Global Phasing's Perspective re. MXCuBE: Optimising and Driving Synchrotron Experiments as a Third Party

Gerard Bricogne,
Claus Flensburg, Peter Keller,
Wlodek Paciorek, Clemens Vonrhein

Global Phasing Ltd., Cambridge, UK

Subtext: Introducing Transferable Expertise into Automated Data Collection

- The ever-increasing speed of MX beamline instrumentation is leading to ever-stronger emphasis being placed on **brevity of execution** as the main design goal for data collection protocols, **often to the exclusion of** other criteria that would aim at **achieving higher data quality**. This can be counter-productive, especially, but not only, for phasing experiments.
- Global Phasing, among others, has been interested in bucking that trend by creating combined capabilities for the **fast design of optimal strategies and the direct supervision of their execution** on an actual beamline.
- Our approach has been to aim for a full "**third-party design and control**" capability rather than for separate add-on programs that would need to be invoked by local software on each specific beamline or group of beamlines running under the BCS.
- To **make this capability as transferable as possible** across the huge diversity of beamline instruments and BCSs, **finding the correct level of abstraction** for all the components and processes involved is of paramount importance.

What expertise do we need to capture?

- Achieving completeness in spite of all impediments
  - e.g “filling the cusp” for low-symmetry samples
- Achieving maximum resolution and uniform data quality
  - e.g. making fullest use of each available sample
  - using a multi-axis goniometer to collect data in multiple orientations
- Eliminating systematic errors
  - e.g. taking advantage of partial cancellation of RD effects by using interleaved strategies and multi-axis goniometry

Biggest impediments to capturing expertise and invoking it in automation

- Time and high-throughput pressures
- Expediency: urgency vs. importance
- Need for multiple improvements at consecutive steps before the benefits of any one of them can become visible
- Fragmentation of the necessary efforts by lack of uniformity across beamlines

An opportunity worth waiting for

- We had been interested in these areas of development for the past 10 years and were involved in several collaborative efforts with synchrotrons (BIOXHIT, EDNA).
- However these were strongly oriented towards ever greater speed and automation for very simple data collection protocols, rather than towards advanced, expert experiments (phasing, but also conventional).
- The invitation from DLS to a co-funded collaboration on I23 finally gave an opportunity to implement our ideas, with the prospect of driving actual experiments.

Overview of I23 Collaboration

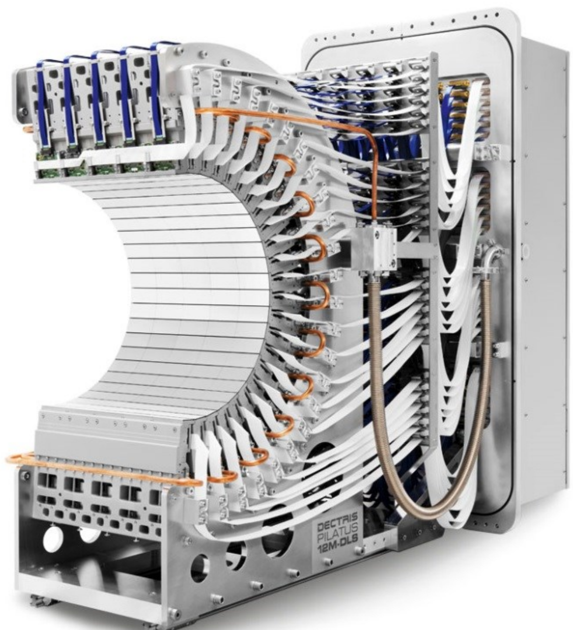
- **Area 1**: Contribution to the development and testing of a version of XDS supporting data collection by the rotation method on the semi-cylindrical Pilatus 12M detector

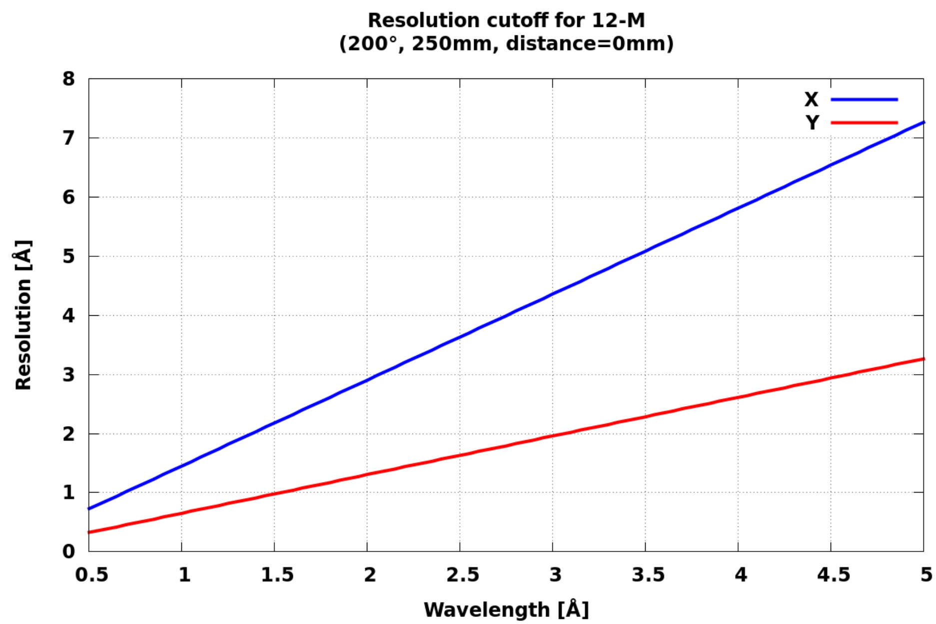
- **Area 2:** Incorporation of the semi-cylindrical geometry of the Pilatus 12M detector and of the collision and shadowing maps for the sample holder and the ATC kappa goniometer into the design of optimal single-sample single-wavelength data collection strategies

- **Area 3:** Connecting the Strategy Design Program ("SDP") to the GDA experiment control software (assumed to have been upgraded to deal with the Pilatus 12M detector and the ATC kappa goniometer) to direct single-wavelength single-sample experiments, to be invoked by user selection

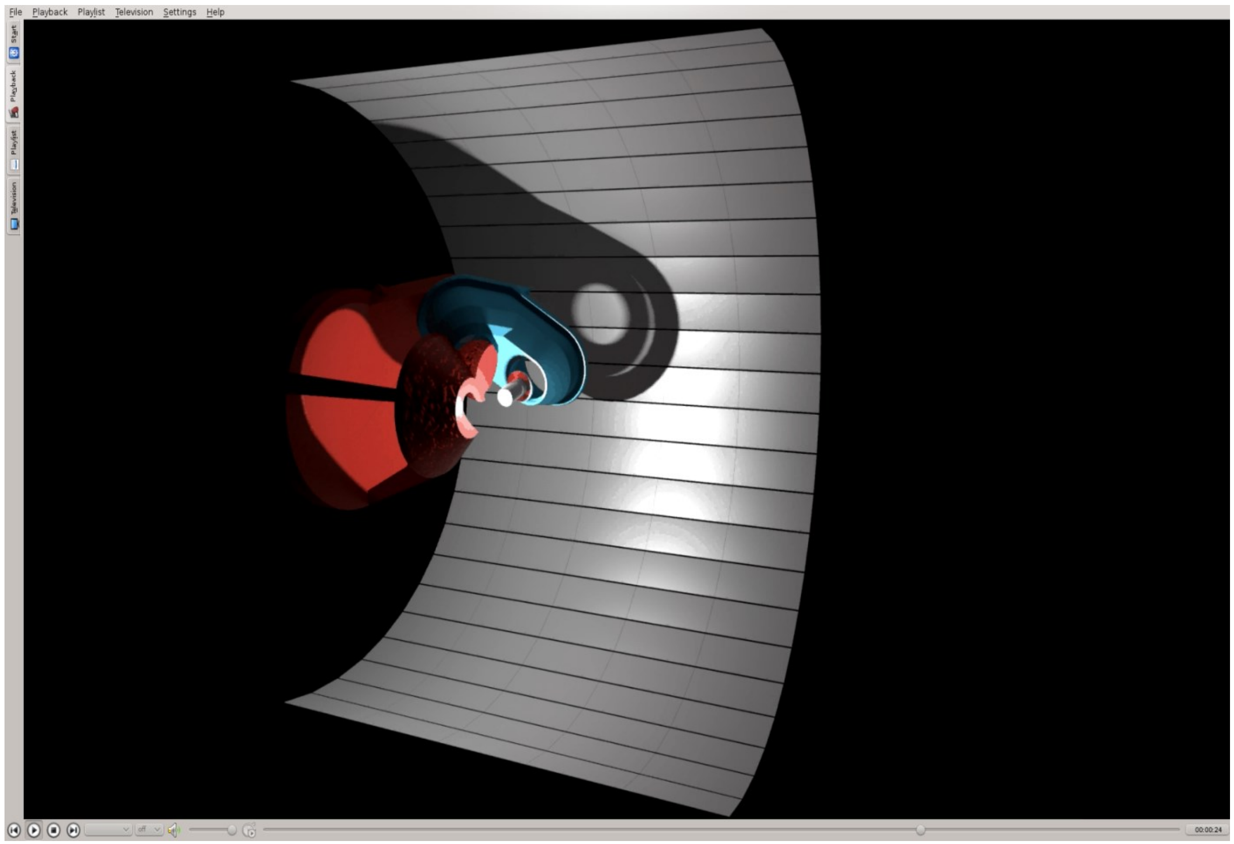
- **Area 4:** Incorporation of interleaved multi-wavelength protocols and of multiple-sample management into the SDP and its interface to GDA

The Pilatus 12M aspect ratio is a hindrance to achieving completeness to a given resolution: multi-orientation strategies are mandatory





Moving shadows are unavoidable



Intercepting shadowed reflections before they confuse scaling/merging

Only in INTEGRATE.HKL (marked as MISFITS in XDS_ASCII.HLK)   In FILTER.HKL   Output in XDS_ASCII.HKL (excluding MISFITS)

EFFECT OF FILTER.HKL ON THE PROCESSING OF 3NOS SHADOWED IMAGES

Without filtering of shadowed reflections:

===== Final processing of data with XDS

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
6.40	11434	2196	2286	96.1%	26.7%	46.5%	11412	2.08	30.1%	74.3*	-11	0.447	1703
4.53	19686	3822	4010	95.3%	35.0%	46.8%	19660	1.85	39.4%	66.5*	-12	0.526	2927
3.70	21972	4512	5183	87.1%	37.7%	46.6%	21933	1.71	42.4%	64.2*	-36	0.399	3211
3.21	22167	4736	6124	77.3%	38.9%	46.7%	22110	1.68	43.6%	67.5*	-50	0.322	3175
2.87	22109	4859	6916	70.3%	37.5%	48.0%	22051	1.55	42.1%	69.3*	-52	0.313	3028
2.62	21959	4991	7628	65.4%	41.5%	54.8%	21909	1.22	46.5%	68.9*	-55	0.296	2858
2.43	21850	5071	8299	61.1%	46.3%	77.0%	21783	0.82	51.8%	68.6*	-50	0.286	2673
2.27	23300	5345	8901	60.0%	55.7%	153.8%	23244	0.44	62.8%	67.4*	-40	0.262	2762
2.14	23244	5434	9483	57.3%	99.0%	443.2%	23140	0.13	112.9%	53.5*	-17	0.233	2578
total	187721	40966	58830	69.6%	34.9%	48.5%	187242	1.18	39.2%	75.1*	-36	0.340	24915

With filtering of shadowed reflections:

===== Final processing of data with XDS

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
6.40	9711	2066	2286	90.4%	0.3%	0.3%	9592	310.15	0.3%	100.0*	-8	0.702	1343
4.53	16804	3762	4010	93.8%	0.4%	0.4%	16656	246.18	0.5%	100.0*	-4	0.757	2326
3.70	18978	4480	5183	86.4%	0.6%	0.6%	18878	188.49	0.6%	100.0*	2	0.793	2750
3.21	19292	4702	6124	76.8%	1.0%	1.0%	19196	108.24	1.1%	100.0*	0	0.818	2906
2.87	19357	4834	6916	69.9%	2.0%	2.0%	19283	52.80	2.4%	100.0*	-1	0.819	2845
2.62	19225	4978	7628	65.3%	4.5%	4.4%	19151	24.51	5.2%	99.8*	-3	0.806	2697
2.43	19190	5055	8299	60.9%	9.8%	9.7%	19101	11.69	11.5%	99.0*	1	0.816	2557
2.27	20559	5330	8901	59.9%	24.7%	24.8%	20497	5.13	28.7%	94.5*	-3	0.776	2657
2.14	21103	5411	9483	57.1%	75.3%	76.9%	20985	1.74	87.6%	66.0*	-2	0.691	2461
total	164219	40618	58830	69.0%	1.1%	1.1%	163339	83.54	1.2%	100.0*	-1	0.782	22542

Current status

- The **goal of providing of a complete third-party workflow** for single-sample single-wavelength experiments with optimal kappa/P12M strategy design **has been achieved** (although: sign-off by emulation)
- Funding has been approved by Diamond to support a **continuation and extension of our collaboration** towards
  - Multi-sample multi-wavelength experiments
  - Full-instrument calibration **active on I04**
  - Deployment on other Diamond MX beamlines **active on I04**
  - Adoption and further elaboration of our Abstract Beamline Interface within GDA itself
  - Implementation by GPhL of ray-tracing based integration in DIALS
- The joint paradigms of **third-party control of experiments**, of **development by emulation** and of an **Abstract Beamline Interface** have been validated and have formed the basis of our parallel (although, so far, small) contribution to the MXCuBE collaboration.

Outreach: beyond I23 and Diamond

- We have conducted these developments from the start with the deliberate imperative in mind that they should allow **maximum transferability**, in particular
  - simulation, strategy design and data processing make use of **generic descriptions of all the components** that are readily adaptable to describe other instruments, e.g. other kappa-capable beamlines at Diamond with Pilatus 6M detectors
  - specificities in the communication with BCS are carefully confined to whatever is the equivalent of the MessageBus to GDA (e.g. the XML-RPC server in MXCuBE) by reference to an **"Abstract Beamline" model**
- Global Phasing has **close working connections with most other European synchrotrons**, through the MXCuBE collaboration (comprising ESRF, SOLEIL, PETRA (EMBL, DESY), BESSY, ALBA and MAX-IV) and directly with the SLS (joint post-doc for 2 years starting on July 1st). **Initial contacts have been made with IMCA-CAT and the ALS**
- Both factors will **ease the dissemination of our I23 developments**

Conclusions

- The goal of capturing expertise in transferable form can be achieved through
  - Instrument-generic applications (cf EEC Workshop) for simulation, prediction and strategy design
  - A Workflow capable of controlling the execution of a designed strategy by communicating with the BCS in a generic manner thanks to the Abstract Beamline Interface
  - A development methodology using Emulation
- The actual transfer of these developments requires a minimal but non-zero contribution from each family of synchrotron beamlines sharing a common BCS

Acknowledgements within Global Phasing and for funding

- Wlodek Paciorek and Claus Flensburg (simulation, prediction of shadowing, strategy design)
- Peter Keller (persistence layer, workflow, message bus)
- Clemens Vonrhein (feature integration, testing, processing)
- The Global Phasing Consortium for funding
- The Diamond Light Source for supplementary funding and for the opportunity to build a complete workflow for I23

External acknowledgements
for synchrotron experiments and related matters

- Meitian Wang, Ezequiel Panepucci, Vincent Olieric, Sandro Waltersperger (SLS)
- Andrew Thompson, Pierre Legrand (SOLEIL)
- Wolfgang Kabsch for pre-release versions of XDS, prompt response to enquiries and feedback, and the “filter” feature
- Armin Wagner and his I23 beamline co-workers
- The MXCuBE collaboration