Charge-Flipping

Eine neue Methode zur Lösung des Phasenproblems.

Internes Seminar, 9.1.2007, CR

Inhalt

- 1. Einleitung
- 2. Bekannte Lösungswege
 - 2.1. Patterson-Synthesen
 - 2.2. Direkte Methoden
- 3. Charge-Flipping
 - 3.1. Idee und Algorithmus
 - 3.2. Programme
 - 3.3. Superflip: Programmtechnisches und Demo
- 4. Zusammenfassung

1. Einleitung

1.1. Das Phasenproblem

Basics:

$$F_{hkl} = \sum_{j=1}^{N} f_j e^{2\pi i (hx_j + ky_j + lz_j)}$$
(1)

bzw. in Vektorform:

$$F_{\vec{h}} = \sum_{j=1}^{N} f_j e^{2\pi i (\vec{h} \vec{x_j})}$$
 (2)

mit

- $F_{\vec{h}}$: Strukturfaktor
- j = 1 ... N: Atome in der Elementarzelle
- f_j : Atomformfaktoren = f(Gesamtelektronenzahl des Atoms, λ , Θ)
- e^{\dots} = Phase \mapsto Strukturinformation $\vec{x_j}$ (relative Anordnung der Streuzentren zueinander)

alternativ zu (2) als Integral über das Volumen der Elementarzelle:

$$\mathsf{F}_{\vec{\mathsf{h}}} = \int_{\mathsf{VEZ}} \rho_{\vec{\mathsf{x}}} \mathsf{e}^{2\pi \mathsf{i}\vec{\mathsf{h}}\vec{\mathsf{x}}} \mathsf{d}\mathsf{V} \quad \textbf{(3)}$$

Phasenproblem:

- $F_{\vec{h}}$ nicht messbar, sondern nur Reflex-Intensitäten $I_{\vec{h}} = F_{\vec{h}}^2$
- How Phaseninformation (bei zentrosymmetrischen Strukturen Vorzeicheninformation) geht verloren

1.2. Prinzip der Fouriertransformation

eindimensional: (Mathe I/II):

• periodische Funktion f(x) (Periode L) \mapsto als Fourier-Reihe entwickelt:

$$f(x) = \sum_{n} A_{n} e^{2\pi i \frac{nx}{L}}$$
 (4 a)

• mit Fourier-Koeffizienten:

$$A_n = \frac{1}{L} \int f(x) e^{-2\pi i \frac{nx}{L}} dx \quad (3 a)$$

Übertragung:

• Strukturfaktor F \mapsto periodische Funktion der Elektronendichte $\rho_{\vec{x}}$:

$$\mathsf{F}_{\vec{\mathsf{h}}} = \int \rho_{\vec{\mathsf{x}}} \mathsf{e}^{2\pi \mathsf{i} \vec{\mathsf{h}} \vec{\mathsf{x}}} \mathsf{d} \mathsf{V} \quad (3)$$

- Vergleich mit Prinzip der Fourier-Synthese
 - ◇ $F_{\vec{h}}$ sind die Fourierkoeffizienten der periodischen Funktion ρ_x : $\rho_{\vec{x}} =$

$$= \frac{1}{V} \sum_{\vec{h}} \mathsf{F}_{\vec{h}} \mathsf{e}^{-2\pi \mathsf{i} \vec{h} \vec{x}}$$
(4)

- Elektronendichte = Fourierreihe der F-Werte
- F-Werte = Fouriertransformierte der Elektronendichte

Anwendung:

- für bekannte $F \mapsto Elektronendichtekarte mit (4) berechenbar$
- → Basis jeder Strukturverfeinerung

1.3. Übersicht: realer – reziproker Raum

Raum	reziprok	real
Ort (Koord.)	$\vec{h} = h,k,l$	$\vec{x} = x, y, z$
Amplitude	Strukturfaktor F	Elektronendichte ρ
Symmetrie	11 Laueklassen	32 Punktgruppen
	81 Beugungssymbole	230 Raumgruppen
	aus F ²	
	keine Translationssymmetrie	translationssymmetrisch
	$F_{\vec{h}} = \sum_{j=1}^{N} f_j e^{2\pi i (\vec{h} \vec{x_j})}$ (2)	
	$F_{\vec{h}} = \int \rho_{\vec{x}} \mathrm{e}^{2\pi \mathrm{i}\vec{h}\vec{x}} \mathrm{dV}$ (3)	$ ho_{ec{x}} = rac{1}{V} \sum_{ec{h}} F_{ec{h}} e^{-2\pi i \vec{h} \vec{x}}$ (4)

2. Bekannte Lösungswege

2.1. Patterson-Methode

- älteste Methode, auch ohne Rechner verwendbar
- Idee
 - ♦ analog (4) für ρ_{x} :

$$\rho_{\vec{\mathsf{x}}} = \frac{1}{\mathsf{V}} \sum_{\vec{\mathsf{h}}} \mathsf{F}_{\vec{\mathsf{h}}} \mathsf{e}^{-2\pi \mathsf{i}\vec{\mathsf{h}}\vec{\mathsf{x}}} \tag{4}$$

werden statt F einfach F² (Messung!) eingesetzt \mapsto neue Funktion (Patterson-Funktion) (mit Ortskoordinaten u, v, w = \vec{u}):

$$\mathsf{P}_{\vec{\mathsf{u}}} = \frac{1}{\mathsf{V}} \sum_{\vec{\mathsf{h}}} \mathsf{F}_{\vec{\mathsf{h}}}^2 \mathsf{e}^{-2\pi \mathsf{i} \vec{\mathsf{h}} \vec{\mathsf{u}}} \tag{5}$$

o durch Einsetzen von (3) für F

$$\mathsf{F}_{\vec{\mathsf{h}}} = \int \rho_{\vec{\mathsf{x}}} \mathsf{e}^{2\pi \mathsf{i}\vec{\mathsf{h}}\vec{\mathsf{x}}} \mathsf{d}\mathsf{V} \quad (3)$$

in (5) folgt für $P_{\vec{u}}$:

$$\mathsf{P}_{\vec{\mathsf{u}}} = \frac{1}{\mathsf{V}} \int_{\mathsf{V}} \rho_{\vec{\mathsf{x}}} \rho_{\vec{\mathsf{x}}+\vec{\mathsf{u}}} \mathsf{d}\mathsf{V} \quad (6)$$

 $\diamond \mapsto \mathsf{P}_{\vec{u}}$ ist Amplitude in einem Vektorraum \vec{u} (Patterson-Raum)

2.1. Übersicht: realer – reziproker – Patterson-Raum

Raum	reziprok	real	Vektor
Ort (Koord.)	$\vec{h} = h,k,l$	$\vec{x} = x, y, z$	$\vec{u} = u, v, w; u = x_1 - x_2$
Amplitude	Strukturfaktor F	Elektronendichte ρ	Pattersonfunktion P
Symmetrie	11 Laueklassen	32 Punktgruppen	24 Pattersongruppen
	81 Beugungssymbole	230 Raumgruppen	Harker-Geraden;
	aus F ²		Harker-Schnitte
	keine Translationssymmetrie	translationssymmetrisch	keine Translationssymmetrie
	$F_{\vec{h}} = \sum_{j=1}^{N} f_j e^{2\pi i (\vec{h} \vec{x_j})}$ (2)		
	$F_{\vec{h}} = \int \rho_{\vec{x}} e^{2\pi i\vec{h}\vec{x}} dV$ (3)	$ \rho_{\vec{x}} = \frac{1}{V} \sum_{\vec{h}} F_{\vec{h}} e^{-2\pi i \vec{h} \vec{x}} $ (4)	$P_{\vec{u}} = \frac{1}{V} \sum_{\vec{h}} F_{\vec{h}}^2 e^{-2\pi i \vec{h} \vec{u}}$ (5)
			$P_{\vec{u}} = rac{1}{V} \int_{V} \rho_{\vec{x}} \rho_{\vec{x}+\vec{u}} dV$ (6)

2.1. Patterson-Funktion: Anwendung

- anschaulich:
 - > Patterson-Maxima an den Orten u, wo Kombinationsvektoren 2-er Atome liegen (Vektorraum!)
 - d.h. erkennbar werden Abstände vom jeweiligen Atom aus gesehen
 - in vielen Programmen implementiert (z.B. SHELXS-97: PATT)
- Nachteile:
 - sehr breites Maximum bei P_{0,0,0}

 - Maxima sind <u>breiter</u> als bei der Fouriersynthese der F-Werte (Elektronendichtekarten)
 - <u>sehr viele</u> Peaks, da die Maxima bei Interkombinationsvektoren liegen
- daher i.A. nur geeignet für
 - Schweratomstrukturen (z.B. Metallkomplexe) (1 Vektor dominiert P)
 - nach Schweratom-Ersatz auch f
 ür Proteine

2.1. Beispiel: SHELXS-97-Input

TITL SrIn2 CELL 0.71070 5.009 5.009 8.036 90.0 90.0 120.0 ZERR 2 0.0020 0.0020 0.0040 0.0 0.0 0.0 P 63/M M C,NR.194*CB* LATT 1 SYMM -Y, X-Y, Z SYMM Y-X, -X, Z SYMM X-Y, -Y, .5+ZSFAC SR IN UNIT 2 4 OMIT 2 C TREF FMAP 10 HKLF 4 1



2.1. Beispiel: SHELXS-97-Output (Auszug I)

Super-sharp Patterson for SrIn2

Maximum = 999.10, minimum = -69.32 highest memory used = 9320 / 3956

0.1 seconds CPU time

Rms Patterson density excluding points close to the origin or an equivalent lattice point is 34.45

	Х	Y	Z	Weight	Peak	Sigma	Length
1	0.0000	0.0000	0.0000	24.	999.	29.00	0.00
2	0.0000	0.0000	0.4092	12.	347.	10.07	3.29
3	0.6667	0.3333	0.5000	12.	343.	9.97	4.95
4	0.6667	0.3333	0.2076	6.	253.	7.34	3.34
5	0.6667	0.3333	0.2945	6.	224.	6.49	3.74
6	0.6667	0.3333	0.0870	6.	150.	4.35	2.98
7	0.5000	0.0000	1.0000	8.	48.	1.41	2.50
8	0.5000	0.0000	0.5000	8.	47.	1.36	4.73
9	0.9364	0.1840	0.0000	2.	42.	1.21	1.12

2.1. Beispiel: SHELXS-97-Output (Auszug II)

Patterson vector superposition minimum function for SrIn2 Patt. sup. on vector 1 0.6667 0.3333 0.2076 Height 253. Length 3.34 Maximum = 217.13, minimum = -123.52 highest memory used = 12281 / 7798 58 Superposition peaks employed, maximum height 49.5 and minimum height Heavy-Atom Location for SrIn2 109 reflections used for structure factor sums Solution 1 CFOM = 81.78 PATFOM = 99.9 Corr. Coeff. = 90.5 SYMFOM = Shift to be added to superposition coordinates: 0.3326 1.1656 0.3485 Name At.No. x y z s.o.f. Minimum distances/PATSMF (self first IN1 52.4 0.6667 0.3333 0.4569 0.1667 2.97 160.2 IN2 48.1 1.0000 0.0000 0.2500 0.0833 4.02 3.34 169.6 247.6 SR3 13.3 0.6667 0.3333 0.6546 0.1667 1.53 1.59 2.99 0.0 64.3 0.0

2.2. Direkte Methoden: Prinzip

- Brute force-Methode zur Phasenbestimmung
- Idee/Grundprinzipien
 - die Elektronendichte darf an keinem Ort \vec{x} negativ sein ($\rho_{\vec{x}} > 0$)
 - $\circ \rho_{\vec{x}}$ ist an den Atompositionen konzentriert
- \mapsto Reflexstatistiken
 - Harker-Kasper-Ungleichungen (1948)
 - Sayre-Gleichung (1952) (Triplett-Beziehung, TPR)

 $\vec{s(h_1)s(h_2)s(h_1+h_2)}\approx 1$

mit:

- s: Vorzeichen
- \sim ist mit einer bestimmten Wahrscheinlichkeit (abhängig vom F-Wert)
- Quartett-Beziehungen,
- statistische Aussagen zu Beziehung zwischen den Vorzeichen von mindestens 3 Reflexen
- bei nichtzentrosymmetrischen Strukturen komplizierter

2.2. Direkte Methoden: Anwendung (Programm-Prinzip)

- 1. geschickte Wahl starker Reflexe (große E-Werte)
- 2. zufällige Wahl von Phasen/Vorzeichen für diese Reflexe
- 3. Test auf Konsistenz mit den statistischen Aussagen (TPR usw.) \mapsto z.B. CFOM
- 4. \mapsto schlechte Übereinstimmung \mapsto 2
- 5. \mapsto gute Übereinstimmung \mapsto Berechnung der Phasen aller Reflexe aus den TPRs usw.
- 6. Fouriersynthese der E-Werte \mapsto Elektronendichte-Karten
- 7. Zuordnung der Maxima der Elektronendichte zu Atomen

2.2. Direkte Methoden: Beispiel mit SHELXS-97: Input

TITL SrIn2 CELL 0.71070 5.009 5.009 8.036 90.0 90.0 120.0 ZERR 2 0.0020 0.0020 0.0040 0.0 0.0 0.0 P 63/M M C,NR.194*CB* LATT 1 SYMM -Y, X-Y, Z SYMM X-Y, -Y, .5+Z SFAC SR IN UNIT 2 4 OMIT 2 TREF FMAP 10

HKLF 4 1

2.2. Beispiel: SHELXS-97-Output (Auszug I)

SUMMARY OF PARAMETERS FOR SrIn2

ESEL Emin 1.200 Emax 5.000 DelU 0.005 renorm 0.700 axis 0 OMIT s 2.00 2theta(lim) 180.0 INIT nn 7 nf 16 s+ 0.800 s- 0.200 wr 0.200 PHAN steps 10 cool 0.900 Boltz 0.400 ns 24 mtpr 40 mnqr 10 TREF np 256. nE 24 kapscal 0.900 ntan 2 wn -0.950 FMAP code 10 PLAN npeaks -5 del1 0.500 del2 1.500 MORE verbosity 1 TIME t 9999999.

19 Reflections and 95. unique TPR for phase annealing

19 Phases refined using 95. unique TPR
19 Reflections and 95. unique TPR for R(alpha)
0 Unique negative quartets found, 0 used for phase refinement

ONE-PHASE SEMINVARIANTS

h	k	l	E	P+	Phi
-2	4	0	1.865	1.00	
0	2	2	1.656	0.32	
0	4	2	1.693	0.25	
-2	4	4	1.663	0.86	
Expec	ted	valu	ae of Sig	ma-1 =	0.837

Following phases held constant with unit weights for the initial 4 weighted tangent cycles (before phase annealing):

h	k	l	E	Phase/Comment	
-1	2	0	1.604	random phase	
-2	4	0	1.865	0 sigma - 1 = 0.997	
-2	л	Л	1 663	0 sigma - 1 - 0.859	
	Т	I	T.000	$5 \pm 9 = 0.057$	

All other phases random with initial weights of 0.200 replaced by 0.2*alpha (or 1 if less) during first 4 cycles - unit weights for all phases thereafter 128 Parallel refinements, highest memory = 414 / 5520

Try Ralpha Nqual Sigma-1 M(abs) CFOM Seminvariants 1420309. 0.756 0.000 0.623 0.573 0.756 ++++ 597829. 0.035 0.000 1.000 1.230 0.035 +--+ 1702605. 0.035 0.000 1.000 1.230 0.035 +--+

CFOM Range Frequency 0.000 - 0.020 0 0.020 - 0.040 232 0.040 - 0.060 0 0.140 - 0.160 0 0.600 0.000 24

2.2. Beispiel: SHELXS-97-Output (Auszug III)

E-Fourier for SrIn2 Maximum = 646.79, minimum = -152.38 highest memory used = 8680 / 1741 Heavy-atom assignments: x y z s.o.f. Height IN1 0.6667 0.3333 0.0456 0.1667 646.8 Peak list optimization RE = 0.290 for 1 surviving atoms and 41 E-values Highest memory used = E-Fourier for SrIn2 Maximum = 634.84, minimum = -153.68 highest memory used = 8696 / 1741 Peak list optimization RE = 0.134 for 2 surviving atoms and 41 E-values Highest memory used = E-Fourier for SrIn2 Maximum = 600.29, minimum = -141.74 highest memory used = 8720 / 1741 IN1 0. 0.6667 0.3333 0.0456 0.167 3.14 0 1 3.326 0 2 3.326 120.8 0 3 2.342 49.0 87.9 0 4 2.605 30.2 145.1 77.0 0 5 1.866 74.4 87.4 108.2 1 IN1 2.983 164.6 43.8 121.2 487. 1.0000 0.0000 0.2500 0.083 4.12 0 IN1 3.326 1 0 3 2.514 44.6

2.2. Direkte Methoden: Vor/Nachteile

Vorteile:

- gute Chance f
 ür die meisten Strukturen mit gleichen Streuern (reine Organik, Intermetallisches, Li-Borid-Carbide, ...)
- bei Anwendung von Quartetts recht robust (z.B. bzgl. Symmetriefehlern)
- heute die am meistverwendete Methode bei Standard-Strukturanalysen
- in SHELXS-97 umfassend implementiert (Quartett-Beziehung, Phase-Annealing usw.)

Nachteile:

- Probleme mit zentrosymmetrischen Strukturen (wie auch bei Patterson)
- Lösung von Strukturen
 - mit mehr als 3 Dimensionen (modulierte Strukturen),
 - mit extremer dynamischer Fehlordnung von Schweratomen,
 - ◊ aus Pulvern,
 - ◊ von Quasikristallen

nicht möglich

3. Charge Flipping

3.1. Idee und Algorithmus

- G. Oszlanyi, A. Sütő (2004)
- Ansatz/Idee
 - die Elektronendichte darf an keinem Ort \vec{x} negativ sein ($\rho_{\vec{x}} > 0$)
- Algorithmus
 - 1. Start: statistische Phasenzuordnung $\phi_{\vec{h}}^n$ zu allen beobachteten Reflexen \vec{h} (alle anderen Reflexe: $\phi^n=0$; $\phi_{\vec{O}}^n$ kritisch!)
 - 2. $\rho_{\vec{x}}^{n}$ mittels inverser Fourier-Transformation (Gl. 4) berechnen (in einem Volumen-Grid von N1xN2xN3 Pixeln)
 - 3. negative Dichten $\rho_{\vec{x}}^n$ durch positive gleicher Höhe austauschen \mapsto <u>Charge-Flipping</u> \mapsto modifizierte Dichte ($\rho_{\vec{x}}^m$)
 - 4. aus dieser modifizierten Dichte $\rho_{\vec{x}}^m$ mittels Fourier-Transformation (Gl. 3) temporäre Strukturfaktoren $G_{\vec{h}}^m$ berechnen
 - 5. durch Kombination der experimentellen Amplituden mit den Phasen von $G_{\vec{h}}^{m}$ die neuen Strukturfaktoren Fⁿ⁺¹ für den nächsten Zyklus ermitteln (alle nicht gemessenen Reflexe: $\phi = 0$)
 - 6. GOTO 2 (Wiederholung der Schritte 2-5 bis zur Konvergenz)

3.1. Charge-Flipping: Flowchart



3.2. Programme

Superflip:

- Superflip = Charge Flipping in Superspace
- Autoren: L. Palatinus, Lausanne
- Fortran 90 (+ FFTW-Library)
- Input:
 - ◇ Datei *.inflip (Instructions inkl. hkl-F-Daten) \Downarrow

Flipper:

- Unterprogramm von PLATON
- Bunt!
- Autor: T. Spek ?
- nur 3D ?
- ????



3.3. Superflip: Input

```
title SrIn2
perform CF
######## Keywords influencing the form of the files
outputfile srin2.m80
outputformat m80
expandedlog yes
coverage yes
######## Basic crystallographic information
dimension 3
voxel 80 80 80
cell 5.009 5.009 8.036 90.00 90.0 120.00
symmetry
         x2 x3
     x1
endsymmetry
####### Keywords influencing the algorithm
delta AUTO
weakratio 0.000
biso 0.000
randomseed AUTO
searchsymmetry average
######## List of reflections
reflectionlist complete
dataformat intensity
fbeqin
  -1
      0 0 127.67 7.74
                             0
```

3.3. Superflip: Output-Dateien

Dateien:

- Protokoll-Datei *.sflog
- Elektronendichtekarte, wahlweise im
 - *.m81-Format (binär, Dichten für JANA)
 - \diamond *.m80-Format (ascii, F_o + ϕ für JANA oder DRAWxtl)

0	0	0	0	1 0.13078E+04	0.13078E+04	0.13078E+04 0.13078E+04 0.00000E+00
0	0	0	2	1 0.29614E+02	0.29614E+02	0.29614E+02-0.29614E+02-0.19621E-05
0	0	0	4	1 0.59498E+01	0.59498E+01	0.59498E+01-0.59498E+01-0.78840E-06
-2	0	0	-4	1 0.64498E+01	0.64498E+01	0.64498E+01 0.64498E+01-0.92795E-06
-2	0	0	-2	1 0.79498E+01	0.79498E+01	0.79498E+01-0.79498E+01 0.61705E-06
-2	0	0	0	1 0.85922E+02	0.85922E+02	0.85922E+02 0.85922E+02-0.97639E-06
-2	0	0	2	1 0.53184E+02	0.53184E+02	0.53184E+02-0.53184E+02-0.29193E-05
-2	0	0	4	1 0.18844E+02	0.18844E+02	0.18844E+02 0.18844E+02 0.22829E-05

Demos:

- SrIn₂ ohne Symmetrie, Ausgabe als *.m81-Datei \mapsto JANA
- SrIn₂ mit Symmetrie, Ausgabe als *.m80-Datei \mapsto DRAWxtl
- Lösung einer modulierten (4-dim.) Struktur, Ausgabe als *.m81-Datei → JANA

4. Zusammenfassung

Zusammenfassung

- Lösung von 3D-Strukturen
 - Charge-Flipping (Programm: Superflip) als Alternative/Ergänzung zu
 - Patterson-Methode (z.B. SHELXS-97)
 - Direkte Methoden (z.B. SHELXS-97)
- Lösung modulierte Strukturen/Superspace
 - Alternative zu viel Handarbeit

Literatur

- Giocovazzo (ed.): Fundamentals in Crystallography, Oxford.
- W. Massa: Einführung in die Kristallstrukturanalyse, Teubner.
- L. Palatinus, Acta Crystallogr. A60, 604-610 (2004).
- G. Oszlanyi, A. Sütő, Acta Crystallogr. A60, 134-141 (2004).
- G. Oszlanyi, A. Sütő, Acta Crystallogr. A61, 147-152 (2005).
- L. Palatinus: SUPERFLIP User Manual (2006).