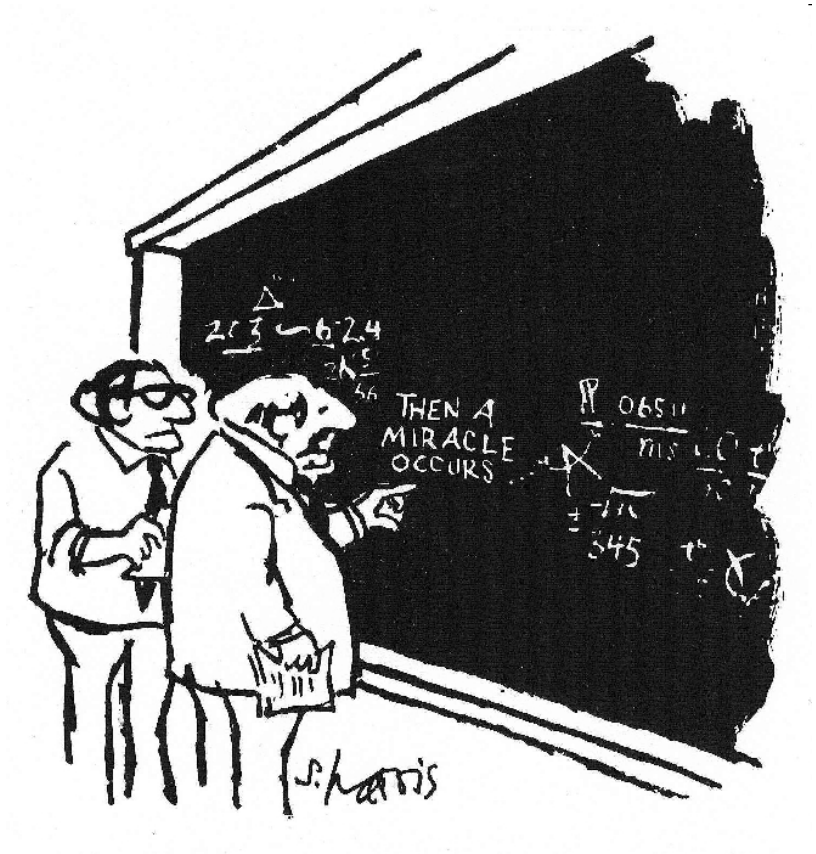
Computational Needs and Requirements for X-ray Fluorescence Microspectroscopy Users Now and Into the Future

A. Lanzirotti, M. Newville, M. Rivers , S. Sutton,
The University of Chicago, Center for Advanced Radiation Sources
S. Vogt,
Advanced Photon Source
M. Marcus,
Advanced Light Source
D. Dale
Cornell High Energy Synchrotron Source
S. Webb
Stanford Synchrotron Radiation Laboratory
D. Paterson, C. Ryan
Australian Synchrotron, CSIRO

or…What do Microspectroscopy **Users** Want (and Need)

- Analysis of materials (mostly natural samples, earth, environmental, and biological) compositionally heterogeneous at the sub-micrometer scale.
- That natural heterogeneity means samples analyzed are incredibly diverse (size, thickness, density, attenuation and self-absorption, spectral complexity).
- Requires versatility in:
  - How users collect their data
  - How users process their data
- Many microprobe users have never been to a synchrotron.



"I think you should be more explicit here in step two."

- *It's natural for programmers to want to develop software tools that output "results" to the user*
- *For users, this can be very bad thing*

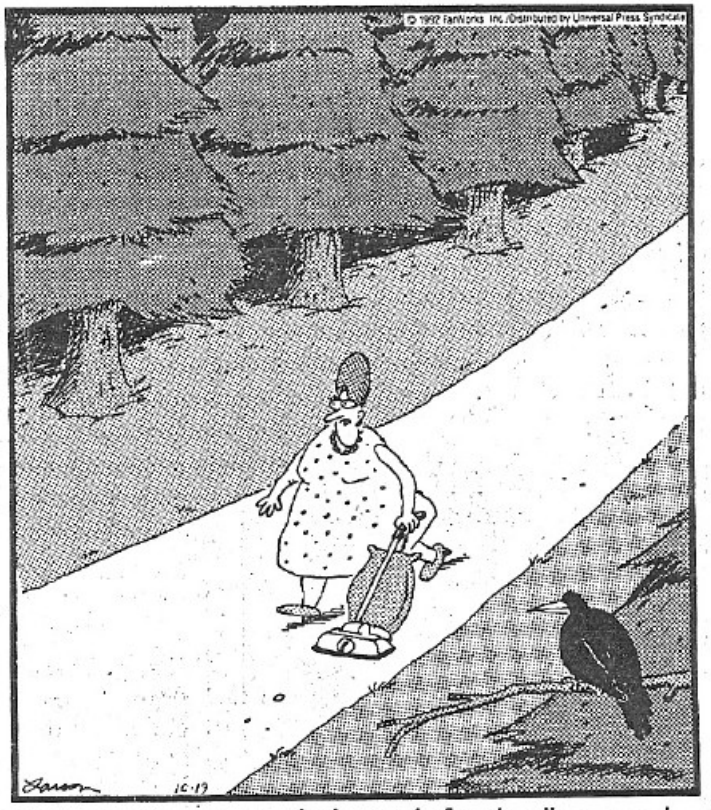
or…What do Microspectroscopy **Users** Want (and Need)

- Analysis of materials (mostly natural samples, earth, environmental, and biological) compositionally heterogeneous at the sub-micrometer scale.

- That natural heterogeneity means samples analyzed are incredibly diverse (size, thickness, density, attenuation and self-absorption, spectral complexity).

- Requires versatility in:
  - How users collect their data
  - How users process their data

- Many microprobe users have never been to a synchrotron.



The woods were dark and foreboding, and Alice sensed that sinister eyes were watching her every step. Worst of all, she knew that Nature abhorred a vacuum.

- *Meaningful results require understanding of sample state and careful evaluation of spectra.*
- *Very easy to "get it wrong"…what do users go home with?*

Hard X-ray Microprobe Datasets

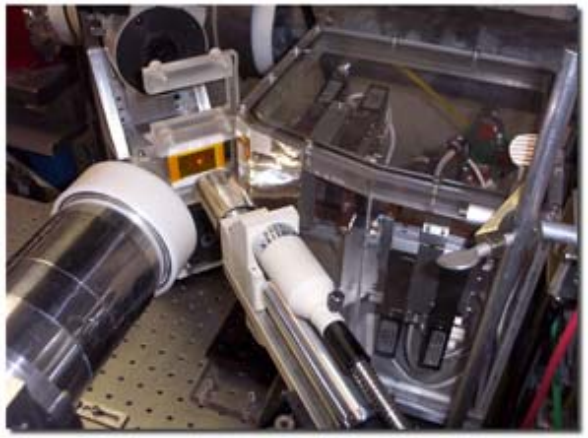
Techniques

- **µXRF:** Spot XRF analyses of low abundance elemental composition
- **Element Mapping:** 2D compositional imaging (along with phase/absorption contrast)
- **µXAFS:** Spot and 2D XANES and EXAFS determinations of oxidation state and speciation
- **µXRD:** phase identification and correlation with elemental and speciation information
- **Fluorescence Microtomography:** Internal elemental imaging 2D and 3D

Virtually all users do this in some form, but few come to do only this.

Principal Science Programs

- **Metal Speciation in the Environment**
- **Contaminant Transport**
- **Igneous Petrology**
- **Cosmochemistry**
- **Roles of Metals in Human Disease**
- **Metal homeostasis in organisms**



Hard X-ray Microprobe Datasets

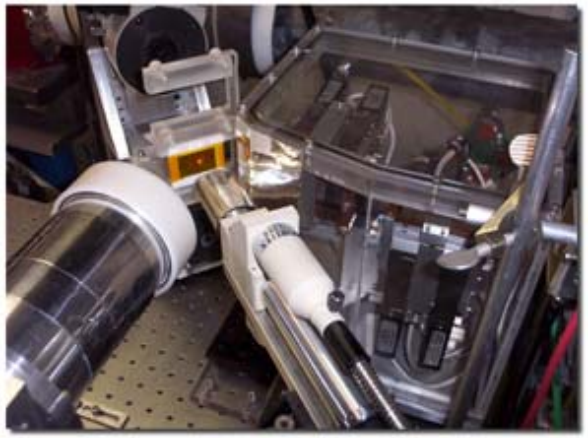
Techniques

- **µXRF:** Spot XRF analyses of low abundance elemental composition
- **Element Mapping:** 2D compositional imaging (along with phase/absorption contrast)
- **µXAFS:** Spot and 2D XANES and EXAFS determinations of oxidation state and speciation
- **µXRD:** phase identification and correlation with elemental and speciation information
- **Fluorescence Microtomography:** Internal elemental imaging 2D and 3D

A core technique of these instruments and vital for identifying specific components for further analysis

Principal Science Programs

- **Metal Speciation in the Environment**
- **Contaminant Transport**
- **Igneous Petrology**
- **Cosmochemistry**
- **Roles of Metals in Human Disease**
- **Metal homeostasis in organisms**



Hard X-ray Microprobe Datasets

Techniques

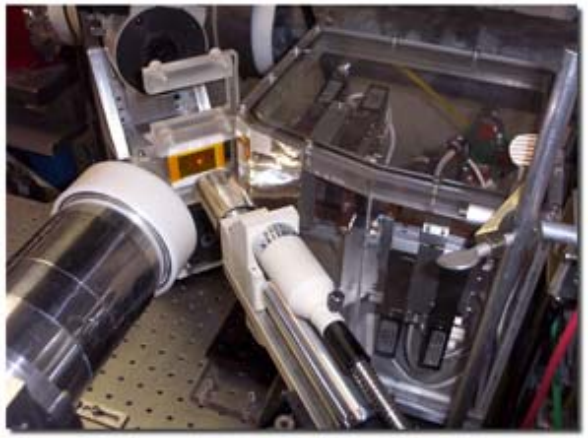
- **µXRF:** Spot XRF analyses of low abundance elemental composition
- **Element Mapping:** 2D compositional imaging (along with phase/absorption contrast)
- **µXAFS:** Spot and 2D XANES and EXAFS determinations of oxidation state and speciation
- **µXRD:** phase identification and correlation with elemental and speciation information
- **Fluorescence Microtomography:** Internal elemental imaging 2D and 3D

More than 60% of experiments on KB instruments

Perhaps ~30% on ZP based instruments

Principal Science Programs

- **Metal Speciation in the Environment**
- **Contaminant Transport**
- **Igneous Petrology**
- **Cosmochemistry**
- **Roles of Metals in Human Disease**
- **Metal homeostasis in organisms**



Hard X-ray Microprobe Datasets

Techniques

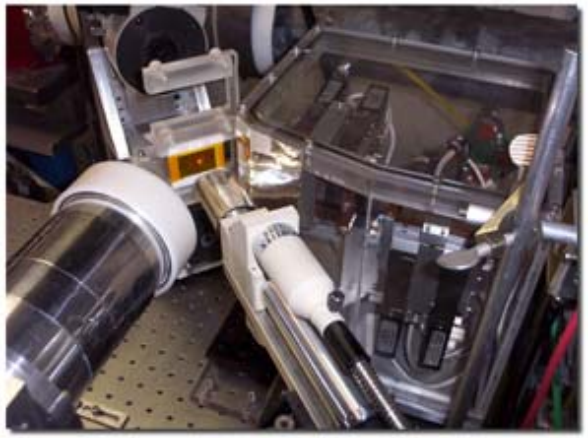
- **µXRF:** Spot XRF analyses of low abundance elemental composition
- **Element Mapping:** 2D compositional imaging (along with phase/absorption contrast)
- **µXAFS:** Spot and 2D XANES and EXAFS determinations of oxidation state and speciation
- **µXRD:** phase identification and correlation with elemental and speciation information
- **Fluorescence Microtomography:** Internal elemental imaging 2D and 3D

Approximately 25% of experiments on KB instruments

Single point diffraction dominantly, but 2D imaging modes developing

Principal Science Programs

- **Metal Speciation in the Environment**
- **Contaminant Transport**
- **Igneous Petrology**
- **Cosmochemistry**
- **Roles of Metals in Human Disease**
- **Metal homeostasis in organisms**



Hard X-ray Microprobe Datasets

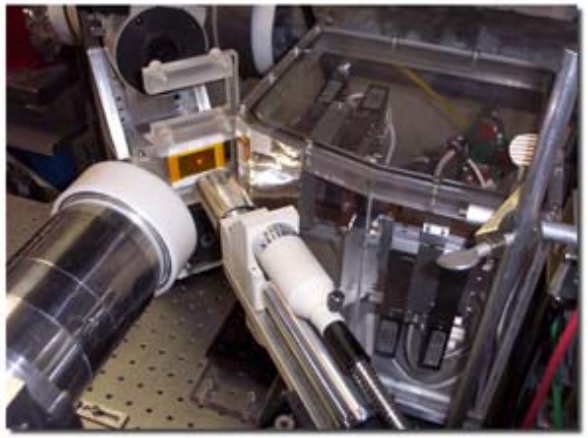
Techniques

- **µXRF:** Spot XRF analyses of low abundance elemental composition
- **Element Mapping:** 2D compositional imaging (along with phase/absorption contrast)
- **µXAFS:** Spot and 2D XANES and EXAFS determinations of oxidation state and speciation
- **µXRD:** phase identification and correlation with elemental and speciation information
- **Fluorescence Microtomography:** Internal elemental imaging 2D and 3D

Relatively small program currently, but very rapid growth in demand

Principal Science Programs

- **Metal Speciation in the Environment**
- **Contaminant Transport**
- **Igneous Petrology**
- **Cosmochemistry**
- **Roles of Metals in Human Disease**
- **Metal homeostasis in organisms**



Microprobe Control Software - End User Data Collection

- Robust, easy to use software to control experiments.
- Many solutions exist, most home-built and specific to particular beamlines or facilities. EPICS provides maximum coverage of beamlines and hardware.
- Experimental Diversity:
    - wide range of “techniques” (µXRF, µXAFS of several elements, µXRD)
    - wide range of scan types (motor scans, energy scans, environmental scans)
    - wide range of detector types within a single experiment
- Flexible enough to allow end-users to do all of these measurements easily (GUI)
- Flexible enough to do relatively complicated experiments, must be fully scriptable so that "macros" can be run.
- Well-defined low-level API so that anything from a macro to a new and different GUI could be written. These modifications are usually done by beamline scientists. You don’t want to wait on facility software support.
- Remote access is **very highly desirable**. Working through firewalls should be considered up front.

Microprobe Control Software – Varying Scan types

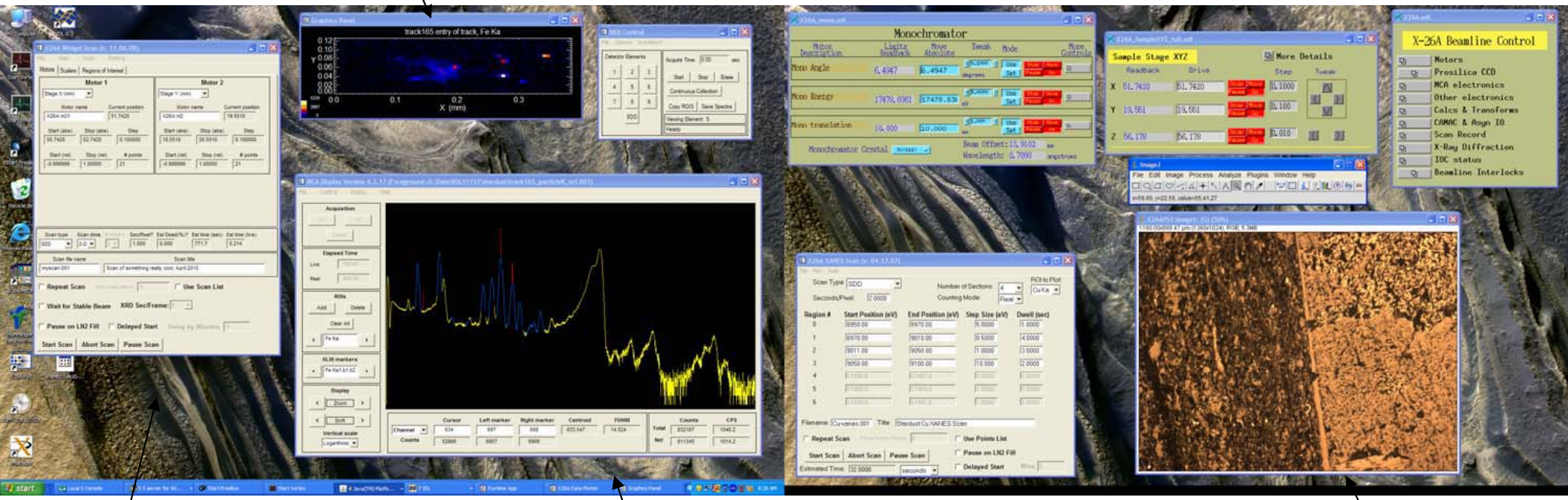
Microprobe experiments often consist of a sequence of differing scans. Optimally users should be able to transition from one scan type to another seamlessly:

- Single point XRF analysis (what have I got, where am I)
- Simple 1 dimensional step scans of motors (line scans…XRF, Scalers, XRD, etc)
- Simple 2 dimensional step scans of motors (maps)
- Three-dimensional scanning (tomography)
- XAFS scans in "step mode" (complicated set of energy points and dwell times)
- Predefined sequences of all the above (lists of sequential maps or points)
- Motor and XAFS scans in "slew mode" & fast-scans with hardware (eg Struck)
- Support for multi-element MCA for XRF spectra, saved at each point in scan.
- Support for video image as data (using EPICS areaDetector NDPluginROI you can save specific pixels of the video image to make a visible image overlay!).
- Being able to "Save Positions" or Settings of various beamline components or parameters.
- GUI client that can combine video images of sample with beamline control and data collection (again, easy with EPICS areaDetector, but many scans outside field of view).
- Quasi-live display of scan data during collection!!

Microprobe Control Software – Example Layout

Live ROI count display, 1D or 2D

MEDM

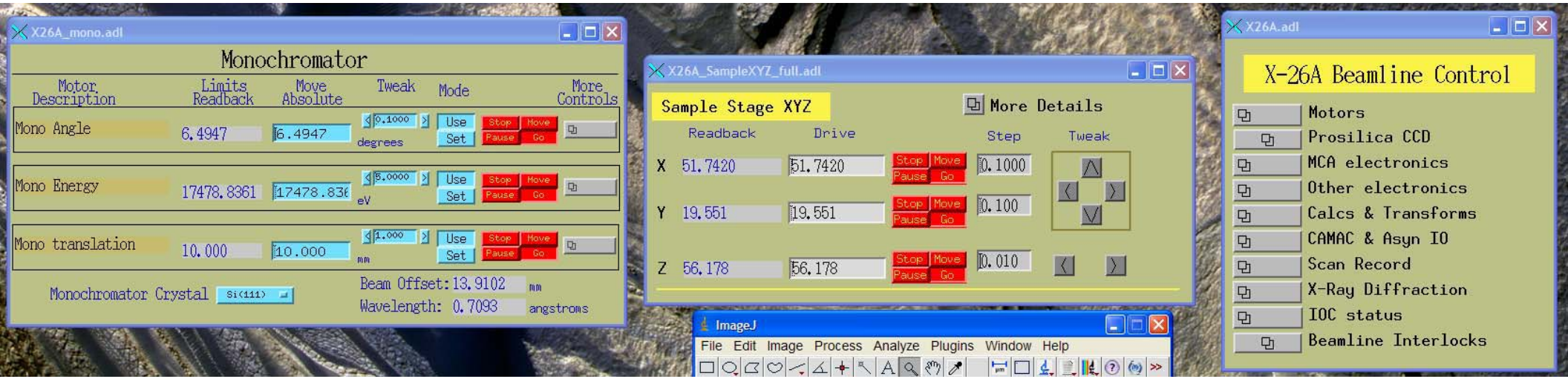


1D,2D,3D scan control

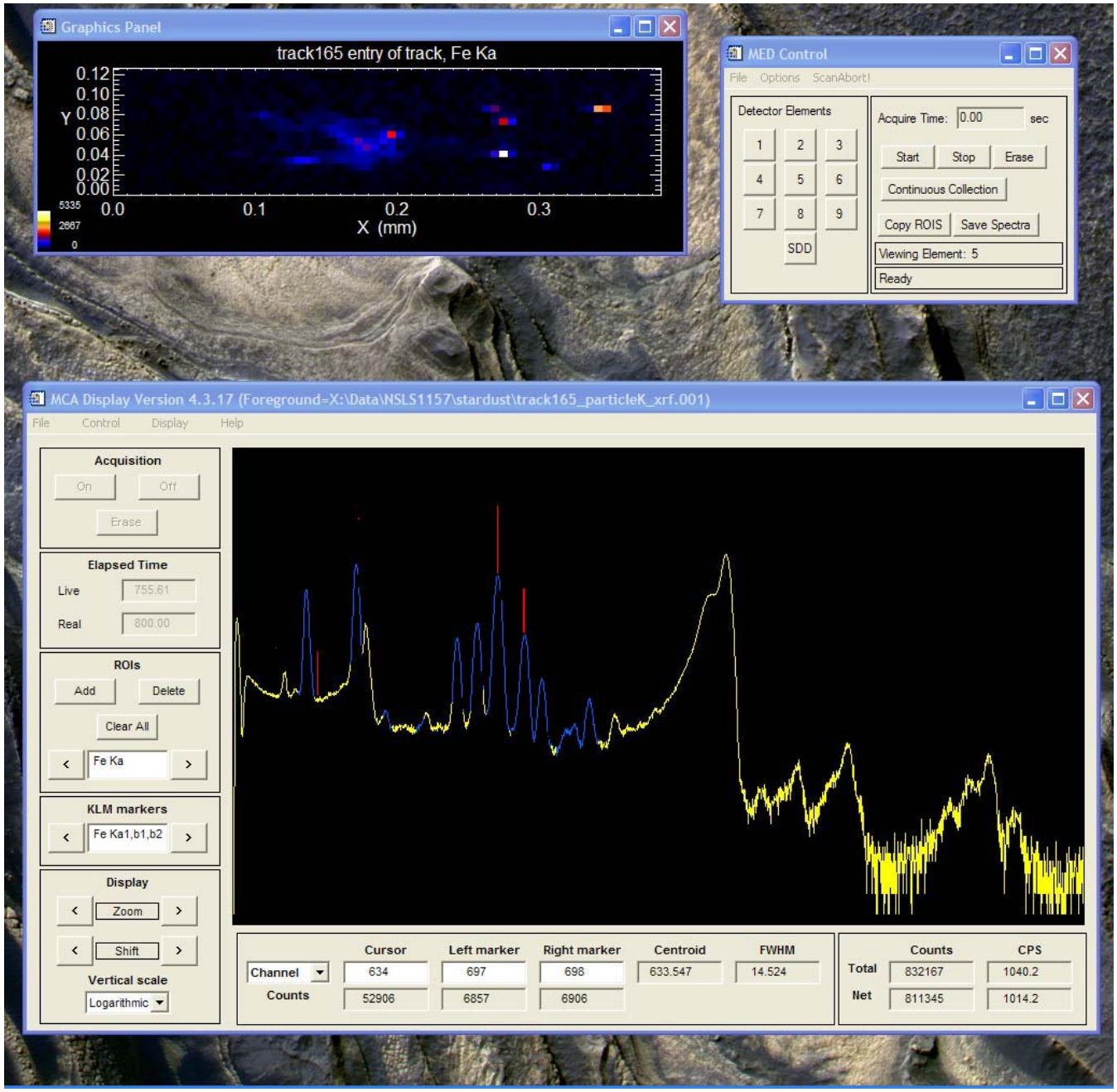
Spectroscopy control

Live energy dispersive spetra

Live digital video



Standard set of MEDM controls, for users this is mostly motor control. Simpler the better.



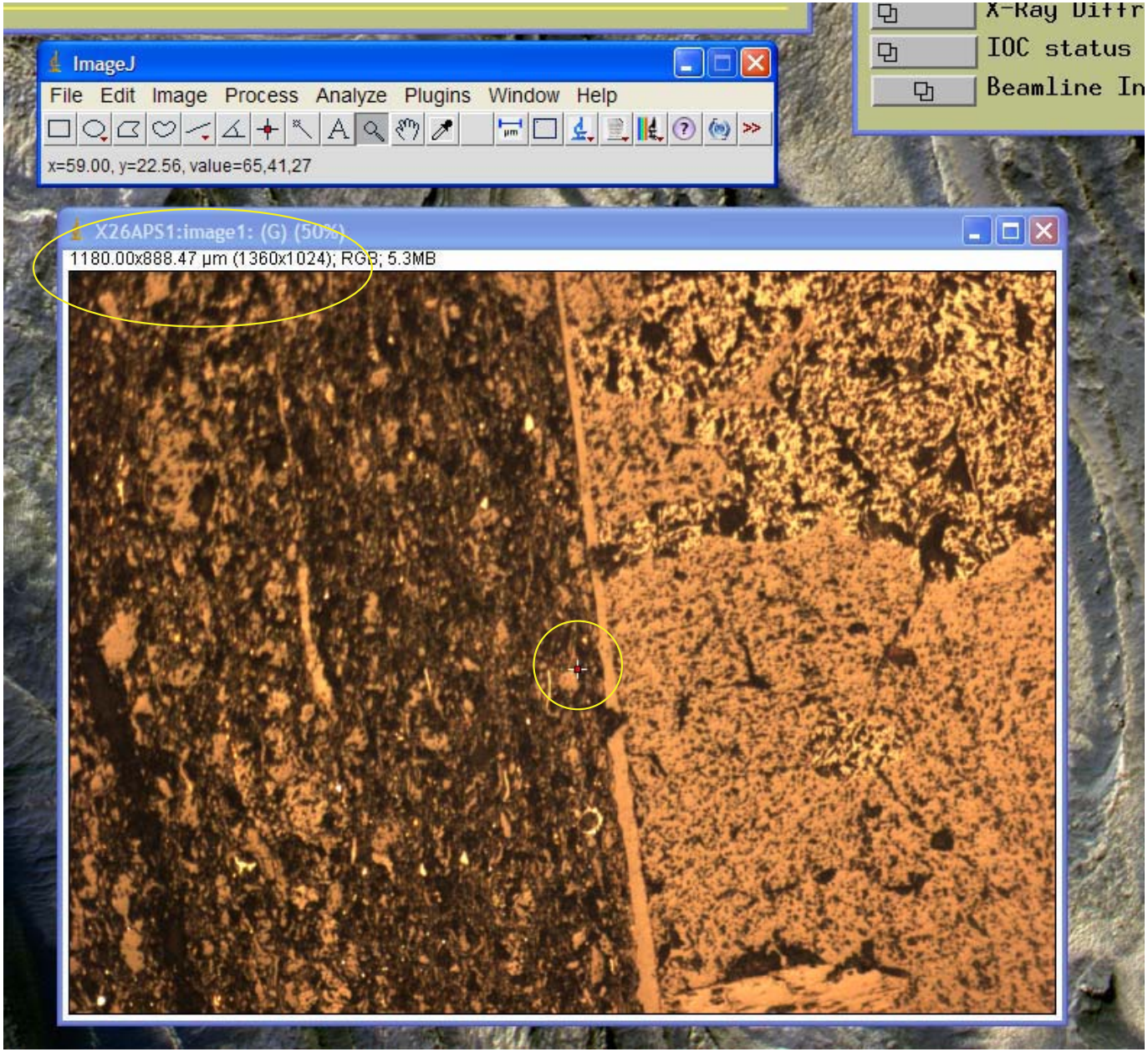
MED control in IDL through channel access

Live energy dispersive spectra with ROI's defined

Outputs roi data to screen as scan progresses, but all roi's and full EDS saved.

Simple GUI based scan setup routines. Front end to channel access, can scan any motor (more than 2), use any combinations of detectors

Support for automated lists of scans, pause on event, delayed starts



Visible light image crucial

areaDetector enables the use of generic image display clients that obtain their data via EPICS Channel Access

This is a plugin written by Mark Rivers for the popular ImageJ Java-based program

Note that the image is calibrated in distance units. Also note the beam position marker, this is a separate ImageJ file.

HDF as a data storage format for x-ray microprobes

- **Hierarchical Data Format (HDF)** is a data file format designed by the National Center for Supercomputing Applications (NCSA) to assist users in the storage and manipulation of scientific data across diverse operating systems and machines.

- HDF supports a variety of data types: scientific data arrays, tables, and text annotations, as well as several types of raster images and their associated color palettes

- HDF is a platform independent file format. It can be used on many different computers, regardless of the operating system that machine is running.

- HDF5 support is built-in to many scripted languages like IDL, Matlab, and Python, and trivially downloadable for C on nearly any computing platform.

HDF as a data storage format for x-ray microprobes

- HDF files are also **self-describing**
  - Each data object in the file has predefined tags that identify such information as the type of data, the amount of data, its dimensions, and its location in the file
  - Thus you can define the structure and contents of a file just from the information stored in the file itself
  - Self-description also means that many types of data can be bundled in an HDF file
  - Self-description also means that additional groups can be added in the future as needed without having to rewrite the software that reads the data, since specific calls are made to structures by name
- HDF supports a **low-level compression** interface, which allows any individual data-object within the HDF file to be compressed using a variety of algorithms
- HDF 5 functionality is **fully built into IDL**. It contains an HDF browser utility as part of the program and a full complement of commands to write and read data to HDF format

X26A/X27A HDF Data Format

- 12 Global attributes
  - Beamline ID
  - Image Title
  - Date and Time
  - Number of detectors
  - Number of columns
  - Number of rows
  - Motor n start
  - Motor n stop
  - Motor n+1 start
  - Motor n+1 stop
  - Elapsed real time
  - Elapsed live time

- 12 Scientific datasets (SD)
  - MCA offset
  - MCA slope
  - MCA quad
  - MCA units
  - MCA two_theta
  - I0
  - ICR/OCR
  - X_dist
  - Y_dist
  - Data_title
  - Image_data
  - MCA_data

Detector calibrations

all rois, scalers, motor pos

Several free HDF viewers (browsers) are available. So relatively simple for someone to, independent of the beamline, extract data

Creating and retrieving HDF data in IDL

```
;create an HDF file
mca_id = HDF_SD_START(sd.file_name+'.hdf',/create)
;Copy info about the scan to the HDF
HDF_SD_ATTRSET, mca_id, 'Image title', 'Title: '+sd.title
HDF_SD_ATTRSET, mca_id, 'Date & Time', SYSTIME(0)
HDF_SD_ATTRSET, mca_id, 'Number of columns', n_colsL
HDF_SD_ATTRSET, mca_id, 'Number of rows', n_rowsL
HDF_SD_ATTRSET, mca_id, 'Motor 1 Start', md(0).start[0]
HDF_SD_ATTRSET, mca_id, 'Motor 1 Stop', md(0).stop[0]
HDF_SD_ATTRSET, mca_id, 'Motor 2 Start', md(1).start[0]
HDF_SD_ATTRSET, mca_id, 'Motor 2 Stop', md(1).stop[0]
HDF_SD_ATTRSET, mca_id, 'Preset Real Time (msec)', sd.dwell_time*1000.
HDF_SD_ATTRSET, mca_id, 'Beamline', sd.beamline

;create an id to store the ion chamber counts in the HDF file
i0_id=HDF_SD_CREATE(mca_id,"i0",[n_colsL*n_rowsL],/float)

;create the MCA data volume to store in the HDF file
data_id=HDF_SD_CREATE(mca_id,"mca data",[n_colsL*n_rowsL,2048],/long)
;Set the mca_data and image_data id's for gzip compression
HDF_SD_SETCOMPRESS,data_id,4
HDF_SD_SETCOMPRESS,imagedata_id,4

  ; Add the data to the HDF
  HDF_SD_ADDDATA, i0_id, i0vol
  HDF_SD_ADDDATA, data_id, datavol

  ; Close HDF Access
  HDF_SD_ENDACCESS, data_id
  HDF_SD_ENDACCESS, i0_id
  HDF_SD_ENDACCESS, xdist_id
  HDF_SD_ENDACCESS, ydist_id
  HDF_SD_ENDACCESS, datatitle_id
  HDF_SD_ENDACCESS, imagedata_id
  HDF_SD_END, mca_id
```

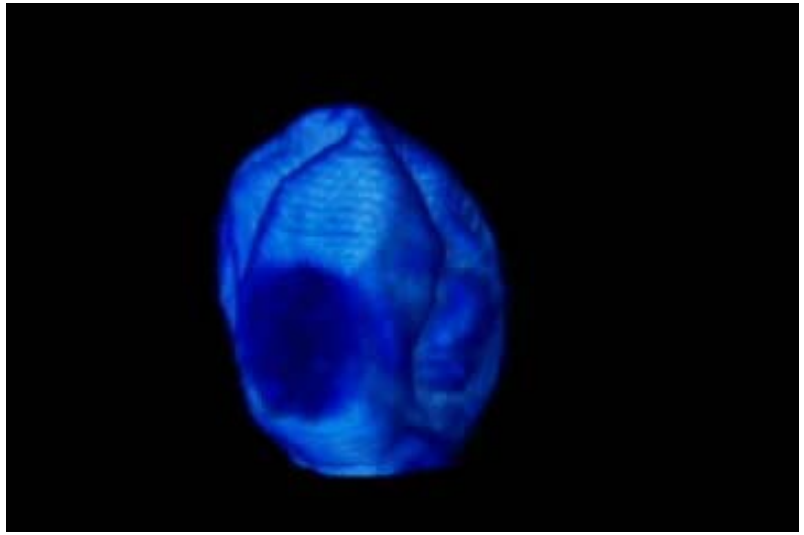
```
hdf_id = HDF_SD_START(info.filename,/read)

title_id = HDF_SD_ATTRFIND(hdf_id, 'Image title')
HDF_SD_ATTRINFO, hdf_id, title_id, DATA=title

i0_id = HDF_SD_NAMETOINDEX(hdf_id, 'i0')
spectra_id = HDF_SD_NAMETOINDEX(hdf_id, 'mca data')

sdsI0 = HDF_SD_SELECT(hdf_id, i0_id)
sdsSPECTRA = HDF_SD_SELECT(hdf_id, spectra_id)

HDF_SD_GETDATA, sdsI0, i0
HDF_SD_GETDATA, sdsSPECTRA, mca_spectra

HDF_SD_END, hdf_id
```

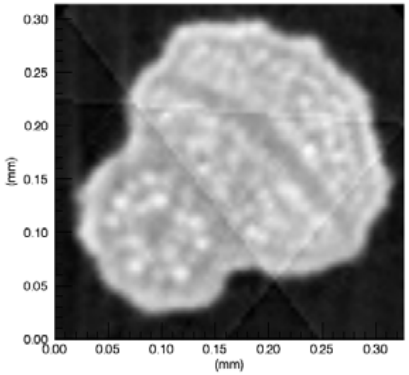
Future (now) Data Collection for X-ray Microprobes

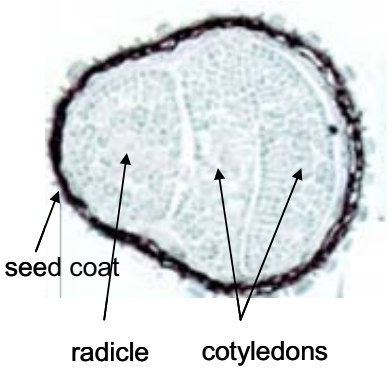
Future (now) Data Collection for X-ray Microprobes

- **3D!**

  - **Ionomics:** the study of how genes regulate ions in cells.
  - Was a need to identify what key genes are involved in iron uptake in plants have been identified. Armed with this knowledge, it would be possible to engineer or breed plants with improved iron uptake abilities and in more bioavailable forms.
  - We used synchrotron XFCT to image and assign functions to metal homeostasis genes *in-situ/in-vivo*. These phenotypes could not be observed using volume-averaged metal analysis techniques.
  - Transition metal abundances imaged at concentrations at ~70 ppm three-dimensionally.

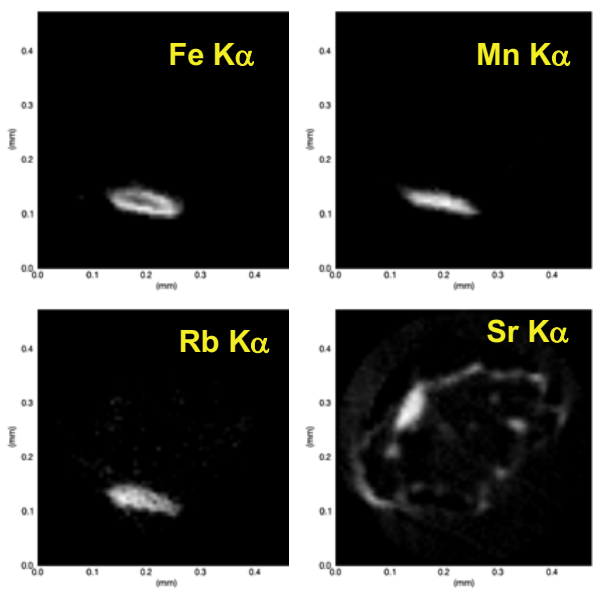


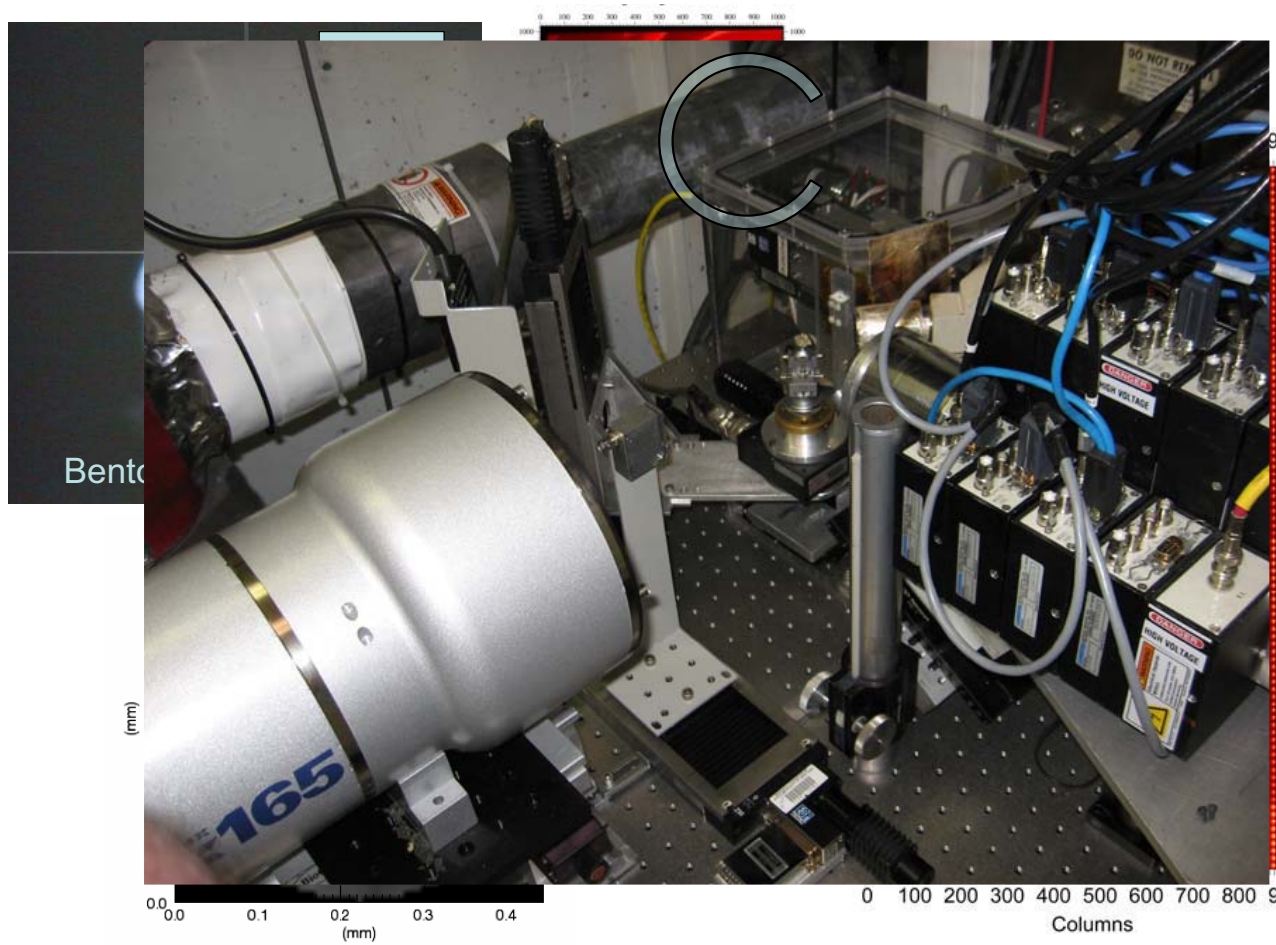




Example – Coupled XRF/XRD Tomography

- Fluorescence CMT can be coupled with other x-ray microbeam tomographies
- For example, x-ray diffraction data can be collected simultaneously using CCD area detectors, allowing for the reconstruction of a given mineral reflection along with the fluorescence data.





Tomographic reconstruction of Montmorillonite {110} reflection

Sinogram consisting of 1500 CCD area detector frames (& XRF, etc)

~3 Gb of data

Future (now) Data Collection for X-ray Microprobes

- 3D!
- NSLS-II will push the limits with respect to spatial resolution

Future (now) Data Collection for X-ray Microprobes

- **3D!**
- **NSLS-II will push the limits with respect to spatial resolution**
  - Beam stability an issue

Future (now) Data Collection for X-ray Microprobes

- **3D!**
- **NSLS-II will push the limits with respect to spatial resolution**
  - Beam stability an issue
  - Radiation dose needs to be minimized

Future (now) Data Collection for X-ray Microprobes

- **3D!**
- **NSLS-II will push the limits with respect to spatial resolution**
  - Beam stability an issue
  - Radiation dose needs to be minimized
  - Detectors are getting faster

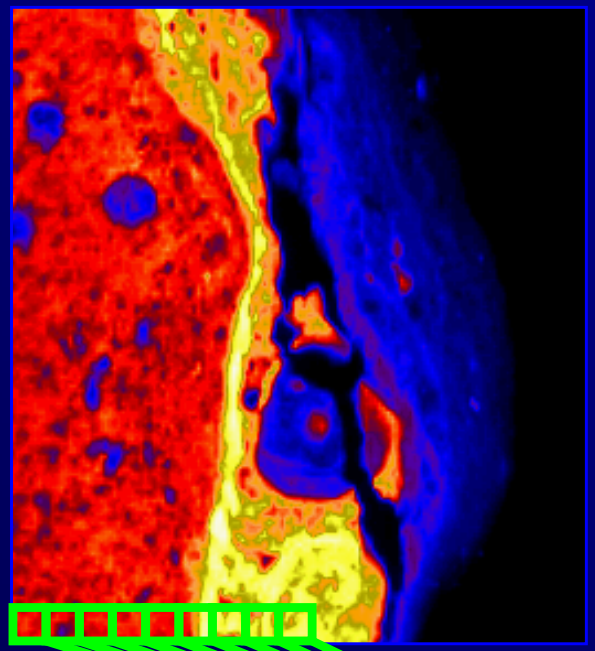
Future (now) Data Collection for X-ray Microprobes

- **3D!**
- **NSLS-II will push the limits with respect to spatial resolution**
  - Beam stability an issue
  - Radiation dose needs to be minimized
  - Detectors are getting faster
- **Spectroscopy on an undulator will likely best be done in a scanning mode**

Future (now) Data Collection for X-ray Microprobes

- **3D!**
- **NSLS-II will push the limits with respect to spatial resolution**
  - Beam stability an issue
  - Radiation dose needs to be minimized
  - Detectors are getting faster
- **Spectroscopy on an undulator will likely best be done in a scanning mode**
- **All this means that continuous scan modes will probably evolve to be the standard mode of operation**

Full Spectral Data Collection: Raster sample through beam

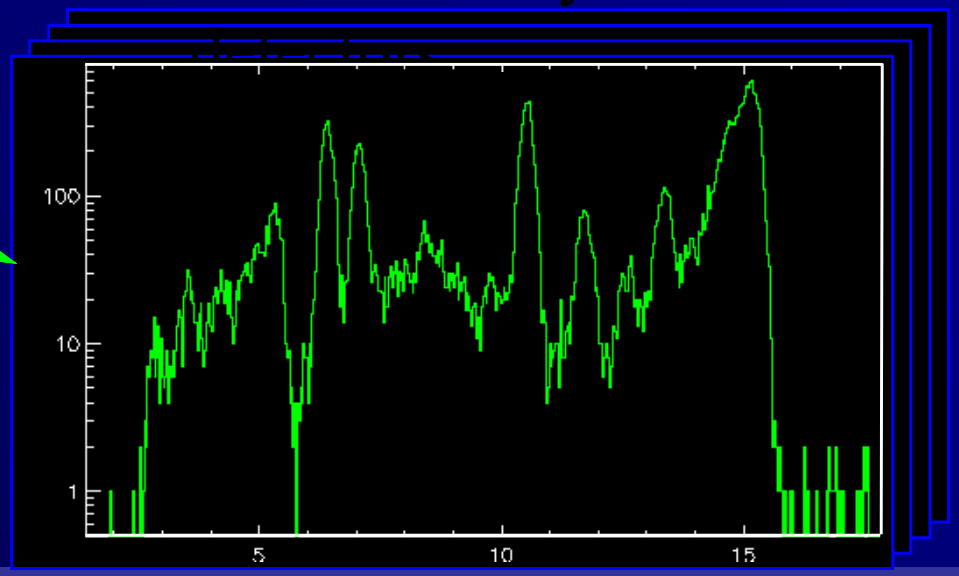


Conventional synchrotron approach:

Read-out *N* full spectra at each pixel (~1 sec)

- 150 x 150 pixels → ~6-7 hours
- 15 minutes → ~30 x 30 pixels

Detector array: *N*

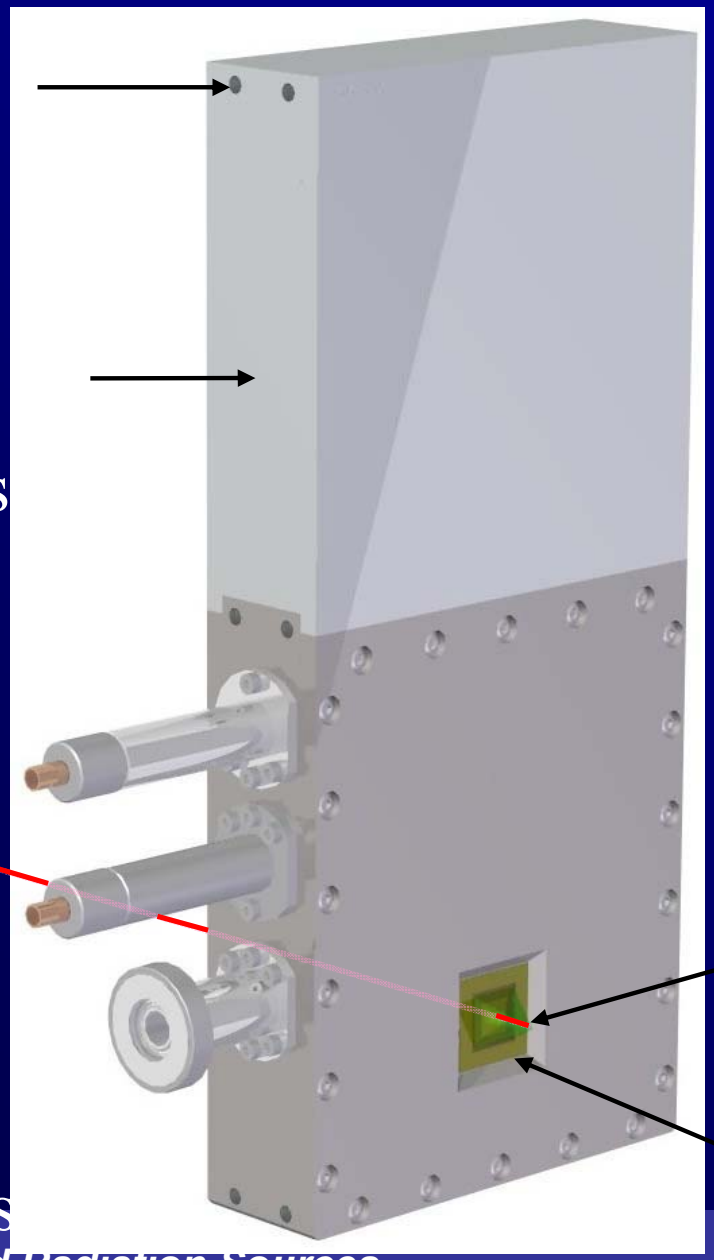


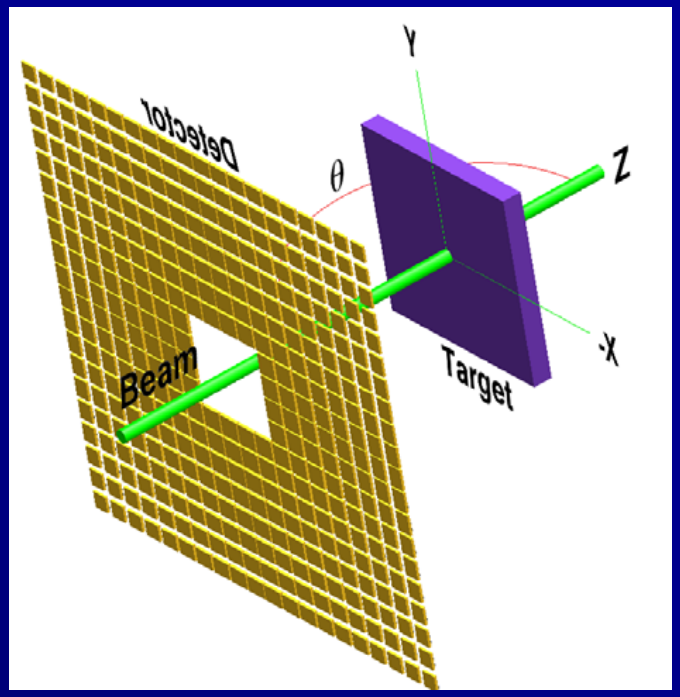
Raster sample in X,Y through microbeam

Maia detector

Mounting points

Electrical/ optical data connections



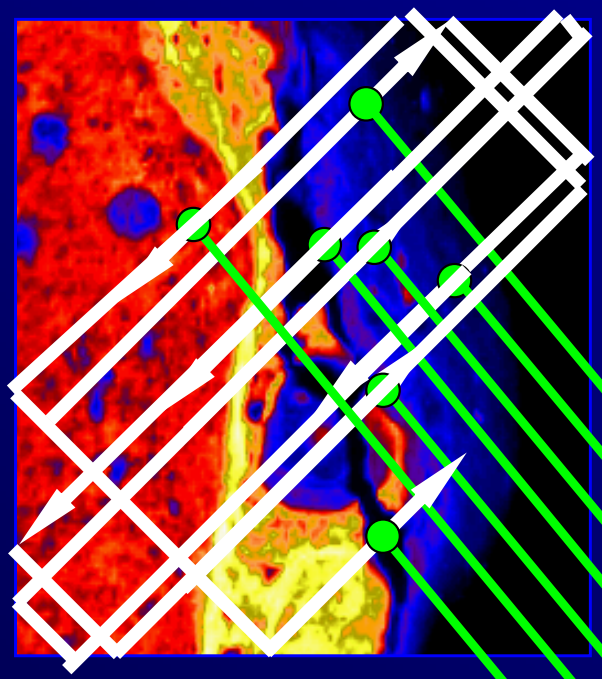


Cooling/ vacuum connections

Optimum sample position
- 1 mm from front face
- 10 mm from detector wafer
- Peltier cooled to -35 ºC

Beryllium window

Full Spectral Data Collection: Event-by-event processing



Nuclear Physics Approach:

Sample X, Y for each detected X-ray event

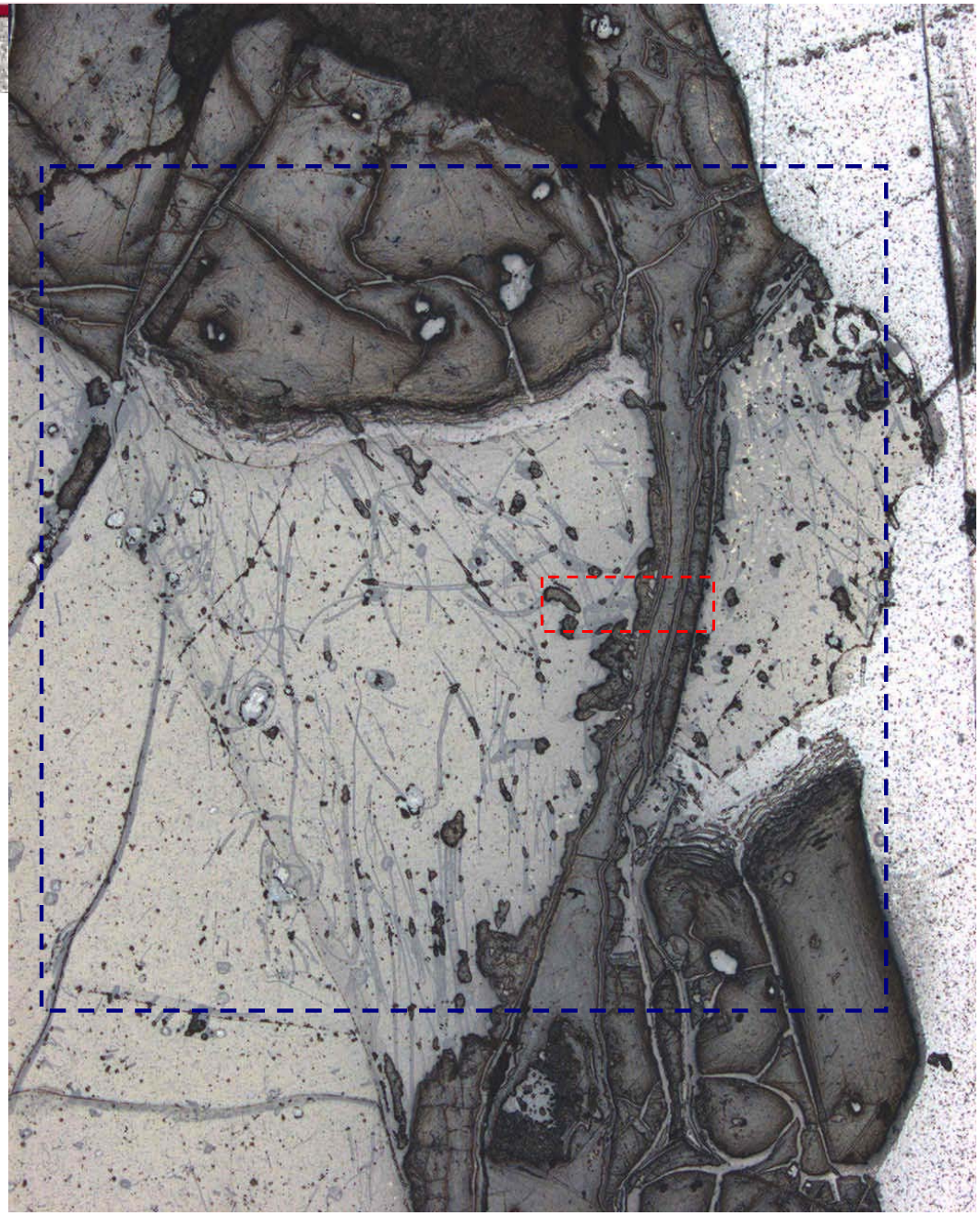
- Freedom to use high scan rates
- Real-time processing of event stream

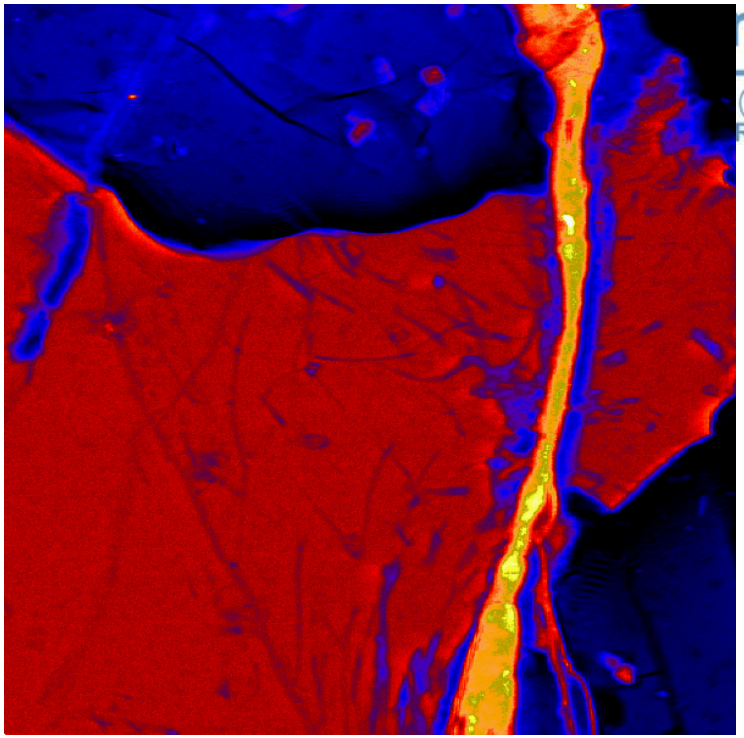
List-mode data stream:

X_1, Y_1, E_1, n_1
X_2, Y_2, E_2, n_2
X_3, Y_3, E_3, n_3
X_4, Y_4, E_4, n_4
X_5, Y_5, E_5, n_5
X_6, Y_6, E_6, n_6
X_7, Y_7, E_7, n_7

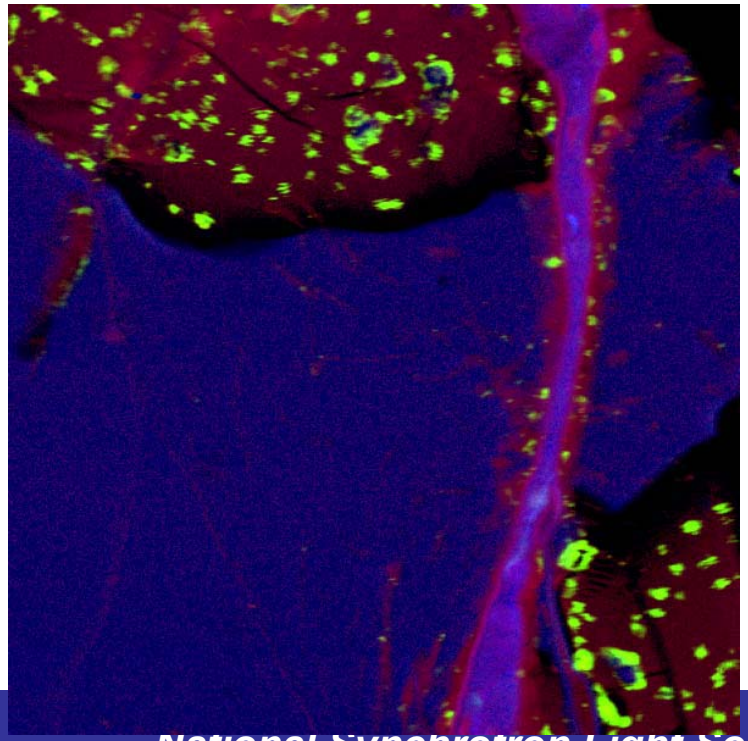
X_i X coordinate
Y_i Y coordinate
E_i Energy
n_i detector #

Used for Nuclear Microprobe and new SXRF development





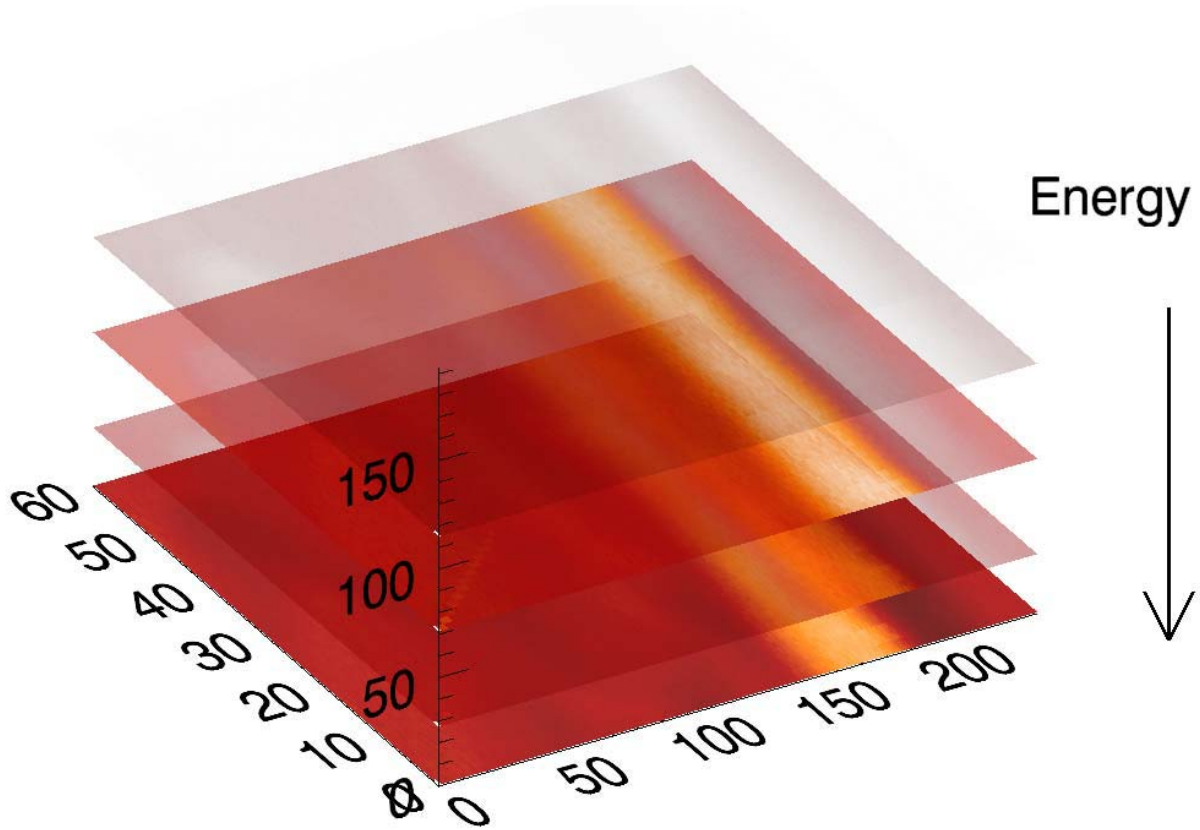
Fe abundance

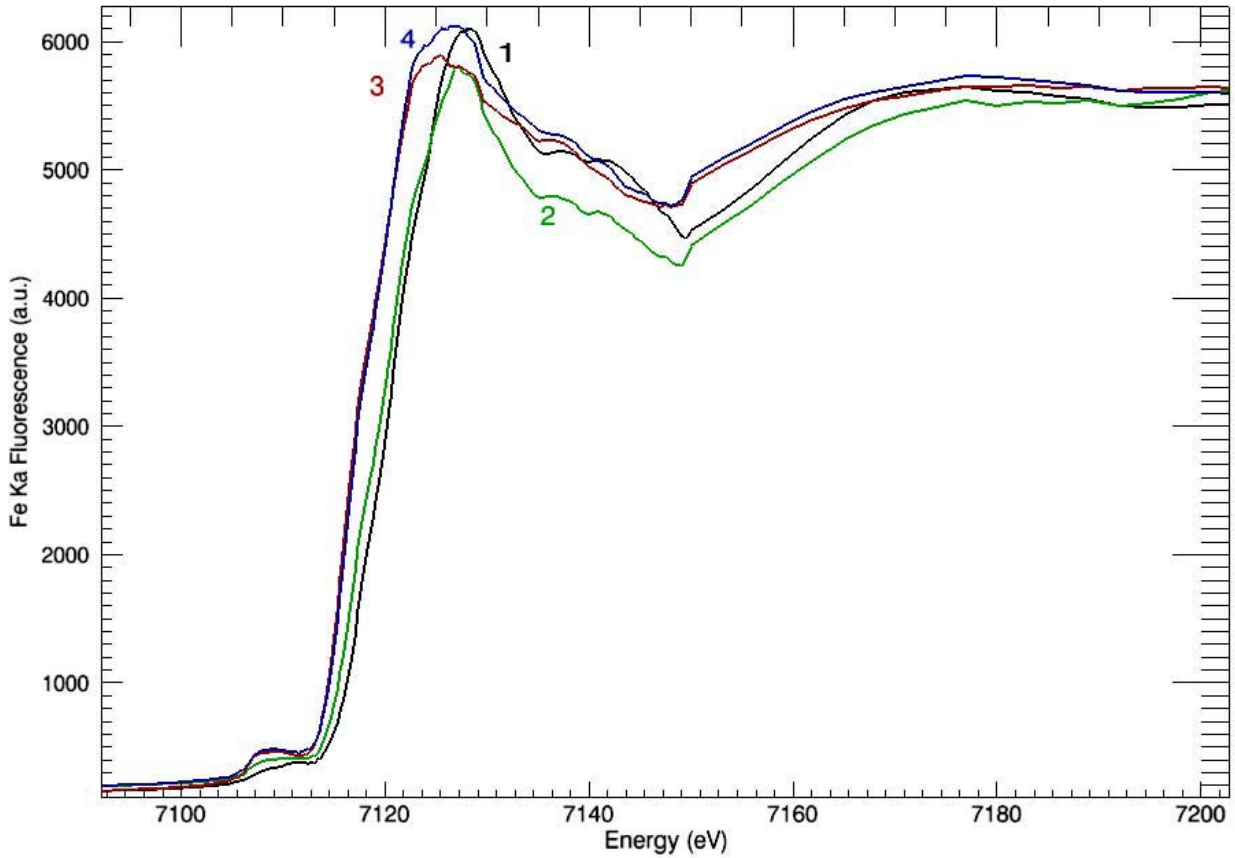


Fe (B) Ti (G) Mn (R)

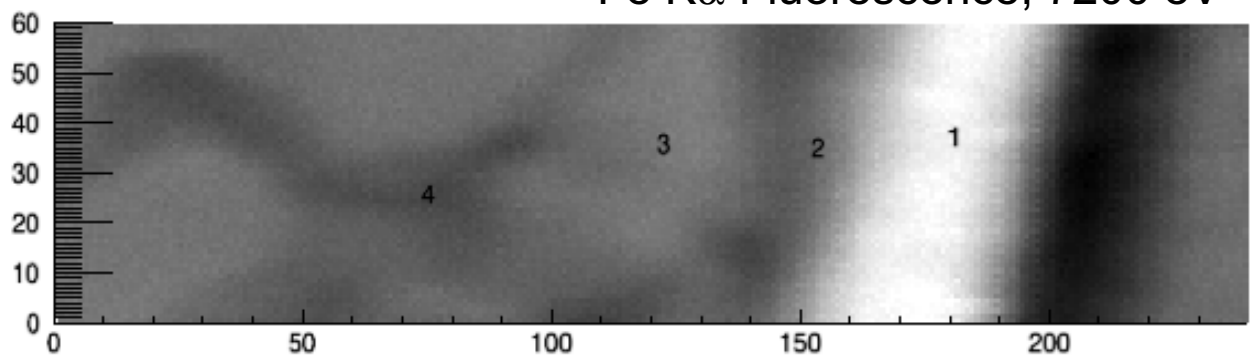
1,440,000 pixels, 16 msec dwell time
1.5 x 1.5 mm, would take 17 days at 1 sec!

E= 7030.00 eV


Energy
150
100
50
60
50
40
30
20
10
0
0
50
100
150
200



Fe Kα Fluorescence, 7200 eV



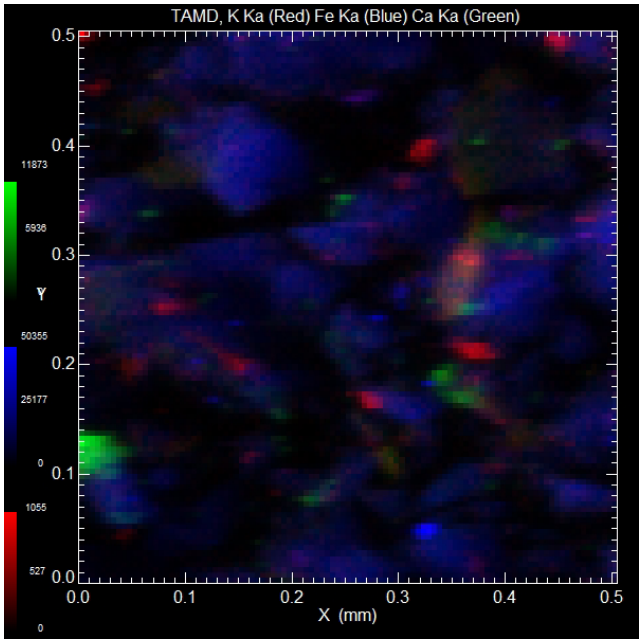
14,400 pixels
16 msec dwell time
197 energy steps

Future (now) Data Collection for X-ray Microprobes

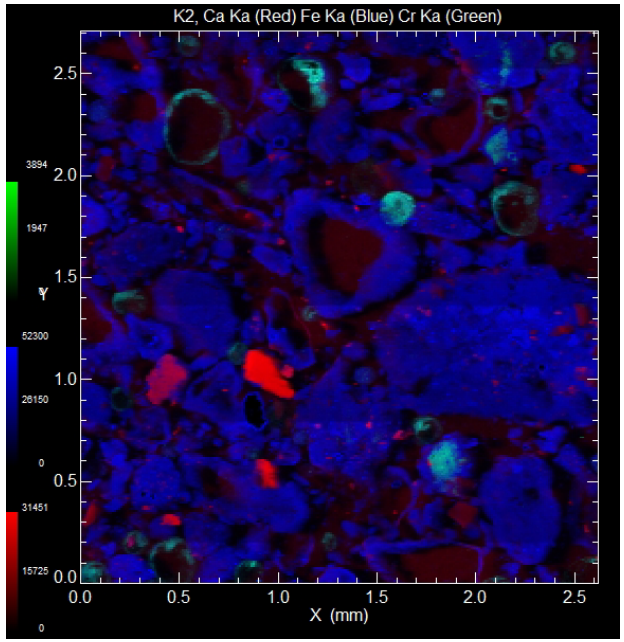
- **3D!**
- **NSLS-II will push the limits with respect to spatial resolution**
  - Beam stability an issue
  - Radiation dose needs to be minimized
  - Detectors are getting faster
- **Spectroscopy on an undulator will likely best be done in a scanning mode**
- **All this means that continuous scan modes will probably evolve to be the standard mode of operation**
- **Consequences for data storage (easily hundreds of Gb per user), transfer (you'll need fast fiber networks and you won't be getting data to users by internet), and retention (should be forever!)**

HDF File Size

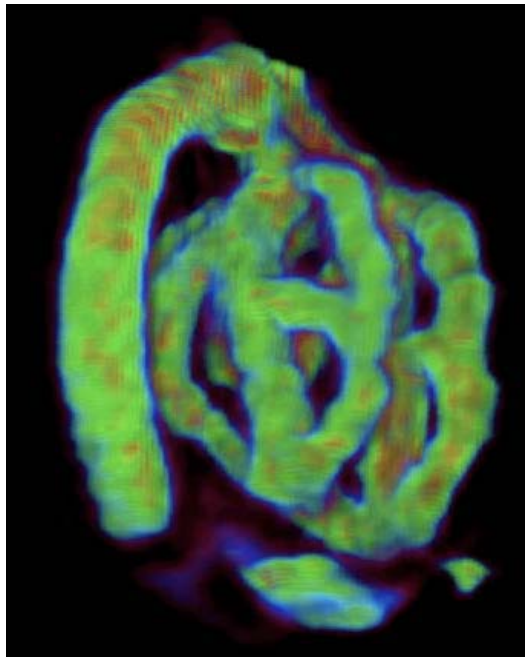
- We currently save the entire energy dispersive data for every pixel, however for array data we sum all detector channels into a single 2048 channel array



101x101, 10201 pixels
8 Mb



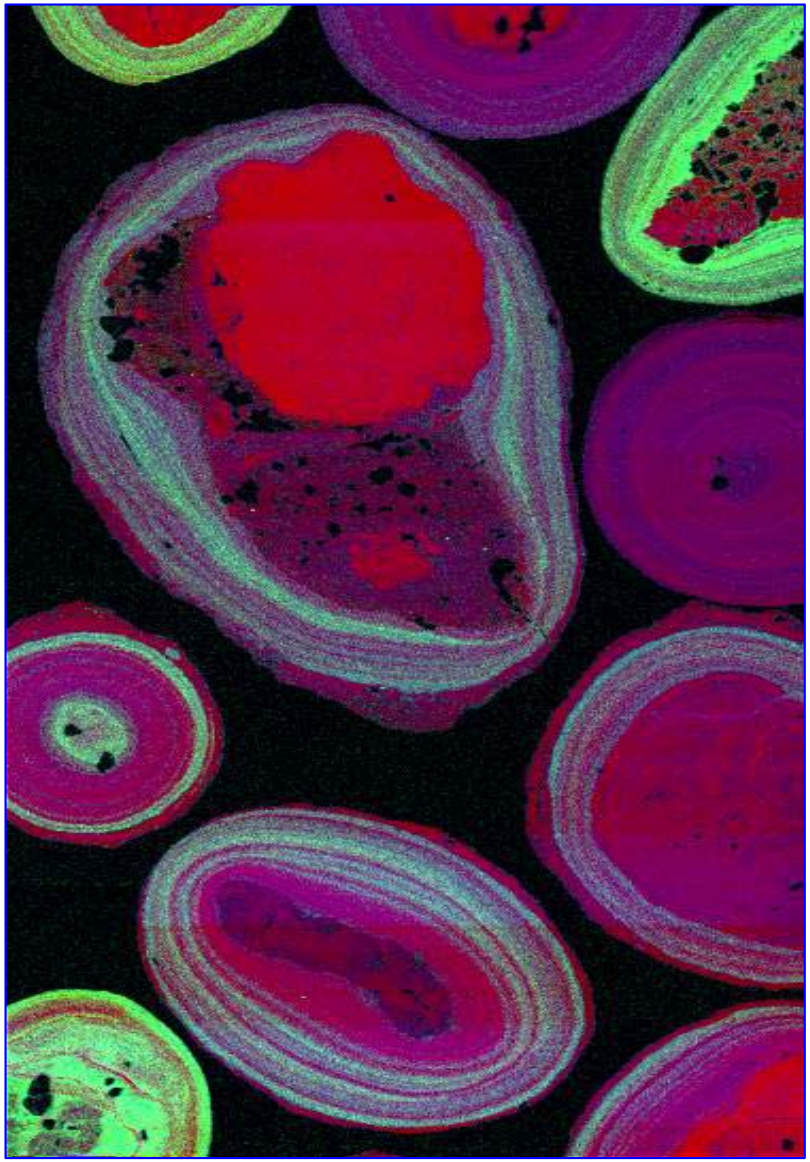
262x271, 71002 pixels
103 Mb



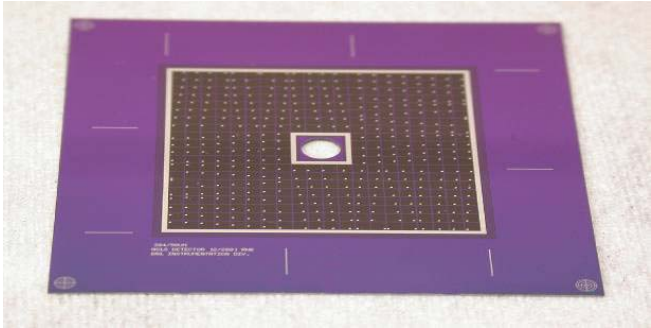
97x164x200,
3181600 pixels
3.7 Gb

Add XRD area detector data
~140 Gb

Siddons/Ryan MAIA Detector



Fe-Y-Cu RGB composite (1500 x 2624 pixel images, 13 x 21 mm^2)



1500x2624

w/ 1 MCA channel ~ 3 Gb
w/32 channels ~ 99 Gb
w/ 100 channels ~ 310 Gb

(Ryan GeoPIXE II software also written in IDL and supports multiple synchrotron microprobe data formats. Commercial software)

Future (now) Data Collection for X-ray Microprobes

- **3D!**
- **NSLS-II will push the limits with respect to spatial resolution**
  - Beam stability an issue
  - Radiation dose needs to be minimized
  - Detectors are getting faster
- **Spectroscopy on an undulator will likely best be done in a scanning mode**
- **All this means that continuous scan modes will probably evolve to be the standard mode of operation**
- **Consequences for data storage (easily hundreds of Gb per user), transfer (fast fiber networks and you won't be getting data by internet), and retention (should be forever!)**
- **This has consequences for how data is collected**

Maybe we should rethink the problem

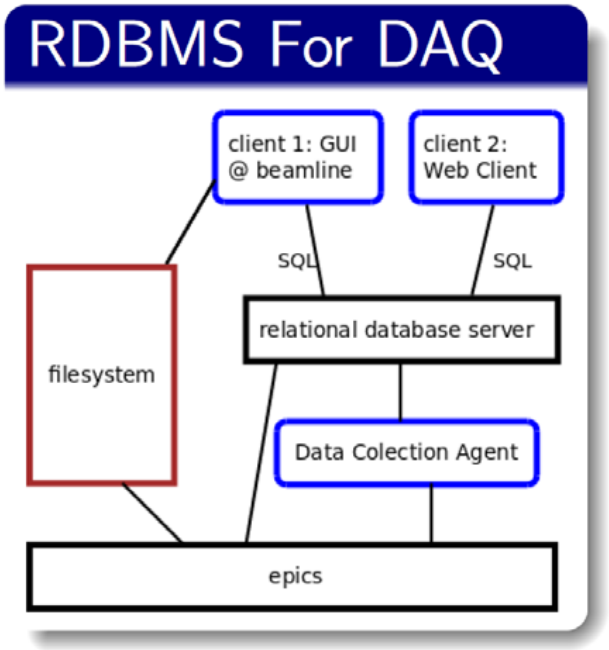
- There are many types of data for an x-ray beamline: XRF spectra, scaler data, XRD detector images, Video images, sample information, external environmental data (ring current, room temperature) and stuff we haven't thought about yet
- There's actually a lot of information about an experiment we'd like to track (user info, sample provenance, measurement conditions)
- How do we relate differing measurements made on a single sample? For example, a series of XAS spectra collected from points identified on an XRF compositional map?
- Continuous scanning modes work best when NOT streaming data to a single file. Whether using the XIA XMap continuous scan modes or the Maia list scan mode, an experiment is defined by a series of files. You can even "fill-in" data later to, for example, improve statistics or fill in data loss.
- Do we need to think about an entire experiment as a single set of data, or must each scan and map stand on it's own?
- I want my users to walk away with all this data, tied together

Maybe we should rethink the problem

- There are many types of data for an x-ray beamline: XRF spectra, scaler data, XRD detector images, Video images, sample information, external environmental data (ring current, room temperature) and stuff we haven't thought about yet)
- There's actually a lot of information about an experiment we'd like to track (user info, sample provenance, measurement conditions)
- How do we relate differing measurements made on a single sample? For example, a series of XAS spectra collected from pits identified on an XRF compositional map?
- Continuous scanning modes work best when NOT streaming data to a single file. Whether using the XIA XMap continuous scan modes or the Maia list scan mode, an experiment is defined by a series of files. You can even "fill-in" data later to, for example, improve statistics or fill in data loss.
- Do we need to think about an entire experiment as a single set of data, or must each scan and map stand on it's own?
- I want my users to walk away with all this data, tied together

It's starting to look like we need a database!

Relational Database for Metadata and Data Collection



- Could we use a relational database at the heart of data acquisition?

- Relational Databases (MySQL, Postgres, Oracle) are robust data storage and retrieval systems, and solve many issues for modern data acquisition:
  - ❑ Multi-client access to database during experiment.
  - ❑ Remote Access.
  - ❑ Security can be built in.
  - ❑ Automated logging.
  - ❑ After experiment, DB can be converted to single database (SQLite).

- What would go into the database for acquisition?
  - ❑ Configuration of detectors, monochromators, motors, other instruments. Facility parameters. These could be continuously logged and then you query the data that corresponds to the timestamps of the experiment.
  - ❑ Queue of Macro commands to be run.
  - ❑ State information about what is being done, what has been done.

- When a user is finished the database can be queried to produce a single file for the user (in HDF5). We could package it how they like…run number, sample number, day, user id, etc.

Closing Thoughts

- Lots of details need to be considered. What is the data retention policy, what are beamline and facility storage needs, network speeds need to be fast.
- User interface does need most work, what exists is good but there is no uniformity (which is not necessarily bad). But these will likely outdated as we move to an even more ingrained "imaging" mode in collecting data. A hyperspectral format for microprobe experiments will be standard.
- But it's easy to over-engineer these solutions. Flexibility is important. An imposed solution that strives to be "facility-wide" can leave individual beamlines stuck when new technologies become available.
- Common data formats are a good idea, but they must take into account the variabiltiy in hardware (what is i0? Where is the detector? What is the airpath?). Without these details you can't quantify properly.
- There is a critical need for remote access (with all the headaches of navigating firewalls)
- I haven't even touched on data processing, every microprobe has their own solution and generally they're **ALL GOOD**. Ask the users what they want. Just remember, for x-ray microprobes getting from XRF counts to elemental abundance not trivial. It's very easy to get it wrong.
- But bear in mind, data processing tools based on robust commercial applications (IDL, MATLAB) give users endless options on how they can visualize their data. Silly to reinvent the wheel.