Structure Determination and Refinement using **TOPAS**

Arnt Kern
Structure Determination and Refinement with TOPAS - Overview
DIFFRAC\textsuperscript{plus} TOPAS
TOtal Pattern Analysis Solutions

- Generalized software for profile and structure analysis
- Seamless integration of all currently employed profile fit techniques and related applications
  - Single Line Fitting
  - Indexing (LSI, LP-Search)
  - Whole Powder Pattern Decomposition (Pawley, Le Bail)
  - Structure determination (Simulated Annealing, Charge Flipping, 3D Fourier Analysis)
  - Structure refinement (Rietveld refinement, Two-Stage Method)
  - Quantitative Rietveld analysis

- Current user base: >3000
TOPAS Users

- User's base: >3000 users as of 12/2009
- About 1400 structure determination papers using TOPAS as of 12/2009
The Classic SDPD Process

- Peak Finding
- Indexing

- Intensity Extraction
  - Le Bail, Pawley

- Structure Determination using $F^2(\text{obs})$

- Structure Refinement using $y_i(\text{obs})$ or $F^2(\text{obs})$

$F^2(\text{obs})$: Observed structure factors
$y_i(\text{obs})$: Observed step intensity data
Structure Refinement using $F^2(\text{obs})$: Two-Stage Method (Will, 1979)
The Classic SDPD Process

- Peak Finding
- Indexing

- Intensity Extraction
  - Le Bail, Pawley
- Structure Determination using $F^2(\text{obs})$
- Structure Refinement using $y_i(\text{obs})$ or $F^2(\text{obs})$
SDPD Processes in TOPAS

F²(obs) or yᵢ(obs)

- Peak Finding
- Indexing

- Intensity Extraction
  - Le Bail, Pawley
- Structure Determination using F²(obs)
- Structure Refinement using yᵢ(obs) or F²(obs)

- "Profiling"
  - Le Bail, Pawley
- Structure Determination AND Refinement using yᵢ(obs)

TOPAS Approach
Coelho (2000)

Suited for
- Simulated annealing
- Charge Flipping

Suited for
- Simulated annealing
TOPAS
Structure Determination Features

- Indexing: LSI and LP-Search methods
- Structure determination: Simulated Annealing, Charge Flipping, Fourier Analysis
- Simultaneous refinement on any number of powder and single crystal data sets (lab and synchrotron X-ray data, CW and TOF neutron data)
  - Refines on any number of structures per diffraction pattern.
- For a given multiphase pattern, all profile fitting techniques supported by TOPAS can be used simultaneously to describe individual phase contributions to the full pattern
  - Structure determination in the presence of additional phase(s) with known or unknown structure
  - Successful structure determination of 2 phases simultaneously (Simulated Annealing)
- Flexible macro language
  - Support of user-defined refinement parameters / refinement models
  - Computer algebra system for function minimization and for the application of linear and non-linear restraints
TOPAS
Structure Determination Features

- Choice between predefined and user-defined
  - linear and non-linear restraints; can be combined with penalty functions, e.g. anti-bump, parabola, lattice energy minimization, ...
  - minimization schemes, e.g. standard least squares, "robust refinement" (David, 2001),...
  - weighting schemes
  - ...
- Rigid bodies
  - All parameters can be refined / restrained (lengths, angles, Bs, Occs, ...)
  - Cartesian, fractional or internal (z-matrix notation) coordinates
- Rigid body editor for graphical creation of rigid bodies
- Spherical harmonics to account for preferred orientation
- ...

Bruker AXS
Structure Determination

Indexing
TOPAS
LSI and LP-Search

TOPAS introduces two unique ab-initio powder pattern indexing methods

- **LSI**
  - Iterative use of least squares
  - Operates on d-values extracted from reasonable quality powder diffraction data

- **LP-Search**
  - Monte-Carlo Based Whole Powder Pattern Decomposition
  - Independent of d-spacing extraction and line profile shape and therefore suited for indexing of poor quality powder data

! No d-values required !
Method
1. LSI Iterative Process
   • hkls assigned using present (random) lattice parameters
   • Reciprocal lattice relationship solved using least squares for all hkl

\[ X_{hh} h^2 + X_{kk} k^2 + X_{ll} l^2 + X_{hk} h l + X_{hl} h l + X_{kl} k l = 1 / d_{hkl}^2 \]

2. Monte-Carlo approach to searching parameter space
   • Randomize lattice parameters
   • Execution of the LSI iterative process until convergence
TOPAS
LSI

Most important features:

- Seamless integration into TOPAS
- Zero-point error consideration
- Automatic determination of possible spacegroups
- Highly tolerant to impurity peaks, missing high d-spacings, extreme lattice parameter ratios as well as large d-spacing and zero point errors (> 0.05° 2θ)
- Particularly strong in indexing of very large cells (>> 100.000 Å³) and dominant zone problems
- Weighting of reflections using observed peak intensities or user-defined weights
- Fully automated Pawley or Le Bail fitting of all or user-selected solutions
- Goodness-of-fit versus volume plots
Indexing of Difficult Cells with LSI
Example 1: 4-Methoxy 3 Nitro Benzaldehyde Form II

Data courtesy of P. Stephens, Stony Brook, USA. To be published.
The first 18 observed peaks are fit by a single zone (H0L).

- **Spacegroup C2/c**
- **a = 62.424 Å, b = 3.849 Å, c = 14.180 Å, β = 104.4°**
Example 2: Six-peptide sequence with Zn atom

Data courtesy of P. Stephens, Stony Brook, USA. To be published.
The first 20 observed peaks are fit by a single zone (H0L)

- Spacegroup P2_1
- \(a = 23.497 \, \text{Å}, \ b = 4.773 \, \text{Å}, \ c = 21.113 \, \text{Å}, \ \beta = 103.6^\circ\)
TOPAS
LSI: Large Unit Cells

- Example 1: Tetragonal Hen Egg White Lysozyme (HEWL)

Data courtesy of B. Von Dreele, Argonne, USA.
HEWL
Spacegroup P4₁₂₁₂
a = 78.61 Å
b = 38.525 Å
V = 238063 Å³
TOPAS
LSI: Large Unit Cells

- HEWL: Pawley Fit

\[ R_{WP} = 1.8\% \]
TOPAS
LSI: Large Unit Cells

- Example 2: T3R3 Human Insulin-Zinc Complex

Data courtesy of B. Von Dreele, Argonne, USA.
TOPAS
LSI: Large Unit Cells

- T3R3 Human Insulin-Zinc Complex
  - Spacegroup R3
  - $a = 81.301 \text{ Å}$
  - $c = 73.052 \text{ Å}$
  - $V = 418173 \text{ Å}^3$
TOPAS
LSI: Large Unit Cells

- T3R3 Human Insulin-Zinc Complex: Pawley Fit

R_{WP} = 2.9%
Indexing of powder diffraction patterns by iterative use of singular value decomposition

A. A. Coelho

LP-Search is a Monte-Carlo based Whole Powder Pattern Decomposition approach

It minimizes on a new figure of merit function that gives a measure of correctness for a particular set of lattice parameters

\[
\text{FOM} = \sum_{j} \sum_{i} I(2\theta_i) \left| 2\theta_i - 2\theta_{0,j} \right|
\]

The figure of merit function assigns parts of the diffraction pattern to calculated peak positions and then sums the absolute values of the products of the diffraction intensities multiplied by the distance to the calculated peak positions.

LP-Search avoids difficulties associated with extracting d-spacings from complex patterns comprising heavily overlapped lines.
TOPAS
LP-Search

- Generate sets of lattice parameters and calculate d-values
  - For each solution, for each calculated d-value
    - define pattern segments
    - sum the absolute values of (step intensities * distance to the d-value)

- Refine the best solution

- Reiterate
TOPAS
LP-Search

Most important features:
- Seamless integration into TOPAS
- Independent of 2θ or d-spacing extraction
- Independent of line profile shape
- Zero-point error consideration
- Highly tolerant to large zero point errors (> 0.05° 2θ)
- Particularly suited for indexing of poor quality powder data, where reliable 2θ or d-spacing extraction is difficult or even impossible
Particularly suited for indexing of poor quality powder data:
How many peaks are there?

- Peak overlap?
- Anisotropic line broadening?
- 2+ phases?
TOPAS
LP-Search: LT-ZrMo$_2$O$_8$

- Data are easily indexed with LP-Search
- LP-Search profile fit reveals strong anisotropic line broadening

LT-ZrMo$_2$O$_8$
$a = 5.879$ Å
$b = 7.329$ Å
$c = 9.130$ Å
D8 ADVANCE, K$_{\alpha_1}$
Allen et al. (2003)
TOPAS
LP-Search: LT-ZrMo$_2$O$_8$

- Final Pawley fit taking anisotropic line broadening into account
- Spherical harmonics function used to model excess broadening

LT-ZrMo$_2$O$_8$

$\begin{array}{c}
a = 5.879 \text{ Å} \\
b = 7.329 \text{ Å} \\
c = 9.130 \text{ Å}
\end{array}$

D8 ADVANCE, K$_{\alpha_1}$
Allen et al. (2003)
Discussion of the indexing algorithms within TOPAS
A. Coelho & A. Kern (2005)
CPD Newsletter No. 32, 43-45
Structure Determination
Simulated Annealing
Structure Determination
Simulated Annealing

- Simulated annealing is a direct space approach where adjustable parameters lie in direct rather than reciprocal space.
- Procedure:
  1. A trial crystal structure is constructed by randomly positioning and orienting individual atoms, molecular fragments or complete molecules taking into account (known or guessed) space group information.
  2. After calculating diffraction data and comparing it against the measured diffraction data, the variable parameters of the model are adjusted in order to maximise the level of agreement between the observed and calculated data (i.e., minimize $\chi^2$).
- This procedure is typically applied to observed structure factors, $F^2(\text{obs})$, but has been extended to step intensity data, $y_i(\text{obs})$.

⇒ TOPAS (Coelho, 2000)
Whole-profile structure solution from powder diffraction data using simulated annealing

A. A. Coelho


Topas

Simulated Annealing

Techniques and methods to facilitate the solution of structures by simulated annealing have been developed from the starting point of a space group and lattice parameters. The simulated annealing control parameters have been systematically investigated and optimal values characterised and determined. Simulated annealing in the inclusion of electron-potential energy functions in a non-linear least-squares Rietveld refinement procedure. The large-range electron-potential functions are calculated using a general real-space minimisation which can be used for all space groups. In addition, a general weighting scheme for potential functions requires the need to determine weighting schemes for the electron-potential energy function and for the Rietveld method. It has also been shown that simulated annealing procedures can be used to refine structural parameters. The behaviour and structure of the technique have been studied on 11, 55 and 76, using the whole-profile powder data against the known structures of Al36Cu5, Fe36Cu5, and 99.5% pure Al36Cu5.

1. Introduction

The fact in the crystal structure minimum energy, as determined by the forces of physics, help explain and predict atomic behavior. Atoms align themselves into a low-energy configuration and then arranged in a low-energy configuration that would resemble structures observed in nature. Also, using the presence of a system states random motion in time and spatial space, the system will tend to find the lowest energy state. The result is a new minimum energy configuration that would also yield scattering intensity that closely matched the low-energy state as determined by X-ray and electron diffraction.

In the thermodynamic process of annealing, materials available over time as the cooling of the system, the possibility of observing a particular energy state E, relates to a given energy state E, given by the Boltzmann distribution or (BTD), where T is the temperature and A(T) = e^(E/kT) is the change in energy from the previous state. Atoms that change this process are known as terminal annealing (TA). The computer determines the probability of bonds to the particular energy state E, given by the Boltzmann distribution or (BTD), where T is the temperature and A(T) = e^(E/kT) is the change in energy from the previous state. The probability of change between bond (i) and bond (j) in the probability of observing a bond change of A, the parameter T can be determined after simulated annealing, or quickly to simulate quenching.

An algorithm consists of a number of essential parameters that influence the search for the global minimum; these include the initial temperature T, the size of the initial random search, the starting configuration, the energy of initial configuration, and the energy of the model. Associated atom movements can be approximated with a change in the energy of the model. The energy of the model is determined such that 80% of the energy of the atom is accepted according to the criterion of 0.05 (Hartley & Smith, 1986).

The problem associated with implementing TA is in the choice of the search for a global solution is very slow. To overcome this, the local minimum of the algorithm, the program EMOD (Le Bail et al., 1993) implements the so-called Monte Carlo method (Metropolis, 1953), which has no derivative information and consequently the expected computational effort is large when attempting to solve complex structures. On the other hand, a structure solution method described here use derivative information to locate local minimum of an objective

E = \phi(E)

\phi(E) = \begin{cases} 
E, & \text{if } \Delta < 0 \\
\tilde{E}, & \text{if } \Delta > 0
\end{cases}

\Delta = E - \bar{E} = E - \phi(E)

A. A. Coelho - Simulated Annealing - 979
Hugo M. Rietveld, 1967/1969

The basic principle of the method is a description of all data points of a powder pattern using analytical functions

The parameters of these functions, consisting of crystal structure, sample, instrument and background parameters, are refined simultaneously using least squares methods

\[ \text{Chi}^2 = \sum_i w_i [y_i(\text{obs}) - y_i(\text{calc})]^2 \rightarrow \text{min} \]
Important Key Features:

- Step intensity data instead of structure factors, $F_2^{\text{obs}}$, are used. Each data point is an observation.
  - No attempts are made to deconvolute overlapped peaks, avoiding problems associated with intensity partitioning.
- A preconceived (at least partial) structure model is required, "with its parameters reasonably close enough to the final values".
- This automatically raises the question:

  "How far off the position of an atom may be and the refinement still brings it in?"
"A correctly formulated global optimisation approach may be regarded as a Global Rietveld Refinement" (K. Shankland, 2004)

- For step intensity powder data, repeated Rietveld refinements of trial structures are performed: after convergence a new Rietveld refinement is initiated with parameter values changed according to a temperature regime (⇌ simulated annealing)

- Using step intensity data for structure determination has important and obvious advantages:
  - No preceeding intensity extraction required
  - No problems associated with peak overlap (intensity partitioning)
  - Structure determination from poor quality powder data
Diffraction data types supported by TOPAS

- Structure factors, $F^2(\text{obs})$
  - good data quality needed
  - single crystal data can be used
  - fast

- Step intensity data
  - no preceeding intensity extraction required
  - avoids problems associated with peak overlap (intensity partitioning)
  - structure solution from poor quality powder data
  - slow

- "Peak maximum intensities"
  - step intensity data set comprising only data at calculated peak positions; data in between are discarded
  - fast
Example
Structure Determination of Cimetidine


- 17 (non-H) atoms
- 9 torsion angles
Example
Structure Determination of Cimetidine

- From step intensity data to "peak maximum intensities"

~6 times faster
Example (1 GHz PIII, 250,000 iters)
Structure Determination of Cimetidine

Individual atoms
- Nr. of DoFs : 51
- Nr. of solutions : 11
- Time : 2090 sec.
- Success rate : 190 sec / solution

Individual atoms, S "boxed"
- Nr. of DoFs : 51
- Nr. of solutions : 29
- Time : 2118 sec.
- Success rate : 73 sec / solution
Example (1 GHz PIII, 250,000 iters)
Structure Determination of Cimetidine

Rigid body, all torsions refined
- Nr. of DoFs: 15
- Nr. of solutions: 13
- Time: 1718 sec.
- Success rate: 132 sec / solution

Ideal rigid body
- Nr. of DoFs: 6
- Nr. of solutions: 70
- Time: 1490 sec.
- Success rate: 21 sec / solution
Example
Structure Determination of Mo$_2$P$_4$O$_{15}$

- One of the largest structures solved with TOPAS (simulated annealing)
- Single crystal data (Bruker AXS SMART 6000)
  - SG: Pn (7)
  - a = 24.1134(6) Å
  - b = 19.5324(5) Å
  - c = 25.0854(6) Å
  - $\beta = 100.015(1)^\circ$
  - V = 4450.9 Å$^3$
  - 441 atoms in asymmetric unit

Lister et al., Chem. Commun., 2004, 2540
Structure Determination
Charge Flipping
Charge Flipping
Oszlányi and Sütő, 2004

- Iterative algorithm
- Requires only lattice parameters and reflection intensities
- No use of chemistry / trial structure models
- The output is an approximate scattering density of the structure sampled on a discrete grid
- Charge flipping is very fast
  - The grid size determines the calculation speed
Charge Flipping
Oszlányi and Sütő, 2004

1. Take $|F_{hkl}|$
   Guess phases

2. Calculate electron density $\rho(r)$

3. If $\rho(r) < \text{value}$ "flip charge"
   $\rho(r) = -\rho(r)$

4. Calculate $|F_{hkl}|_{\text{new}}$
   and new phases from new $\rho(r)$

5. Keep new phases and replace by $|F_{hkl}|$
Charge Flipping
Memory Lane

- The beginning:
- Superspace solutions:
- Powder diffraction:
  Wu, Leinenweber, Spence & O'Keeffe
- Histogram matching:
- Tangent formula, symmetry consideration, determination of origin, atom picking and assignment:
A charge-flipping algorithm incorporating the tangent formula for solving difficult structures

A. A. Coelho

Example
Structure Determination of Mo$_2$P$_4$O$_{15}$

- One of the largest structures solved with TOPAS (simulated annealing)
- Single crystal data (Bruker AXS SMART 6000)

- SG: Pn (7)
- $a = 24.1134(6)$ Å
- $b = 19.5324(5)$ Å
- $c = 25.0854(6)$ Å
- $\beta = 100.015(1)^\circ$
- $V = 4450.9$ Å$^3$
- 441 atoms in asymmetric unit

Lister et al., Chem. Commun., 2004, 2540
Example
Structure Determination of $\text{Mo}_2\text{P}_4\text{O}_{15}$

Charge Flipping
- "Default" run
- Typically very high proportion of 441 atoms correctly identified (>99%)?

~15 sec.
Structure Determination
3D Fourier Analysis
TOPAS
3D Fourier Analysis

- 3D visualisation of electron density distributions
  - Observed Fourier maps
  - Calculated Fourier maps
  - Difference Fourier maps
  - User-defined maps
- Atom picking capabilities with recognition of special positions
- Allows simultaneous display of electron densities, picked atoms, and crystal structures
- The ideal tool for structure completion, if Simulated Annealing or Charge Flipping methods only deliver partial structure models
TOPAS
3D Fourier Analysis

- PbSO$_4$

Difference Fourier analysis to locate missing oxygen positions

Final structure after atom picking
Structure Determination
Simulated Annealing vs. Charge Flipping
### SDPDRR-3
Structure Determination by Powder Diffractometry Round Robin - 3

**Organized by:**
A. Le Bail and L.M.D. Crammick
February 2008

**CLOSED RESULTS AVAILABLE**

**IES Commission for Powder Diffraction Chairmen's Message, June 2007**

In a decade since Alain Le Bail and Liesbeth Crammick issued a challenge to the powder diffraction community to solve two crystal structures from powder diffraction data alone. Despite a generous timetable of several months, there were less than a handful of correct solutions. Would the situation be different this time?

**Introduction**

After the first SDPDRR (Structure Determinations by Powder Diffractometry Round Robin - 3)

<table>
<thead>
<tr>
<th>Sample 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random formula:</strong> CaCo3(OH)4Cl2PO4</td>
</tr>
<tr>
<td>- cell parameters: a = 5.22 Å, b = 10.43 Å, c = 6.24 Å, α = 105.37, β = 107.5, γ = 94.34 degrees, symmetry: triclinic</td>
</tr>
<tr>
<td>- CIF file corresponding to the previously known crystal structure (two determinations) providing the molecular formula</td>
</tr>
<tr>
<td>- Reflection laboratory X-ray data from a PANalytical diffractometer, CuKα</td>
</tr>
<tr>
<td>- Data: sample1.txt</td>
</tr>
<tr>
<td>- See the patterns</td>
</tr>
</tbody>
</table>

**Sample 2**

| **Random formula:** CaCo3(OH)4Cl2PO4 |
| - cell parameters: a = 5.22 Å, c = 32.60 to 33.65 Å due to composition variation, symmetry: triclinic |
| - CIF file according to the ICD-DDPS code 032-8402 and 032-8408 |
| - Optical synthesis data, BRUKER D8 |
| - Powder data from the EIL DDS, low resolution mode, wavelength = 1.5444 Å |
| - Data: sample2.txt |
| - See the patterns and the solution patterns |

**Cumulative histogram**

After 12 weeks, 167 data downloads and 7 solutions submitted by 3 participants.
SDPDRR-3
Charge Flipping - "LaWO"

SDPDRR-3, sample 2:

Sample info provided:
- Probable formula close to $\text{La}_{14}\text{W}_8\text{O}_{45}$ or $\text{La}_8\text{W}_5\text{O}_{27}$
- Symmetry: hexagonal
- $a = 9.039 \text{ Å}$, $c = \text{from 32.60 to 33.65 Å}$ due to composition variation

Proposed solution (organizers):
- $\text{La}_{18}\text{W}_{10}\text{O}_{57}$, $Z = 2$
- P-62c (No. 190)
- W6 half occupied site (very short W4-W6 interactomic distance 2.42 Å)
SDPDRR-3
Charge Flipping - "LaWO"

~4 sec.
SDPDRR-3
Charge Flipping - "LaWO"

~4 sec.
SDPDRR-3
Charge Flipping - "Tartrate"

SDPDRR-3, sample 1:

Sample info provided:
- Probable formula: CaC$_4$H$_4$O$_6$·4H$_2$O
- Symmetry: Triclinic
- cell parameters:
  \[ a = 8.222 \, \text{Å}, \quad \alpha = 105.97° \]
  \[ b = 10.437 \, \text{Å}, \quad \beta = 107.51° \]
  \[ c = 6.249 \, \text{Å}, \quad \gamma = 94.94° \]

Proposed solution (organizers):
- CaC$_4$H$_4$O$_6$·4H$_2$O, Z = 2
- P-1 (No. 2)

March-Dollase PO(101): ~0.9
(obtained from final Rietveld refinement)

Charge Flipping is successful here despite significant preferred orientation!
SDPDRR-3
Charge Flipping - "Tartrate"

- Key to success: Low Density Elimination (Shiono & Woolfson, 1992)
SDPDRR-3
Charge Flipping - "Tartrate"

Key to success: Low Density Elimination (Shiono & Woolfson, 1992)
Simulated Annealing vs. Charge Flipping

Conclusions

Simulated Annealing:
- Requires a trial structure model, which can be partial or random
- Performs better on poor quality data. **Important advantage!**
- Comparatively slow

Charge Flipping:
- No use of chemistry / trial structure models. **Important advantage!**
- Requires high quality data
- Even if the structure doesn’t solve completely, heavy atoms and / or molecular fragments can often be found very quickly, which greatly assists subsequent simulated annealing structure determination
- Very fast; structures can be (partially) solved in seconds up to a few minutes, i.e. faster than one typically can create a start model / rigid body for simulated annealing
SDPD Processes in TOPAS
Methods of Solution

- Poor quality (powder) data
- Trial (random) structure model required

Simulated Annealing

3D Fourier Analysis

Partial solution

Charge Flipping

Structure solved

- High quality (powder) data
- No structure model required
Structure Determination
Data Quality Issues
SDPD
Variable Counting Time

Constant Counting Time vs. Variable Counting Time

I ~ LP * thermal vibration * f^2

Boehmite (Madsen, 1992)

The gain in data quality is obvious
A VCT strategy can greatly enhance the chances of success of SDPD but has always significant benefits for structure refinement.

Atomic coordinates, occupancy factors and (anisotropic) thermal parameters are better determined, especially in the case of light atoms.

Refinement of atomic coordinates and thermal parameters of very light atoms, is more likely to be stable with VCT.
Conclusions
Conclusions
Structure Determination with TOPAS

- Structure determination using direct space and charge flipping methods can be considered routine for many powder diffraction problems as emphasised by the significant increase in the number of published structures solved in this way.

- The major limitations are related to the well known ambiguities related to systematic and accidental peak overlap in powder diffraction. Profound crystallographic knowledge is required to deal with these limitations.
  - The maximum size of structures that can be solved with TOPAS is thus mainly limited by data quality.

- Charge Flipping is generally suggested to start with from the beginning due to its ease of use and speed. Chances are high to find at least a partial solution, which may then greatly assist to create a better trial structure for subsequent Simulated Annealing runs.