You can collect useful data from tiny samples - so maximises data from minimal number of samples.

Useful data possible from under 1 million unit cells (of lyozyme…) CPV - XFEL needed 20,000+ samples, ED needed 20. ED also sensitive to H positions, as the electrons are scatted by the electric fields, and for the same reason you can see the charges on side chains etc. as well as determining absolute structures.

The wavelength is very short so the Ewald sphere is very flat - 1/40 Å or so equivalent wavelength - also have to consider multiple scattering, no longer in the kinematic diffraction regime, need to consider dynamic diffraction, but at the moment we are ignoring the problem and just accepting the errors.

Data collection hardware also not designed for these experiments - continuous rotation is not done well, unlike X-ray goniometers. Is very much work in progress now, e.g. for HEXI beamline at Diamond. Lots of data format headaches too. Things are improving though.

Further discussion of experimental issues, stability. Being able to use Friedel pairs to determine the beam centre is nice.

RR: It sounds like the beam centre could be determined very simply by inverting the diffraction image and then doing a phased translation function to see where the original and inverted images superimpose best

Several features added to DIALS to address some of these challenges e.g. "threshold.algorithm=radial_profile" and a truncated distribution spot finding method. Discussion on scaling for dynamic diffraction etc.

2. Feature presentations

a. xia2 for stills (James Beilstein)

James Beilsten-Edmands xia2 for stills. Tools for processing still data are much less well established than rotation crystallography. Lots of work ongoing - aim of this project is to develop some new tools / pipelines for end users to process SSX data - not aiming at XFEL right now. Two parts to the work - make a set if dials programs so you can follow the usual sequence for rotation data, but for stills. Given these tools you can then automate the processing in the style of xia2 the pipeline also gives a framework for testing new ideas later. Experimental pipeline exists xia2.ssx which is not yet released.

Main changes w.r.t. usual dials processing are replacing dials.index with dials.ssx_index and the same with integrate. The names may change before release but these are the working names. The integration includes "stills" algorithms (same as still process) and the "ellipsoid" method based on James Parkhurst's Ph.D. research - assumes a 3D Gaussian profile, which includes a ML estimate of the spot shape, to give a better estimate of partiality etc. - should give improved estimates, which was the motivation for getting started with the development of the pipeline.

For data reduction we already have the tools from multiplex - cosym, scale & merge - though defaults may need to be different. Providing a reference PDB file can help make the scaling more stable.

The pipeline is currently split into indexing, integration as the first part and then reduction as the second. The pipeline refines the geometry on a subset of the data, before proceeding to re-process all of the data with this improved geometry - which means that the bulk processing is nicely parallel. By default the processing

works on blocks of 1000 images. Retains a focus on minimal amount of user tweaking, just a few key parameters. Data reduction: pre-filtering, reindexing, scaling - including cosym if any chance of ambiguity. Does not currently automatically determine the symmetry. Current challenges - cosym with huge data sets, more filtering, dealing with noisy data and improving the integration, though ellipsoids works pretty well.

Discussed some test cases - VMXi massive grid scans, i24 data - wall clock time 10 minutes - couple of hours - decent results from MR + refinement. So the basic structure works, feeding in CuNiR data which has indexing ambiguity, working in batches (resolve within batches then between batches - handling all data in one go is work in progress). Obviously more challenging data sets show scope for further improvement.

Questions: Mike H - any chance of really quick and dirty processing for real-time feedback? Just off spot finding could be useful. Richard: we are looking at this pretty actively right now.

Ben: what if you have different crystal forms? Examples would make that easier to investigate :-)

b. coot v.1 (Paul Emsley))

Coot updates - 1.0 version! 18th birthday of the program … with a new interface and new rendering. The upgrade to Python 3 made this necessary. This also triggered change to GTK3 and openGL 3.3 and was a substantial body of work … while keeping the 0.9 series alive, requiring more threading etc. and sometimes merges of branches.

Undertaken were Python 3 porting, Unicode, modules, code reviewing and a good deal of compilation fun. Changing the way the GUI generation works had impacts on the wider GUI code. Updated graphics from 1990's to something a little more recent. In retrospect, greenfield development might have been better.

Many tutorials have been crated, with more in the works (https://www.youtube.com/watch?v=Xhonm4K1y0c). Coot has become more game-like … also integrated with Blender. Working on improving performance for MacBook machines. On Pauls Computer Coot 1.0 is a lot faster than the 0.9 series.

There is a general layout change, moving from individual windows to full-screen. A lesson learned from users not following the tutorial instructions is to write output to console. The code write output to the console is easy, changing to internal logging and using warning, status-bar or in-graphics representation of those messages is a lot harder. Now adding console etc. to application itself.

Improved the lighting, surfaces etc. in the rendering. Can also add models for things to make it more accessible (colour blindness). Graphics are much better for big complicated molecules now. New electron density is more pretty too, solvent cavities are much better visible etc.

Dynamic Ramachandran plots were integrated.

Discussion on using Blender for rendering pretty pictures for ribbon diagrams. Updating density maps as you change the models is very useful. Qt and web assembly, so make coot headless and then put web interface over top.

KSW: when will it be in ccp4?
PE: CCP4 will be able to add this version to the CCP4 Suite now. However, currently the performance on Mac OS is preventative. Discussions with Charles ensue.

3. DUI (Luis Funetes-Monetero).

Plan for the development of DUI2. The most important change is the chronological order of features that will be programmed in DUI2. An updated version (DUI2) should still be released by November 2022. But instead of adding the server-client feature will be prioritized the reciprocal lattice viewer, this way the new version will

be capable of doing everything that the previous version did plus the multi-crystal and multi-tasking features. The server-client feature should be finished by the end of 2023.

Short demonstration of DUI2 running in local mode. The new more visible differences between DUI1 and DUI2 shown in this demo are:

- A new control graph capable of forking like the old one but also capable of merging
- Ability to run a new command without waiting for the current command to finish
- Peculiarities of image viewer as it should be capable of rendering images remotely

After the demo there were questions about how to handle the responsiveness of the GUI while the same computer is busy processing data, continued by some discussion about how to transfer images through the network and the speed of data transfer.

4. **Take note of the date of the next meeting.**

Proposed date: proposal 25th May 2022, online.

5. **AOB**

None.