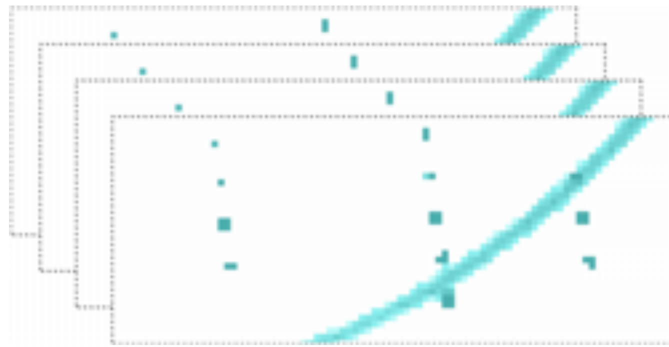


Managing Crystallographic Data in Facilities Using Integrated CIF, HDF5 and NeXus

Herbert J. Bernstein
Dowling College, Shirley, NY, USA

providing an update on work previously reported as

"**Coping with BIG DATA Image Formats: Integration of CBF, NeXus and HDF5**,"
by H. J. Bernstein, J. M. Sloan, G. Winter, T. S. Richter, NIAC, COMCIFS, poster T-16, meeting of the American Crystallographic Association, Honolulu, HI, USA, 20 – 24 July 2013

Improving the Management of Experimental Data, 25 Aug 2013

Room B2.02, Sciences Building, University of Warwick, Coventry, UK

Crystallographic Information and Data Management

A Satellite Symposium of the 28th European Crystallographic Meeting

25 - 29 August 2013

Work supported in part by NIGMS, DOE, NSF, PaNdata ODI (EU 7th Framework Programme)

Introduction

- The BIG DATA demands of the new generation of X-ray pixel array detectors necessitate the use of new storage technologies as we meet the limitations of existing file systems. In addition, the modular nature of these detectors provides the possibility of more complex detector arrays which in turn requires a complex description of the detector geometry.

Introduction

- The BIG DATA demands of the new generation of X-ray pixel array detectors necessitate the use of new storage technologies as we meet the limitations of existing file systems. In addition, the modular nature of these detectors provides the possibility of more complex detector arrays which in turn requires a complex description of the detector geometry.
- Taken together these give an opportunity to combine the best of CBF/imgCIF (the Crystallographic Binary File), NeXus (a common data format for neutron, x-ray and muon science) and HDF5 (Hierarchical Data Format, version 5) for the management of such data at synchrotrons.

Introduction

- The BIG DATA demands of the new generation of X-ray pixel array detectors necessitate the use of new storage technologies as we meet the limitations of existing file systems. In addition, the modular nature of these detectors provides the possibility of more complex detector arrays which in turn requires a complex description of the detector geometry.
- Taken together these give an opportunity to combine the best of CBF/imgCIF (the Crystallographic Binary File), NeXus (a common data format for neutron, x-ray and muon science) and HDF5 (Hierarchical Data Format, version 5) for the management of such data at synchrotrons.
- Discussions are in progress between COMCIFS (the IUCr Committee for the Maintenance of the CIF Standard) and NIAC (the NeXus International Advisory Committee) on an integrated ontology.

Introduction

- The BIG DATA demands of the new generation of X-ray pixel array detectors necessitate the use of new storage technologies as we meet the limitations of existing file systems. In addition, the modular nature of these detectors provides the possibility of more complex detector arrays which in turn requires a complex description of the detector geometry.
- Taken together these give an opportunity to combine the best of CBF/imgCIF (the Crystallographic Binary File), NeXus (a common data format for neutron, x-ray and muon science) and HDF5 (Hierarchical Data Format, version 5) for the management of such data at synchrotrons.
- Discussions are in progress between COMCIFS (the IUCr Committee for the Maintenance of the CIF Standard) and NIAC (the NeXus International Advisory Committee) on an integrated ontology.
- A proof-of-concept API based on CBFlib and the HDF5 API is being developed in a collaboration among Dowling College, Brookhaven National Laboratory and Diamond Light Source. A preliminary mapping and a combined API are under development.

Introduction (ctd)

- The new generation of high performance x-ray detectors requires integration of HDF5, NeXus and CBF.
- The DECTRIS workshop in Baden, Switzerland in January 2013 established the parameters of the integration.
- A collaboration has been working on specifications and code.

Introduction (ctd)

- The new generation of high performance x-ray detectors requires integration of HDF5, NeXus and CBF.
- The DECTRIS workshop in Baden, Switzerland in January 2013 established the parameters of the integration.
- A collaboration has been working on specifications and code.
- CBFlib 0.9.2.12
  - Can store arbitrary CBF files in HDF5 and recover them.
  - Supports use of all CBFlib compressions in HDF5 files.
  - Provides minicbf2nexus to convert sets of minicbf files to a single NeXus file.
- A draft concordance between MX CBF and NeXus has been prepared.
- Updated CBF dictionary has been prepared.
- There is much work still to be done -- collaborators welcome.

Interactions among CBF, HDF5 and NeXus

- CBF remains CBF, HDF5 remains HDF5, NeXus remains NeXus
- Each gains in functionality from interoperable mappings
- New dictionaries and extensions to existing dictionaries will help in documenting the mappings.
- DDLm is useful in providing clear documentation and portable implementations of mappings, such as between CBF laboratory coordinate system and NeXus McStas coordinate system
- Applications gain from extension to the APIs, starting with CBFlib.
- HDF5 and NeXus users gain CBFlib compressions.
- CBF users gain HDF5 compressions.

Where to Find Software and Documentation

- Draft imgCIF/CBF version 1.7 dictionary that now includes information on going from CBF to NeXus:
  https://www.sites.google.com/site/nexuscbf/home/cbf-dictionary

- PDF summary of the concordance:
  https://www.sites.google.com/site/nexuscbf/mapping-draft

- CBFlib kit:
  http://downloads.sf.net/cbflib/CBFlib-0.9.2.12.tar.gz
  includes both Jonathan Sloan's utility minicbf2nexus to convert sets of minicbfs into a single NeXus file and a plugin filter that supports the full set of CBFlib compressions in HDF5

Data Rates, Formats & High Performance X-ray Detectors

- CCD X-ray detectors provide images at a moderate data rate of one every few to several seconds. Current higher performance X-ray detectors, such as the DECTRIS Pilatus are capable of collecting six-megapixel images at 10 -- 25 frames per second [True12], while the newest Pilatus3 6M instruments can operate at 100 frames per second.

Data Rates, Formats & High Performance X-ray Detectors

- CCD X-ray detectors provide images at a moderate data rate of one every few to several seconds. Current higher performance X-ray detectors, such as the DECTRIS Pilatus are capable of collecting six-megapixel images at 10 -- 25 frames per second [True 12], while the newest Pilatus3 6M instruments can operate at 100 frames per second.
- The coming next generation of high performance X-ray detectors for MX such as the DECTRIS Eiger will be capable of collecting 16+ megapixel images at more than 125 frames per second [Will 11, page 6] [John 12].

Data Rates, Formats & High Performance X-ray Detectors

- CCD X-ray detectors provide images at a moderate data rate of one every few to several seconds. Current higher performance X-ray detectors, such as the DECTRIS Pilatus are capable of collecting six-megapixel images at 10 -- 25 frames per second [True 12], while the newest Pilatus3 6M instruments can operate at 100 frames per second.
- The coming next generation of high performance X-ray detectors for MX such as the DECTRIS Eiger will be capable of collecting 16+ megapixel images at more than 125 frames per second [Will 11, page 6] [John 12].
- The ADSC DMPAD [Haml 12] is also expected to produce 900 fine-sliced images in steps of two-tenths of a degree at 125 frames per second.

Data Rates, Formats & High Performance X-ray Detectors (ctd)

- Typical sustained data rates for detectors used for MX at NSLS, Diamond Light Source, etc. compared to expected rates from Eiger, expressed in terms of the typical data rate for an inexpensive USB disk of 25 MB/sec = 200 Mb/sec.

Detector	Raw Image Size (MB)	Frame Rate (Hz)	Comp Rate (Gb/sec)	USB Disk Data Rate (%)
ADSC Q315 (2x2 binned)	18	0.37	0.13	7
Pilatus 2 6M	24	10	0.48	240
Pilatus 2 Fast 6M	24	25	1.2	600
Pilatus 3 6M	24	100	4.8	2400
Eiger 16M	72	125	18	9000

Data Rates, Formats & High Performance X-ray Detectors (ctd)

• Typical sustained data rates for detectors used for MX at NSLS, Diamond Light Source, etc. compared to expected rates from Eiger, expressed in terms of the typical data rate for an inexpensive USB disk of 25 MB/sec = 200 Mb/sec.

Detector	Raw Image Size (MB)	Frame Rate (Hz)	Comp Rate (Gb/sec)	USB Disk Data Rate (%)
ADSC Q315 (2x2 binned)	18	0.37	0.13	7
Pilatus 2 6M	24	10	0.48	240
Pilatus 2 Fast 6M	24	25	1.2	600
Pilatus 3 6M	24	100	4.8	2400
Eiger 16M	72	125	18	9000

• Today for MX alone Diamond Light Source employs one Pilatus 2M, three Pilatus 6M fast and one Pilatus 3 6M, giving a combined data rate of over 1 GB/sec and over 200 files/sec.

•These new detectors are creating the need to manage hundreds of thousands of images being received at rates from sixty megapixels to 2.5 gigapixels per second and beyond.

Data Rates, Formats & High Performance X-ray Detectors (ctd)

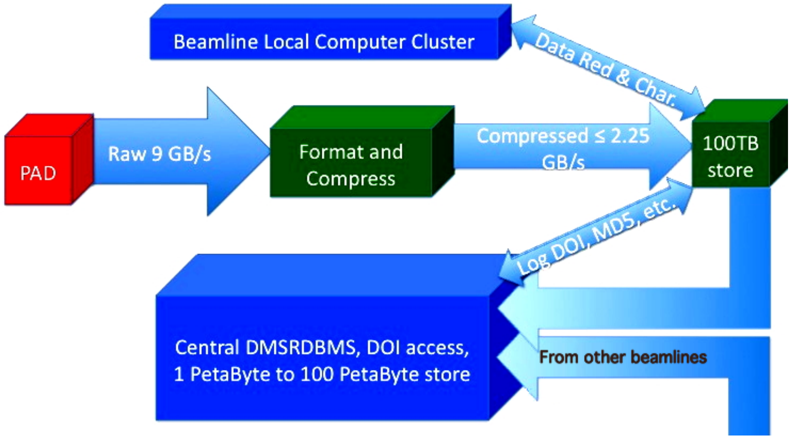
- Today for MX alone Diamond Light Source employs one Pilatus 2M, three Pilatus 6M fast and one Pilatus 3 6M, giving a combined data rate of over 1 GB/sec and over 200 files/sec.
- These new detectors are creating the need to manage hundreds of thousands of images being received at rates from sixty megapixels to 2.5 gigapixels per second and beyond.

Detector	Raw Image Size (MB)	Frame Rate (Hz)	Comp Rate (Gb/sec)	USB Disk Data Rate (%)
ADSC Q315 (2x2 binned)	18	0.37	0.13	7
Pilatus 2 6M	24	10	0.48	240
Pilatus 2 Fast 6M	24	25	1.2	600
Pilatus 3 6M	24	100	4.8	2400
Eiger 16M	72	125	18	9000

- For the Advanced Beamlines for Biological Investigations with X-rays (ABBIX) that are being built for NSLS-II [Hend 12], just two of the beam lines, the Frontier Macromolecular Crystallography (FMX) beamline and the Automated Macromolecular Crystallography (AMX) beamline [Schn 12], are expected to produce an aggregate of more than 94 terabytes per operational half day, 660 terabytes per week or 38 petabytes per year. The anticipated beamline flux is 10^{13} photons per second for FMX and $2*10^{13}$ photons per second for AMX, approximately 50 times the NSLS X25 and X29 fluxes. One subtle effect of these high fluxes is that there will be more photons per pixel in images, making them more difficult to compress.

Data Rates, Formats & High Performance X-ray Detectors (ctd)

- A data management system designed for very large numbers of files as well as for very large data volumes and data rates is needed.
- Efficient recording of metadata coordinated with the data is needed, and database access to information about images and experimental runs is needed.



Major data flows from the beamline advanced pixel array detectors (PADs) to the data management system relational database (DMSRDBMS).

- In order to be manageable, the raw 9 gigabyte per second data flow from each PAD needs to be compressed locally by at least 4:1, before going into a beamline 100 terabyte store for beamline local computer cluster access for up to a week for data reduction and characterization.

Data Rates, Formats & High Performance X-ray Detectors (ctd)

- The required bandwidth of the pipes from the beamlines to the DMSRDBMS depends on the compression used. If no further compression is used, 2.25 gigabyte (18 gigabit) per second per beamline network connections are required. If a combined lossless/lossy compression is used, then 225 -- 450 megabyte (1.8 to 3.6 gigabit) per second per beamline network connections will suffice. The flows for transfers to user home institutions are not shown.
- For MX, these data rates, data volumes, numbers of distinct images and numbers of distinct experiments argue for a very organized, high performance infrastructure.
- HDF5 and NeXus provide the necessary organization of the raw data.
- CBF provides the necessary organization of the associated metadata for subsequent processing
- as well as contributing useful compressions.

Data Rates, Formats & High Performance X-ray Detectors (ctd)

- A final issue, in addition to the actual recording of data, is that of automated processing. At Diamond Light Source and elsewhere there has been a push towards the automated analysis of diffraction data, as interactively processing diffraction data at the current rate of typically 20 data sets per hour is unsustainable. This however places an increased strain on the file system, as typically the same data are read as many as six times in order to be processed, resulting in over 1000 file access operations per second.

- With the storage of many frames per file, as planned for NeXus, the rate of file operations would decrease substantially.

HDF5 and NeXus

- The Hierarchical Data Format Version 5 (HDF5) is a self-describing file format with a robust, well documented API capable of handling (and routinely used to handle) multi-gigabye files of data. It has a diverse user community covering a wide range of disciplines and is fully supported [Doug 09].

HDF5 and NeXus

- The Hierarchical Data Format Version 5 (HDF5) is a self-describing file format with a robust, well documented API capable of handling (and routinely used to handle) multi-gigabye files of data. It has a diverse user community covering a wide range of disciplines and is fully supported [Doug 09].
- NeXus [Filg 01] is a tree-oriented view of HDF5 (and XML and HDF4) of importance in managing neutron and x-ray data. NeXus is a convenient thin layer over HDF5 that is widely used at many physics research centers, including at synchrotrons. Together they provide a portable, extensible and efficient format for the storage and management of data.

HDF5 and NeXus

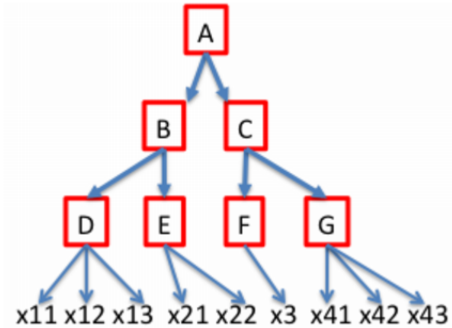
- The Hierarchical Data Format Version 5 (HDF5) is a self-describing file format with a robust, well documented API capable of handling (and routinely used to handle) multi-gigabye files of data. It has a diverse user community covering a wide range of disciplines and is fully supported [Doug 09].
- NeXus [Filg 01] is a tree-oriented view of HDF5 (and XML and HDF4) of importance in managing neutron and x-ray data. NeXus is a convenient thin layer over HDF5 that is widely used at many physics research centers, including at synchrotrons. Together they provide a portable, extensible and efficient format for the storage and management of data.
- HDF5 is particularly well suited to the management of very large volumes of complex scientific data and has been adopted as the primary data-management framework in a wide range of disciplines (http://www.hdfgroup.org/HDF5/users5.html) and provides the "inner workings" of important formats, such as NetCDF [Rew 04] and NeXus.

HDF5 and NeXus

- The Hierarchical Data Format Version 5 (HDF5) is a self-describing file format with a robust, well documented API capable of handling (and routinely used to handle) multi-gigabye files of data. It has a diverse user community covering a wide range of disciplines and is fully supported [Doug 09].
- NeXus [Filg 01] is a tree-oriented view of HDF5 (and XML and HDF4) of importance in managing neutron and x-ray data. NeXus is a convenient thin layer over HDF5 that is widely used at many physics research centers, including at synchrotrons. Together they provide a portable, extensible and efficient format for the storage and management of data.
- HDF5 is particularly well suited to the management of very large volumes of complex scientific data and has been adopted as the primary data-management framework in a wide range of disciplines (http://www.hdfgroup.org/HDF5/users5.html) and provides the “inner workings” of important formats, such as NetCDF [Rew 04] and NeXus.
- While HDF5 is tree-oriented, which is a very powerful and useful characteristic allowing file-system-like nesting of groups of data within groups of data, in order for information to be easily, reliably and efficiently searched, tables are more useful so that the information can be loaded into a relational database management system [Codd 70].

HDF5 and NeXus

- The Hierarchical Data Format Version 5 (HDF5) is a self-describing file format with a robust, well documented API capable of handling (and routinely used to handle) multi-gigabye files of data. It has a diverse user community covering a wide range of disciplines and is fully supported [Doug 09].
- NeXus [Filg 01] is a tree-oriented view of HDF5 (and XML and HDF4) of importance in managing neutron and x-ray data. NeXus is a convenient thin layer over HDF5 that is widely used at many physics research centers, including at synchrotrons. Together they provide a portable, extensible and efficient format for the storage and management of data.



Node	Parent
A	.
B	A
C	A
D	B
E	B
F	C
G	C

Node	Child
A	B
A	C

Node	Child
B	D
B	E

Node	Child
C	F
C	G

Node	Data
D	x11
D	x12
D	x13

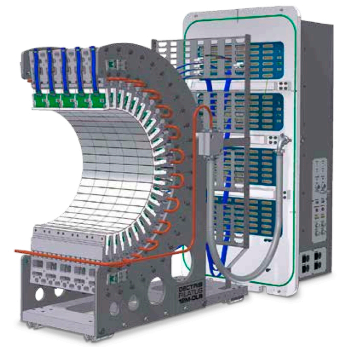
Node	Data
E	x21
E	x22

Node	Data
F	x3

Node	Data
G	x41
G	x42
G	x43

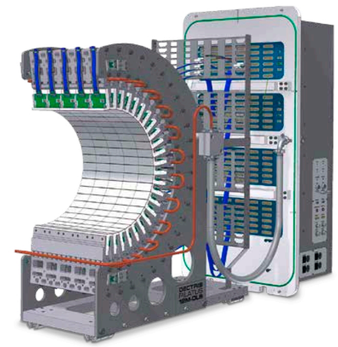
- HDF5 is particularly well suited to the management of very large volumes of complex scientific data and has been adopted as the primary data-management framework in a wide range of disciplines (http://www.hdfgroup.org/HDF5/users5.html) and provides the “inner workings” of important formats, such as NetCDF [Rew 04] and NeXus.
- While HDF5 is tree-oriented, which is a very powerful and useful characteristic allowing file-system-like nesting of groups of data within groups of data, in order for information to be easily, reliably and efficiently searched, tables are more useful so that the information can be loaded into a relational database management system [Codd 70].

CIF Metadata and Database Friendly Tables



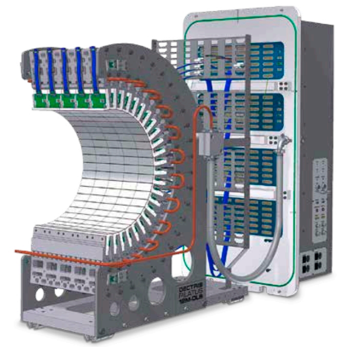
- For this project, organizing data and metadata according to the conventions of the IUCr Crystallographic Information File [Hall 91] using imgCIF [Bern 05] and its open source supporting software CBFlib [Elli 05] provides a database-friendly tabular structure. The imgCIF ontology provides the metadata needed for diffraction images and is supported by all the major detector manufacturers. This aspect is particularly important for instruments with complex geometries, *e.g*. a detector being constructed by DECTRIS for the long wavelength beamline I23 at Diamond Light Source.

CIF Metadata and Database Friendly Tables



- For this project, organizing data and metadata according to the conventions of the IUCr Crystallographic Information File [Hall 91] using imgCIF [Bern 05] and its open source supporting software CBFlib [Elli 05] provides a database-friendly tabular structure. The imgCIF ontology provides the metadata needed for diffraction images and is supported by all the major detector manufacturers. This aspect is particularly important for instruments with complex geometries, a detector being constructed by DECTRIS for the long wavelength beamline I23 at Diamond Light Source.
- The embedding of CIF tables in HDF5 files was demonstrated at the "HDF5 as hyperspectral data analysis format" workshop in January 2010. [Gotz 10] The workshop recommendation was, in part, "... Adopt at as much as possible from imgCIF and sasCIF". There is no conflict with the use of HDF5 in this approach.

CIF Metadata and Database Friendly Tables



- For this project, organizing data and metadata according to the conventions of the IUCr Crystallographic Information File [Hall 91] using imgCIF [Bern 05] and its open source supporting software CBFlib [Elli 05] provides a database-friendly tabular structure. The imgCIF ontology provides the metadata needed for diffraction images and is supported by all the major detector manufacturers. This aspect is particularly important for instruments with complex geometries, *e.g.* a detector being constructed by DECTRIS for the long wavelength beamline I23 at Diamond Light Source.
- The embedding of CIF tables in HDF5 files was demonstrated at the "HDF5 as hyperspectral data analysis format" workshop in January 2010. [Gotz 10] The workshop recommendation was, in part, "... Adopt at as much as possible from imgCIF and sasCIF". There is no conflict with the use of HDF5 in this approach.
- Tables are easily embedded into trees. Going in the other direction is more difficult. There is serious effort required to make general trees into tables suitable for use in a relational database management system, involving a process known as "normalization." [Codd 72].

.

CIF Metadata and Database Friendly Tables

- One of the tasks of this project is to extend the imgCIF ontology to ensure workable database access to metadata in the HDF5 tree that has not already been normalized into CIF categories. For example, DOIs and SHA2 or SHA3 checksums from multiple experiments will need to be brought forward into a common table for searching for forensic validation service support.

- **The database, the DOIs and checksums are important elements in dealing with the issue of fraud in crystallography. See John Helliwell's talk at this meeting.**

CBF and Database Access

- The Crystallographic Binary File (CBF) format is a complementary format to the Crystallographic Information File (CIF), supporting efficient storage of large quantities of experimental data in a self-describing binary format [imgCIF], with a sophisticated description of the experimental geometry.
- For large pixel-array detector (PAD) images, the raw binary CBF format is heavily used both within laboratories and for interchange among collaborating groups. When dealing with large numbers of independent experiments producing large numbers of CBF/imgCIF image files, HDF5 provides the necessary virtual file system needed to manage the massive data flow.
- While it is feasible to simply encapsulate the CBF/imgCIF image files as opaque objects within an HDF5-based data-management system, active management of the data can be done more efficiently when the imgCIF tags are made visible in the HDF5 tree, a capability demonstrated in 2010.
- With the anticipated throughput of NSLS II beamlines, and the current capabilities of MX beamlines equipped with pixel array detectors, the management of data and the possibility of interrogating the data files for experimental information becomes critical.

CBF and Compression

- Long-standing issues in Crystallography
  - High speed, high compression ratio compression is a critical issue for the next generation of detectors.
  - Some compressions raise license issues.
  - Some popular compressions are slow or inefficient or both.
  - Some compressions can be handled in processing programs such as XDS if license and language issues can be addressed.

CBF and Compression

- Long-standing issues in Crystallography
  - High speed, high compression ratio compression is a critical issue for the next generation of detectors.
  - Some compressions raise license issues.
  - Some popular compressions are slow or inefficient or both.
  - Some compressions can be handled in processing programs such as XDS if license and language issues can be addressed.
- Low pixel density fine-slicing with clean backgrounds makes some compressions more effective.
- CBFlib provides useful compressions.
- A plugin has been written to allow HDF5 to read and write CBFlib compressions.

CBF and Compression (ctd)

Tests on fifty X4_lots_M1S4_1_0???.cbf files from DLS , of low to moderate pixel density. Total size 1.2 gigabytes. Relative times are shown. Multiply by 3 for the number of cores needed to keep up with the compression workload.

Compression Method	Compression Ratio	Relative Time
external bzip2 compression	20.4:1	5.6
HDF5 CBFlib canonical compression	15.7:1	3.9
HDF5 CBFlib nibble offset compression	11.5:1	2.9
HDF5 CBFlib packed V2 compression	11.0:1	2.8
HDF5 zip compression	9.7:1	2.4
external LZ4 compression (C1 one pass)	8.7:1	2.2
HDF5 CBFlib packed compression	8.6:1	2.2
external LZ4 compression (C0 two passes)	5.2:1	1.3
HDF5 CBFlib byte offset compression	4.0:1	1.0

The fifty files are from a set of 900 images recorded by Graeme Winter at Diamond Light Source beamline I02 as part of routine development work, and come from a crystal of DNA (TCGGCGCCGA) bound to a large ligand (lambda-[Ru(TAP)2(dppz)]2+). The aggregate of fifty was chosen to produce an uncompressed file small enough (1.2 GB) to be acceptable at user sites. Larger aggregates could be used for sites able to accommodate large files.

Jan 2013 DECTRIS Eiger Workshop and Followup

- The attendees at the January 2013 DECTRIS Workshop agreed on use of an HDF5-based NeXus format for DECTRIS Eiger pixel array detector.
- The workshop charged Herbert J. Bernstein with followup on mapping additional terms to the new format.
- Tobias Richter, Jonathan Sloan and Herbert J. Bernstein have worked on a CBF-NeXus concordance and supporting software based on CBFlib and HDF5 with the cooperation of Bob Sweet, Graeme Winter and Mark Koennecke.
- Discussions with NIAC and COMCIFS will have taken place prior to this meeting, but after preparation of these slides.
- Eiger Workshop attendees have been sent an updated CBF dictionary and draft CBF-NeXus-HDF5 concordance, as well as a link to a CBFlib-0.9.2.12 kit with minicbf2nexus and a plugin to make CBFlib compression available in HDF5.

mincbf2nexus

- Takes some minicbf files and axis configuration settings for them and creates a NeXus file containing the same data. As this is an early version of the program it should not be assumed to be stable.
- -z (or --compression) {zlib|none}
- At present zlib is supported. CBFlib compressions will follow shortly.
- -c (or --config) CONFIGFILE
- This takes a single argument giving the file name of a configuration file which describes how the axes of the minicbf file relate to each other.
- -o (or ,--output) OUPUTFILE
- Other arguments are interpreted as file names describing the minicbf files to be packed into the new NeXus file.

CBFlib compressions in HDF5

- Starting with CBFlib release 0.9.2.11, a plugin module is provided to allow access to CBFlib compressions from HDF5 1.8.11 and later. For general documentation on HDF5 dynamically loaded filters, see http://www.hdfgroup.org/HDF5/doc/Advanced/DynamicallyLoadedFilters/HDF5DynamicallyLoadedFilters.pdf.
- The filter has been registered with the HDF5 group as 32006, and cbf.h includes the symbolic name for the filter CBF_H5Z_FILTER_CBF.
- The source and header of the CBFlib filter plugin are cbf_hdf5_filter.c and cbf_hdf5_filter.h, respectively. To use the filter in C applications, you will need to include cbf_hdf5_filter.h in the application and have the cbflib.so library in the search path used by HDF5 1.8.11.
- Each compressed image in the HDF5 is in the same format as the MIME-headed compressed images in a CBF, so the Fortran image search logic used in XDS can be used directly on these files.

CBFlib compressions in HDF5

- Starting with CBFlib release 0.9.2.1...

Typical code:

```
unsigned int cd_values[CBF_H5Z_FILTER_CBF_NELMTS];
    hsize_t chunk[3];
    hid_t valspace;
    chunk[0] = 1;  chunk[1] = dimmid; chunk[2] = dimfast;
    cd_values[CBF_H5Z_FILTER_CBF_COMPRESSION] = compression;
    cd_values[CBF_H5Z_FILTER_CBF_RESERVED]    = 0;
    cd_values[CBF_H5Z_FILTER_CBF_BINARY_ID]   = id;
    cd_values[CBF_H5Z_FILTER_CBF_PADDING]     = padding;
    cd_values[CBF_H5Z_FILTER_CBF_ELSIZE]      = (bits+7)/8;
    cd_values[CBF_H5Z_FILTER_CBF_ELSIGN]      = sign;
    cd_values[CBF_H5Z_FILTER_CBF_REAL]        = realarray;
    cd_values[CBF_H5Z_FILTER_CBF_DIMFAST]     = dimfast;
    cd_values[CBF_H5Z_FILTER_CBF_DIMMID]      = dimmid;
    cd_values[CBF_H5Z_FILTER_CBF_DIMSLOW]     = dimslow;
    valprop = H5Pcreate(H5P_DATASET_CREATE);
    H5Pset_chunk(valprop,3,chunk);
    H5Pset_filter(valprop,CBF_H5Z_FILTER_CBF,H5Z_FLAG_OPTIONAL,
        CBF_H5Z_FILTER_CBF_NELMTS,cd_values);
```

Nexus - CBF Concordance

Mapping from CBF to NeXus

- Almost all CBF MX-related categories have proposed slots in the NeXus tree.
- If desired, all CBF data blocks, categories, columns and values will be preserved in tabular form in the NeXus tree for database use or faithful recovery of the original CBF.

Mapping from NeXus to CBF

- All NeXus base classes have proposed slots in CIF categories.
- Handling of the DECTRIS-proposed Eiger HDF5 format is in the concordance.

This concordance will require some relaxation of current NeXus name practices.

- ``CBF_'' prefixes have been proposed.

References

[Bern 05] H. J. Bernstein and A. P. Hammersley. “Specication of the Crystallographic Binary File (CBF/imgCIF)". In: S. R. Hall and B. McMahon, Eds., International Tables For Crystallography, Chap. 2.3, pp. 37 { 43, International Union of Crystallography, Springer, Dordrecht, NL, 2005.

[Codd 70] E. F. Codd. “A relational model of data for large shared data banks". Communications of the ACM, Vol. 13, No. 6, pp. 377 { 387, 1970.

[Codd 72] E. F. Codd. Courant Computer Science Symposium 6, Chap. “Further Normalization of the Data Base Relational Model”, pp. 33 -- 64. Prentice-Hall, 1972, 1972.

[Doug 09] M. T. Dougherty, M. J. Folk, H. J. Bernstein, F. C. Bernstein, K. W. Eliceiri, W. Benger, E. Zadok, and C. Best. “Unifying Biological Image Formats with HDF5”. Communications of the ACM, Vol. 52, No. 10, pp. 42 -- 47, 2009.

[Elli 05] P. J. Ellis and H. J. Bernstein. “Definition and Exchange of Crystallographic Data”, International Tables For Crystallography, Chap. CBFlib: an ANSI C library for manipulating image data, pp. 544 -- 556. International Union of Crystallography, Springer, Dordrecht, NL, 2005.

[Filg 01] U. Filges. “The new NeXus API based on HDF5". In: VITESS Workshop Berlin, 25 -- 27 June 2001, 2001.

References

[Gotz 10] A. Götz, V. Sole, C. Madonna, and A. F. Maydew. “ELISA VEDAC Workshop Report, Workshop Title: HDF5 as hyperspectral data exchange and analysis format, Grenoble, January 11th to January 13th, 2010.". January 2010. http://vedac.esrf.eu/public-discussions/hdf5-workshop/workshop-report.

[Hall 91] S. R. Hall, F. H. Allen, and I. D. Brown. “The Crystallographic Information File (CIF): a new standard archive file for crystallography". Acta Crystallographica Section A: Foundations of Crystallography, Vol. 47, No. 6, pp. 655 { 685, 1991.

[Haml 12] R. Hamlin, T. Hontz, and C. Nielsen. “The New Dual Mode Pixel Array Detector”. in: Meeting of the American Crystallographic Association, Boston, MA, 28 July -- 1 August 2012, American Crystallographic Association, 2012. Abstract 11.01.1151.

[Hend 12] W. A. Hendrickson. “NSLS-II - Status of the Life Sciences Program". Tech. Rep., Brookhaven National Laboratory, X6A Science Advisory Committee, February 2012. protein.nsls.bnl.gov/mediawiki/images/e/e3/Hendrickson_2012.pdf.

[John 12] I. Johnson, A. Bergamaschi, J. Buitenhuis, R. Dinapoli, D. Greienberg, B. Henrich, T. Ikonen, G. Meier, A. Menzel, A. Mozzanica, et al. “Capturing dynamics with Eiger, a fast-framing X-ray detector". Journal of synchrotron radiation, Vol. 19, No. 6, 2012.

[Rew 04] R. Rew, B. Ucar, and E. Hartnett. “Merging netCDF and HDF5". In: 20th Int. Conf. on Interactive Information and Processing Systems, 2004.

References

[Schn 12] D. K. Schneider, B. M. Sweet, and J. Skinner. “Projection of MX ABBIX needs at AMX and FMX for Data Acquisition, Data Processing, Software, Data Archiving and Networking". February 2012. Private Communication.

[True 12] P. Trueb, B. Sobott, R. Schnyder, T. Loeliger, M. Schneebeli, M. Kobas, R. Rassool, D. Peake, and C. Broennimann. “Improved count rate corrections for highest data quality with PILATUS detectors". Journal of Synchrotron Radiation, Vol. 19, No. 3, pp. 347 -- 351, 2012.

[Will 11] P. Willmott. “An Introduction to Synchrotron Radiation: Techniques and Applications”. John Wiley and Sons, Chichester, UK, 2011. page 6.

Acknowlegments

BNL PXRR Group

Robert M. Sweet, Dieter Schneider, Howard Robinson, John Skinner, Matt Cowan, Leonid Flaks, Richard Buono

Diamond Light Source

Alun Ashton, Bill Pulford

Dowling College ARCiB Lab Group

Mojgan Asadi, Kostandina Bardhi, Maria Karakasheva, Ming Li, Limone Rosa

DECTRIS, BIOIHDF, HDF Group

Frances C. Bernstein

Work funded in part by NIGMS, DOE, NSF, PaNdata ODI (EU 7th Framework Programme)