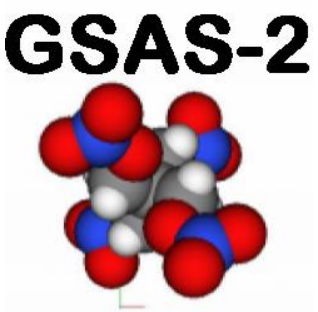
Argonne NATIONAL LABORATORY

INTRODUCTION TO GSAS-II: GSAS TO GSAS-II



ROBERT VON DREELE
APS/ANL(retired)
vondreele@anl.gov

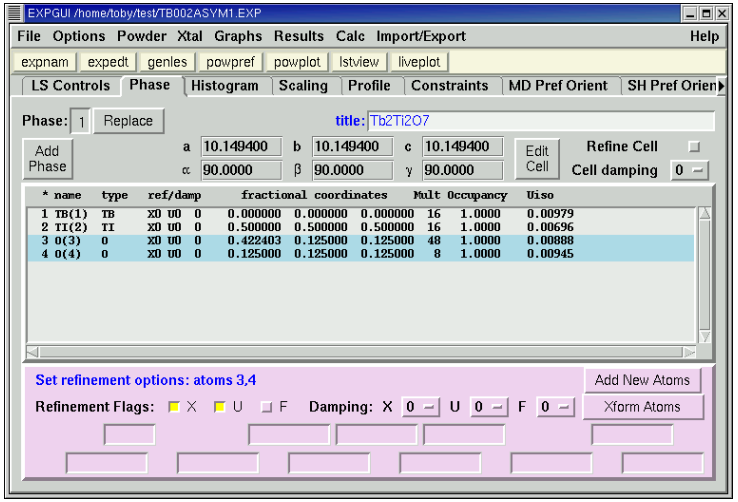
Saskatoon
June, 2024

GSAS (WITH A.C. LARSON)

GENERAL STRUCTURE ANALYSIS SYSTEM GSAS

- 1982-1986 Initial development – for multidata TOF neutron powder & single crystal data – A.C. Larson & R.B. Von Dreele
- VAX Fortran; ISAM file structure; batch process calculations
- Multidata/multiphase – complex input file (experiment file)
- Menu-driven editor – EXPEDT - All about taming the control file
  - state of the art in 1980's
- Later EXPGUI (B.H. Toby) – more modern interface (1990's)
- Fundamental crystallographic calculations – refinement, Fourier maps, structure drawings & geometry (dist, angle, etc.) – no solution tools

```
EXPEDT data setup option (<?>,D,F,K,L,P,R,S,X) >
EXPEDT data setup options:
 <?> - Type this help listing
 D   - Distance/angle calculation set up
 F   - Fourier calculation set up
 K n - Delete all but the last n history records
 L   - Least squares refinement set up
 P   - Powder data preparation
 R   - Review data in the experiment file
 S   - Single crystal data preparation
 X   - Exit from EXPEDT
```

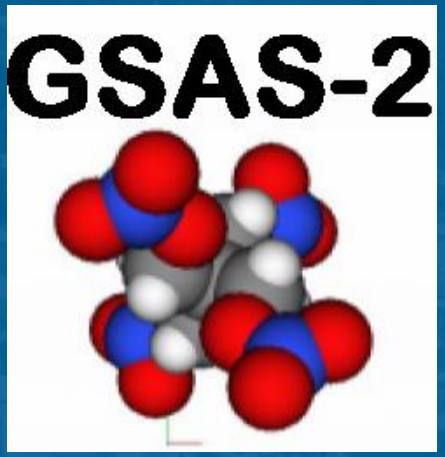


GSAS:

Extensions 1986-2004

- VAX → IRIX (ISAM emulator) → PC (→ OSX)
- CW neutron & X-ray (X-ray energy dispersive)
- New profile functions as developed (mustrain, size, strain, etc.)
- Texture – spherical harmonics
- Proteins (powders!)
- Easily scripted
- Documented – LAUR 86-748 “GSAS Manual”
- Widely accepted by community – by 2024 ~12,000 citations to LAUR 86-748
- Still cited ~400/yr!
- But: reached design limits on expansion (9 phases/99 histograms)
- Dated interface (both EXPEDT & EXPGUI) & steep learning curve
- Lacked powder indexing & structure solution tools
- Modern 2D detector data needed proper preprocessing
- Need new tool for modern crystallography→ GSAS-II

GSAS-II (WITH B.H. TOBY)



GSAS-II: A MODERN ANALYSIS PACKAGE FOR ALL ASPECTS OF CRYSTALLOGRAPHY

GSAS-II is intended to more than replace GSAS & EXPGUI with a new, modern, extensible, and *open-source* crystallographic analysis

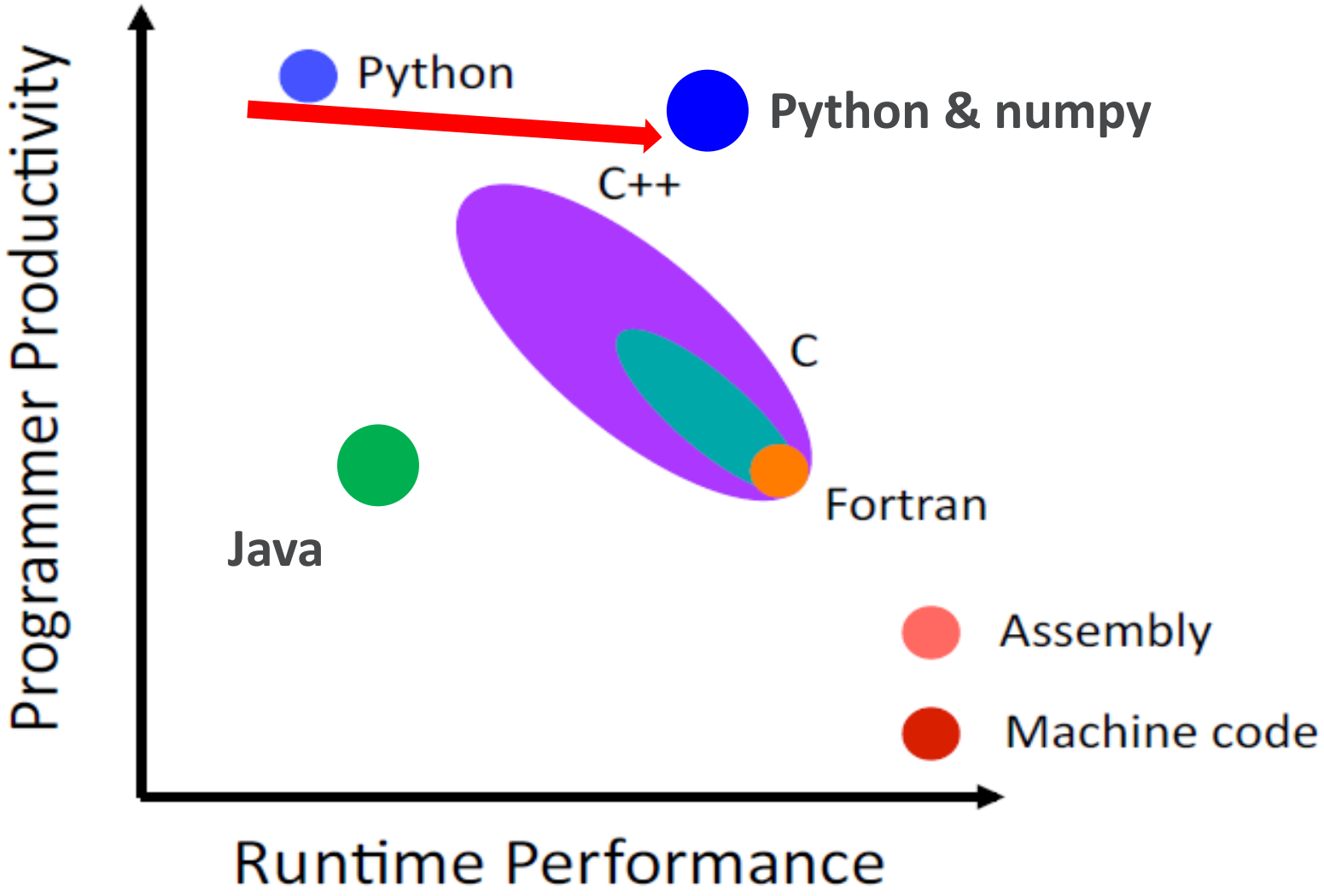
- Support all aspects of diffraction data analysis (from raw data to publication), including capabilities not in GSAS/EXPGUI
- Sequential processing of large numbers of similar datasets
- Written with modern code (Python)
- Incorporates extensive visualization
- Use parameters that “make sense”
- Designed around GUI
- Design goal: Novice friendly, but expert efficient

GSAS-II reads powder diffraction images from all appropriate synchrotron beamlines, as well as the Curiosity Rover on Mars!

Started work ~Nov., 2008

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)**.

WHY PYTHON? – CHOICE OF LANGUAGES (~LOG SCALES!)



WHY PYTHON?

Code snippet – charge flipping all inside a "while" loop

NB: CEhkl is F_{hkl} expanded over full sphere & zero filled out to 1/resolution limit as an array
Start with random phases for CEhkl

```
CErho = np.real(fft.fftn(fft.fftshift(CEhkl)))*(1.+0j)                    #fft Fhkl → ρ(xyz)
CEsig = np.std(CErho)                                                     #get σ(ρ)
CFrho = np.where(np.real(CErho) >= flipData['k-factor']*CEsig,CErho,-CErho)    #CF ρ → ρ'
CFrho = np.where(np.real(CErho) <= flipData['k-Max']*CEsig,CFrho,-CFrho)       #U atom CF!
CFhkl = fft.ifftshift(fft.ifftn(CFrho))                                   #fft ρ(xyz) → F'(hkl)
CFhkl = np.where(CFhkl,CFhkl,1.0)                                         #avoid divide by zero
phase = CFhkl/np.absolute(CFhkl)                                          # get ϕ(hkl) from F'
CEhkl = np.absolute(CEhkl)*phase                                          #apply ϕ  to F
Ncyc += 1                                                                 #count tries
sumCF = np.sum(ma.array(np.absolute(CFhkl),mask=Emask))                        #Σ F
DEhkl = np.absolute(np.absolute(Ehkl)/sumE-np.absolute(CFhkl)/sumCF)      #ΣDF
Rcf = min(100.,np.sum(ma.array(DEhkl,mask=Emask)*100.))                        #R-value for CF
```

NB: the 4D version is almost identical except that F_{hklm} is used

This stuff is fast! ~1s/cycle for 500K reflections/map points

ANOTHER EXAMPLE – OMIT MAP CALCULATIONS

OMIT map – a kind of Fourier map (T.N. Bhat, J. Appl. Cryst. 21, 29-281, 1988)

Usefulness:

- Least biased electron density map for rebuilding structure (mostly for macromolecular structures)

Algorithm:

- Compute density from structure factors & phases
- Unit cell divided into boxes
- In turn – flatten each box, do IFFT to make phases, compute new density & save box (toss the rest).
- Finally assemble all boxes into new “OMIT” map.

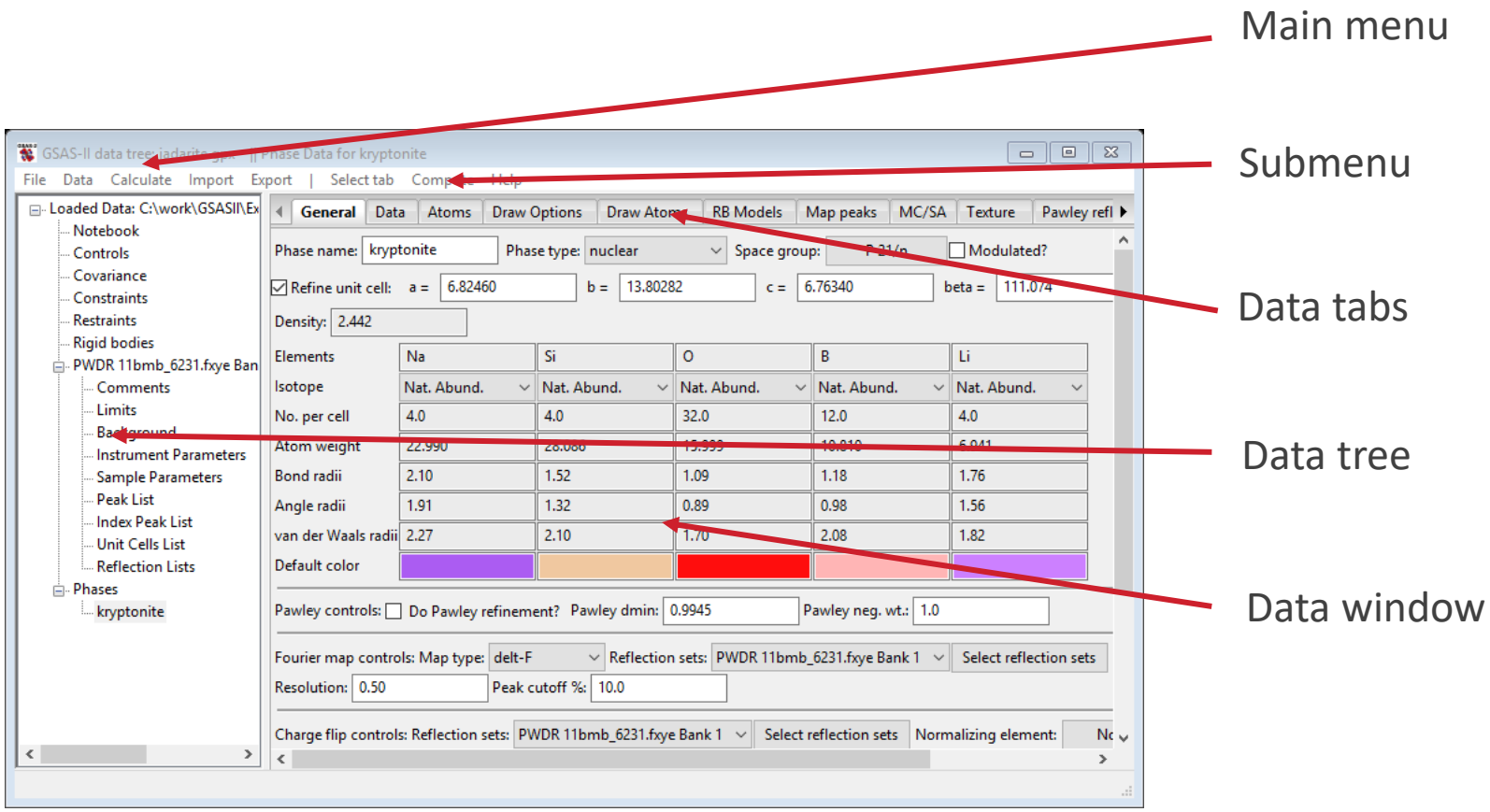
Coding:

- Old CCP4 code – covers 114 pages of printout, ~ 80 lines/page (~9000 lines of Fortran & some c)! Not readily available for nonprotein problems
- GSAS-II python – computed in 65 lines & there is another 80 lines to export it in CCP4 format. Universal for all crystal structures.

Which is easier to understand and maintain?

GSAS-II DESIGN

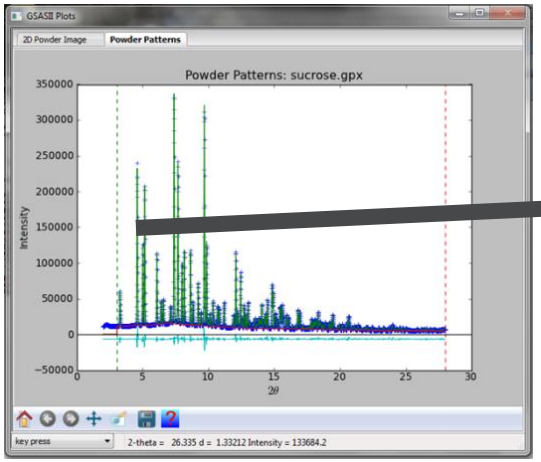
GSAS-II: MODERN GUI – 2 FRAME LAYOUT + CONSOLE

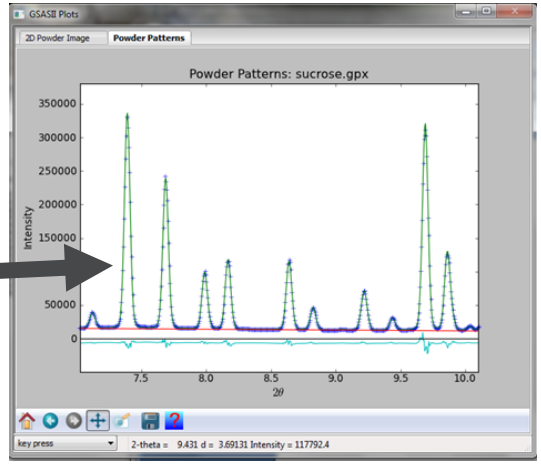


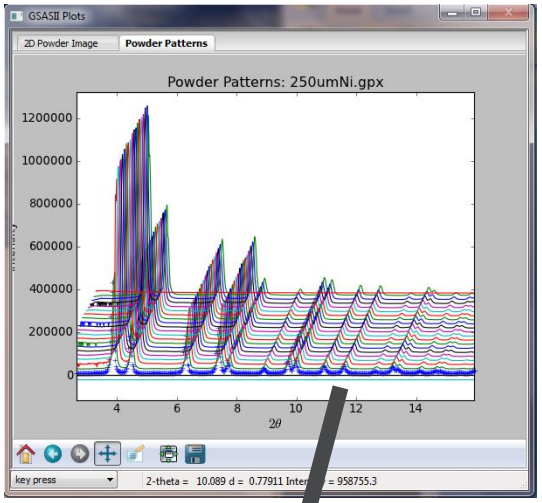
Plot & console in separate frames

THE PLOTS - ADVANCED VISUALIZATION

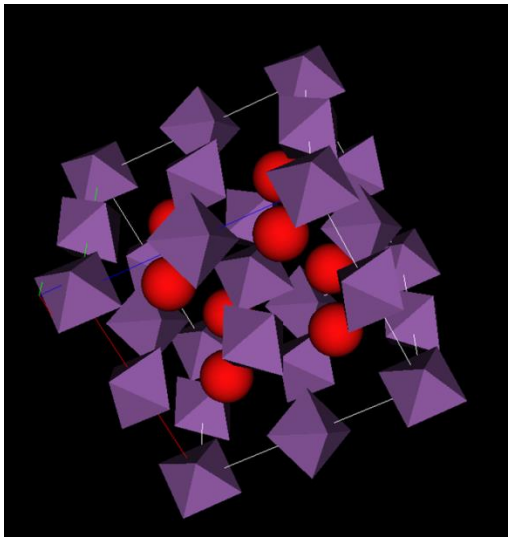
Powder profile – easy zoom



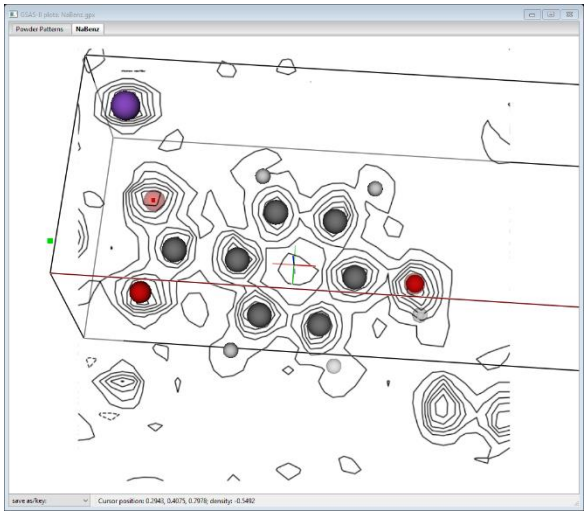




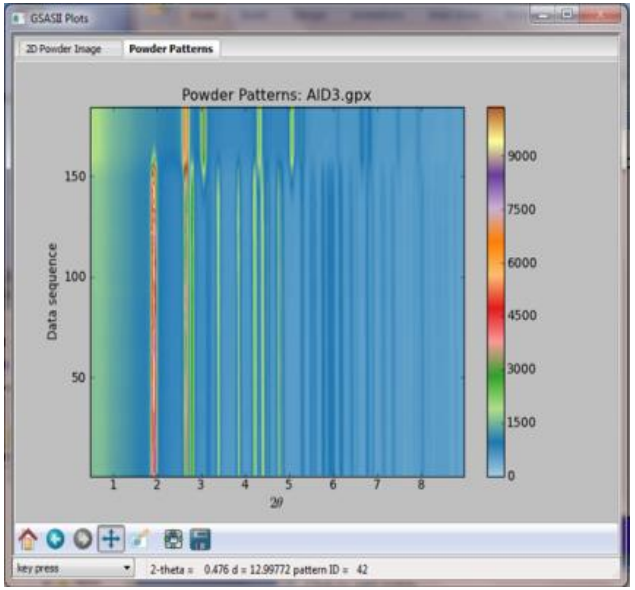
Waterfall plot



Structure drawing



contoured density thru any plane

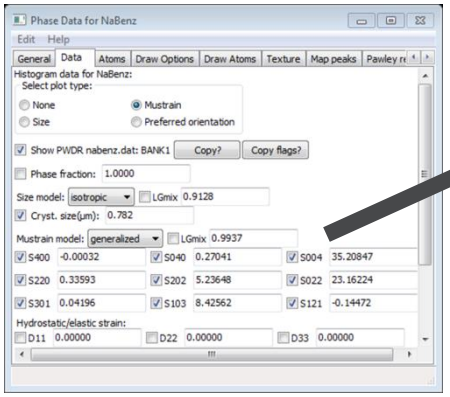


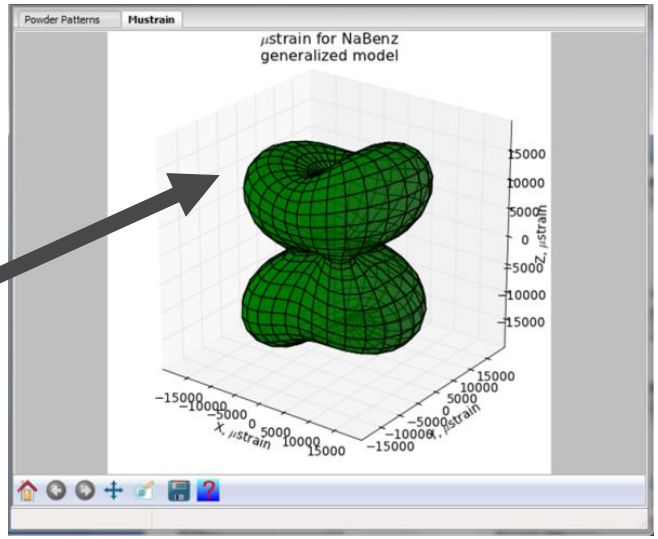
Contour plot

Argonne NATIONAL LABORATORY

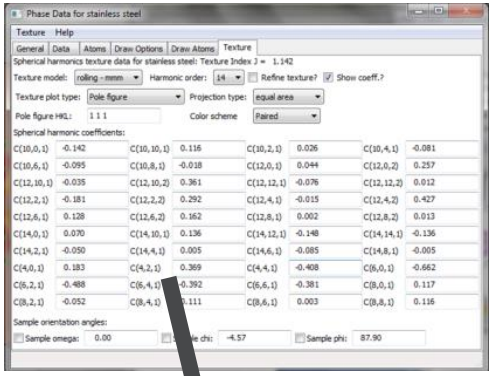
ADVANCED VISUALIZATION IN GSAS-II: NUMBERS AS PICTURES

μstrain surface

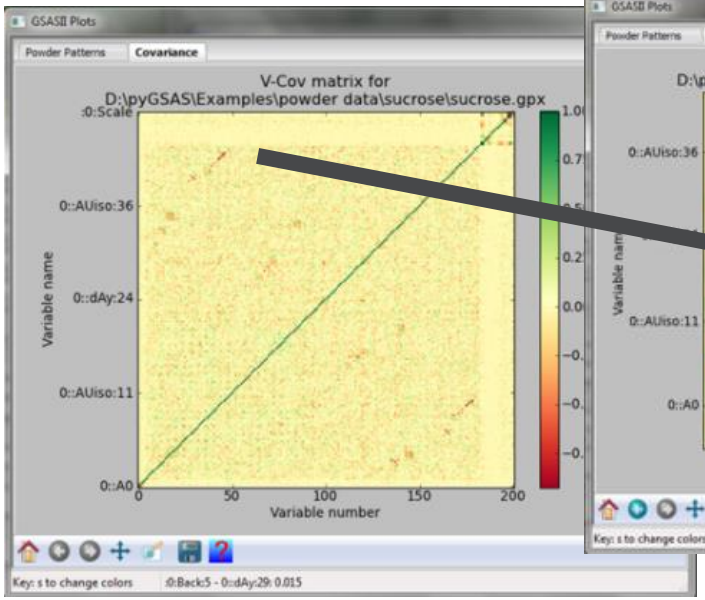


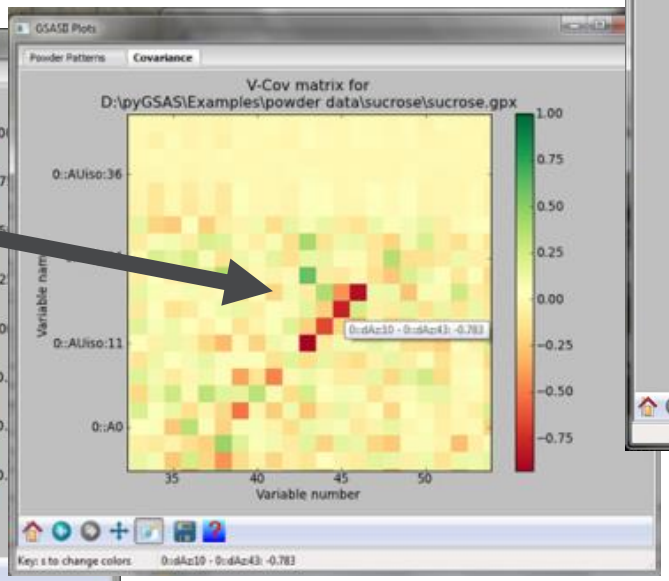


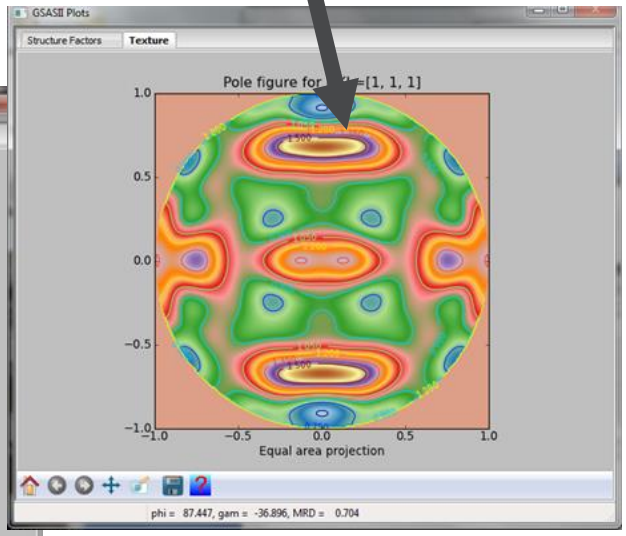
Texture – sph. harmonics



v-cov matrix

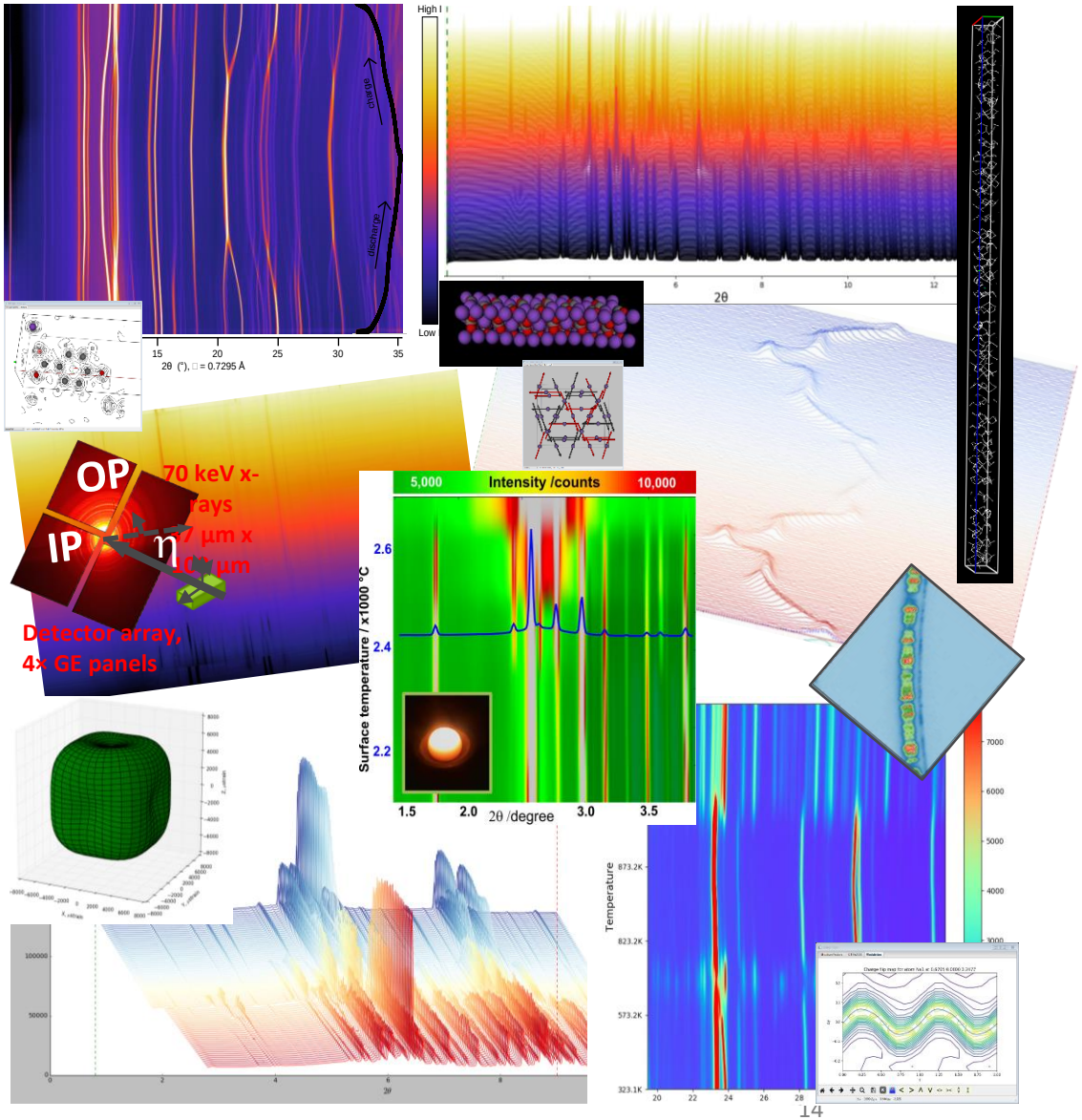






Argonne NATIONAL LABORATORY

GRAPHICS FROM GSAS-II USERS



Recent ACA RefleXions cover:

Large numbers of data sets
>>100 patterns

Visualize phase changes

203Å c-axis mineral –
indexed & solved by CF

Incommensurate structures

Magnetic structures

2D images

Argonne NATIONAL LABORATORY

RIETVELD REFINEMENT IN GSAS-II

Argonne NATIONAL LABORATORY

HISTORY – H.M. RIETVELD

Hugo Rietveld; neutron powder diffractometer, Petten, Netherlands

Papers: H.M. Rietveld, Acta Cryst. 22, 151-2(1967)

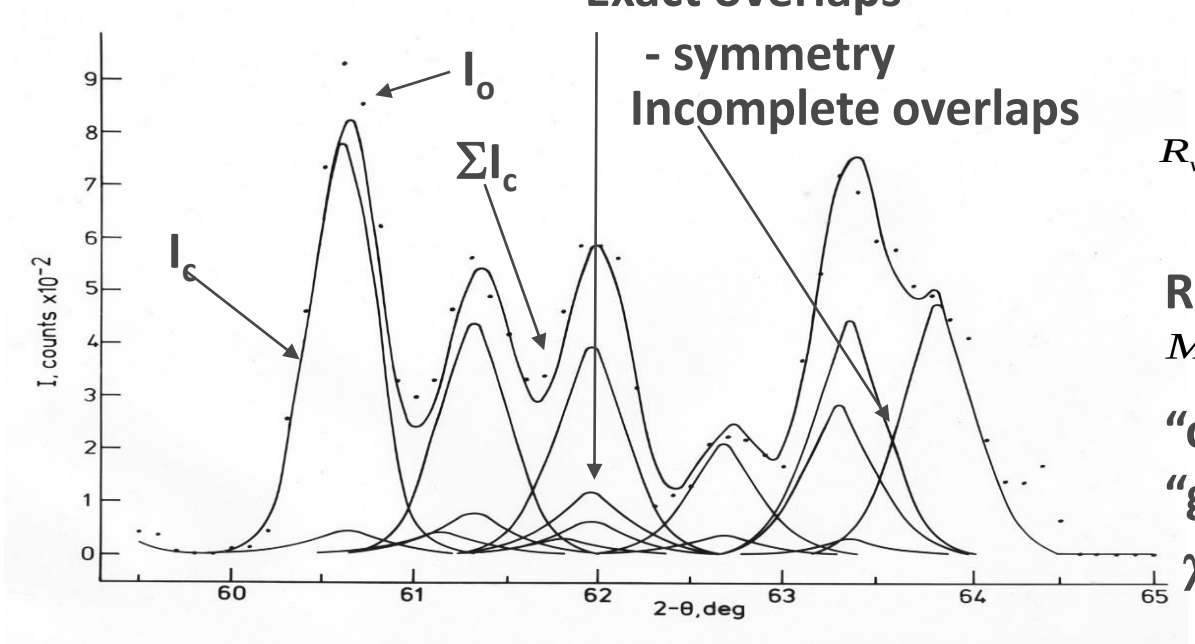
H.M. Rietveld, J. App. Cryst., 2, 65-71 (1969)

Multi-parameter, nonlinear LS curve fitting

Exact overlaps

- symmetry

Incomplete overlaps



Residuals:

$$R_{wp} = \sqrt{\frac{\sum w(I_o - I_c)^2}{\sum w I_o^2}}$$

Rietveld Minimize

$$M_R = \sum w(I_o - I_c)^2$$

"chi-squared" or "goodness-of-fit"

$\chi^2 = M_R/(n-p)$

NON-LINEAR LEAST SQUARES THEORY

Problem - $I(p_i)$ is nonlinear & transcendental (sin, cos, etc.) so can't solve directly

Expand $I(p_i)$ as Taylor series & toss high order terms

$$I_c(p_i) = I_c(a_i) + \sum_i \frac{\partial I_c}{\partial p_i} \Delta p_i$$

a_i - initial values of p_i

$\Delta p_i = p_i - a_i$ (shift)

Normal equations - one for each Δp_i

$$\sum w \left[\Delta I - \sum_i \frac{\partial I_c}{\partial p_i} \Delta p_i \right] \frac{\partial I_c}{\partial p_j} = 0 \qquad \Delta I = I_o - I_c(a_i)$$

Outer sum over observations

Solve for Δp_i - shifts of parameters, <u>NOT</u> values

LEAST SQUARES THEORY - CONTINUED

Rearrange

$$\sum w \frac{\partial I_c}{\partial p_1}\left(\sum_{i=1}^{n} \frac{\partial I_c}{\partial p_i}\Delta p_i\right) = \sum w \Delta I \frac{\partial I_c}{\partial p_1}$$

$$\vdots$$

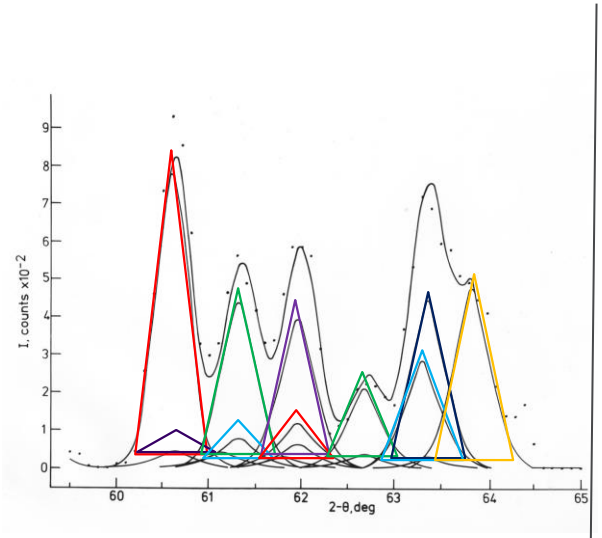
$$\sum w \frac{\partial I_c}{\partial p_n}\left(\sum_{i=1}^{n} \frac{\partial I_c}{\partial p_i}\Delta p_i\right) = \sum w \Delta I \frac{\partial I_c}{\partial p_n}$$

Matrix form: Ax=v

$$a_{i,j} = \sum w \frac{\partial I_c}{\partial p_i}\frac{\partial I_c}{\partial p_j} \quad x_j = \Delta p_j \quad v_i = \sum w(\Delta I)\frac{\partial I_c}{\partial p_i}$$

Solve: $x = A^{-1}v = Bv$; $B = A^{-1}$ This gives set of Δp_i to apply to "old" set of a_i; repeat until Δp_i small.

GSAS-II ALGORITHM



GSAS – process point-by-point to make I_c (value) & $\partial I_c/\partial p_i$ (vector)

$\partial I_c/\partial p_i$

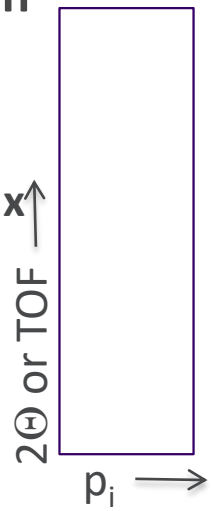
$\partial I_c/\partial p_j$

$\Sigma w[\partial I_c/\partial p_i]\cdot[\partial I_c/\partial p_j]$

Hessian matrix - A

GSAS-II – process reflection-by-reflection to make I_c (vector) & $\partial I_c/\partial p_i$ (Jacobian matrix)

Jacobian matrix - J



Σ

$wJ^TJ = A$

NB: GSAS-II – needs large memory!

Argonne NATIONAL LABORATORY

REFINEMENT VIA MODIFIED LEVENBERG/MARQUARDT-SVD ALGORITHM

Steps:

1. Compute $A_{ij} = \sum w \frac{\partial I_c}{\partial p_i} \frac{\partial I_c}{\partial p_j}$
2. Normalize $A'_{ij} = A_{ij} / \sqrt{A_{ii} A_{jj}}$
3. compute $\chi^2(p)$
4. Select λ (=0.001, “damping factor”)
5. Modify $A''_{ii} = A'_{ii}(1 + \lambda)$
6. **Make SVD inversion of A”**
7. Solve for δp (unnormalized!) & compute $\chi^2(p+\delta p)$
8. If $\chi^2(p+\delta p) > \chi^2(p)$ then $\lambda * 10$ go to 5
9. Else apply δp to p & go to 1 (new cycle)
10. Quit when $\chi^2(p) - \chi^2(p+\delta p) / \chi^2(p) < 0.0001$

SLOW step

FAST steps

NB: all in ~40 lines of python; all double precision

NB^2: this thing is exceedingly robust – no user damping factors needed

SVD – SINGULAR VALUE DECOMPOSITION

Singularities & near singularities – see Mathematical Recipes 2.9

LS matrix: solve for x $Ax=b$ by $x=A^{-1}b$; x are the parameter shifts

SVD: replace $A = UwV$ where U & V are such that $U^{-1} = U^T$ & $V^{-1} = V^T$

& w – diagonal matrix; all same size as A

Then: $A^{-1} = V(1/w_{ii})U^T$

The trick: what to do if $w_{ii} \sim 0$? (singularity) → make $1/w_{ii} = 0$! (instead of ∞)

Then: $x = V(1/w_{ii})U^Tb$ does away with ill-conditioned terms

Have to choose tolerance on $w_{ii} \sim 0$ (typically 10^{-6} but 10^{-3} for proteins works well)

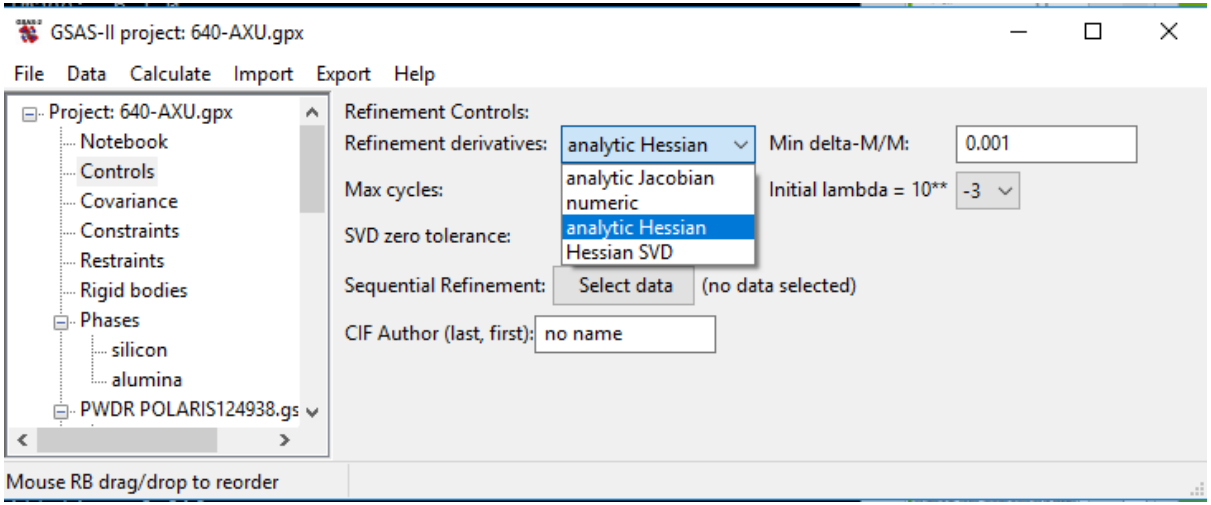
SVD is in python library as numpy.linalg.svd

& uses LAPACK _gesdd routine (fortran – code in MR 2.9)

NB: all double precision in python; downside is w_{ii} not 1:1 to parameters so id of failures difficult.

LEAST SQUARES ALGORITHMS IN GSAS-II

Useful choices – found in Controls



Analytic Hessian – **default** Levenberg-Marquardt SVD from Hessian & computed derivatives
Downside: hard singularities hard to find → "linear algebra errors" cause failures

Analytic Jacobian – uses Jacobian matrix (not Hessian) no SVD; identifies singularities & Removes them from LS refinement; always runs to convergence

Hessian SVD – no Levenberg-Marquardt (might be better for single crystal data)
Same downside as Analytic Hessian

Numeric – no derivatives & slow – mostly for testing purposes.

LEAST SQUARES THEORY - CONTINUED

Error estimates (mostly from W.C. Hamilton)
Given observations n > m parameters
with distributions that have finite 2nd moments
(no need to be "normal" although usually are for powders)
Then LS gives parameter estimates (shifts in our case)
with the minimum variance in any linear combination
The error estimates ("esd's") are

$$\sigma_i = \sqrt{b_{ii} / \chi^2} \qquad \chi^2 = \frac{\sum w(I_o - I_c)^2}{n - m}$$

b_{ii} - diagonal elements of the inverted A matrix
Note: There is little justification for additional scaling of the σ_i **NB: systematic errors will bias results beyond σ_i.**

GSAS-II INSTALLATION

Argonne NATIONAL LABORATORY

GSAS-II INSTALLATION

Search web for "GSAS-II" → only thing out there:

See GSAS-II "home page" https://subversion.xray.aps.anl.gov/trac/pyGSAS.

Better is: https://advancedphotonsource.github.io/GSAS-II-tutorials/install.html

GSAS2full installers for: Windows 10/11, MAC OSX & Linux

- Includes
  - Installation instructions – includes a 1-step for python/GSAS-II
  - Tutorials - ~50 of these; direct access via Help menu for GSAS-II.
  - **NB: Must have internet access!**
  - Usually installs in your personal space so no "authorizations" usually needed (unless you institution forbids it!)
  - Can be installed anywhere (even a memory stick – 5GB or more)
  - Don't use the resulting GSAS-II directory for your project files; work elsewhere
  - **Update often!** Versions can change almost daily as we add/fix things. We work under the continuous software improvement model; not the fixed update schedule model commonly used by commercial software.

GSAS-II CAPABILITIES

Argonne NATIONAL LABORATORY

GSAS-II CAPABILITIES & EXAMPLES

Powder data

- 2D Images:
  - calibration & integration → 1D patterns
  - Direct strain fitting → 3 strain tensor elements*
- 1D patterns
  - Peak picking & fitting*
  - Indexing & space group selection → make new phase
- Multidata X-ray/neutron, CW/TOF → all combinations possible
- Structure solution
  - Stochastic – Monte Carlo/Simulated Annealing
  - Deterministic – Charge Flipping (3D & 4D)
- Structure Refinement – Rietveld Method*
  - Pawley & LeBail refinement (needed for Structure Solution)
  - (3+1) Incommensurate structures
  - Constraints & restraints
  - Rigid bodies (2 kinds)
  - Texture Analysis → spherical harmonics*
- Stacking Faults → DIFFaX simulations (NB: no refinement)
- Pair Distribution Function → data transformation (e.g. make PDF)
- PDF fitting – Large Box & Small Box modelling
- Cluster analysis – Powder data & PDF patterns (New)

* Can use Sequential Analysis

GSAS-II CAPABILITIES & EXAMPLES

Single Crystal Data

- Multidata X-ray/neutron → all combinations possible
- Structure solution
  - Stochastic – Monte Carlo/Simulated Annealing
  - Deterministic – Charge Flipping (3D & 4D)
- Structure Refinement –Levenberg-Marquardt least squares*
  - (3+1) Incommensurate structures
  - Constraints & restraints
  - Rigid bodies (2 kinds)
  - Merohedral & pseudomerohedral twinning
  - Extinction (Gaussian/Lorentzian Primary & Secondary I & II)
- Stacking Faults → DIFFaX simulations (NB: no refinement)

* Can use Sequential Analysis

GSAS-II CAPABILITIES & EXAMPLES

Small Angle & Reflectometry Diffraction Data

- 2D Images:
  - calibration & integration → 1D patterns
- 1D Small Angle Data
  - Scaling to glassy carbon standard
  - Size Analysis
    - Maximum Entropy Analysis
    - Total Non-negative Least Squares
  - Model fitting – components*
    - Particle shapes – e.g. spheres, disks, hollow spheres,…
    - Porod scattering
    - Bragg peaks
    - Protein shape determination
- 1D Reflectometry Data
  - Layer sequence definition & repeats
  - Surface “roughness”, i.e. mixing for all surfaces

* Can use Sequential Analysis

SEQUENTIAL DATA ANALYSIS

Multiple data sets – no maximum number

- 2D Images:
  - Direct strain fitting → 3 strain tensor elements
- 1D patterns
  - Peak picking & fitting
  - Structure Refinement – Rietveld Method
  - Small angle data – model fitting
  - Reflectometry data – model fitting
- Results table
  - Parameter plotting vs experiment variable (e.g.. Temperature)
  - Parametric equation modeling & fitting

GSAS-II TUTORIALS

Argonne NATIONAL LABORATORY

BEST WAY TO LEARN GSAS-II IS BY RUNNING SELECTED TUTORIALS

Most major sections of the program are demonstrated by tutorial examples – best for today is to pick one/two for now

Basic GSAS-II tutorials

- Starting GSAS-II describes how the user interface works

- Fitting laboratory X-ray powder data for fluoroapatite
- CW Neutron Powder fit for Yttrium-Iron Garnet
- Combined X-ray/CW-neutron refinement of PbSO4
- Combined X-ray/TOF-neutron Rietveld refinement

MORE ADVANCED TUTORIALS

Parametric Rietveld fitting

- Sequential refinement of multiple datasets (prerequisite for next)
  - Parametric Fitting and Pseudo Variables for Sequential Fits

Structure solution

- Fitting individual peaks & autoindexing (prerequisite for next two)
  - Charge Flipping structure solution for jadarite
  - Charge Flipping structure solution for sucrose
- Charge Flipping structure solution with Xray single crystal data
- Charge flipping with neutron TOF single crystal data
- Monte-Carlo simulated annealing structure determination

MORE ADVANCED TUTORIALS (II)

Stacking Fault Modeling

- Stacking fault simulations for diamond
- Stacking fault simulations for Keokuk kaolinite
- Stacking fault simulations for Georgia kaolinite

Image Calibration/Integration

- Calibration of an area detector
- Integration of area detector data
- Calibration of a Neutron TOF diffractometer

MORE ADVANCED TUTORIALS (III)

Small-Angle Scattering

- Small angle x-ray data size distribution (alumina powder)
- Fitting small angle x-ray data (alumina powder)
- Image Processing of small angle x-ray data
- Sequential refinement with small angle scattering data

Other

- Texture analysis of 2D data
- Rietveld Refinement detail:
  - Fitting the Starting Background using Fixed Points
- Merohedral twin refinements
- Single crystal refinement from TOF data
- Scripting a GSAS-II Refinement from Python
- Strain fitting of 2D data

THANK YOU

Argonne NATIONAL LABORATORY