



## Doctoral Thesis

# Combining charge flipping with other methods to solve increasingly complex crystal structures from powder diffraction data

**Author(s):**

Šišak Jung, Dubravka

**Publication Date:**

2013

**Permanent Link:**

<https://doi.org/10.3929/ethz-a-009987420> →

**Rights / License:**

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

# Combining charge flipping with other methods to solve increasingly complex crystal structures from powder diffraction data

A dissertation submitted to  
ETH ZURICH

for the degree of  
Doctor of Sciences

presented by  
Dubravka Šišak Jung  
Dipl. ing. chem. University of Zagreb  
born 29. April 1980  
citizen of Croatia

Accepted on the recommendation of

Prof. Dr. Walter Steurer, ETH Zurich, examiner  
Prof. Dr. Lukáš Palatinus, Institute of Physics Prague, co-examiner  
Prof. Dr. Christopher J. Gilmore, University of Glasgow, co-examiner  
Dr. Lynne B. McCusker, ETH Zurich, co-examiner

# Abstract

Starting from the premise that the chemical composition of a polycrystalline material is known, the focus of this thesis was on the elucidation of the arrangement of the atoms in the material (solving its crystal structure). This involves solving not only the "phase problem", which is fundamental to X-ray diffraction structure analysis, but also the "overlap problem", which is inherent to powder diffraction data. Because powder samples contain a huge number of crystallites, a diffraction experiment will necessarily result in the superposition of reflections from different crystals. Some of these reflections (those with similar  $d$ -spacing) will overlap and only the net intensity of these reflections can be measured. Determining the individual contributions of the reflections to the net intensity is therefore central to structure analysis from powder diffraction data. In comparison to a single-crystal structure analysis, where data are collected by illuminating one crystal (of suitable size) with X-rays, the analogous powder diffraction data analysis is at least one level more complex.

Most of the methods developed for single-crystal data analysis (Patterson method, direct methods, Shake 'n Bake, maximum entropy method, charge flipping) have been adapted more or less successfully to XPD data. The overall success of these methods is, of course, dependent on how (or if) the overlap problem is treated. On the other hand, new methods, such as direct-space methods were developed to address the XPD data specifically. These methods simplify both the phase and the overlap problems by including chemical information about structural fragments (molecules, inorganic subunits (polyhedra)) in the phasing procedure. The success of these methods depends critically on the correctness of the chemical model.

Particularly difficult problems, however, require methods to be combined before structure determination is possible. Although results obtained from different analytical techniques (spectroscopy, thermogravimetry) can be used in this way, the combination of different diffraction techniques has proven to be very successful. In particular, the complementarity of XPD and electron microscopy has been exploited in the determination of the structures of zeolites and similar compounds.

The objective of this thesis was to investigate further possibilities of how different structure determination methods could be combined to access more difficult structures. The charge flipping algorithm, a relatively new approach to the phase problem, was chosen as the starting point. The program operates in both reciprocal and real space, so combinations in both realms are possible. The charge-flipping structure-solution algorithm was initially developed for single-crystal data, and is implemented in the program *Superflip*. The algorithm starts with random phases, and these are used to calculate an electron density map. Density values below a certain threshold  $\delta$  are multiplied by minus one to produce a perturbed map with new phases, which are then combined with the observed

amplitudes to generate a new electron density map. This original algorithm was reported to be very successful for centrosymmetric structures with heavy scatterers. Light-atom structures, on the other hand, were reported to be difficult, especially if they were non-centrosymmetric. For such cases, the success of algorithm found its limit once the number of atoms in the asymmetric unit exceeded *ca.* 100. The algorithm was adapted to accommodate powder diffraction data by introducing a repartitioning of the overlapping reflection intensities based on a histogram matching procedure. Powder charge flipping (pCF) is proving to be a powerful tool for solving the structures of polycrystalline materials. When this PhD project started, the pCF algorithm had been applied successfully to a number of different materials, but with rather arbitrarily chosen values for the input parameters. Consequently, one of the first tasks was to undertake a systematic study of the effect of the different parameters on the structure solution. Several aspects were investigated:

- (1) the effect of the input parameters themselves
- (2) the effect of data resolution, and
- (3) the effect of including information from other sources.

As it was not considered to be sensible to test all combinations of all parameters, a more pragmatic approach was followed. Only the input parameters that were believed to be relevant for the XPD data were chosen for the tests. The parameters were:

- (1) the reflection overlap factor,  $\kappa$
  - (2) the isotropic atomic displacement factor,  $B_{iso}$ ,
  - (3) the cycle for starting the repartitioning of overlapping reflections and the repeat interval,  $HMstart$  and  $HMinterval$ , and
  - (4) the threshold value for charge flipping,  $\delta$ .
- Then, only certain combinations of input parameters were tested: values of the parameters were chosen to fall into what were thought to be sensible ranges. To investigate possible differences between inorganic and organic materials, the tests were performed on a zirconium phosphate (ZrPOF-Q1) and D-ribose. Results were monitored by comparing the resulting electron density maps with the correct one.

For ZrPOF-Q1, optimal parameters for a data resolution of 0.97 Å were found to be  $\kappa = 0.2-0.3$ ,  $B_{iso} = 0-1$ ,  $HMstart =$  not too early (20th cycles),  $HMinterval =$  frequent (every 5 cycles), and  $\delta = \delta_{AUTO}$ . The optimization procedure was then repeated for lower resolution data, and it was found that the structure could still be solved with a 1.8 Å data resolution. Below that, interpretable maps were still produced, but the effect of the input parameter values was found to be negligible.

This optimized input parameter set was applied to data collected on several samples of unknown structure, three centrosymmetric with heavy scatterers (one zirconium phosphate, ZrPOF-EA, two tricyanomethane derivatives) and four non-centrosymmetric all-light atom structures (cyclo- $\beta$ -peptides). While the former were solved easily, only the simplest of the all-light atom structures could be solved.

Attempts to optimize the *Superflip* input parameters for D-ribose, however, failed. The

second part of this PhD project was then devoted to addressing this problem. Reasoning that starting the *Superflip* calculation with a more realistic set of phases might help, "optimized" starting phases were calculated from structures approximating the correct one. Such approximate structures were generated using the global optimization of a flexible model (chain) using the simulated-annealing algorithm in the program *FOX*. None of these were correct, but it was hoped that they would be in the vicinity of the correct solution. Phases calculated from such an approximate structure were used to generate starting phase sets (by varying the phases by a user-defined percentage - phase change,  $pc$ ). The other input parameters values were then re-optimized. The optimal parameters were found to be  $pc = 10\text{-}20\%$ ,  $\kappa = 0.3$ ,  $\delta = 0.5\delta_{AUTO}$ ,  $HMstart = \text{early}$  (5th cycle) and  $HMinterval = \text{frequent}$  (every 5th cycle). This resulted in easily interpretable electron density maps with the correct symmetry. Using the optimized input parameter set, the structure could still be solved with  $1.18 \text{ \AA}$  data.

In order to investigate whether the degree of reflection overlap caused the failure of *Superflip*, single-crystal diffraction data were simulated from the refined D-ribose structure. Sets of fully interpretable solution were obtained. Similarly, several XPD patterns with various peak widths were simulated from the solved structure, and here it was found that the structure solution was possible only when the degree of overlap is low (*ca.* 40%).

As it became apparent that non-centrosymmetric all-light atom structures present a challenge for *Superflip*, a series of such compounds, both of known and unknown structure (analogous sugars, cyclo- $\beta$ -peptides), were investigated to see what exactly defines a structure as a complex one. Compounds differing in the number of atoms in the unit cell, in the degree of reflection overlap in the XPD pattern, and in symmetry were tested in *Superflip*. First, the standard approach using random phases was applied, and then, if that failed, the approximate structure approach. Here it was found that only the simplest structures (low number of atoms, low degree of overlap) could be solved using the standard approach. The approximate structure approach allows more complex structures to be solved, but appears to be limited by the number of atoms in the unit cell. So far, no all-light atoms, non-centrosymmetric structure with more than 50 atoms in the unit cell has been solved using this approach.

The third part of this PhD project addressed the evaluation of the solutions generated by *Superflip*. As it is a multisolution algorithm, usually more than 100 solutions are generated within a single job. In easier cases (centrosymmetric structures, heavy scatterers) the better solutions are usually easily recognized by the convergence  $R$  value,  $R_{SF}$ . However, in more complex cases, such as D-ribose, this was not found to suffice. This presented a particular problem, because only 14% of the solutions were found to be fully interpretable. Therefore, an attempt to evaluate the maps using the other tools provided in *Superflip* (symmetry recognition, peakiness) was made, but without success.

Then, the solution sets were evaluated using two cluster analyses, one based on a com-

parison of electron density maps, and the other based on a comparison of the phase sets. While the latter was not found to be useful, the former showed that indeed similar (fully interpretable maps) do cluster. However, it was not clear which cluster would contain the good maps. Finally, an evaluation based on entropy was tried. No direct correlation between the entropy value and quality of the solution was apparent, but applying the entropy criterion resulted in the formation of a small group of interpretable solutions. By averaging these electron density maps, a map that was either fully interpretable or a good starting point for recycling in a subsequent *Superflip* run was produced. Recognizing the best solutions from *Superflip* remains a challenge, but both cluster analysis and an entropy-based evaluation of the solutions look promising.

# Zusammenfassung

Der Fokus dieser Arbeit lag darin die Anordnung der Atome in einem Material zu untersuchen (die Kristallstruktur zu lösen), unter der Annahme, dass die chemische Zusammensetzung des polikristallinen Materials bekannt ist. Dies beinhaltet nicht nur das Lösen des „Phasen Problems“, das eines der grundlegenden Themen in der Röntgenbeugungsstrukturanalyse ist, aber auch das „Überlappungs Problem“, ein inherentes Problem der Pulverdiffraktometrie Daten. Ein Pulverdiffraktogramm Experimente ergeben immer eine Überlagerung der Reflexe von verschiedenen Kristallen, da Pulver Proben eine grosse Anzahl an Kristallen enthalten. Einige dieser Reflexe (mit ähnlichen  $d$ -Abständen) werden überlappen und die Netto Intensität dieser Reflexe kann gemessen werden. Das Bestimmen der individuellen Beiträge der Reflexe zu der Netto Intensität ist daher zentral für die Struktur Analyse der Pulverdiffraktogramm Daten. Im Vergleich zu der Einkristall Struktur Analyse, bei der Daten gesammelt werden indem man einen einzigen Kristall mit Röntgen Strahlen bestrahlt, ist die Pulverdiffraktogramm Daten Analyse um einiges komplexer.

Die meisten Methoden für Einkristall Daten Analyse (Patterson method, direct methods, Shake 'n Bake, maximum entropy method, charge flipping) sind mehr oder weniger erfolgreich für XPD Daten adaptiert worden. Im allgemeinen ist der Erfolg dieser Methoden aber stark davon abhängig ob und wie das Überlappungs Problem behandelt wird. Andererseits sind neue Methoden, wie zum Beispiel die Direct-Space Methoden entwickelt worden, spezifisch für die XPD Daten. Diese Methoden vereinfachen das Phasen und das Überlappungs Problem indem die chemische Information über Strukturfragmente (Moleküle, inorganische Untereinheiten (Polyhedra)) in die Phasing Prozedur miteinbezogen werden. Der Erfolg dieser Methoden hängt kritisch von der Richtigkeit des chemischen Modells ab.

Besonders schwierige Probleme erfordern jedoch eine Kombination von Methoden, bevor eine Strukturbestimmung möglich ist. Obwohl die Ergebnisse aus verschiedenen analytischen Techniken (Spektroskopie, Thermogravimetrie), in dieser Weise verwendet werden können, hat sich die Kombination von verschiedenen Röntgen Beugungs Techniken als sehr erfolgreich erwiesen. Insbesondere hat sich die Komplementarität von XPD und Elektronenmikroskopie bei der Bestimmung der Struktur der Zeolithe und ähnlicher Verbindungen erwiesen.

Das Ziel dieser Arbeit war es, weitere Möglichkeiten zu untersuchen, wie verschiedene Strukturbestimmungs Methoden kombiniert werden können, um schwierigere Strukturen zu lösen. Der Charge-flipping Algorithmus, eine relativ neue Herangehensweise an das Phasen Problem, wurde als Ausgangspunkt gewählt. Das Programm arbeitet im reziproken und realen Raum, daher sind Kombinationen in beiden Bereichen möglich. Der Charge-Flipping Struktur-Lösung Algorithmus wurde ursprünglich für Einkristall-Daten

entwickelt und ist im Programm *Superflip* implementiert. Der Algorithmus beginnt mit zufälligen Phasen, und diese werden verwendet, um eine Elektronendichte zu berechnen. Dichte Werte unterhalb einer bestimmten Schwelle  $\delta$  werden mit minus eins multipliziert, um eine gestörte Karte mit neuen Phasen zu erzeugen, die dann mit den beobachteten Amplituden kombiniert werden, um eine neue Elektronendichte zu berechnen. Dieser ursprüngliche Algorithmus war sehr erfolgreich für zentrosymmetrischen Strukturen mit schweren Streuern. Leichte-Atom Strukturen, waren schwierig zu lösen, insbesondere wenn sie nicht zentrosymmetrischen sind. Für solche Fälle stieß Algorithmus an seine Grenzen, sobald die Anzahl der Atome in der asymmetrischen Einheit *ca.* 100 überschritten.

Der Algorithmus wurde an Pulverbeugungsdaten angepasst, durch die Einführung einer Repartitionierung der überlappenden Reflexionsintensitäten basiert auf einer Histogramm Matching Prozedur. Powder Charge-Flipping (PCF) erweist sich als ein leistungsfähiges Werkzeug für die Lösung der Strukturen von polykristallinen Materialien sein. Als dieses PhD-Projekt startete, wurde der PCF-Algorithmus bereits erfolgreich an einer Reihe von verschiedenen Materialien angewendet, aber mit ziemlich willkürlichen gewählten Werten für die Input-Parameter. Folglich war eine der ersten Aufgaben, systematisch den Einfluss der verschiedenen Parameter auf die Struktur Lösung zu untersuchen. Verschiedene Aspekte wurden untersucht:

- (1) der Effekt der Inputparameter
- (2) der Effekt der Dataenauflösung, und
- (3) der Effekt Information von anderen Quellen miteinzubeziehen.

Nicht alle möglichen Kombinationen der Parameter wurde getestet, nur die Inputparameter die relevant für die XPD Daten sind, wurden für die Tests ausgewählt. Diese Parameter sind: (1) der Reflex Überlappungs Faktor,  $\kappa$  (2) der Debye-Waller,-Faktor,  $B_{iso}$ , (3) der Zyklus für das starten der Repartitionierung der Überlappungs Reflexe und das Wiederholungsintervall,  $HMstart$  and  $HMinterval$ , und (4) der Schwellwert für das Charge-Flipping,  $\delta$ .

Nur bestimmte Kombinationen von Eingangsparametern wurden getestet: Die Werte der Parameter wurden so gewählt, dass sie in sinnvollen Bereichen liegen. Um mögliche Unterschiede zwischen anorganischen und organischen Materialien zu untersuchen, wurden die Tests auf einem Zirkoniumphosphat durchgeführt (ZrPOF-Q1) und auf D-Ribose. Die Ergebnisse wurden durch einen Vergleich der resultierende Elektronendichte Karten mit der richtigen überwacht.

Die optimalen Parameter für ZrPOF-Q1, wurden bei einer Auflösung von  $0.97 \text{ \AA}$   $\kappa = 0.2-0.3$ ,  $B_{iso} = 0-1$ ,  $HMstart =$  nicht zu früh (20 Zyklen),  $HMinterval =$  frequent (alle 5 Zyklen), und  $\delta = \delta_{AUTO}$  gefunden. Das Optimierungsverfahren wurde dann mit einer geringeren Auflösung wiederholt, und es wurde festgestellt, dass die Struktur immer noch mit einer bis zu  $1,8 \text{ \AA}$  tiefen Auflösung gelöst werden kann. Darunter wurden immer noch



interpretierbare Karten produziert, aber die Wirkung der Eingabeparameterwerte erwies sich als vernachlässigbar.

Dieses optimierte Eingangsparameterset wurde auf die Daten von mehrere Proben mit unbekannter Struktur angewandt, drei zentrosymmetrische mit schweren Streuer (ein Zirkoniumphosphat, ZrPOF-EA, zwei tricyanomethane Derivate) und vier nicht-zentrosymmetrische Leicht-Atom-Strukturen (cyclo- $\beta$ -peptides). Während erstere leicht gelöst wurde, konnte nur die einfachste der Leicht-Atom-Strukturen gelöst werden.

Versuche, die *Superflip* input Parameter für D-Ribose zu optimieren, schlugen jedoch fehl. Der zweite Teil dieser Dissertation wurde diesem Problem gewidmet. Unter Berücksichtigung, dass das Starten der *Superflip* Berechnung mit einem realistischeren Set von Phasen hilfreich sein könnte, wurden "optimierte" Startphase von Strukturen berechnet, die der richtigen ähnlich sind. Solche ungefähre Strukturen wurden unter Verwendung des globalen Optimierung aus einem flexiblen Modells (Kette) mit dem simulierten Annealing Algorithmus des Programms *FOX*. Keine von diesen war korrekt, aber es wurde gehofft, dass sie in der Nähe des korrekten Lösung liegen würden. Die Phasen dieser approximierten Strukturen wurden benutzt um Phasen Sets zu erzeugen, dies geschah über eine Benutzerdefinierten prozentuelle Phasen Änderung, *pc*. Die anderen Eingabeparameter Werte wurden dann erneut optimiert. Die optimalen Parameter waren *pc* = 10-20 %  $\kappa$  = 0,3  $\delta$  = 0,5  $\delta_{AUTO}$ , *HM start* = early (5. Zyklus) und *HM Intervall* = frequent (jeder 5. Zyklus). Dies führte zu leicht interpretierbaren Elektronendichte Karten mit der richtigen Symmetrie. Mit dem optimierten Eingangsparametersatz könnte die Struktur sogar noch mit 1,18 Å gelöst werden. Um zu untersuchen, ob das Mass an Reflex Überlappung für das Scheitern von Superflip verantwortlich ist, wurde Einkristall Beugungs Daten simuliert von der verfeinerten D-Ribose Struktur. Sets von ganz interpretierbaren Lösungen wurden gefunden. In ähnlicher Weise, mehrere XPD Patterns mit verschiedenen Peakbreiten wurden aus der gelösten Struktur simuliert, und es hat sich ergeben, dass die Struktur Lösung nur dann möglich war, wenn der Grad der Überlappung niedrig (ca.40 %) ist. Als klar wurde, dass nicht-zentrosymmetrischen Leicht-Atom-Strukturen eine Herausforderung für *Superflip* darstellte, wurde eine Reihe solcher Verbindungen, sowohl von bekannten und unbekanntem Struktur (analog Zucker, Cyclo- *beta*-Peptide) untersucht um zu sehen, was genau eine Struktur als komplex definiert. Verbindungen, die sich in der Anzahl der Atome in Einheitszelle, im Grad der Reflex Überlappung im XPD Muster und in Symmetrie unterscheiden, wurden in *Superflip* getestet. Zunächst wurde die Standard-Ansatz mit zufälligen Phasen angewendet und wenn dies nicht funktionierte, wurde der approximierte Struktur Ansatz verwendet. Dabei wurde festgestellt, dass nur die einfachsten Strukturen (geringe Anzahl an Atomen, geringe Überlappung) mit dem Standard-Ansatz gelöst werden konnte. Der approximierte Struktur Ansatz ermöglicht mehr Strukturen zu lösen, aber nur für eine bestimmte Anzahl der Atome in der Zelle. Bisher konnte keine Leicht-Atome, nicht-zentrosymmetrischen Struktur mit mehr

als 50 Atomen in der Zelle mit diesem Ansatz gelöst werden. Der dritte Teil dieser Dissertation behandelt die Auswertung der von *Superflip* generierten Lösungen. Da es ein multiresolution Algorithmus ist, werden in der Regel mehr als 100-Lösung in einem einzigen Job erzeugt wird. In einfacheren Fällen (zentrosymmetrischen Strukturen, schwere Streuer), werden die besseren Lösungen in der Regel leicht durch die Konvergenzwert  $R$ ,  $R_{SF}$  erkannt. Jedoch in komplexeren Fällen, z. B. D-Ribose, genügt dies nicht. Dies stellte ein besonderes Problem dar, weil nur 14 % der Lösungen sich als vollständig interpretierbar erwiesen. Daher wurde ein Versuch gemacht, die Karten mit den anderen Werkzeugen in *Superflip* (Symmetrie Erkennung, peakiness) zu bewerten gemacht, aber ohne Erfolg. Dann wurden die Lösung Sets mit Hilfe zweier Cluster-Analysen evaluiert, eine basierend auf einem Vergleich der Elektronendichte Karten, und die andere basiert auf einem Vergleich der Phase-Sets. Während letztere nicht hilfreich war, zeigte der erste, dass in der Tat ähnliche (vollständig interpretierbare Karten) sich clustern. Allerdings war es nicht klar, welche Cluster die guten Karten enthalten würden. Schließlich wurde eine Bewertung auf Basis der Entropie versucht. Keine direkte Korrelation zwischen dem Entropie Wert und der Qualität der Lösung war offensichtlich, aber die Anwendung des Entropie Kriterium führte zur Bildung einer kleinen Gruppe von interpretierbaren Lösungen. Durch Mittelung diese Elektronendichte Karten, entstand eine Karte, die entweder vollständig interpretierbar oder ein guter Ausgangspunkt für das Recycling in einem anschließenden *Superflip* Lauf war. Die besten Lösungen von *Superflip* zu erkennen bleibt eine Herausforderung, aber die Cluster-Analyse und die Entropie-basierte Auswertung der Lösungen sind vielversprechend.