Meeting Minutes
X-ray Diffraction Methods Subcommittee
ICDD, March 25, 2015

R. Papoular called the meeting to order at 10:12 am.
Appointment of Minutes Secretary: M. Suchomel
Approval of Minutes from 2014 passed without objection.
Attendance list of attendees was distributed.

Directors' Liaison Report by J. Rius
Text below is read

In conjunction with the running Excipient Project & Top 200 Drug Initiative, Leoni moved that the joint measurement(s) of active pharmaceutical ingredients of interest, together with those of their solid excipients that are not represented in the ICDD PDF databases, be added.

Response: *Database Department – The Top 200 Drug Initiative has entered another phase where the lists are updated and expanded. A list of excipients to be added to the database needs to be generated by the Organics and Pharmaceuticals Subcommittee. This list will be used to prioritize excipients that should be added to the Top 200 Drugs project. The linking of APIs with expected excipients will require member input through a task group.*

Additional Comment: *Executive Director – During a recent project review, the editorial team used some authoritative references on excipients to identify candidates and cross verify subfile assignments (i.e. are all excipients labeled as such).*

T. Blanton made a comment:
ICDD board/editorial panel is looking for member input on the topic – experts on which flags and info on modern instruments to include in the database – more discussed later in the meeting.

R. Papoular gave an introduction for C. Gilmore (University of Glasgow)

Meeting proceeded with four short presentations

Presentation # 1

By T. Blanton, titled:
"2D XRD – next generation ability to directly analyze 2D XRD data in ICDD products"

This topic is being worked on at ICDD headquarters by a small team with goals of:

- Displaying 2d data
- Simulating orientation
- 2D Data processing
- Phase ID directly in 2D images

Current focus is in reading image files using data available from Bruker and Rigaku.

There are several things to consider when working with image files: pixel size, sample to detector distance, and corrected data.

ICDD's team currently has made some progress to read 2D images, integrate, and do phase ID. T. Blanton showed examples.

See more detail during J. Blanton's talk Thursday morning: "New Product Developments"

T. Blanton requests a TASK GROUP to define new flags for modern instruments conditions that should be included in the database, especially related to 2D data (parallel beam optics, image plate, *etc.*). Need three to four volunteers.
- R. Papoular, S. Kabekkodu, T. Ida, and Scott Misture

Presentation # 2
B. Toby presented: *"ScanCCD: high resolution and reduced-cost powder diffraction area detection"*

Topics covered:

- How do you get high-resolution powder diffraction?
- Synchrotrons use Analyzer Crystals (i.e. 11-BM)
- Focused beam on Area Detector (limited by size of pixels)
- How do we make smaller pixels?
- Coupled CDD Camera Readout to 2θ arm motion.
- See B. Toby's paper for details on this detector design and its performance

ABSTRACT:

ScanCCD: high-resolution and reduced-cost powder diffraction area detection. Area detectors provide several advantages over point and linear detectors: solid angle of detector can be optimized, increasing sensitivity; systematic error due to texture and sample graininess can be minimized through radial integration; low-angle peak asymmetry is eliminated; and full diffractogram collection can be 0.1 sec or faster. The most significant disadvantage is that the instrumental resolution is far from optimal. This is due to suboptimal focusing geometry choices and more significantly because area detectors typically have large pixels compared to what would be desired. Our detector design couples the line-readout property inherent in a CCD design with motion of the detector to provide data collection times on the scale of a minute and has demonstrated resolution close to that of the best available in the US. Unlike most CCD detectors used for diffraction, there is no need for the expensive demagnification fiber-optic coupler that is the single most costly component of many detectors.

Presentation # 3

B. Toby presented: *"Parametric fits and secondary analysis in GSAS-II "*

Diffraction is now commonly used for in-situ exploration; they wanted to rethink the idea of considering each pattern one-by-one; GSAS-II is trying to add ways for parameterization of sequential refinements.
See more details on tutorial for $CuCr_2O_4$ data from 7 to 300 K.

Added feature

- Fit functions of parameters (for example b/c ratio)
- Fitting equations of state (parametric functions)

ABSTRACT:

Parameter fits and secondary analysis in GSAS-II Parametric Rietveld refinements are used when powder diffraction data are collected as a function of one or more physical or chemical variables, such as temperature or in situ reaction conditions. A number of approaches have been used to perform such analyses by experts, but GSAS-II is the first package to offer this as a general feature and to offer secondary fits with proper treatment of covariance, when a user supplies an equation of state. I will explain the two approaches uses for fitting. I will show how these refinements are handled in GSAS-II and how expressions from refined variables can be computed and how equations are fit.

R. Papoular questioned: Which version of Python is B. Toby using?

Answer = 3.x and 2.x (for the moment you want to start with 2.x to use the full packages)

Anaconda distribution package is okay
Enthought package (includes code editor)
Both have free versions and B. Toby is still testing to see which is best.

Presentation # 4

C. Gilmore (University of Glasgow) presented: "New ideas for databases"

C. Gilmore is thinking about how to explore data more easily (e.g. 100's of patterns).
Define some correlation coefficient to evaluate similarity (or difference) of patterns. This is in the domain of 'cluster analysis.'

Example: NH_4NO_3 (~ 100 datasets) several transitions
Makes tree-root diagrams (dendrogram) and also ways to visualize multimensional data.

Can combine correlation matrices between different data sets: e.g. Raman & PXRD data. Together, this joint clustering shows more information than any individual data does, and enables new search-match capabilities and quantitative analysis.

MOTIONS SUGGESTED (by R. Papoular)

Proposed Motion #1: When a pharmaceutical / organic crystal structure (with its atomic coordinates) is mentioned in the a PDF card it would be useful to mention the worst [non-H - non-H], [non-H - H] and [H – H] intermolecular van der Waals contacts as obtained from e.g., the MERCURY / CCDC software.

Discussion:
T. Fawcett doesn't want the database to highlight the **worst** parts of each pattern; bad for marketing. You would say which is the 'best" part of each entry; but showing the three worst parts is not acceptable.

B. Toby said that perhaps this **criteria** could be used internally at ICDD. But you should also be clear to highlight what is good and useful for a structure (need a balance). Is this part of the internal 'quality mark' evaluation?

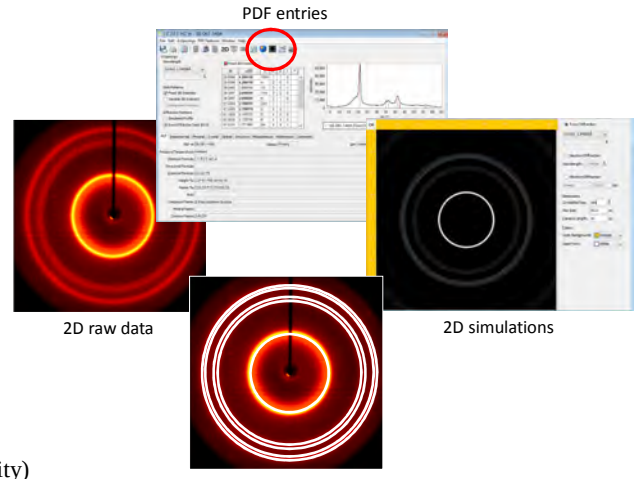
Proposed Motion #2: There is a wide range of X-ray powder diffractometers, from a synchrotron HR beamline to a Bruker bench top. Does inclusion in the PDF database preclude the inclusion of data obtained at the other extreme, provided that the associated data show no appreciable impurity phase and that the diffractograms turn out to be indexable with reasonably good figures of merit?

Time for discussion is cut off; no motions passed in 2015.

Adjournment at 11:15 a.m.

Back to Minutes

2D XRD – next generation ability to directly analyze two-dimensional XRD data



Tom Blanton
Kai Zhong
Stacy Gates
Justin Blanton
Scott Misture (Alfred University)
Bob He (Bruker AXS)

Project Components

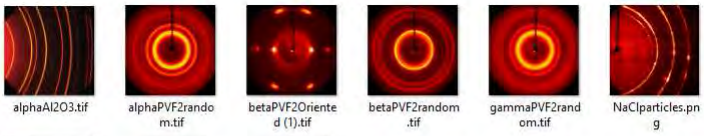
- *Display 2D data*
  - *On center – ability exists today in ICDD data mining software*
  - *Off center*
  - *Combine multiple frames of data*
  - *0-160° image plate data*
- *Simulate Orientation Patterns*
  - *Random orientation – ability exists today in ICDD data mining software*
  - *Preferred orientation (i.e. fiber, biax, planar, in-plane, etc.)*
  - *Single crystal*
- *Data Processing*
  - *Read in 2D image file*
    - *Convert to 1D pattern*
    - *Keep as 2D pattern*
  - *Read in 2D raw data file*
  - *Convert to 1D pattern*
  - *Keep as 2D pattern*
- *Phase identification using 2D data*
  - *Random orientation samples*
  - *Preferred orientation samples*
  - *Single crystal samples*

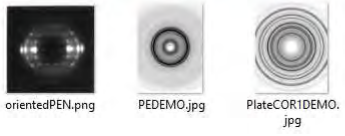
Back to Minutes

Project Objectives

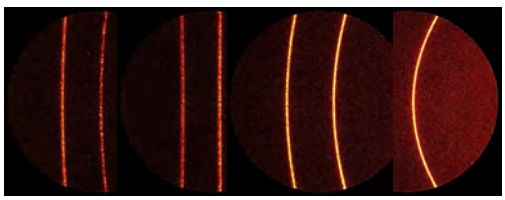
- *Objective 1: 1a Display 2D data on center>3a Read in 2D image file >3ai Convert to 1D pattern>Sleve+ analysis*
- *1a Display 2D data on center>3b Read in 2D raw data file >3ai Convert to 1D pattern>Sleve+ analysis*
- *Objective 2: 1b Display 2D data off center>3a Read in 2D image file >3ai Convert to 1D pattern>Sleve+ analysis*
- *1b Display 2D data off center>3b Read in 2D raw data file >3ai Convert to 1D pattern>Sleve+ analysis*
- *1c Display 2D data combined multiple frames>3a Read in 2D image file >3ai Convert to 1D pattern>Sleve+ analysis*
- *1c Display 2D data combined multiple frames>3b Read in 2D raw data file >3ai Convert to 1D pattern>Sleve+ analysis*
- *1d Display 2D data 0-160° image plate>3a Read in 2D image file >3ai Convert to 1D pattern>Sleve+ analysis*
- *1d Display 2D data 0-160° image plate>3b Read in 2D raw data file >3ai Convert to 1D pattern>Sleve+ analysis*
- *Objective 3: 1a,b,c,or d Display 2D data >3a Read in 2D image file >3aii Keep as 2D pattern>4a Phase id using 2D data random orientation*
- *1a,b,c,or d Display 2D data >3b Read in 2D raw data file >3bii Keep as 2D pattern>4a Phase id using 2D data random orientation*
- *Objective 4: 2a Simulate random orientation 2D patterns*
- *2b Simulate preferred orientation 2D patterns*
- *2c Simulate single crystal 2D patterns*
- *Objective 5: 1a,b,c,or d Display 2D>3a Read in 2D image file >3aii Keep as 2D pattern>4b or c Phase id using 2D data preferred orientation or single crystal*
- *1a,b,c,or d Display 2D>3b Read in 2D raw data file >3bii Keep as 2D pattern>4b or c Phase id using 2D data preferred orientation or single crystal*

Data available March 1, 2015

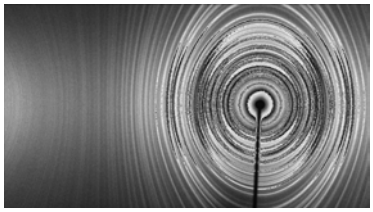




Miscellaneous phases - ICDD



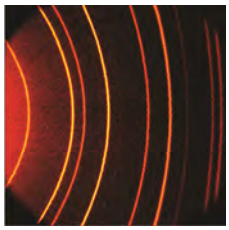
Si (four frames) – Bob He, Bruker
(LaB_6, α-Al_2O_3 also provided)



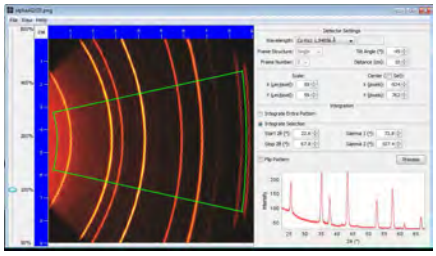
α-Al_2O_3 – Aya Takase, Rigaku

Back to Minutes

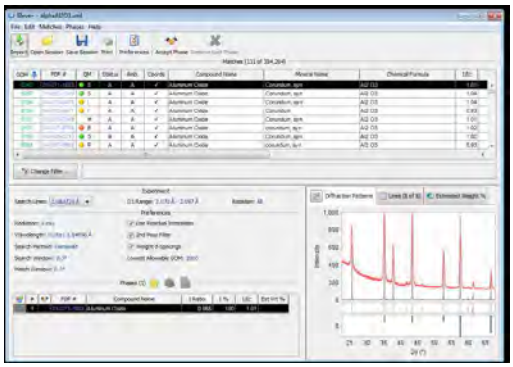
Working with image files



TIFF format image file
α-Al2O3 powder in capillary
2t zero 45deg, Cu Kα
L=10cm, 59μ² image pixel size
Data corrected for flat
detector distortion



2D image file read and converted
to a 1D diffraction pattern
D,I list generated



Sleve+ phase ID – successful

Important to know detector distance, image pixel size, and appropriate data corrections

Instrument Parameters

- We need a task group to help with adding new instrument classifications to the database
- Example we have for instruments:
  - Diffractometer
  - Debye-Scherrer
  - Guinier
  - Gandolfi
  - Other
  - Calculated spacings
- New detectors, optics, are not currently included in the database

Scanning CCD Detector for High Resolution X-ray Powder Diffraction

Timothy Madden

Brian Toby

Robert Von Dreele

Matthew Suchomel

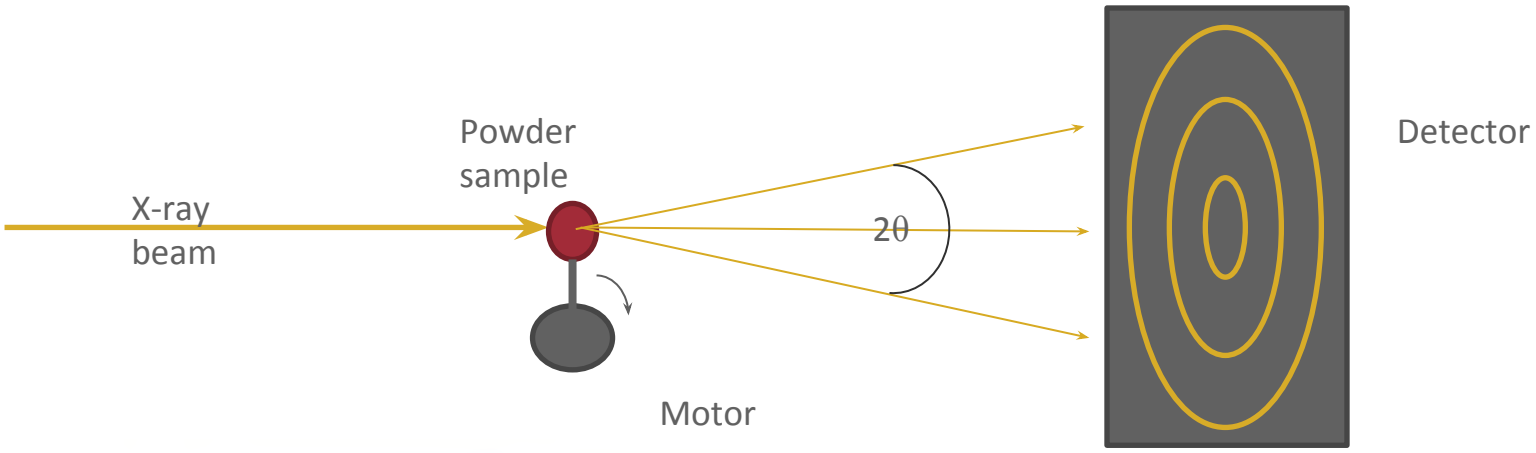
Jonathan Baldwin

Advanced Photon Source, Argonne National Laboratory

Generating High Resolution Powder Diffraction

Back to Minutes

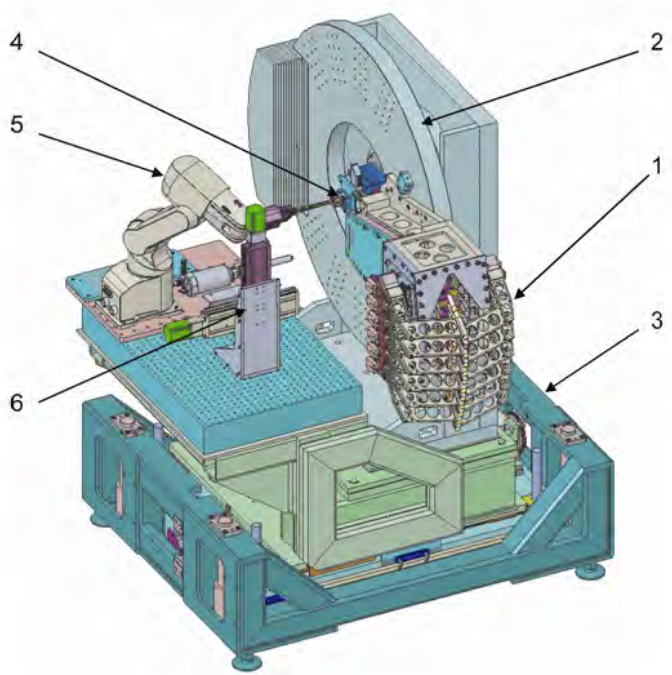
- Step 1: need very small $\Delta\lambda/\lambda$ (x-ray tube or perfect crystal monochromator)
- Step 2: need a small x-ray source with minimal divergence
- Step 3: need a detector that resolves a small solid angle

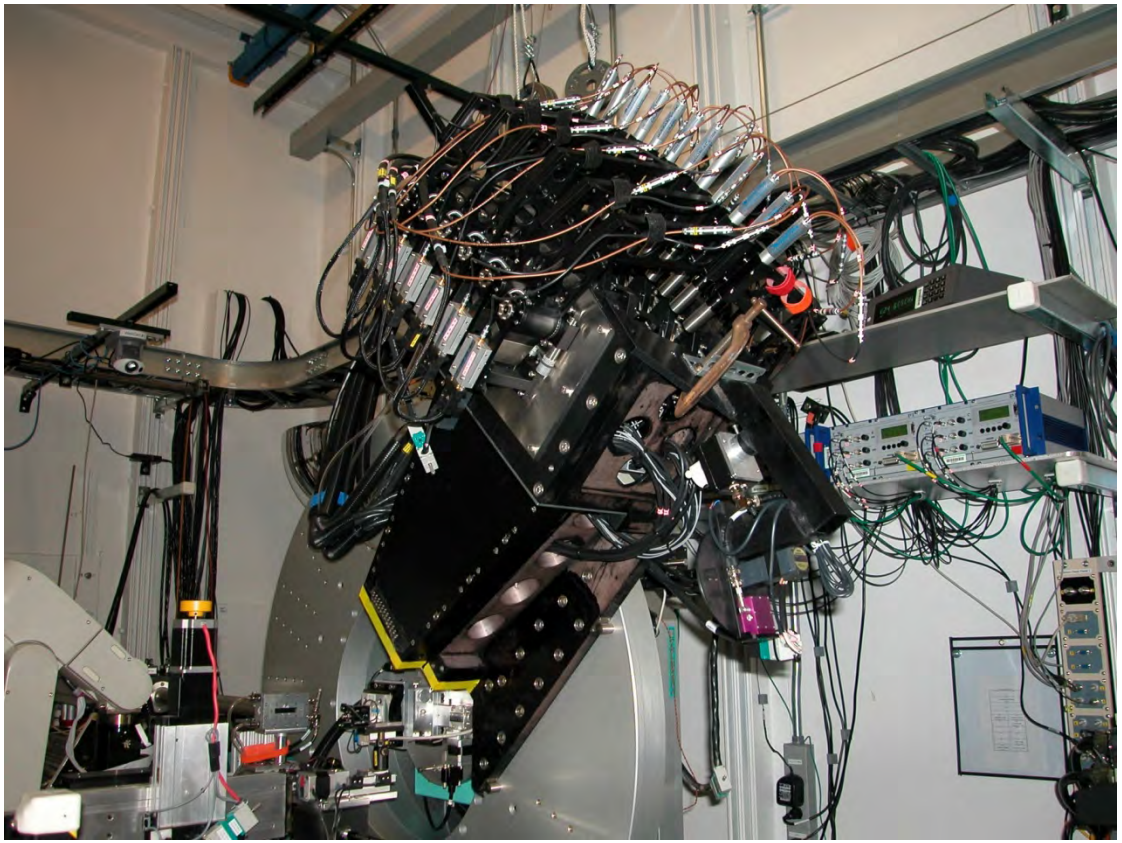


Powder Diffraction at APS Sector 11 (11-BM)

Back to Minutes

- Can take 1 hour to scan.
- Slow, but highest resolution in US.



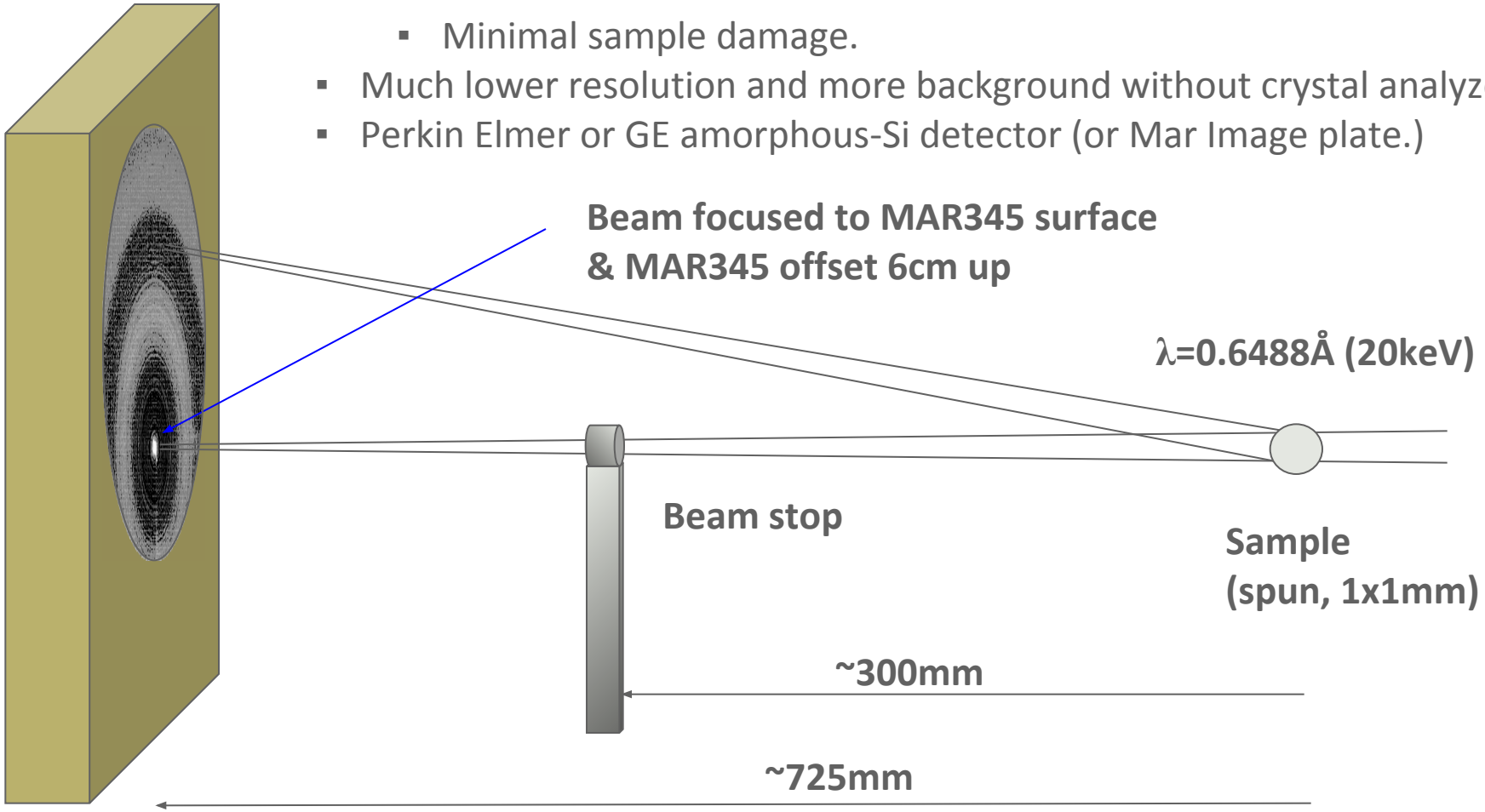


12 detectors with individual analyzer crystals, effectively 12 pixels for each arm position

Powder Diffraction: 1-ID, 6-ID, 11-ID and 17-BM

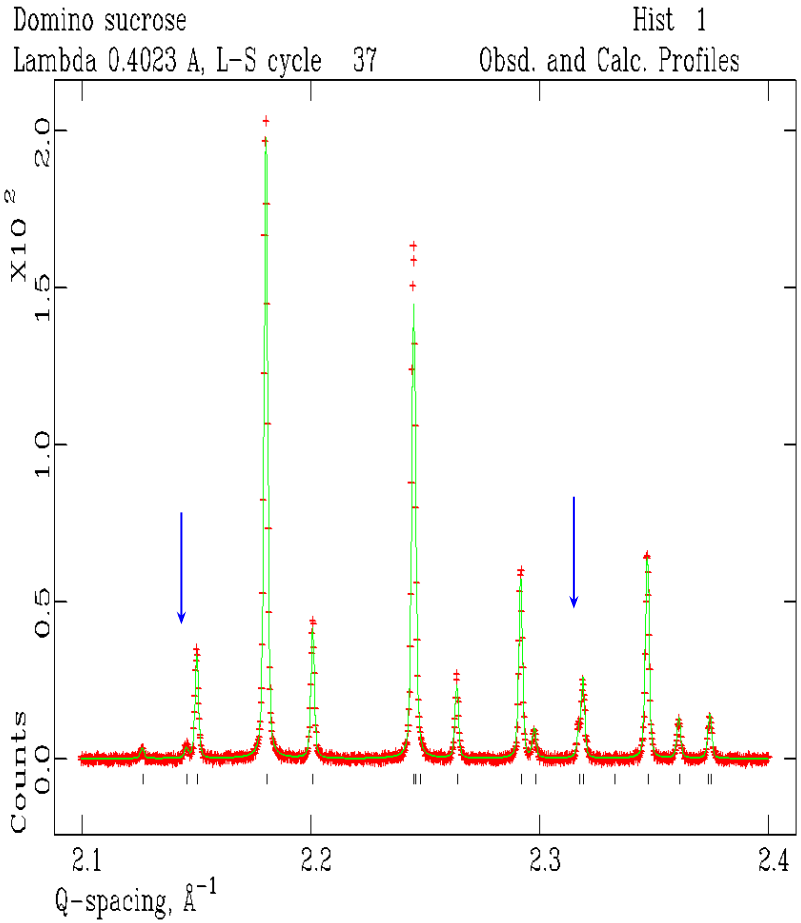
Back to Minutes

- Large Area Detector takes single “picture”.
- Fast: 0.1 to 30 sec to take data.
  - Minimal sample damage.
- Much lower resolution and more background without crystal analyzers.
- Perkin Elmer or GE amorphous-Si detector (or Mar Image plate.)

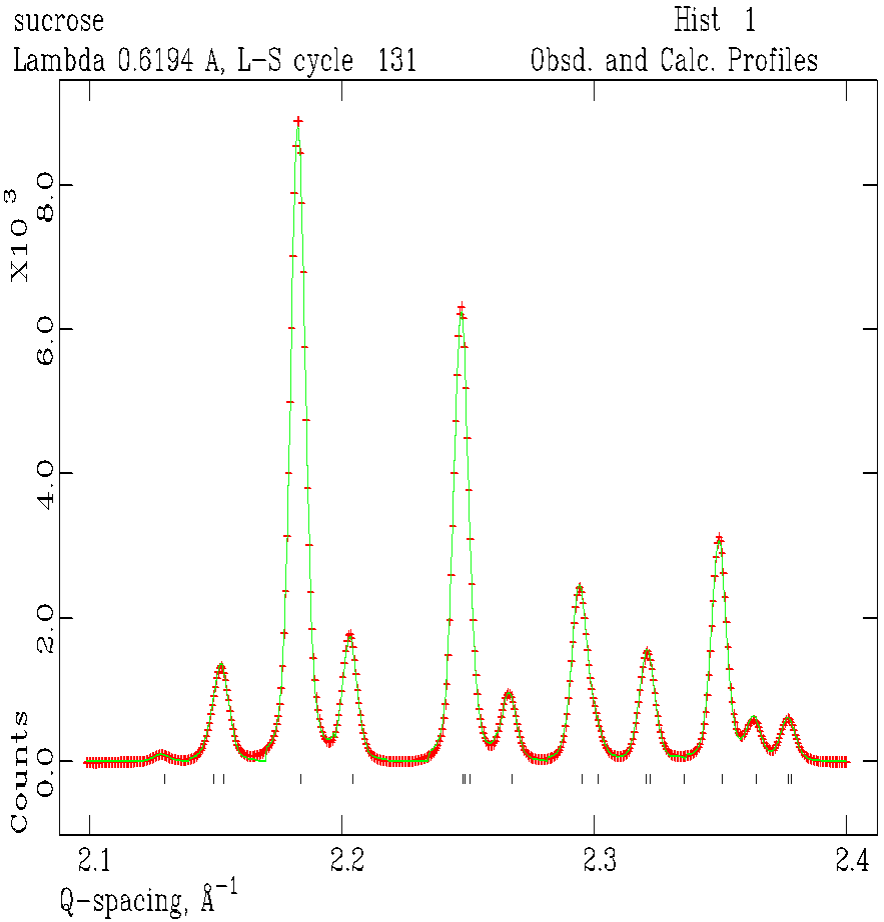


1BM/APS June 2002; thx Peter Lee

Compare image plate with analyzer/detector



11-BM – 10min scan



1-BM/MAR345 – 1sec exposure

3-4x broader peaks!

Problem: pixel size is too large

Back to Minutes

How can we make an area detector with small pixels?

Back to Minutes

- A typical area detector is 1K x 1K (1 Mpixel) to maybe 2K x 2K (4 Mpixel). After integration, will get ~1.5 pixels for each pixel along length
- For high resolution, we need ~40K x 40K pixels:
  - State-of-art: 2K x 2K G.E./Perkin Elmer detector $150K (x 20 = $3M!)
  - If we put 1 detector on an arm, we would need to step scan the arm
    - ~1 minute of motion for every second of data taking!
- Alternate solution: couple the read-out of a CCD to the detector arm motion

How a Charge-Coupled Device (CCD) Works

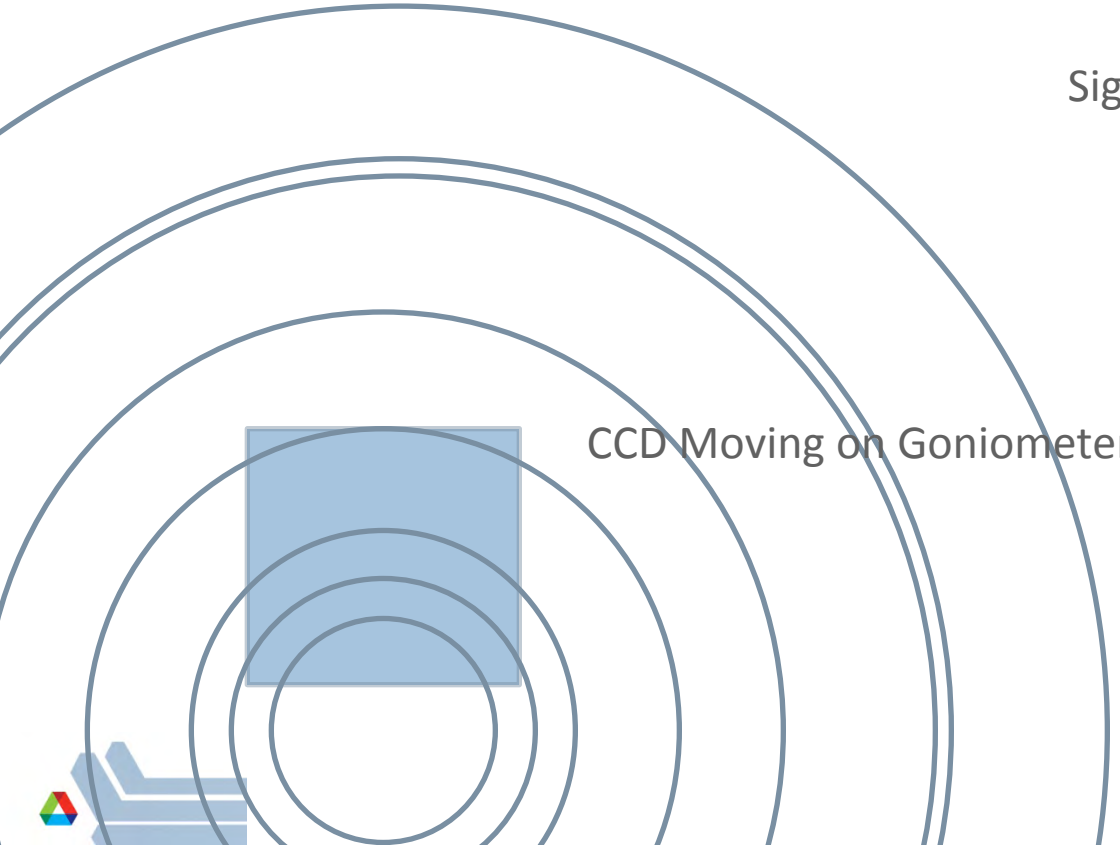
- Image integrates on CCD, creating electrons in the pixels.
- Image is read out:
  - Whole image shifts by one line. Actual electrons are shifted across Silicon wafer.
  - Line is shifted to output amplifier pixel by pixel.
  - Though we shift electrons inches across the chip, the device actually works.

Back to Minutes

Synchronizing CCD Readout with Movement

- CCD moves through diffraction rings projected in space
- CCD readout sync’ ed with movement.
- Area of CCD allows integration of signal as it moves through projected diffraction rings.

Signal on CCD

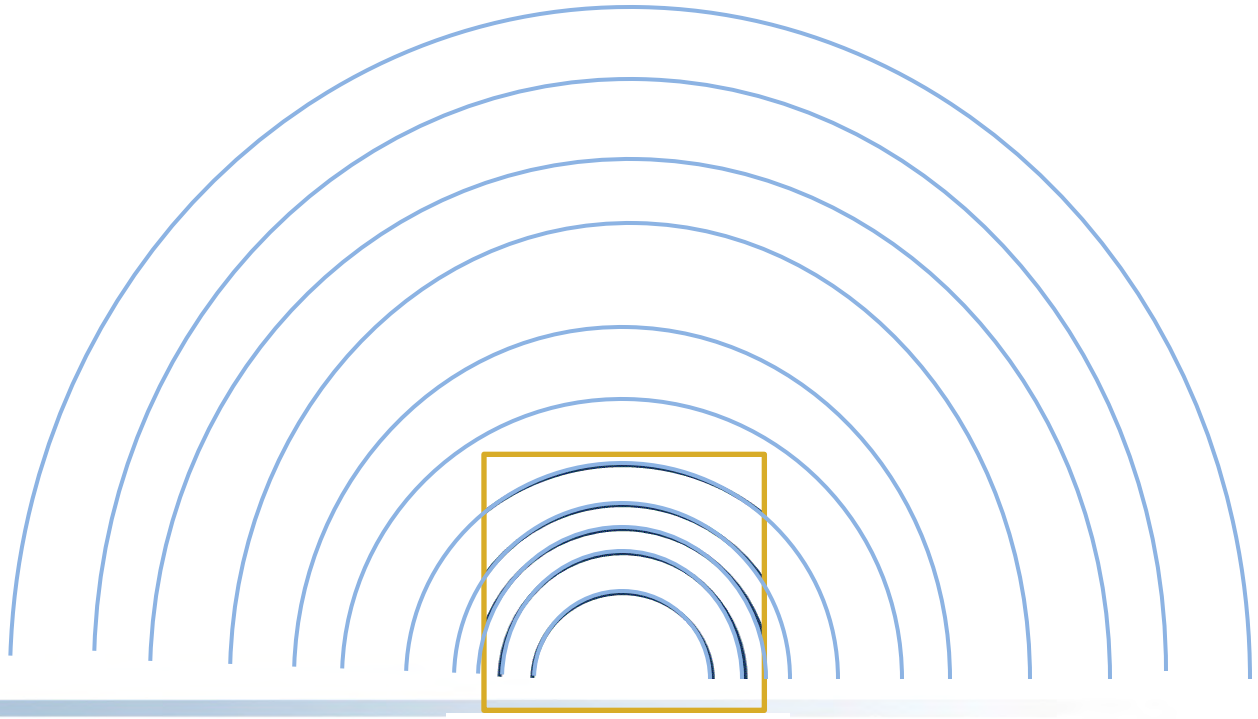


Back to Minutes

Blurring of Image due to shift errors

- The CCD does discrete shifts, the diffraction arm motor scans ~continuously
  - The image moves across a pixel as it is counted
  - The shift is not matched to the angle change from one motor step (requires sync. circuit)

- There will always be blurring
  - The blurring will be on the scale of one pixel (which can be trivially small)



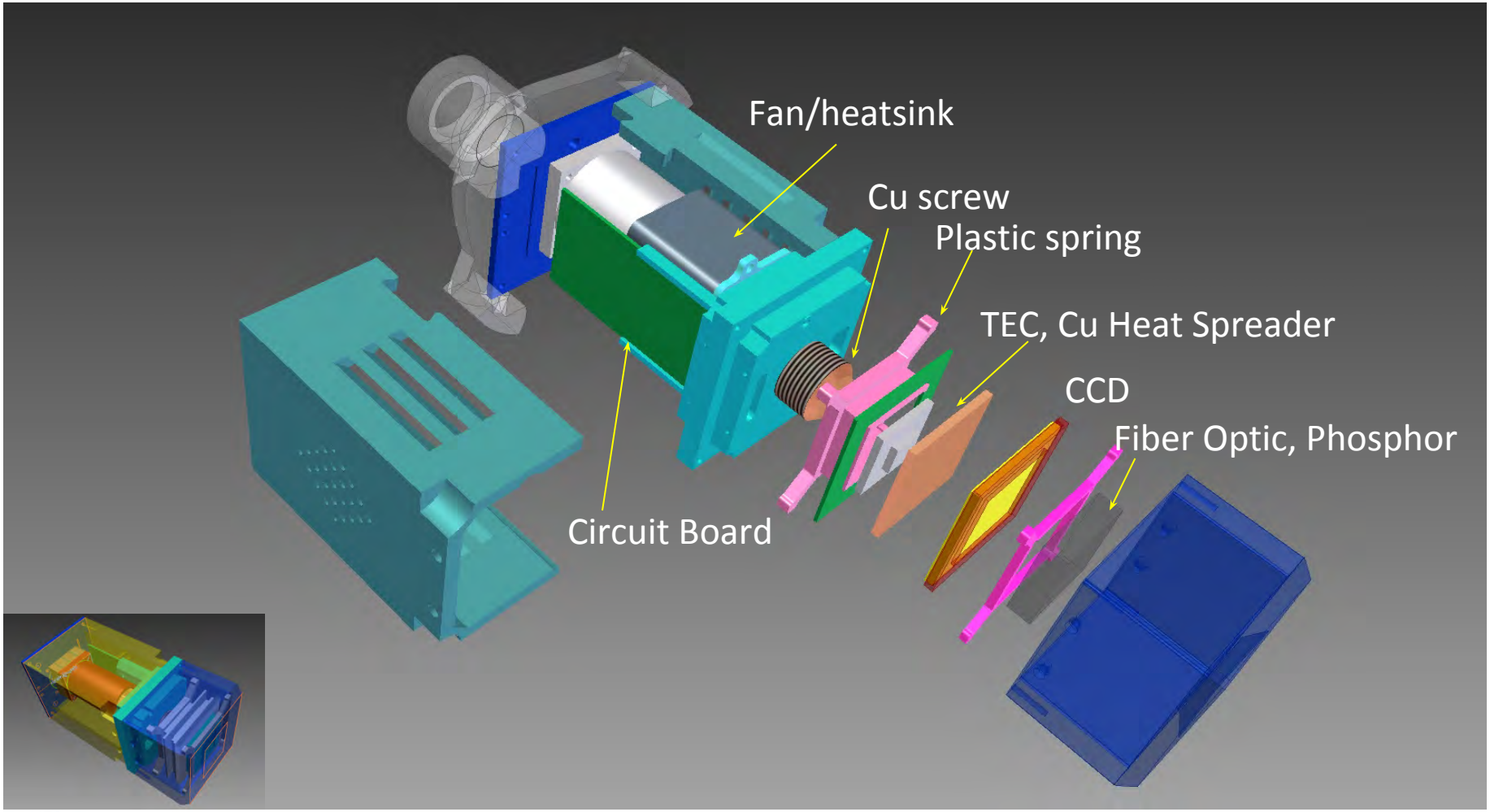
Back to Minutes

Interesting aspects of the Scanning CCD

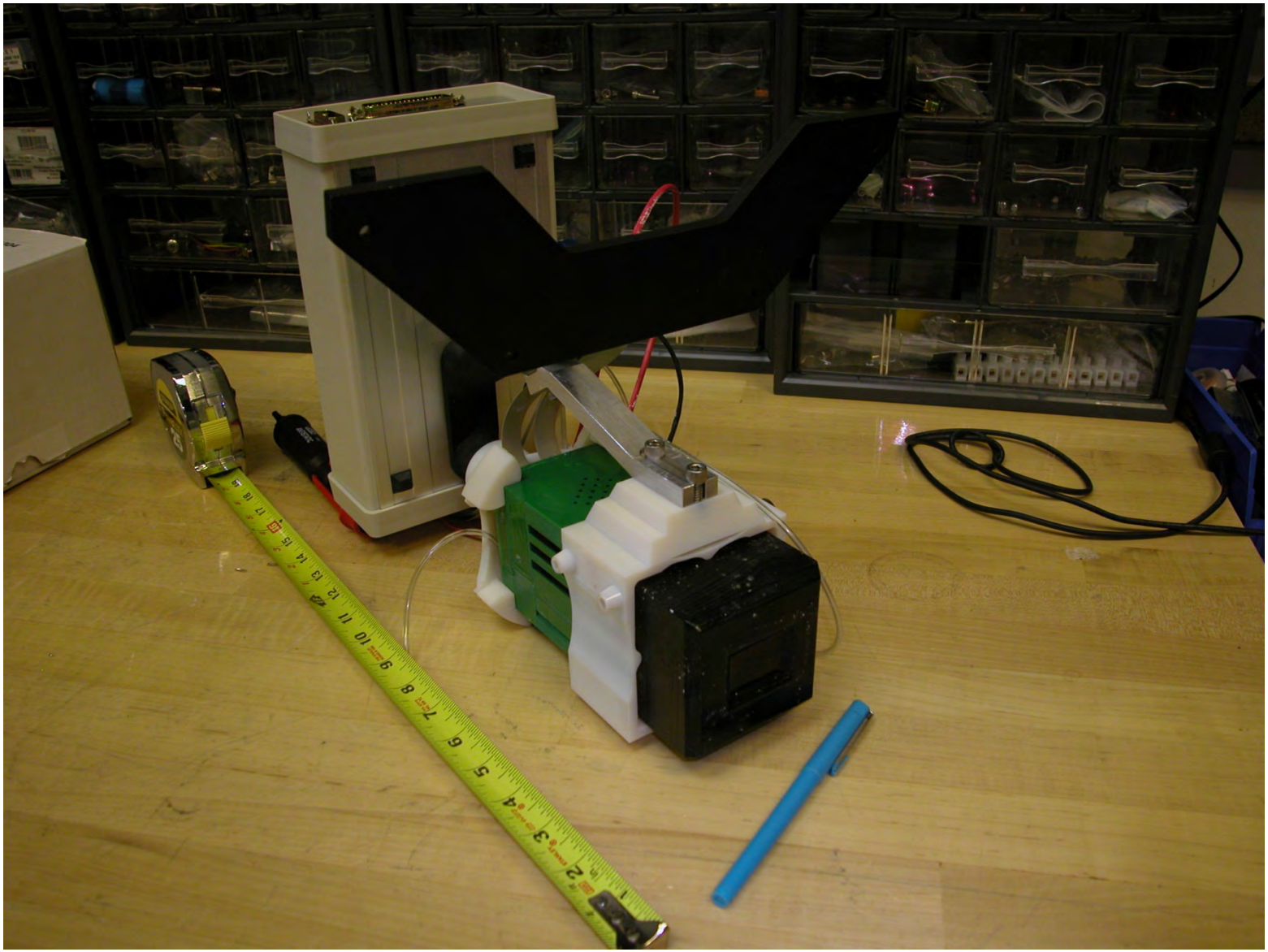
- Want a CCD with very fine pixels to instrumental preserve resolution
- Image integrates on the CCD as it moves. Longer a diffraction line is on the chip, the larger the signal.
- Because CCD moves, variations in the phosphor and chip sensitivity are smeared out.
- Dark current, or thermal noise that accumulates on the CCD, integrates only for the time it takes the CCD to move one CCD length.
- Must use optics on CCD with little or no spatial distortion. No need for demagnifying fiber optic (FO) taper with large spatial distortions. 1:1 FO plate used for detector isolation.
- Need very accurate motor system.
- Need way to reconcile that motor steps are likely not same size as CCD line shifts.

Mechanical Design

Back to Minutes



Finished Detector

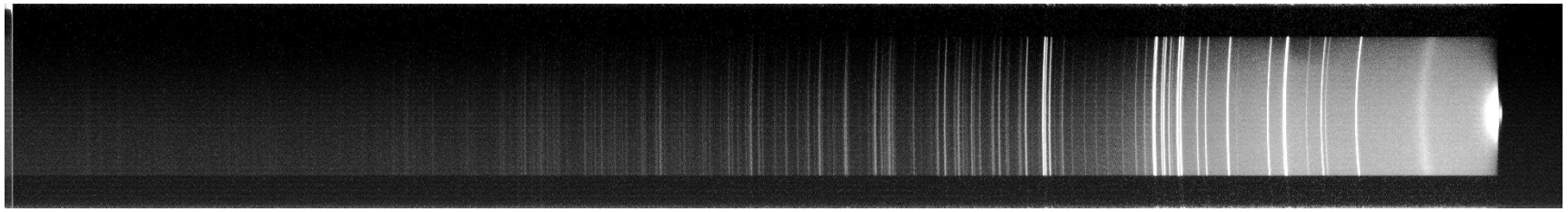


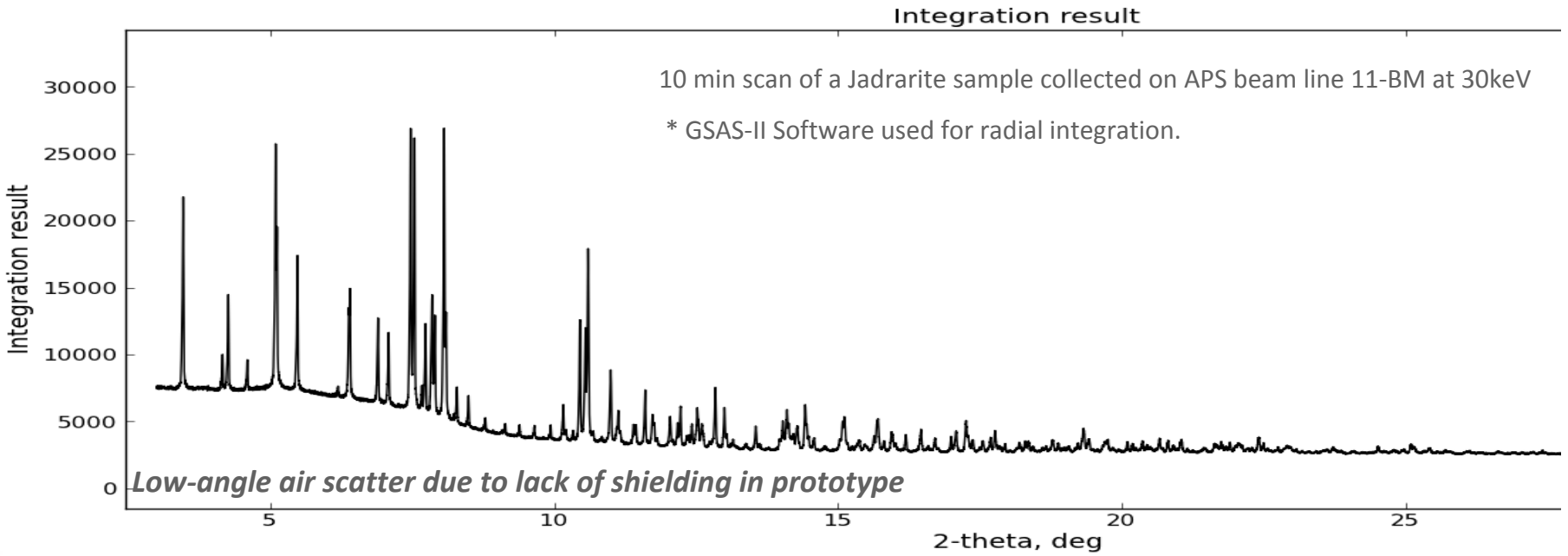
Back to Minutes

Testing of scanCCD at APS 11-BM

Back to Minutes

- Detector returns a series of 4k x 4k pixel images, 16 bit data. 32MB images.
- Images are then stitched into a Debye-Scherrer filmstrip replica.
  - Below, 10 images stitched. Images can be stitched without any fancy math.

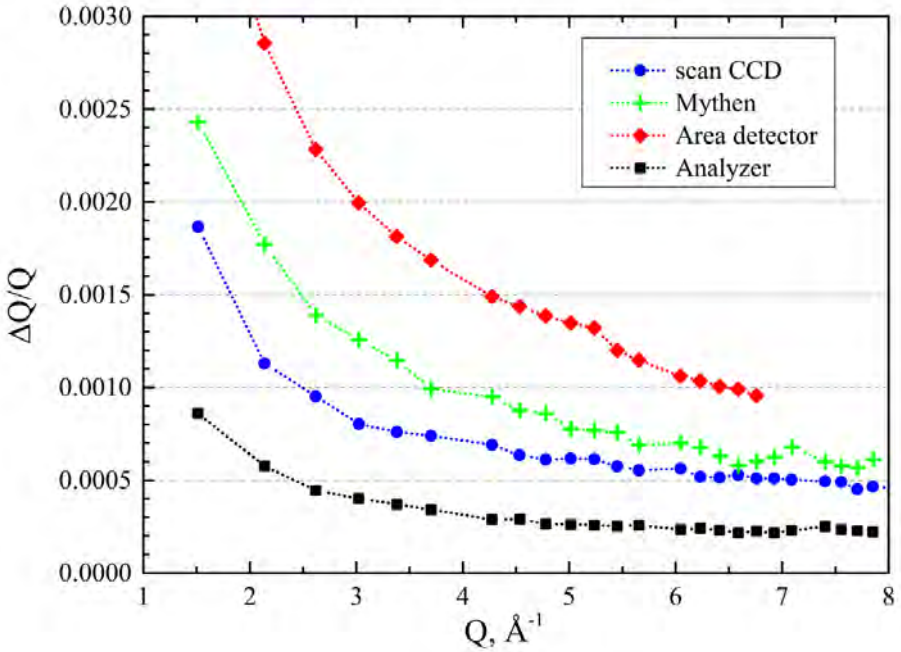


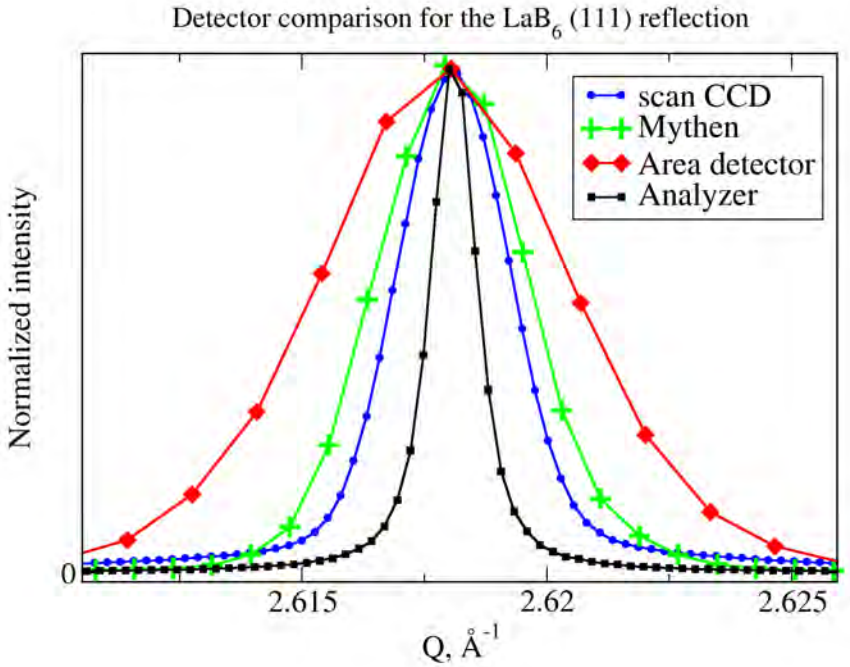


* Toby, B. H. & Von Dreele, R. B. (2013). Journal of Applied Crystallography 46

Comparing Detectors

- Sample: NIST SRM 660a (LaB_6) dusted inside a 0.7 mm diameter polyimide capillary tube.
- Several detectors used to scan sample at 11BM, APS: Perkin Elmer, Mythen, sCCD, 11-BM analyzer detector
- $Q = 2\pi/d$, where d is crystal lattice spacing. $\Delta Q/Q$ measures how accurately we can measure crystal structure. The smaller $\Delta Q/Q$, the better the detector.



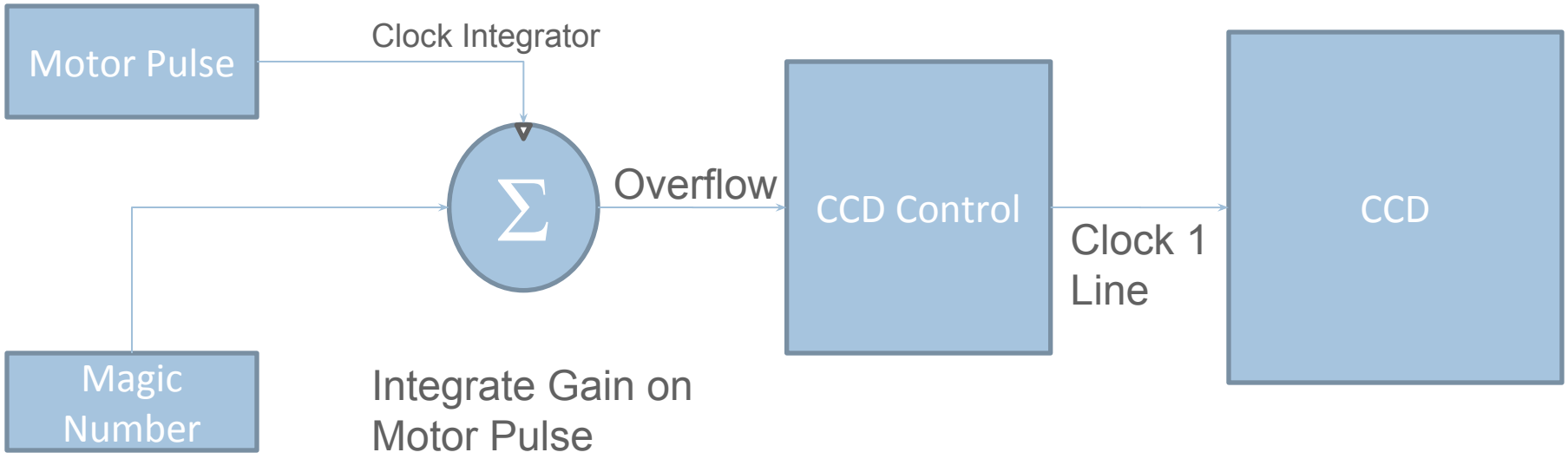


The APS scanCCD is close in resolution to 11-BM detector - analyzer system, but already 10X faster (could be improved at least x20 more)

Sync Algorithm

Back to Minutes

- Pixels are different size than motor steps.
  - G, a “magic number”, is ratio of Pixels P to motor steps M
  - P=floor(M*G)
  - Error of floor() shows up as 1 pixel (9um) of blur in output image.
  - Digital integrator counts motor steps. On integrator overflow we get a pixel shift.
- Currently can shift out lines at ~700Hz. Cabling problem preventing faster readout.

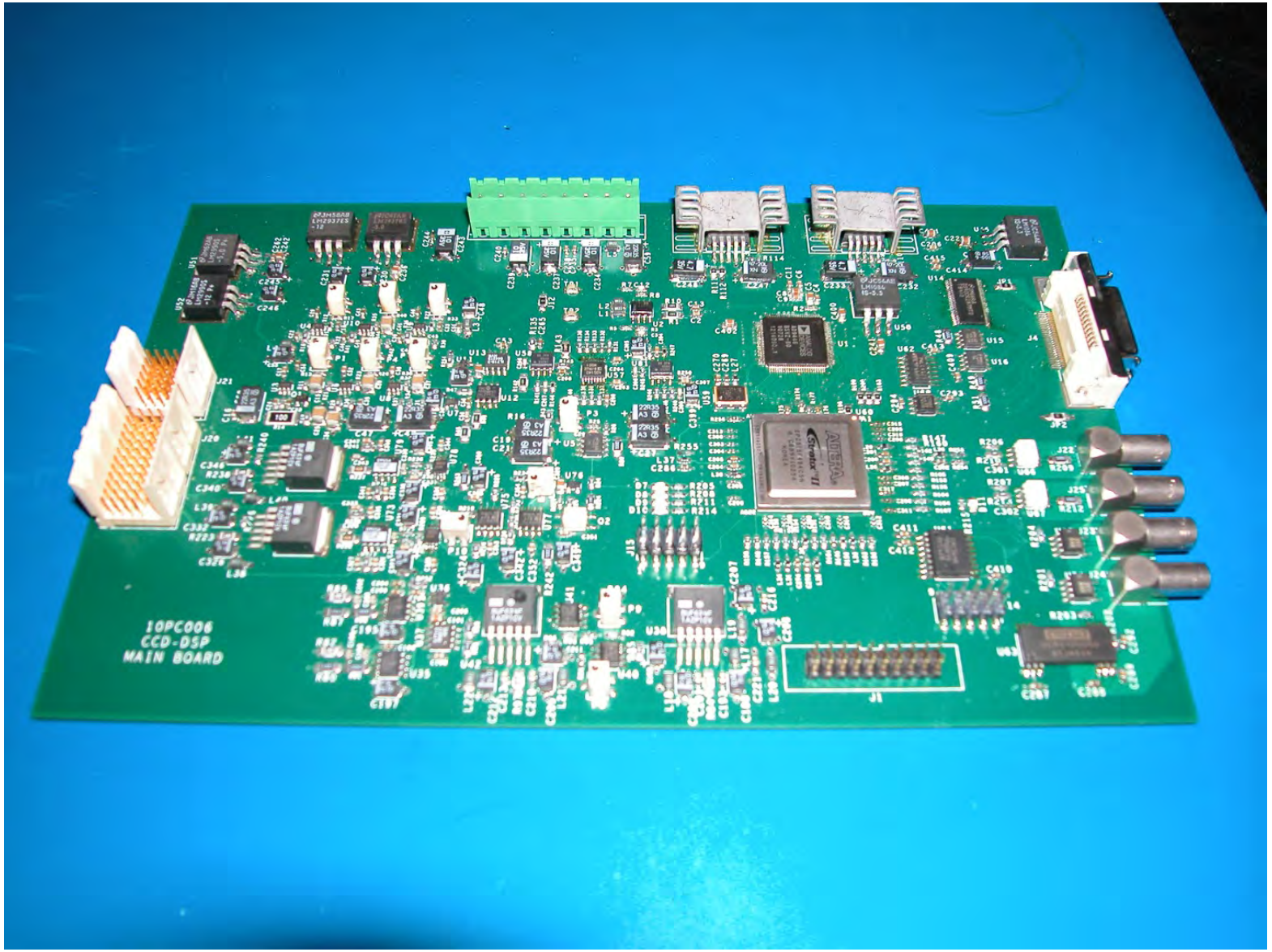


Electronics Design

Back to Minutes

- APS designed electronics (Madden)
- All digital double correlated sampling
  - Needed for 25MHz pixel rate.
- 60MHz 16-bit ADC, digital filtering in FPGA.
  - For controlling noise.(Can run at up to 80MHz)
- Most of electronics removed from CCD by cabling
  - Allows for small CCD head on the arm.
  - Using flex cabling, about 18in. Could be several feet.
- Camera Link interface, EPICS Area Detector
- CCD chilled to -5C with fan-cooled TEC.

Main Electronics Board



Back to Minutes

Mechanical Design

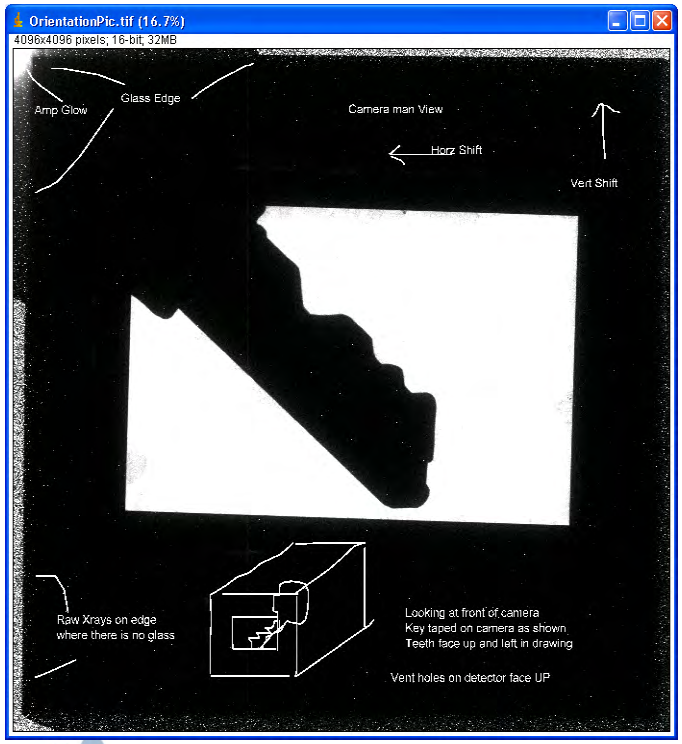
Back to Minutes

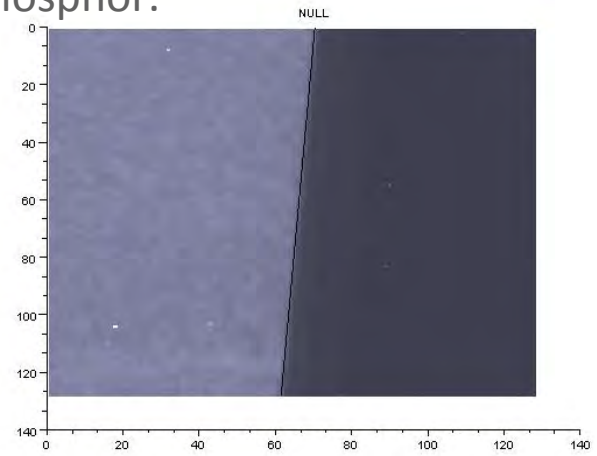
- Most parts printed plastic. Only one expensive machined part. (a few cheap machined parts).
- Cooling with TEC and off-the-shelf fan/heat sink designed for cooling gaming PCs.
- 1:1 Fiber Optic plate to block 30keV x-rays from hitting CCD.
- Scint-X Structured scintillator grease-bonded to taper.
- CCD box back filled with dry N_2. Box sealed with RTV. N_2 is flowed through the box because of leaks.
- Tiny Honeywell pressure sensor detects loss of N_2 gas so cooling can be shut off.
- Plastic parts painted with electrically conductive paint to provide shield.
  - Charge can accumulate on plastic and kill a CCD.
- Main electronics board separated from CCD box by cabling. Allows small CCD box.
- X-ray window printed as part of front cover.

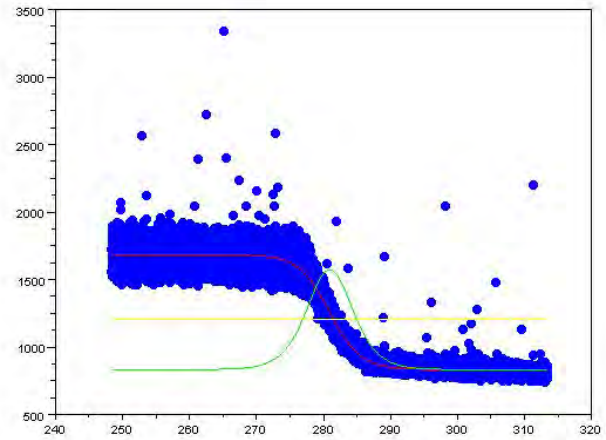
On Spatial resolution

- Scint-X structured phosphor. Scint-X advertises 20um resolution.
- Fiber Optic faceplate with 10um fibers (Incom).
- Camera measured to have around 60um FWHM.
- Phosphor does not cover whole CCD. Around $6k per phosphor.

Back to Minutes







Some Specs

Back to Minutes

Feature	Spec
Area	4096x4096, 1 in by 1 in
Phosphor	Scint-X structured phosphor
CCD Pixel Size	9 µm x 9 µm
QE (CCD)	65%
Readout Speed	650 lines/s
Full Well	~100k e^- per pixel
Typical Scan time	10 min, 70 degrees
Point Spread	60 µm FWHM
Gain	$3e^-$/ADU (middle gain setting), $2e^-$/ADU (high gain setting)
Read Noise	$18e^-$
Dark Current	$15e^-$/pix/sec (25C)

* We plan in fixing read noise and readout speed problems, caused by cabling and non-optimized FPGA algorithms

Why not just use a Pilatus?

Back to Minutes

- Larger pixel size requires very long distance to sample
- Want to use ≥30 keV x-rays
- Loss of exposure time due to need to position and readout detector at multiple fixed positions

Future Work

Back to Minutes

- Optimize electronics to improve noise.
  - No full redesign, but improve FPGA algorithms, increase sample rate, fix cabling problems.
- Try removing FO face plate to improve resolution.
  - CCD harder to work with without faceplate.
- Make metal parts for robustness, improved gas seals.
- Larger phosphor.
- Using binning for faster readout.
  - Probably do not need huge 4k x 4k images.
- Take more data with camera gain set properly.
  - Software bug prevented proper gain setting in original data.
- Use multiple detectors to cut down on scan range and time

Conclusions

Back to Minutes

- Scanning CCD gives resolution that is indistinguishable from 11-BM analyzers for most samples.
  - ScanCCD scan: minutes; 11-BM: hours
  - ScanCCD prototype: $40 K; 11-BM analyzer/detectors $1-2M
  - ScanCCD requires a large die CCD chip, but removes the most expensive component: a high-quality demagnifying fiber-optic coupler
- ScanCCD efficiency similar to a strip detector, since the solid angle coverage is about the same
- But CCD has several advantages:
  - No low angle asymmetry with CCD
  - No missing regions in patterns due to strip detector gaps
  - Possible to increase data quality/decrease scan time by using multiple scanCCD detectors (not possible with strip detectors).

Acknowledgements

Back to Minutes

Use of the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357.

We thank Lynn Ribaud for great assistance with data collection at APS beam line 11-BM. The jadrarite sample was kindly provided by Dr. Pamela Whitfield of the National Research Council Canada.

B.H. Toby, T.J. Madden, M.R. Suchomel, J.D. Baldwin, and R.B. Von Dreele (2013), "A Scanning CCD Detector for Powder Diffraction Measurements". *Journal of Applied Crystallography*. 46(4): p. 1058-63.

Prior art: The idea of a scanning CCD was patented (but not published in the literature) by Rigaku.

T. Taguchi, *X-ray analysis apparatus*, U.S. Patent. 2007, Rigaku Co: USA.

Argonne NATIONAL LABORATORY

Why we are writing GSAS-II in Python

Brian H. Toby

Robert B. Von Dreele

Advanced Photon Source, Argonne National Lab

Back to Minutes

Some Computer languages I have learned

- Fortran
- APL
- PL/1
- Assembly language: IBM 360/370, PDP-8, PDP-11
- Basic
- Forth
- C
- C++
- Tcl/Tk

Python offers the best of all of the above and with the fewest disadvantages

Three approaches to computer languages

- Compiled languages (Fortran, C, C++,…)
  - Compiler produces machine code specific to hardware
  - Produces the fastest code
  - Takes the longest time to write
  - Language capabilities minimal; packages needed for many simple tasks
  - Care needed to keep platform-independent
- Virtual-machine languages (Java)
  - Code is compiled into a generalized (virtual) computer machine language
  - A program converts the generalized instructions into machines specific ones
  - With care, can be machine independent (hard with larger applications)
  - Slower code than compiled, no faster to write
- Interpreted (Scripting) languages (Basic, Tcl/Tk, Perl, Python…)
  - A program runs that interprets and executes the code
  - Sophisticated language capabilities speed code development
  - Code can be tested interactively
  - Tends to give the least efficient code

Back to Minutes

Why Python?

1. Good style: easy to follow
   - Python is one of the easiest languages to learn
   - Very powerful and easy to write code
     - Often hundreds of lines of Fortran or C code replaced by a few lines of Python
2. Use of object-oriented programming is optional
3. Many very well-supported libraries for visualization, graphical interfaces, scientific computations (much more)
4. When needed, computations can be done as fast as in any other language (often not needed)
5. Excellent library of packages for many specialized needs
6. Probably the best overall set of libraries of any language for support of science (math, statistics, graphics,…)
7. Python is free; several free “bundles” that include the most important packages

Back to Minutes

Recent articles on Python

- V. Ayer, S. Miguez, and B.H. Toby (2014), "Why Scientists Should Learn to Program in Python". *Powder Diffraction*. **29**(S2): p. S48-S64.

More credible sources(?):

- J.M. Perkel (2015), "Programming: Pick up Python". *Nature*. **518**(7537): p. 125-6.
- E. Shein (2015), "Python for Beginners". *Communications of the ACM*. **58**(3): p. 19-21.
- H. Shen (2014), "Interactive Notebooks: Sharing the Code". *Nature*. **515**: p. 151-2.

An example of a slow Python computation

$$F_{hkl} = \sum \cos\left[2\pi(hx_j + ky_j + lz_j)\right]$$

(example of "Fortran-style" coding using Python)

```
ref = [[1,0,0], [0,1,0], [0,0,1], [1,1,0], [0,1,1], [0,1,0], [1,1,1]]
atom = [[0,.5,.5], [.5,0,.5],[.5,.5,0], [.5,.5,.5]]
    import math
    F = []
    for j1 in range(len(ref)):
        s = 0
        for j2 in range(len(atom)):
            c = 0
            for i in range(3):
                c += ref[j1][i] * atom[j2][i]
            s += math.cos(2 * math.pi * c)
        F.append(s)
```

Back to Minutes

Previous example, made fast (x~500)

$$F_{hkl} = \sum \cos\left[2\pi(hx_j + ky_j + lz_j)\right]$$

Using NumPy, a package of numerical analysis methods for Python. Provides multi-dimensional arrays, fast math,…

```
ref = [[1,0,0], [0,1,0], [0,0,1], [1,1,0], [0,1,1], [0,1,0], [1,1,1]]
atom = [[0,.5,.5], [.5,0,.5],[.5,.5,0], [.5,.5,.5]]

 import numpy as np
 F = np.sum( np.cos(2. * np.pi * np.inner(ref,atom)), axis=1)
```

How does this work? np.inner computes this:

```
[[ 0. ,  0.5,  0.5,  0.5],
 [ 0.5,  0. ,  0.5,  0.5],
 [ 0.5,  0.5,  0. ,  0.5],
 [ 0.5,  0.5,  1. ,  1. ],
 [ 1. ,  0.5,  0.5,  1. ],
 [ 0.5,  0. ,  0.5,  0.5],
 [ 1. ,  1. ,  1. ,  1.5]]
```

Back to Minutes

So Python is cool, but what does that have to do with diffraction?

GSAS

General Structure Analysis System provides a crystallographic refinement and analysis environment

Fits one model to one combined set of data

- single-crystal and powder diffraction data
- neutron (CW/TOF) and x-ray data
- *Combinations of the above*
- Sophisticated constraints and restraints
- Designed for use by expert crystallographers

GSAS does not provide data reduction, indexing or structure solution tools

EXPGUI

A graphical user interface and a set of add-on tools for GSAS that ease installation and use by both novices and experts

- Reduces learning barrier for non-crystallographers, somewhat.

Works by editing the GSAS input file; no access to GSAS computation library

Some graphics

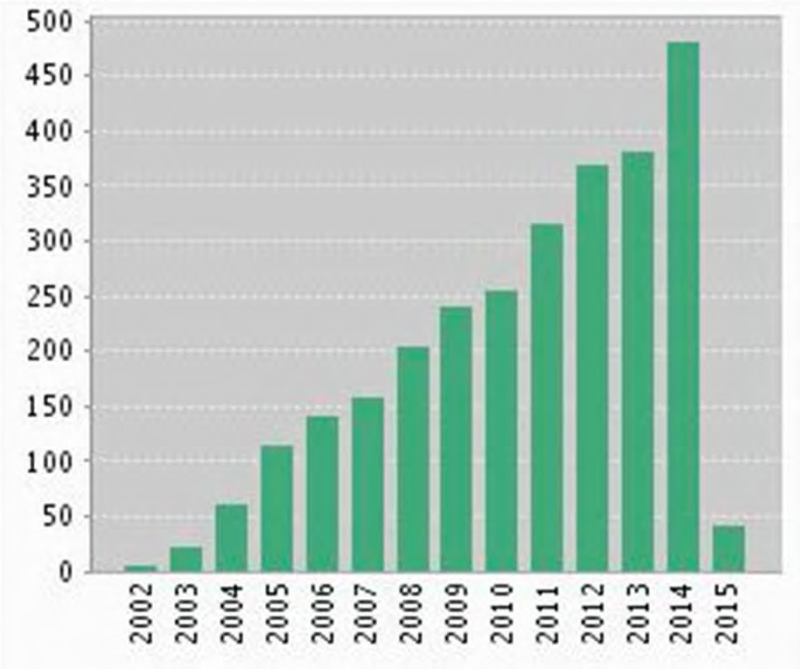
A.C. Larson and R.B. Von Dreele, "*General Structure Analysis System (GSAS)*." Report LAUR 86-748. (1986-2004)

B.H. Toby, "EXPGUI, a Graphical User Interface for GSAS". *Journal of Applied Crystallography*. **34**: 210-3 **(2001)**

Back to Minutes

GSAS and EXPGUI are widely used in many fields of science

- GSAS citations: >6000 (Google Scholar)
- EXPGUI citations: 2841 (Web of Sci), >3000 (Google Scholar)



EXPGUI Citations by year

EXPGUI use: By field

(in last 4 years)

CHEMISTRY PHYSICAL	31.00%
MATERIALS SCIENCE MULTIDISCIPLINARY	30.08%
PHYSICS CONDENSED MATTER	15.36%
CHEMISTRY INORGANIC NUCLEAR	14.86%
CHEMISTRY MULTIDISCIPLINARY	9.55%
PHYSICS APPLIED	8.49%
NANOSCIENCE NANOTECHNOLOGY	8.00%
MINERALOGY	5.95%
ELECTROCHEMISTRY	5.95%
CRYSTALLOGRAPHY	5.66%
ENERGY FUELS	5.45%
METALLURGY METALLURGICAL ENGINEERING	4.67%
GEOCHEMISTRY GEOPHYSICS	4.53%
MATERIALS SCIENCE CERAMICS	3.11%
PHYSICS ATOMIC MOLECULAR CHEMICAL	2.55%
ENGINEERING CHEMICAL	2.34%
PHYSICS MULTIDISCIPLINARY	2.27%
CHEMISTRY APPLIED	2.05%
ENVIRONMENTAL SCIENCES	1.42%

EXPGUI use: Most acknowledged grants

Argonne	DE-AC02-06CH11357	10.97%
Brookhaven	DE-AC02-98CH10886	3.82%
Los Alamos	DE-AC52-06NA25396	1.98%
Lawrence Berkeley	DE-AC02-05CH11231	1.27%
National Basic Research Program of China	2011CB808200	1.27%
Chinese Academy of Sciences	KJCX2-SW-N20	1.20%
Oak Ridge	DE-AC05-00OR22725	1.06%

19% SUF

Back to Minutes

Strengths

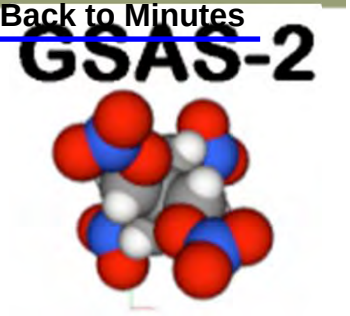
- GSAS:
  - Single package for crystallography, phase quantitation, lattice parameter fitting, texture analysis & microstructure/stress-strain determination (profile analysis)
  - Handles all x-ray and neutron data on planet
  - “Expert efficient”
- EXPGUI:
  - Provides graphical user interface for GSAS (easier to install & learn)
  - Improves on GSAS graphics
  - Import/Export to other software

Weaknesses

- GSAS:
  - “Memory-efficient” Fortran: Code reuse nearly impossible
  - Max: 9 phases & 99 datasets; expansion painful
  - Not easy to learn
  - Lacks many routinely needed tools (data reduction, indexing, structure solution)
  - Clumsy for parametric studies
- EXPGUI:
  - Not easy to keep going: BLT graphics package for Tcl/Tk language not supported

Back to Minutes

GSAS-II: why start again?



- GSAS & EXPGUI are reaching the end of their support lives
  - GSAS: major debugging effort with 64-bit compilers
  - EXPGUI: replace BLT (with what?)

New Start:

- Benefit from decades of experience
- Cleaner code will allow better adaptability
- Open source for community involvement
- New package will allow integrated GUI and visualization
- Treat parametric problems
- Better profile treatment
- Agile programming: write something and then refactor (make it better)

Back to Minutes

What has been done so far:

Like GSAS:

- all scales of problems from perovskites to proteins
- Full range of restraint & constraint functions
- X-ray and neutron; powder & single-crystal data

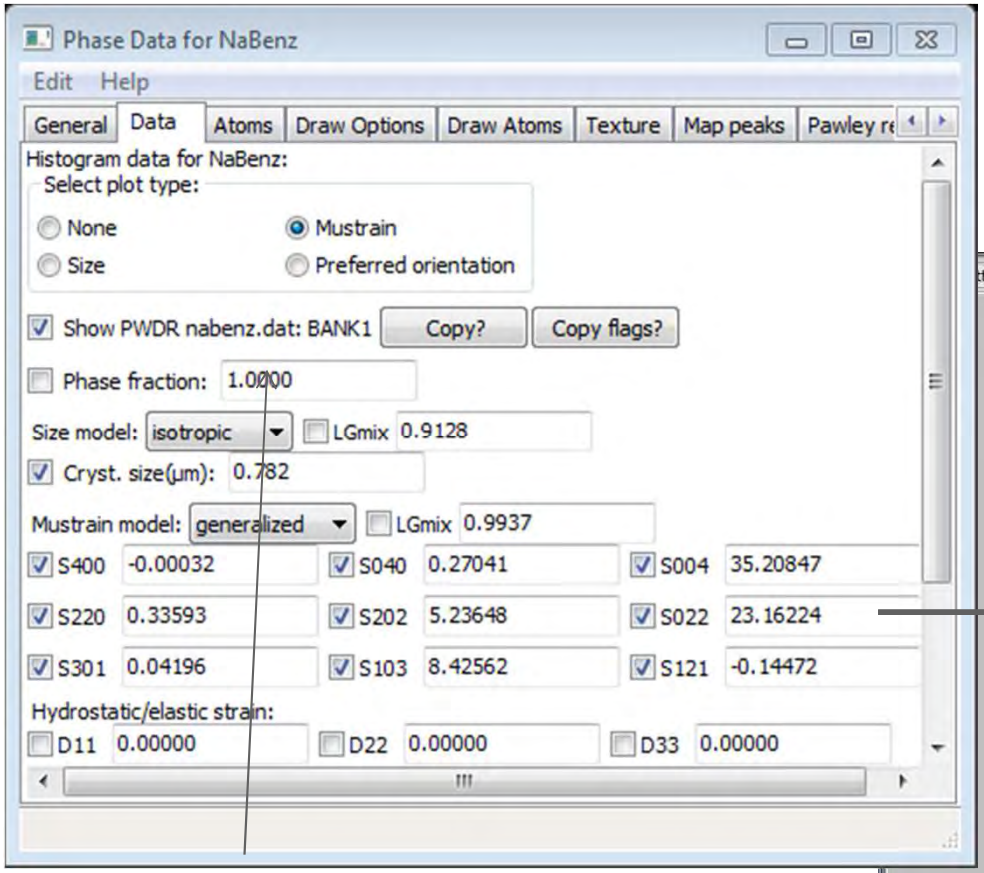
Unlike GSAS, GSAS-II offers:

- Modern platform-compliant GUI
- Excellent graphics
- Area detector integration (w/ proper treatment for offsets)
- Peak fitting, indexing & structure solution (charge-flipping &
- Small-angle scattering analysis
- Allows any complete or incomplete set of linear constraints
- Separation of instrumental and sample profile terms
- Facile computation with any number of phases & datasets (limited by computer memory)

B.H. Toby and R.B. Von Dreele, "GSAS-II: The Genesis of a Modern Open-Source All-Purpose Crystallography Software Package". *Journal of Applied Crystallography*. **46**: p. 544-9 **(2013)**.

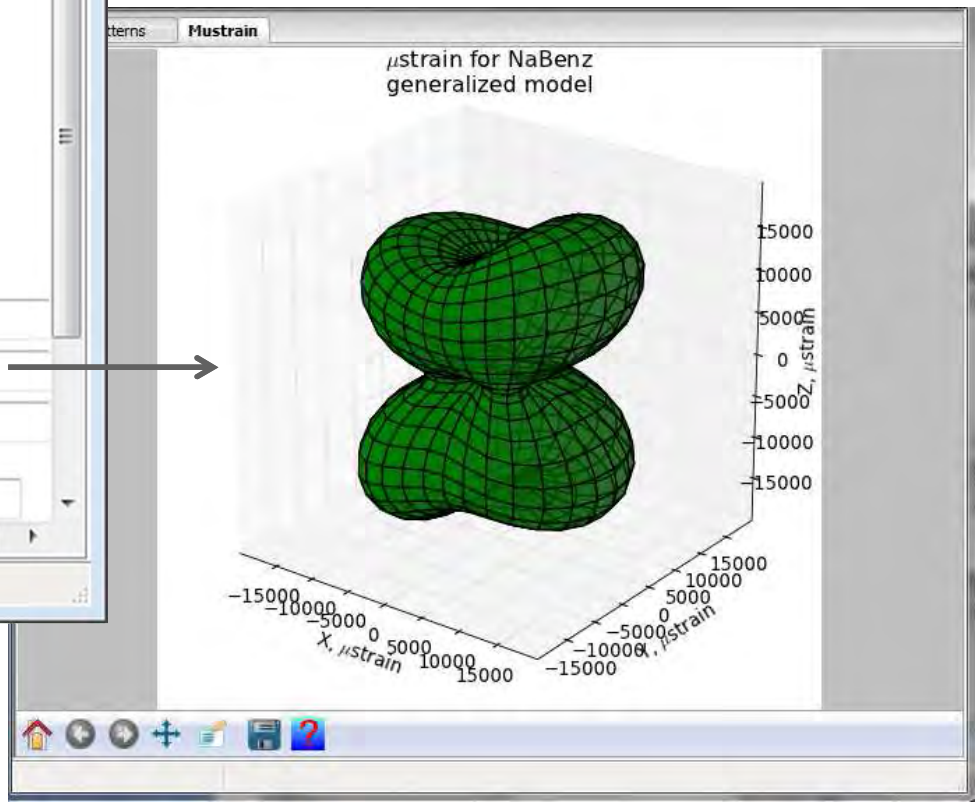
Back to Minutes

GSAS-II: Visualization brings meaning to numbers. Valuable for experts. Vital for novices!



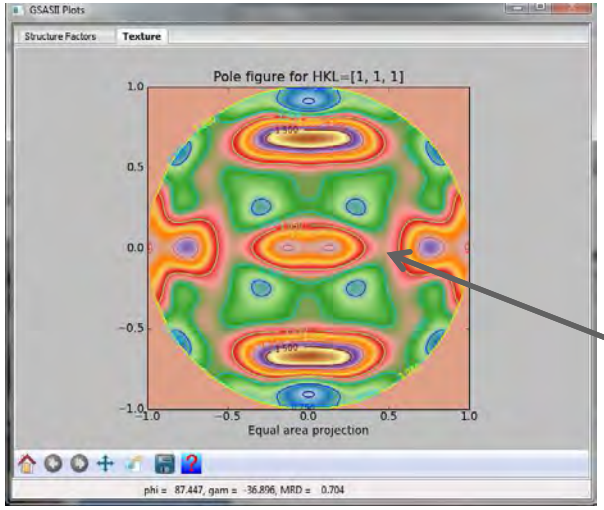
NB: size in μm size

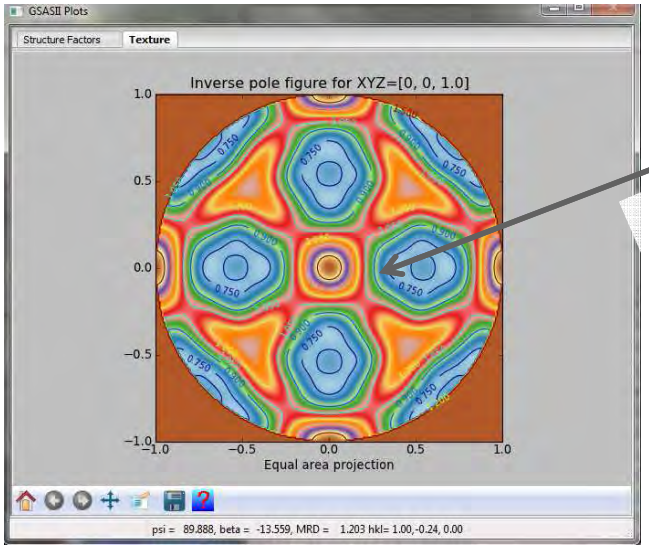
μ-Strain surface



Back to Minutes

GSAS-II: texture from spherical harmonics





Pole figure

Inverse pole figure

Phase Data for stainless steel

Texture Help

General | Data | Atoms | Draw Options | Draw Atoms | Texture

Spherical harmonics texture data for stainless steel: Texture Index J = 1.142

Texture model: rolling - mmm Harmonic order: 14 Refine texture? Show coeff.?

Texture plot type: Pole figure Projection type: equal area

Pole figure HKL: 1 1 1 Color scheme: Paired

Spherical harmonic coefficients:

C(10,0,1)	-0.142	C(10,10,1)	0.116	C(10,2,1)	0.026	C(10,4,1)	-0.081
C(10,6,1)	-0.095	C(10,8,1)	-0.018	C(12,0,1)	0.044	C(12,0,2)	0.257
C(12,10,1)	-0.035	C(12,10,2)	0.361	C(12,12,1)	-0.076	C(12,12,2)	0.012
C(12,2,1)	-0.181	C(12,2,2)	0.292	C(12,4,1)	-0.015	C(12,4,2)	0.427
C(12,6,1)	0.128	C(12,6,2)	0.162	C(12,8,1)	0.002	C(12,8,2)	0.013
C(14,0,1)	0.070	C(14,10,1)	0.136	C(14,12,1)	-0.148	C(14,14,1)	-0.136
C(14,2,1)	-0.050	C(14,4,1)	0.005	C(14,6,1)	-0.085	C(14,8,1)	-0.005
C(4,0,1)	-0.183	C(4,2,1)	0.369	C(4,4,1)	-0.408	C(6,0,1)	-0.662
C(6,2,1)	-0.488	C(6,4,1)	-0.392	C(6,6,1)	-0.381	C(8,0,1)	0.117
C(8,2,1)	-0.052	C(8,4,1)	0.111	C(8,6,1)	0.003	C(8,8,1)	0.116

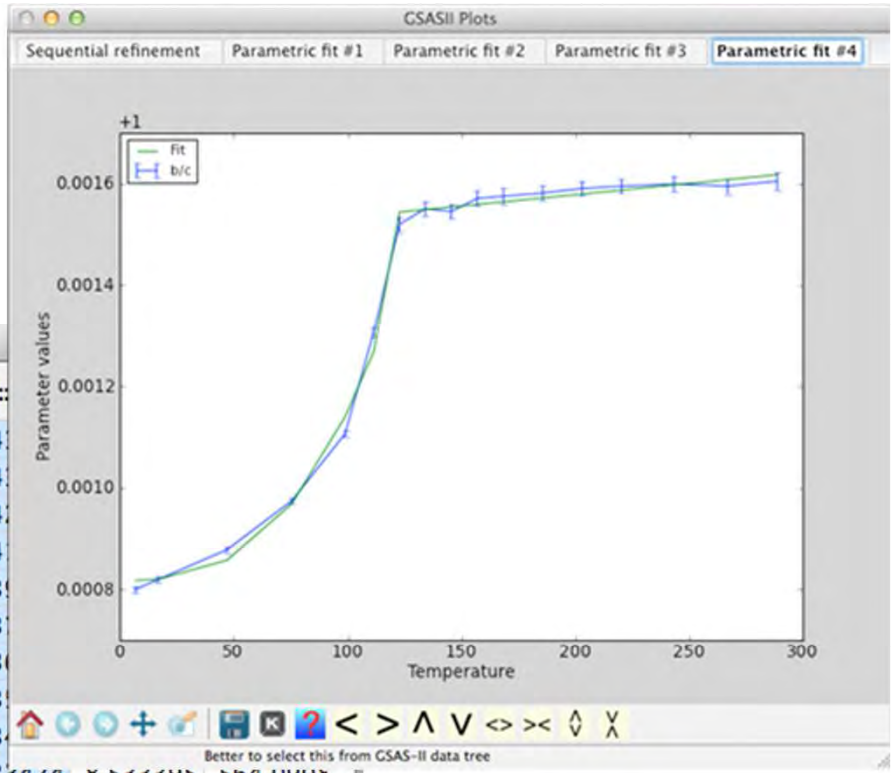
Sample orientation angles:

Sample omega: 0.00 Sample chi: -4.57 Sample phi: 87.90

Back to Minutes

GSAS-II works well for parametric studies

- For example, tens to hundreds of measurements as a function of temperature



Sequential refinement results

	Use	Rwp	$\Delta\chi^2$ (%)	Temperature	0::a	0::		
PWDR OH_00.fxye Bank 1	☑	14.122	0.000101	6.778000	7.712722	8.54		
PWDR OH_04.fxye Bank 1	☑	13.900	0.001334	16.702000	7.712913	8.54		
PWDR OH_09.fxye Bank 1	☑	14.093	0.021350	46.970000	7.713749	8.54		
PWDR OH_14.fxye Bank 1	☑	15.022	0.002720	75.420000	7.716505	8.54		
PWDR OH_19.fxye Bank 1	☑	16.612	0.003825	98.890000	7.720728	8.53		
PWDR OH_22.fxye Bank 1	☑	19.459	0.005316	111.430000	7.725257	8.53		
PWDR OH_24.fxye Bank 1	☑	21.346	0.023153	122.730000	7.730246	8.53		
PWDR OH_26.fxye Bank 1	☑	22.999	0.008724	134.020000	7.735260	8.53		
PWDR OH_28.fxye Bank 1	☑	23.465	0.000859	145.470000	7.741345	8.53		
PWDR OH_30.fxye Bank 1	☑	24.553	0.001952	156.900000	7.746425	8.533474	8.533395	564.0908
PWDR OH_32.fxye Bank 1	☑	25.140	0.000385	168.430000	7.749899	8.532983	8.532774	564.2702
PWDR OH_35.fxye Bank 1	☑	25.851	0.009163	185.670000	7.754554	8.532650	8.532476	564.5674
PWDR OH_38.fxye Bank 1	☑	26.338	0.004873	203.000000	7.758912	8.532247	8.532132	564.8352

Select column to export; Double click on column to plot data; on row for Covariance

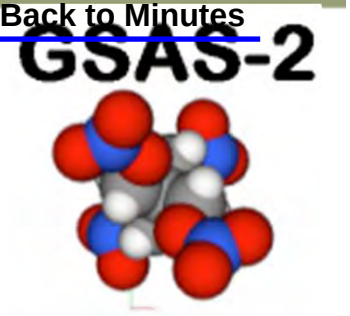
Back to Minutes

GSAS-II: Work in progress

- Learning GSAS-II is best done through worked-through ~21 examples (tutorials)
  - Tutorials take up 3x the space of the program; now downloaded one at a time, by request
- Parametric fits with inhomogeneous parameterization
- Modulated structures (3+1 dimensional space groups)
- Refactor image file reading

Back to Minutes

GSAS-II Future work



- Parametric restraints (not constraints!)
- Magnetism (as a special case of 3+1 modulation)
- Validate PDF computation

Wish list:

- Better packaging & installation kits
- Instructional videos
- Regional instruction workshops
- Local-ordering fitting
- Applications interface (for scripting GSAS-II)
- Textbook

Downloading and Installing Python

What version of Python?

- Two tracks: Python 3.x and *2.x*
- You want to get a "dist" with lots of packages (numpy, scipy, matplotlib,…)
  - Windows: no Python supplied
  - Linux: most dists have Python and packages available as optional OS installs
  - Mac: comes with Python, but not too many packages

- There are a number of good, free Python dists:
  - Python(X,Y): Windows & Linux
  - Canopy (Enthought, Inc): free and "pro" versions for all platforms
  - Anaconda (Continuum, Inc.): : free and "pro" versions for all platforms
  - EPD (older

CONTINUUM
ANALYTICS

Anaconda

Completely free enterprise-ready Python distribution for large-scale data processing, predictive analytics, and scientific computing

- 195+ of the most popular Python packages for science, math, engineering, data analysis
- Completely **free** - including for commercial use and even redistribution
- Cross platform on Linux, Windows, Mac
- Installs into a single directory and doesn't affect other Python installati require root or local administrator privileges
- Stay up-to-date by easily updating packages from our free, online repos
- Easily switch between Python 2.6, 2.7, 3.3, 3.4, and experiment with mu our conda package manager and its great support for virtual environme
- Comes with tools to connect and integrate with Excel

Anaconda

Why Are We Just Giving This Away?

- We want to ensure that Python, NumPy, SciPy, Pandas, IPython, Matplc other great Python data analysis tools can be used everywhere.
- We want to make it easier for Python evangelists and teachers to prom
- We want to give back to the Python community that we love being a pa

Anaconda Add-Ons		
Accelerate	$129.00	Free Trial
IOPro	$79.00	Free Trial
MKL Optimizations	$29.00	Free Trial
All Products are Free for Academic Use		

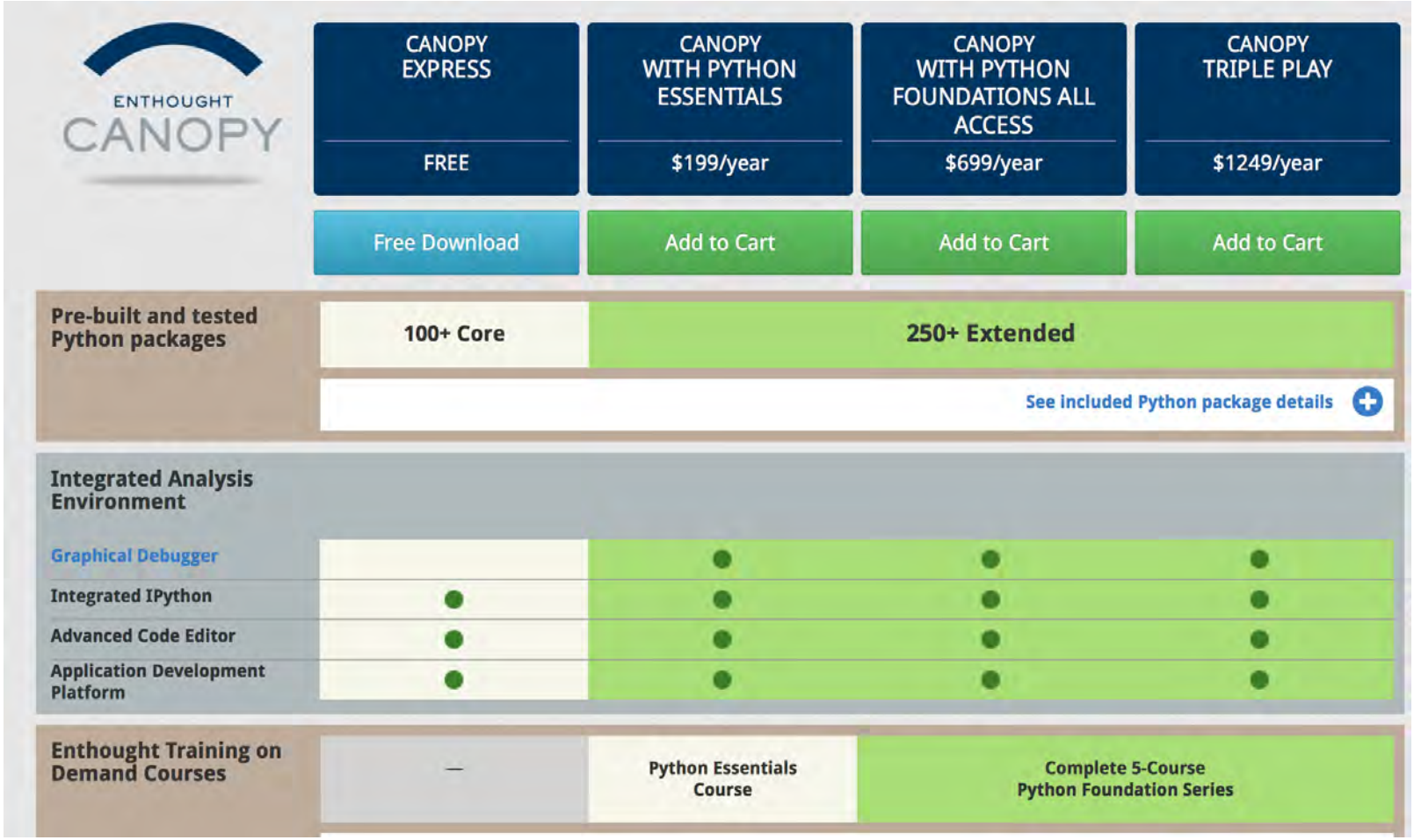
Enthought Canopy

Scientific and Analytic Python Deployment with Integrated Analysis Environment

Enthought Canopy is a comprehensive Python analysis environment that provides easy installation of the core scientific analytic and scientific Python packages, creating a robust platform you can explore, develop, and visualize on. In addition to its pre-built, tested Python distribution, Enthought Canopy has valuable tools for iterative data analysis, visualization and application development including:

- One-Click Python Package Deployment with a Graphical Package Manager
- Code Editor with IPython Notebook Support
- Interactive Graphical Python Code Debugger
- Integrated IPython Prompt
- Convenient Documentation Browser
- Python for Excel with PyXLL (*add-on*)
- Integration with the Intel MKL and Microsoft Python Tools for Visual Studio

Plus, Canopy works seamlessly with the NEW Enthought Python Training on Demand for a hands-on, interactive learning experience.


ENTHOUGHT
CANOPY
CANOPY EXPRESS
FREE
CANOPY WITH PYTHON ESSENTIALS
$199/year
CANOPY WITH PYTHON FOUNDATIONS ALL ACCESS
$699/year
CANOPY TRIPLE PLAY
$1249/year
Free Download
Add to Cart
Add to Cart
Add to Cart
Pre-built and tested Python packages
100+ Core
250+ Extended
See included Python package details
Integrated Analysis Environment
Graphical Debugger
Integrated IPython
Advanced Code Editor
Application Development Platform
Enthought Training on Demand Courses
—
Python Essentials Course
Complete 5-Course Python Foundation Series

Back to Minutes

New Ideas for Databases

Chris Gilmore
Allander Science
and
University of Glasgow

Back to Minutes

The Problem

I have a set of powder patterns from:

- A database.
- Lab measurements on related samples.
- High throughput experiments

I want to simplify what I have to explore the data more easily. (I may have 100s of patterns and they may be poor quality.)

Back to Minutes

The Solution

1. Use all the measured data, not just peaks.
2. Correlate every pattern with every other pattern using the Spearman correlation coefficient, *r*, and the Pearson correlation coefficient *R*.
3. Combine them using Fischer transforms to get a correlation matrix, ρ

$$r_{ij} = \frac{\sum_k \left(R_k - \bar{R}\right)\left(S_k - \bar{S}\right)}{\sqrt{\sum_k \left(R_k - \bar{R}\right)^2 \sum_k \left(S_k - \bar{S}\right)^2}}$$

$$R_{ij} = \frac{\sum_k \left(x_k^2 - \overline{x^2}\right)\left(y_k^2 - \overline{y^2}\right)}{\sum_k \left(x_k^2 - \overline{x^2}\right)^2 \left(y_k^2 - \overline{y^2}\right)^2}$$

Back to Minutes

Then:

Distance matrix

$$\mathbf{d} = 0.5(1 - \boldsymbol{\rho})$$

$$\rho_{ij} = 1 \quad \Rightarrow d_{ij} = 0$$

$$\rho_{i_j} = -1 \Rightarrow d_{ij} = 1$$

Back to Minutes

Example: Ammonium nitrate NH_4NO_3

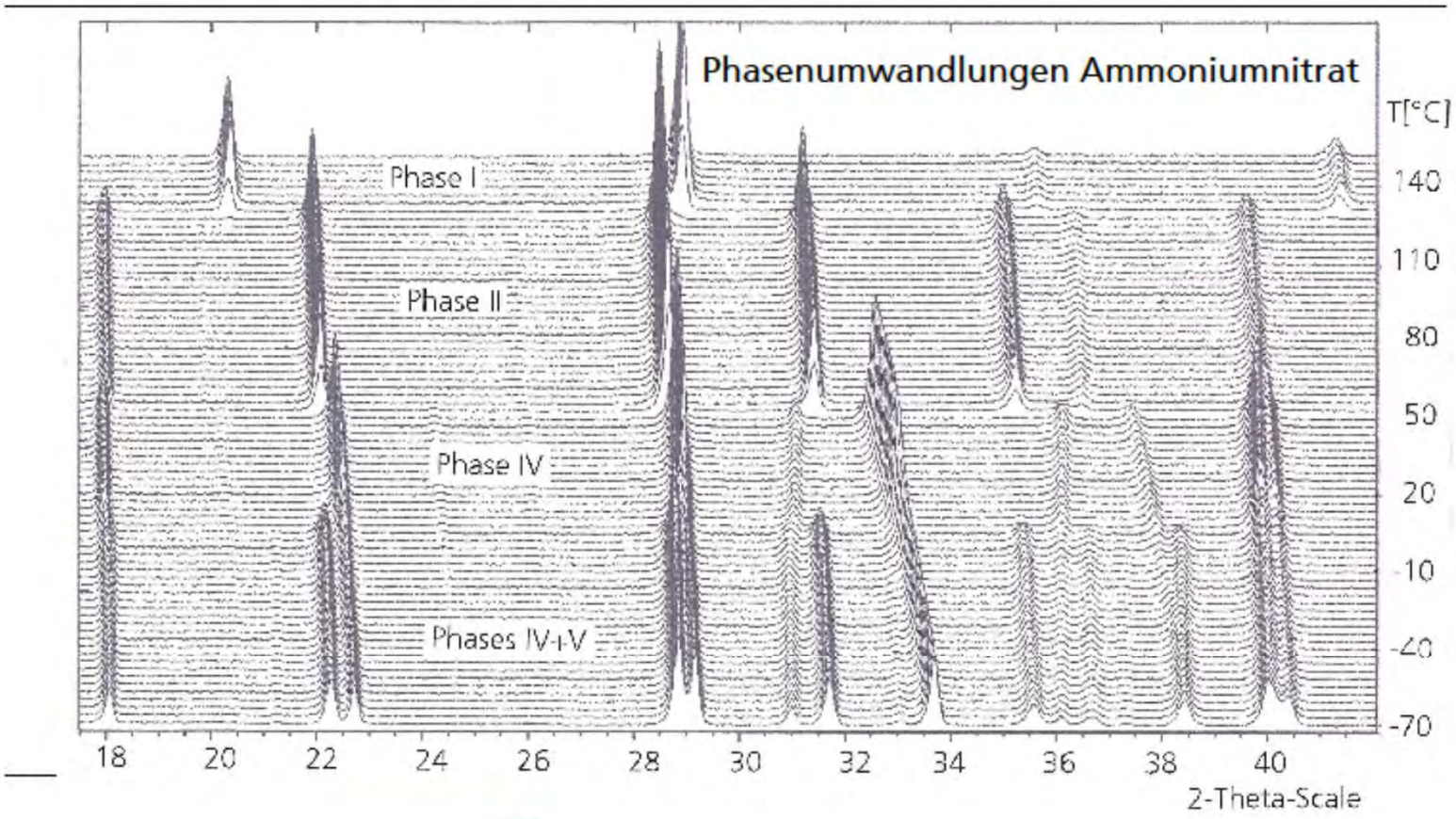
Temperature induced phase transformations:

IV	III	II	I
Pmmn	*Pbnm*	*P-42$_1$m*	*Pm-3m*
256 → 305 K	305 → 357 K	357 → 398 K	> 398 K
Z = 2	Z = 4	Z = 2	Z = 1
a = 5.745 Å	a = 7.14 Å	a = 5.719 Å	a = 4.4 Å
b = 5.438 Å	b = 7.65 Å	c = 4.932 Å	
c = 4.942 Å	c = 5.83 Å		

Back to Minutes

NH_4NO_3

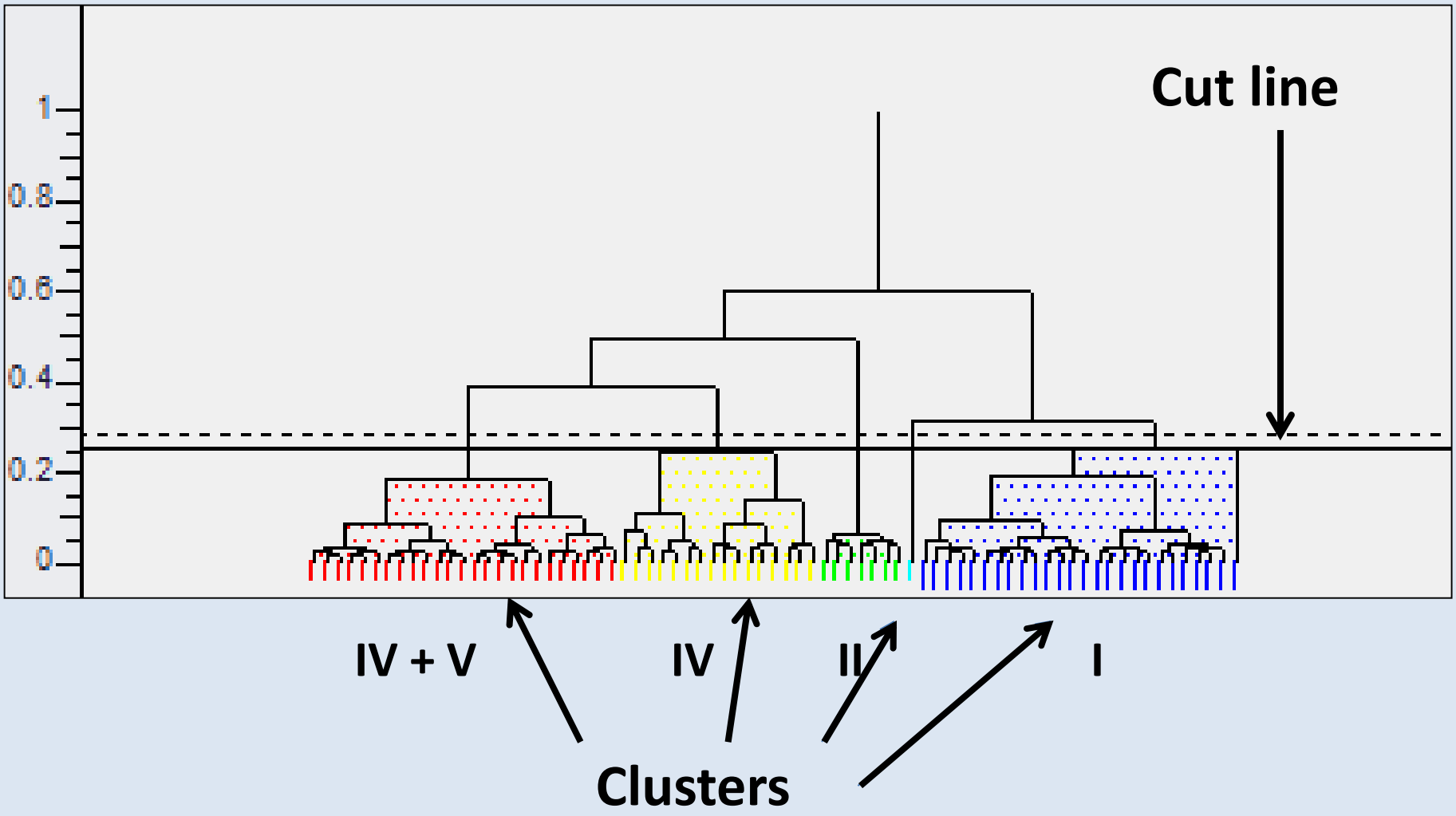
Ammoniumnitrat; Pfinztal 15. März 2007, M. Herrmann



Fraunhofer Institut Chemische Technologie

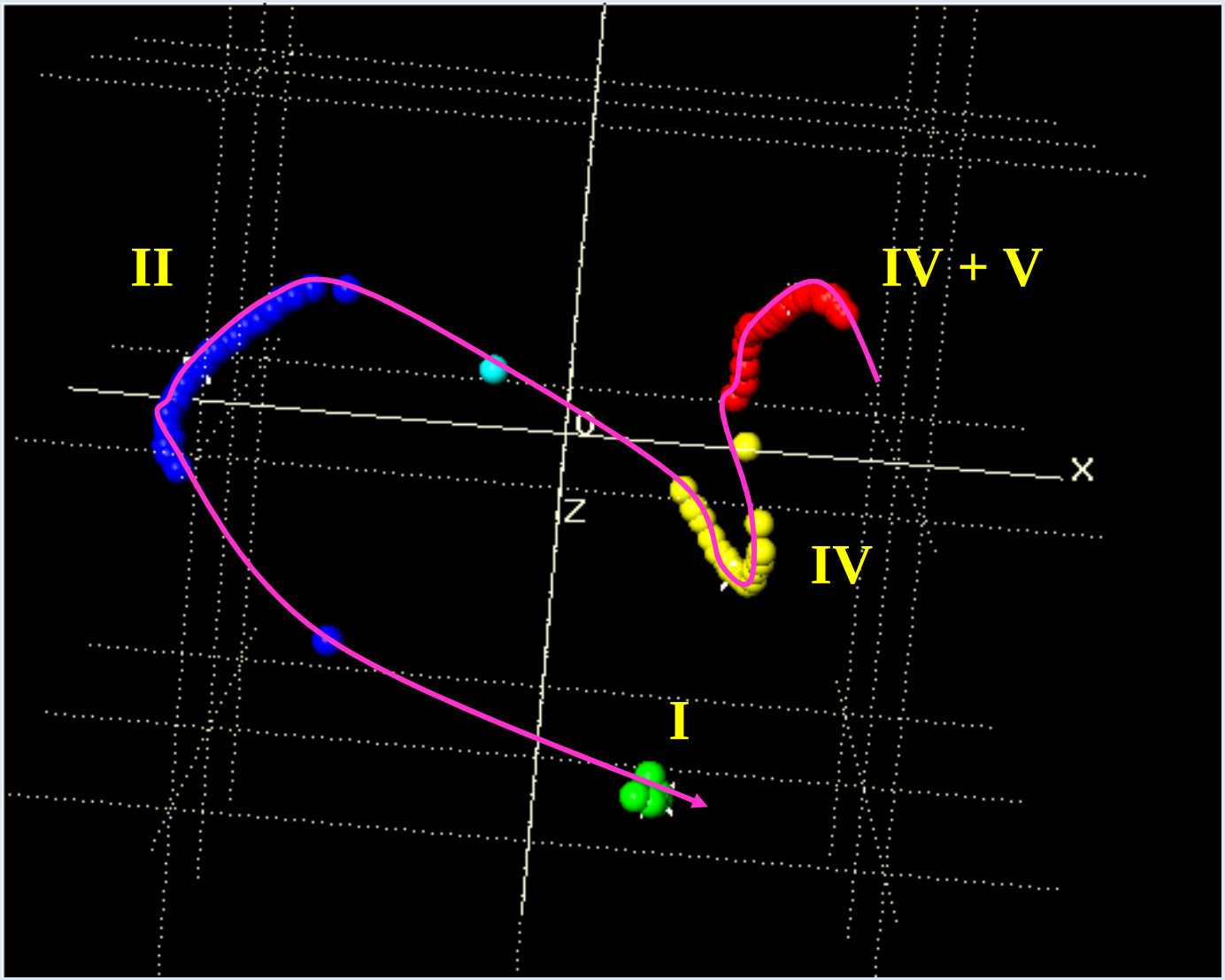
Back to Minutes

Dendrogram



Back to Minutes

Data Visualization: Metric Multidimensional Scaling (MMDS)



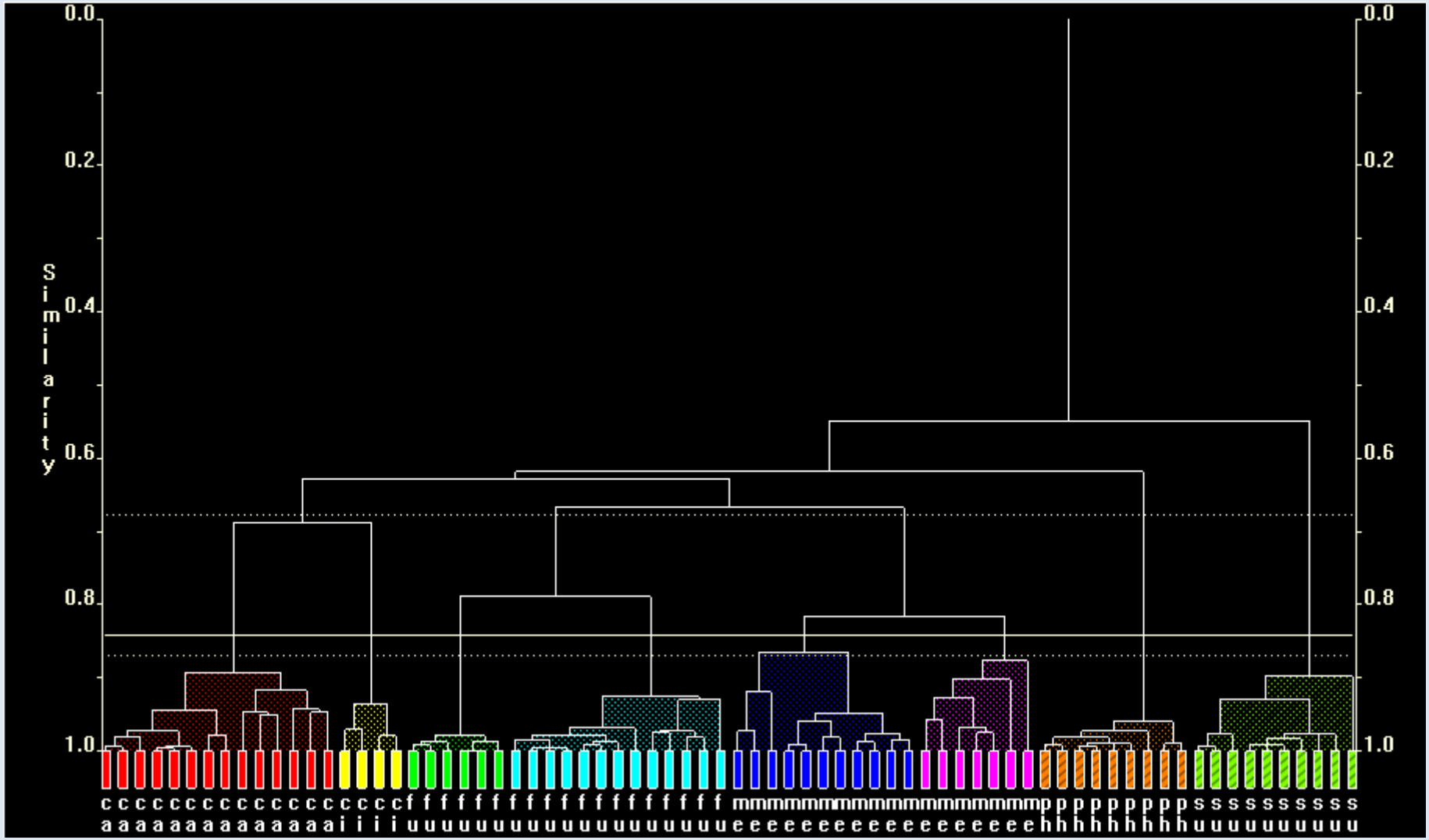
Back to Minutes

Using complementary databases

- **We can use other spectroscopic data *e.g.* Raman either instead of or with PXRD.**
- **Combine correlation matrices using INDSCAL (Independent Scaling of Differences) method (Carroll & Chang, (1970) *Psychometrica* 35, 283-319).**

Back to Minutes

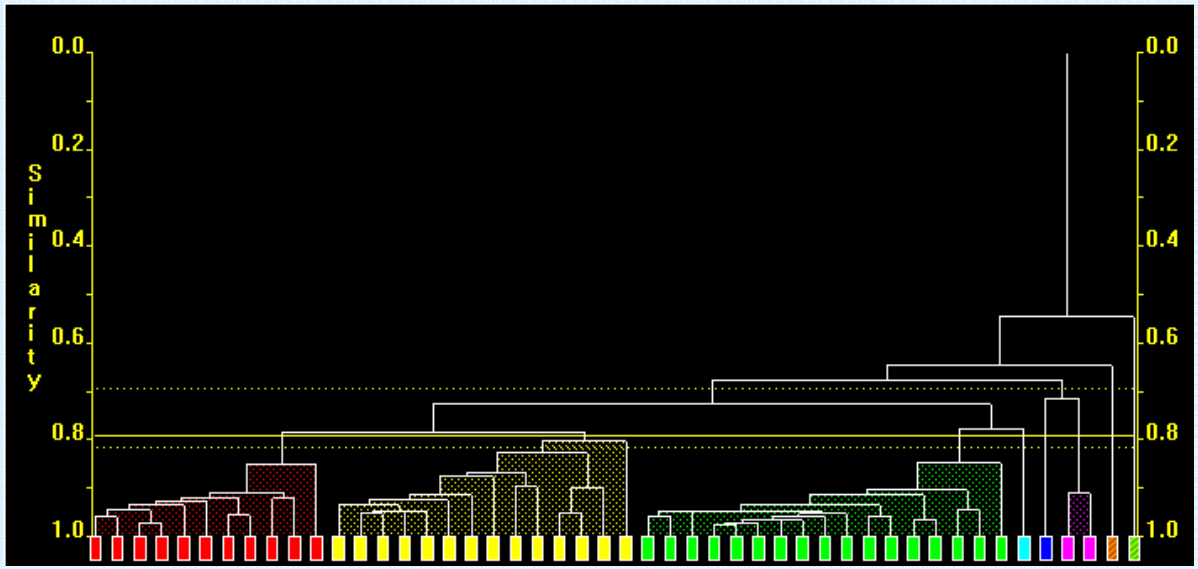
Raman Example

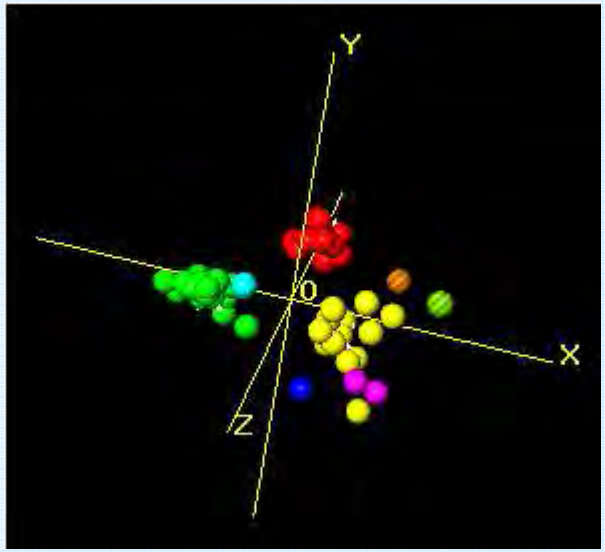


Back to Minutes

Combining PXRD and Raman Datasets I

- **48 patterns of 3 forms of Sulfathiazol (forms 2, 3 and 4).**
- **PXRD and Raman data collected.**
- **PXRD Data only: Splits form 3 into 2 separate clusters:**

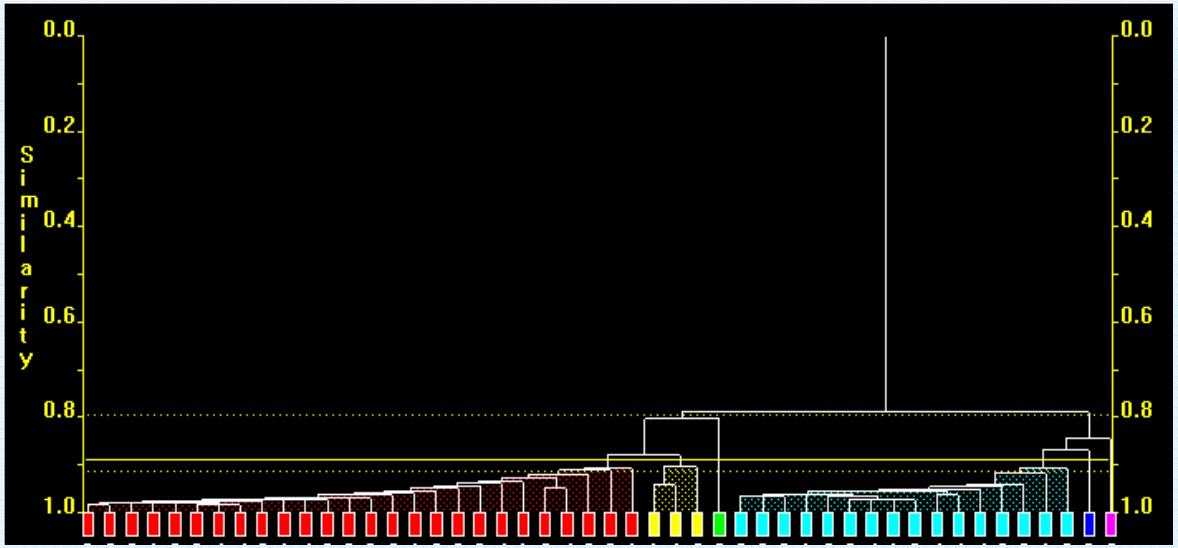




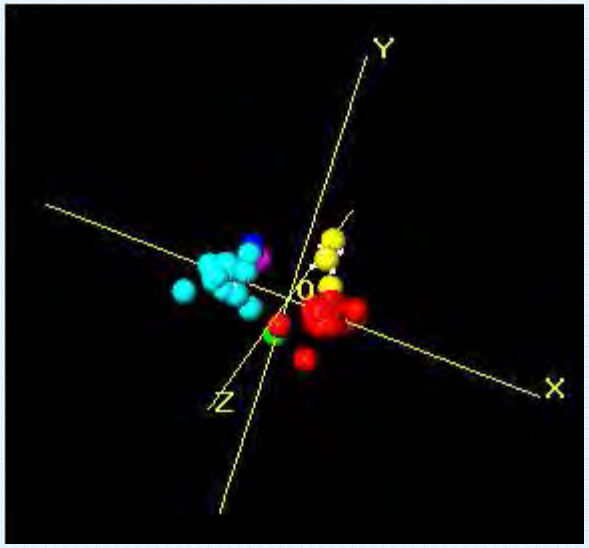
Back to Minutes

Combining PXRD and Raman Datasets II

Raman Data only: Doesn't distinguish between Form 3 and Form 4



Forms 3 & 4

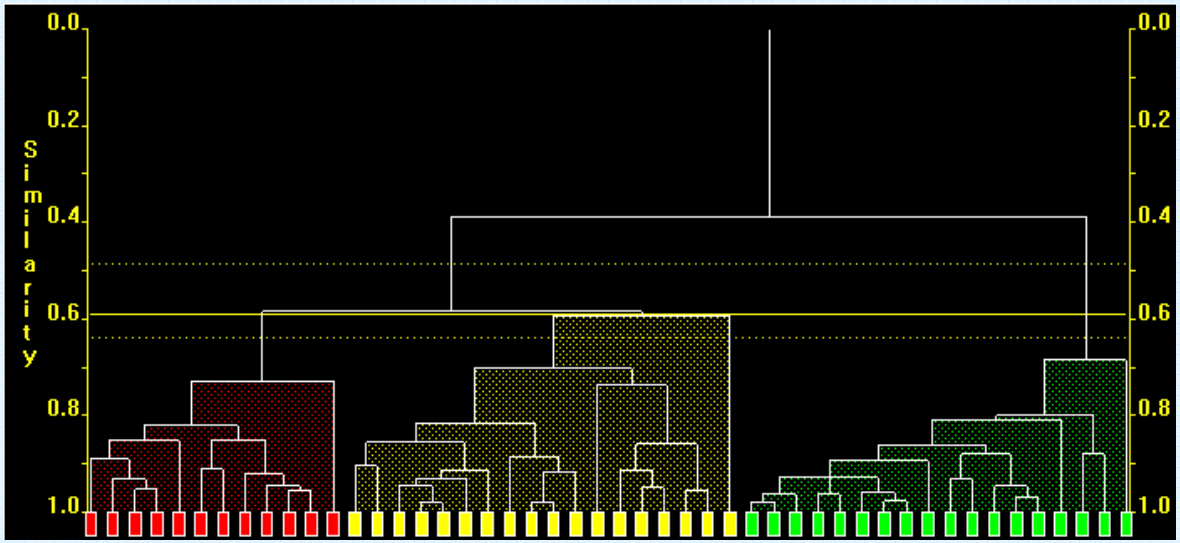


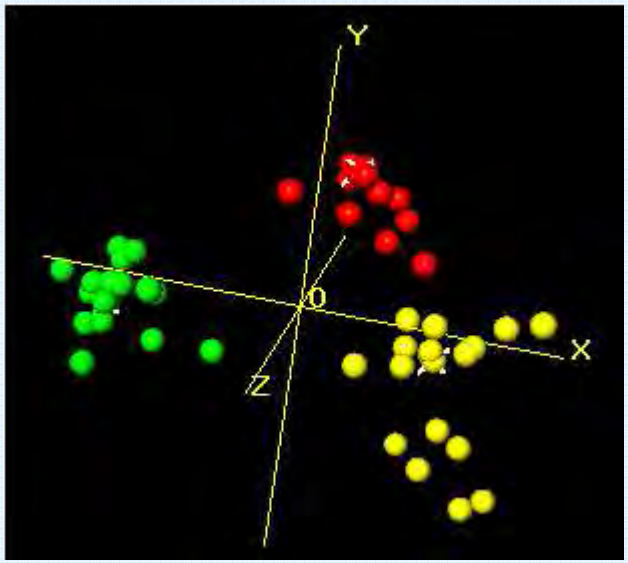
Form 2

Back to Minutes

Combining PXRD and Raman Datasets III

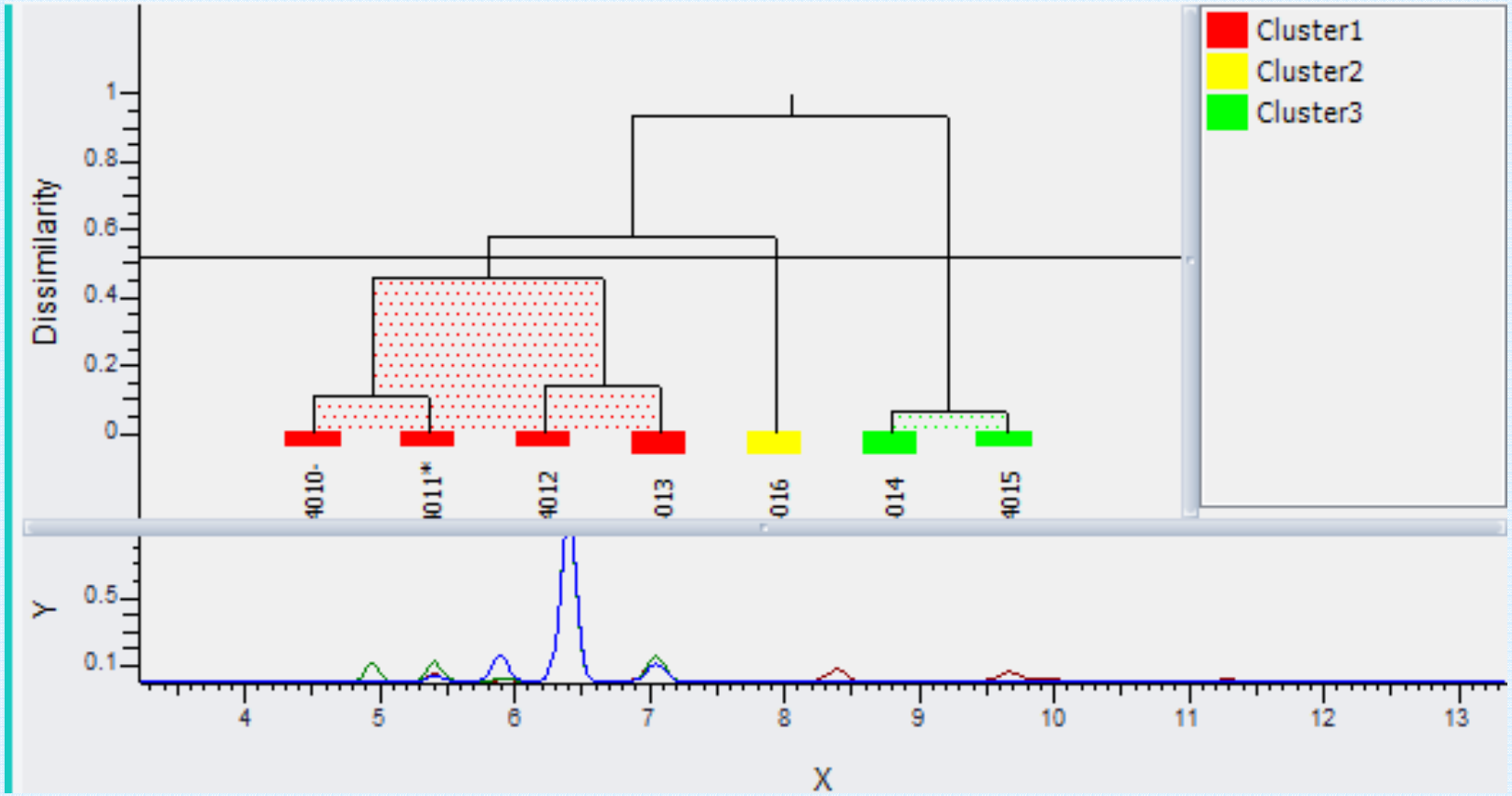
Combined PXRD + Raman using INDSCAL works.





Back to Minutes

EDXRF Works as Well



Back to Minutes

Quantitative Analysis is an Exercise in Linear Algebra

We have a set of reference data and a mixture of these references.

There are *N* phases in the mixture. We want to find the relative fractions $x_1, x_2 \ldots x_N$ of each phase.

- ***n* data points in each reference.**
- **The mixture pattern has data points $b_1, b_2 \ldots b_n$.**

In matrix form:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} \quad \text{or} \quad \mathbf{A} \cdot \mathbf{x} = \mathbf{b}$$

Back to Minutes

Search Match Procedures

We have a mixture. We search a database to find the components of this mixture using linear algebra/multiple regression.

We find the best fit between references in the database and our sample using linear algebra..

Can't use multiple regression because it assumes that all the references are uncorrelated.

Example quantitative analysis on a mixture:

Isopropanol (20%), Ethanol(80%) with 25 references

Result:

Ethanol = 14.8%

Tincture of colocynth = 43.6%

Tincture of gelsemium = 41.6%

Back to Minutes

Quantitative analysis without Rietveld refinement III

Use a combination of:

- **Correlation coefficients as an initial filter.**
- **Ridge regression.**
- **Partial least squares.**

Back to Minutes

IR Example 1

400 references. The test spectrum contains

H0001 85.5 %

H0007 14.5 %

Best solution

Reference	Percent	Error	t-value	PRESS	RSS
H0001	85.5	2.4	4.96	0.0911	0.1806
H0007	14.5	5.1	0.41	0.0911	0.1806

Back to Minutes

IR Example 2

300 references, 1 sample

The test spectrum contains the following reference data:

H0001 22.8 %

H0002 24.1 %

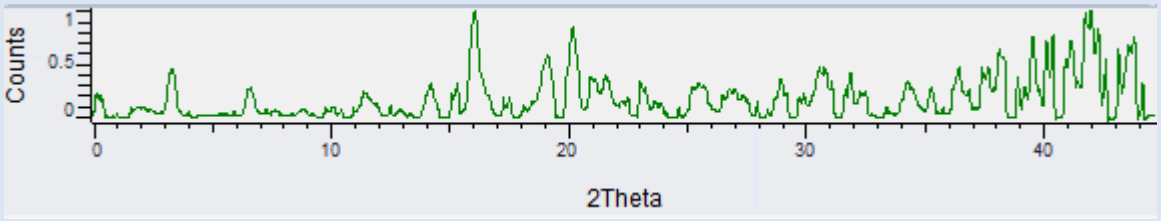
H0003 16.5 %

H0019 36.6 %

Reference	Percent	Error	t-value	PRESS	RSS
H0019	36.5	1.6	7.49	0.0071	0.0140
H0002	24.0	2.2	3.75	0.0071	0.0140
H0001	22.8	1.3	5.95	0.0071	0.0140
H0003	16.6	1.0	5.35	0.0071	0.0140

Back to Minutes

PXRD Example 1



384 references of poor quality, create a 33.3% mixture of 3 components:

Result:

1	33.3	1.04
2	33.3	0.01
3	33.3	0.62

Back to Minutes

PXRD Example 2

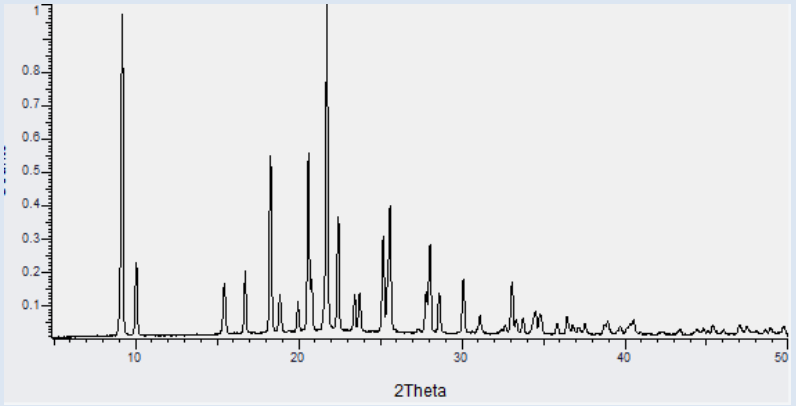
50% 2H2, 45% 6A3, 5% 1A1

Result:

Reference	Percent	Error
2H2	49.9961	0.001
6A3	45.0000	0.005
1A1	5.0004	0.005

Back to Minutes

Experimental Data



Experimental		Results	
% Anhydrous	% Hydrate	% Anhydrous	% Hydrate
99.0	1.0	100.0 ± 0.9	0.0
86.0	14.0	85.4 ± 0.4	14.6 ± 0.5
50.0	50.0	49.5 ± 0.2	50.5 ± 0.3
30.0	70.0	31.1 ± 0.4	68.9 ± 0.5
3.0	97.0	3.0 ± 0.4	97.0 ± 0.4

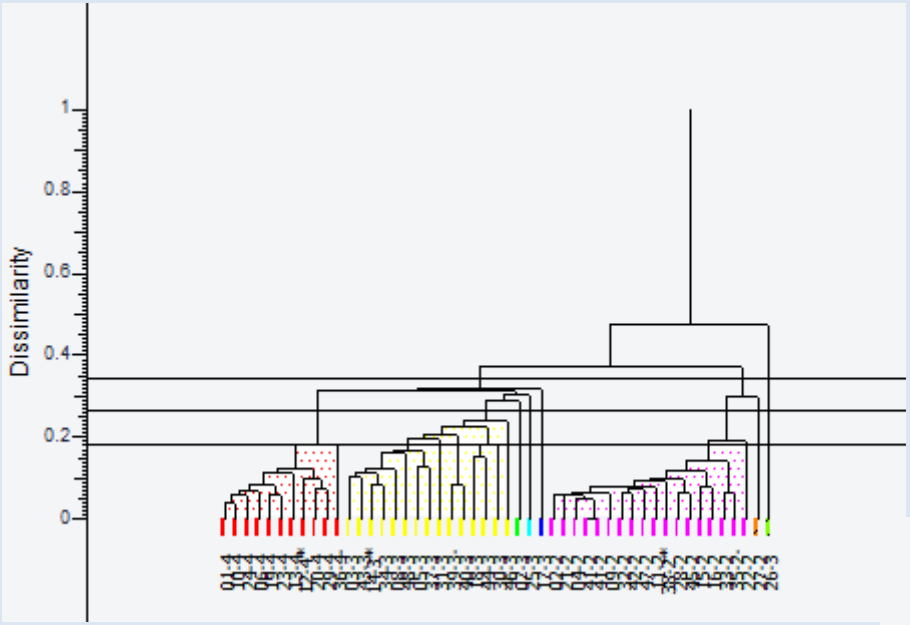
Back to Minutes

Automatically excluding measurement regions that contain no relevant data

- Can improve clustering and quantitative analysis.
- Split diffraction pattern (or spectrum) into ≈10 overlapping regions and use cluster analysis on each region. If you only get one cluster then exclude this region.

Back to Minutes

Example



Can use it in quantitative analysis.

Exclude 17-19 and 35-41 in 2θ