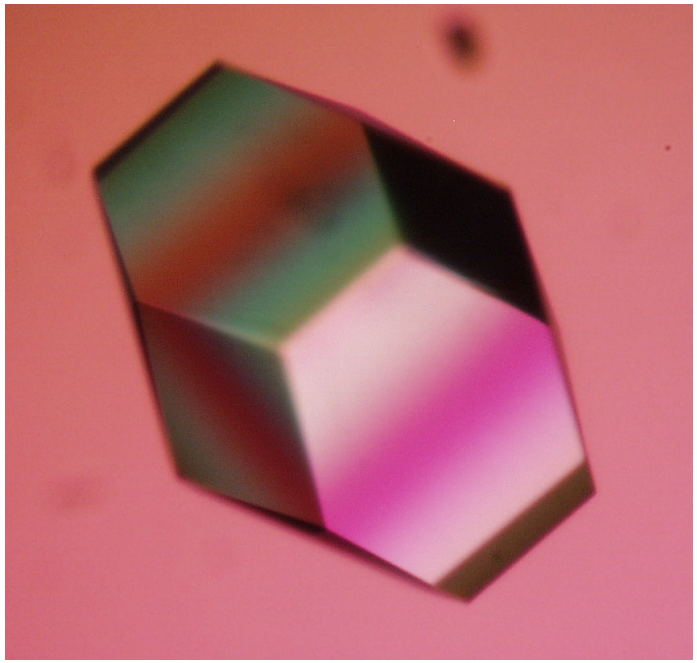
Fundamentals of experimental phasing and SAD phasing in *Phaser*

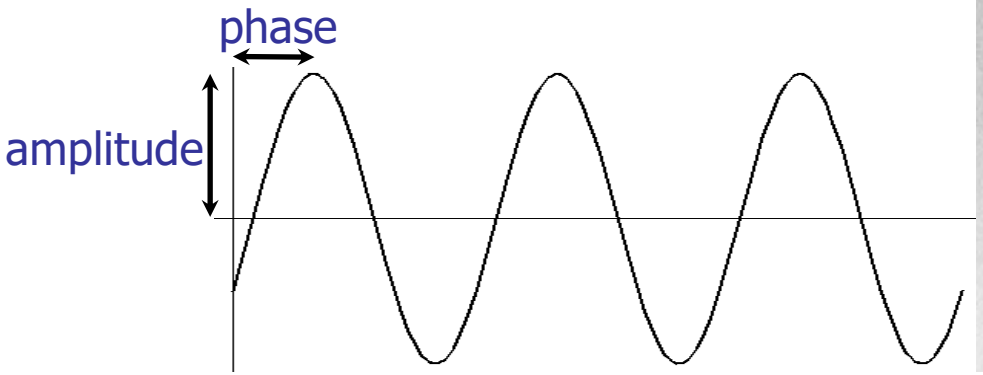
UNIVERSITY OF CAMBRIDGE

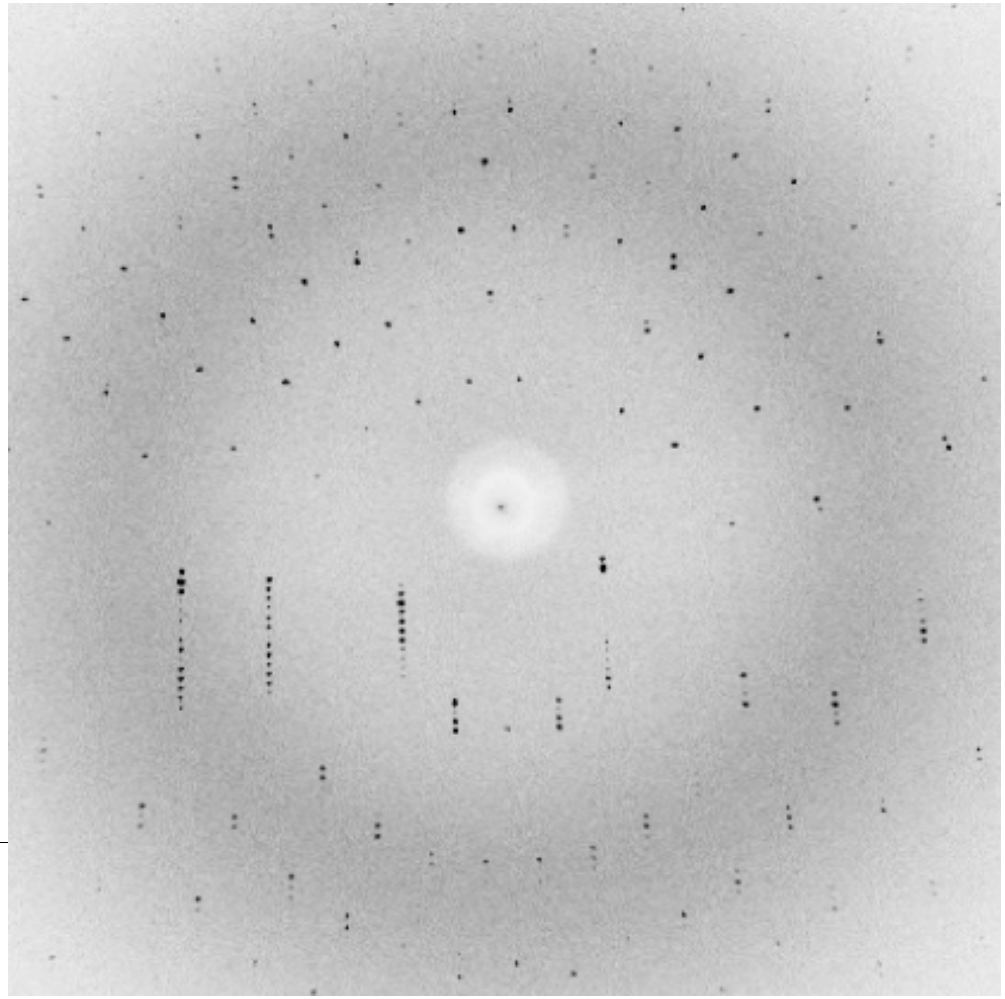
Randy J Read
Department of Haematology

CIMR Cambridge Institute for Medical Research

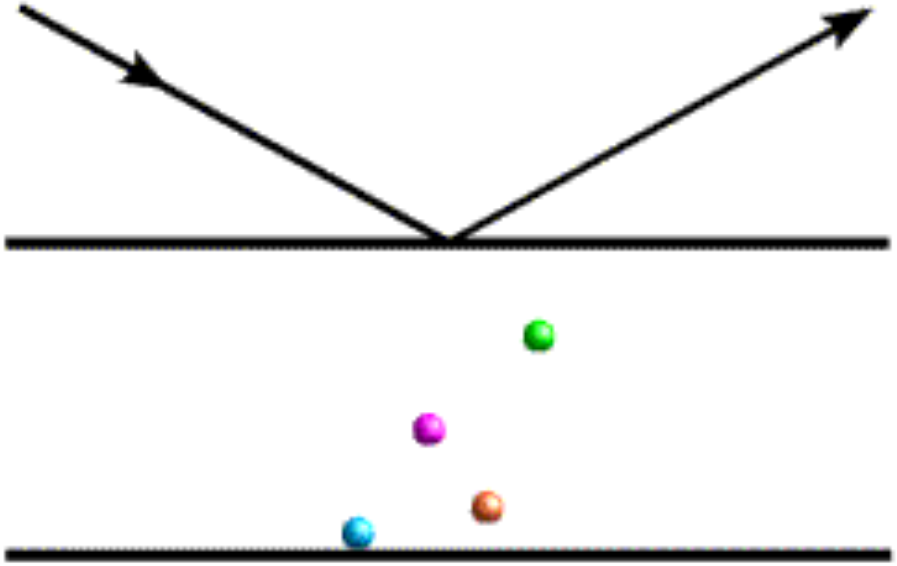
X-ray diffraction

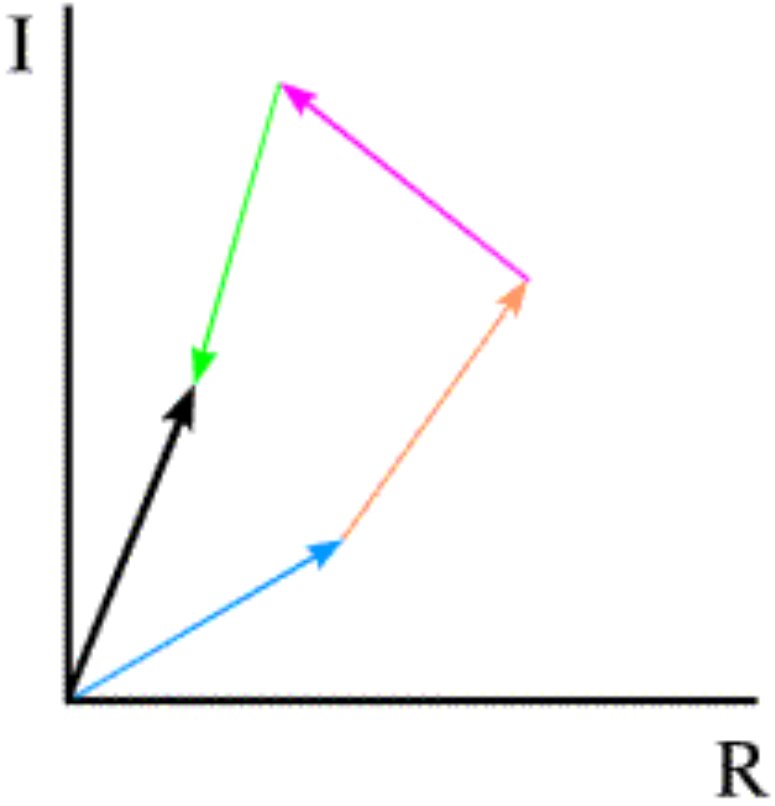




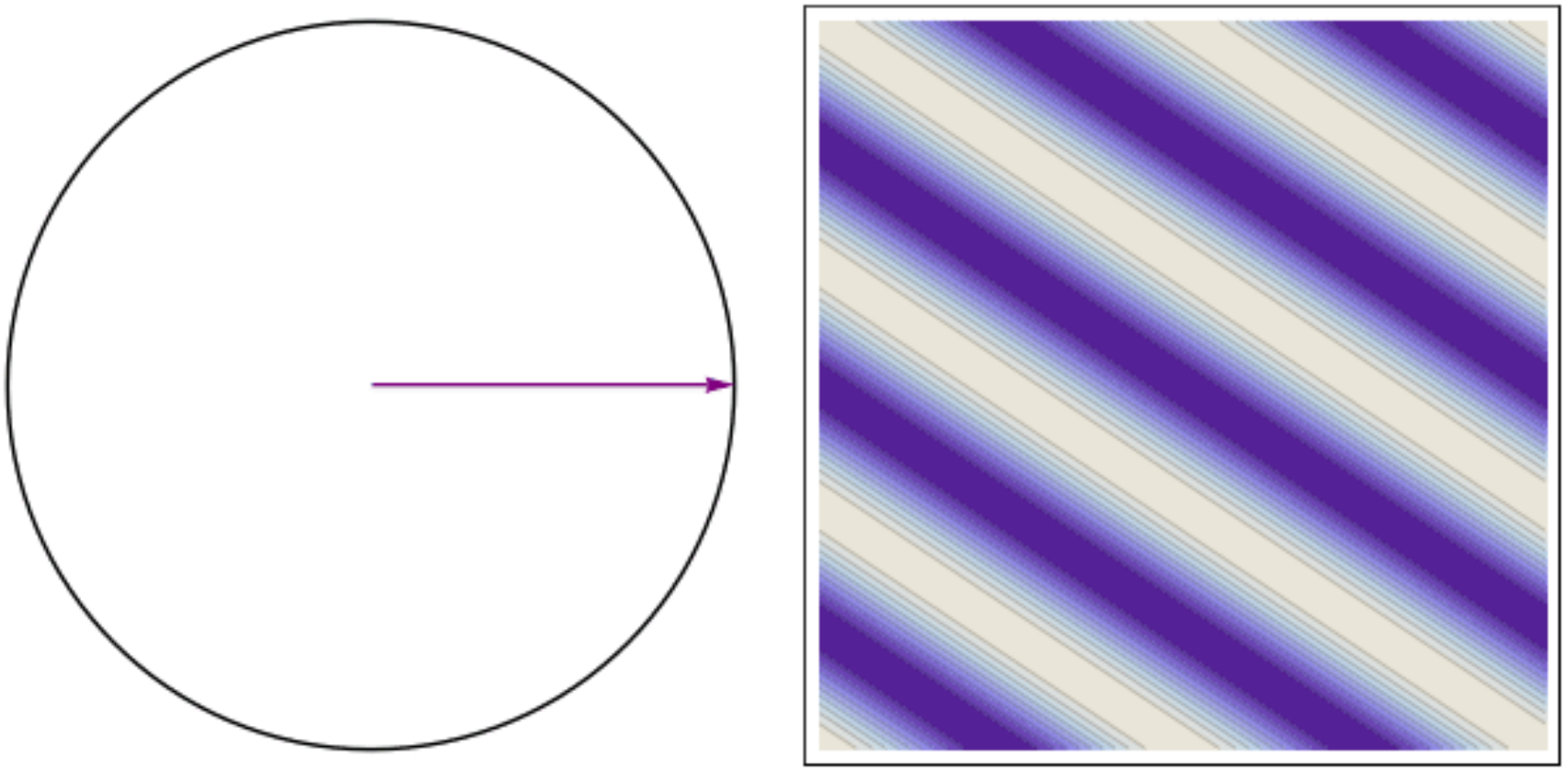


Building up a structure factor





Effect of phase on electron density



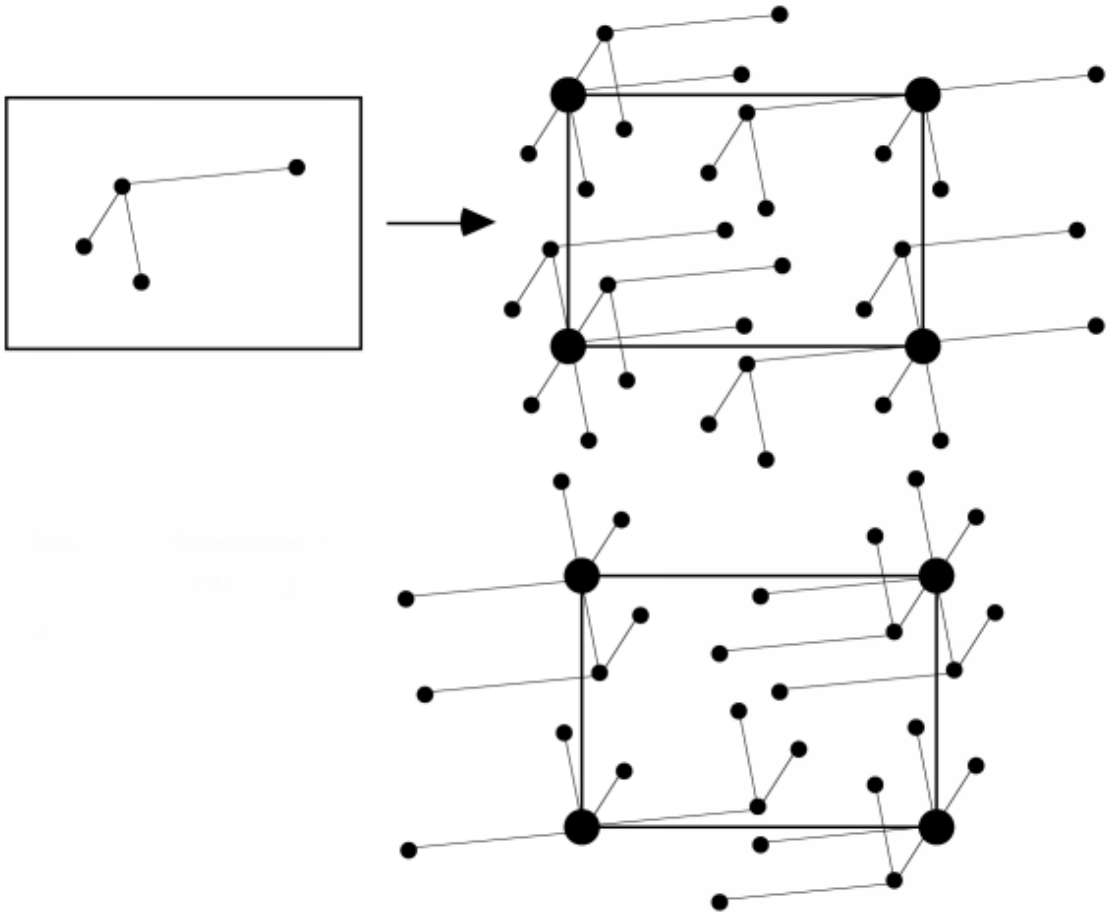
Phasing: *Ab initio* approaches

- Patterson methods
  - Patterson map shows vectors between atoms
- Direct methods
  - based on statistical relationships between phases
- Both methods work more poorly as the number of atoms increases
  - limit of around 1000 atoms for direct methods, even with atomic resolution data

The Patterson function

- Fourier transform of intensities
- Map of vectors between atoms

  = superposition of images
- $n^2 - n$ off-origin peaks



- Mathematical background: www-structmed.cimr.cam.ac.uk/Course/Convolution/convolution.html

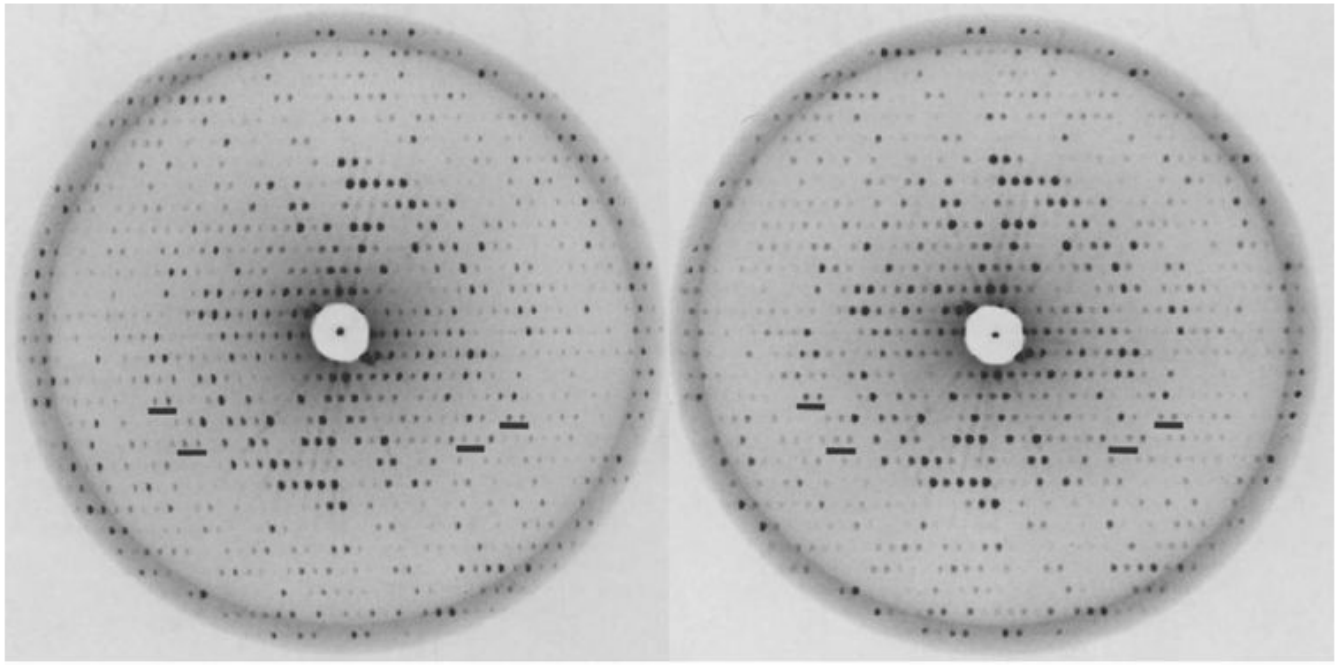
Practical phasing for large molecules

- Need method that works with crystals that diffract to worse than atomic resolution
- Need method that doesn't require finding thousands of atoms at once!
  - divide into problem of finding a few special ("marker") atoms and then using those to work out phase information
    - details depend on nature of the special atoms

Experimental phasing

- Isomorphous replacement (SIR or MIR)
  - "isomorphous" = "same shape"
  - perturb diffraction pattern by adding a few *heavy* atoms
- Anomalous diffraction (SAD or MAD)
  - diffraction pattern perturbed between Bijvoet pairs by a small number of *anomalous scatterers*
- Bootstrap structure determination
  - explain differences in diffraction patterns with a few atoms ("substructure" of more complex structure)
  - solve small substructure by *ab initio* methods

Isomorphous derivative



a

Native
F_{nat}

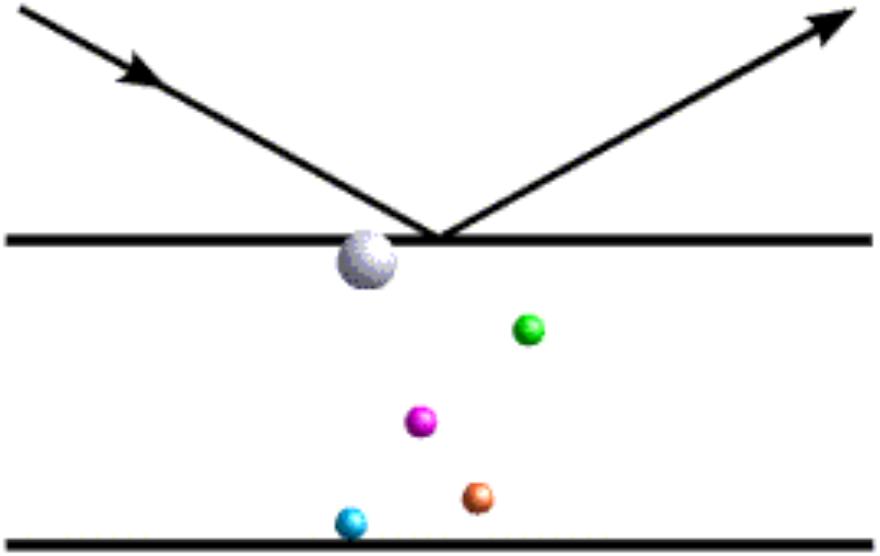
b

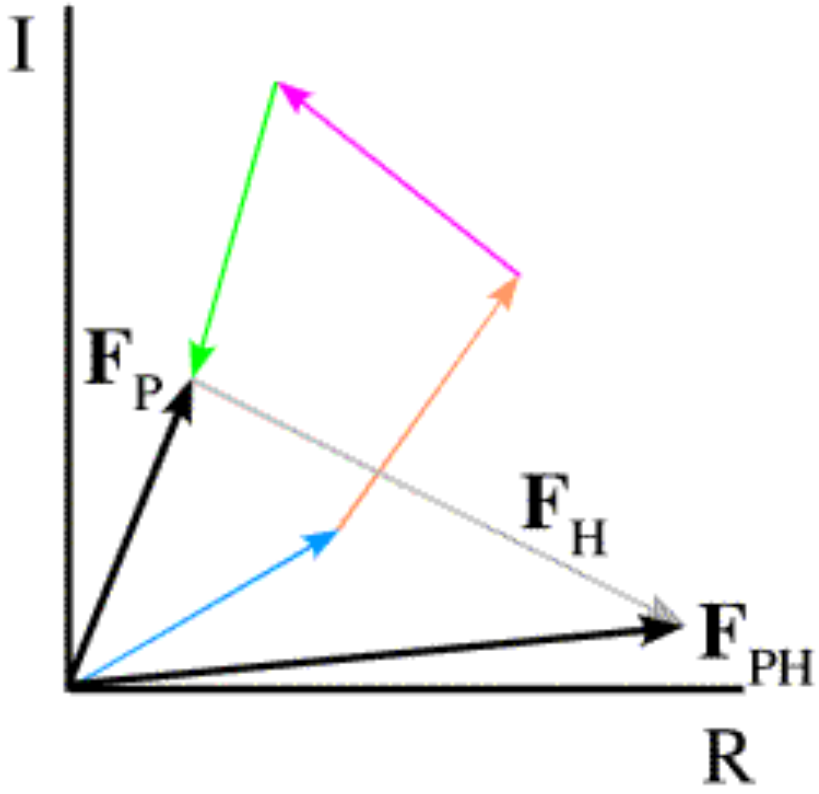
Heavy atom derivative
F_{deriv}

from Gale Rhodes, "Crystallography made crystal clear"

Perturbing diffraction by adding a heavy atom

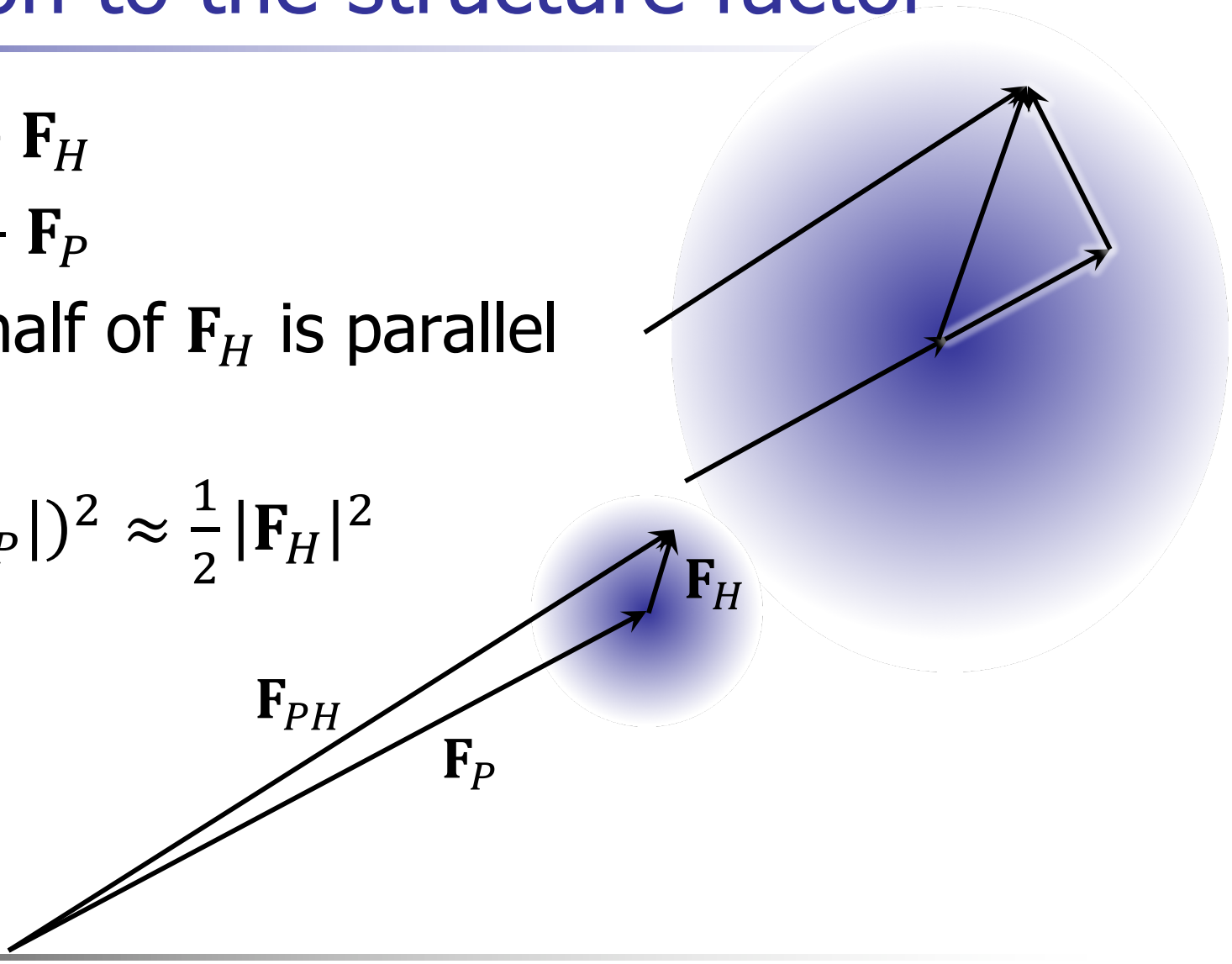
- Heavy-atom contribution is represented by vector addition to protein structure factor





Deducing the size of the heavy atom contribution to the structure factor

- $\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$
- $\mathbf{F}_H = \mathbf{F}_{PH} - \mathbf{F}_P$
- Typically, half of $\mathbf{F}_H$ is parallel to $\mathbf{F}_P$
- $(|\mathbf{F}_{PH}| - |\mathbf{F}_P|)^2 \approx \frac{1}{2}|\mathbf{F}_H|^2$

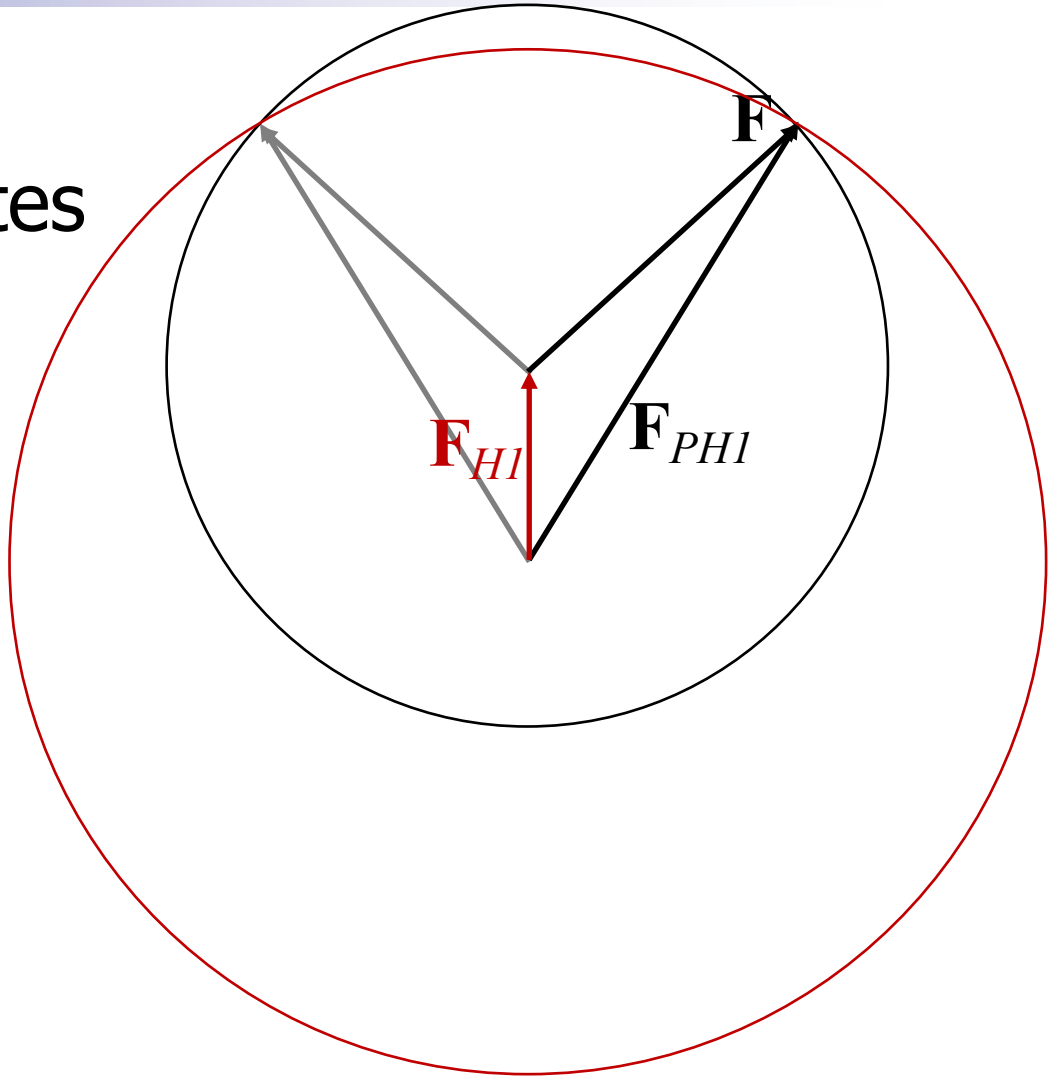


Finding the heavy-atom substructure

- Use the estimate of $|\mathbf{F}_H|$ from the isomorphous difference to solve the substructure by Patterson methods or direct methods
  - relatively few heavy atoms
  - heavy atoms are generally far apart compared to resolution of data

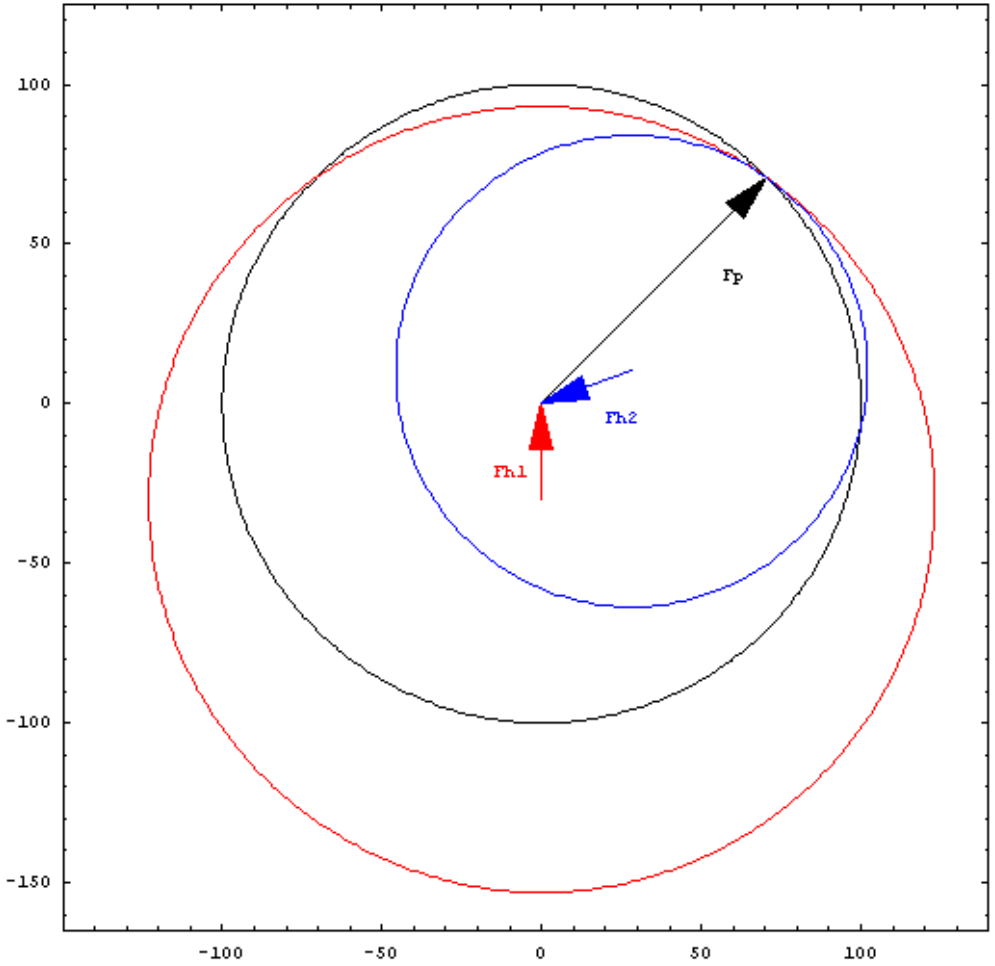
Deducing phase information from isomorphous replacement

- Compute $\mathbf{F}_H$ from substructure coordinates
- Deduce possible phases from Harker construction



The Harker construction for multiple isomorphous replacement (MIR)

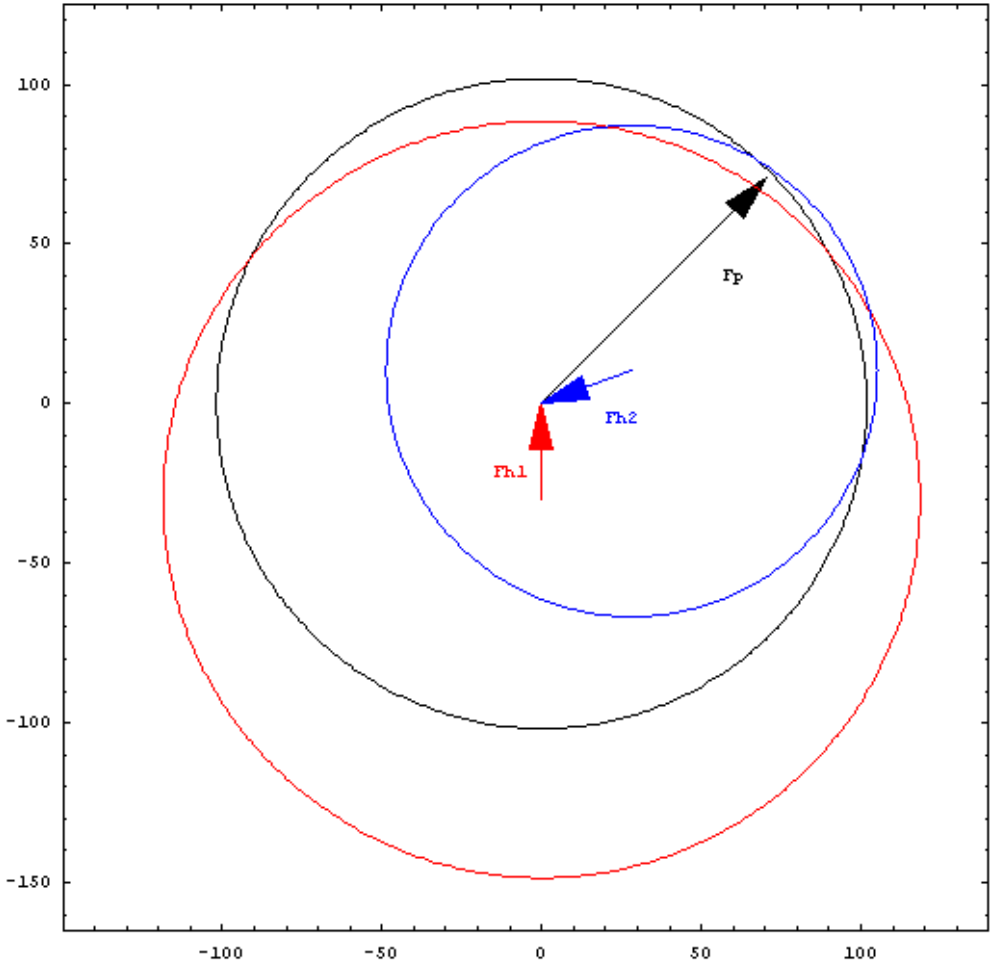
- Two derivatives can give unambiguous phase
  - only in the absence of any errors!
    - measurement
    - heavy-atom model
    - lack-of-isomorphism
  - maximum likelihood!



Lack of isomorphism

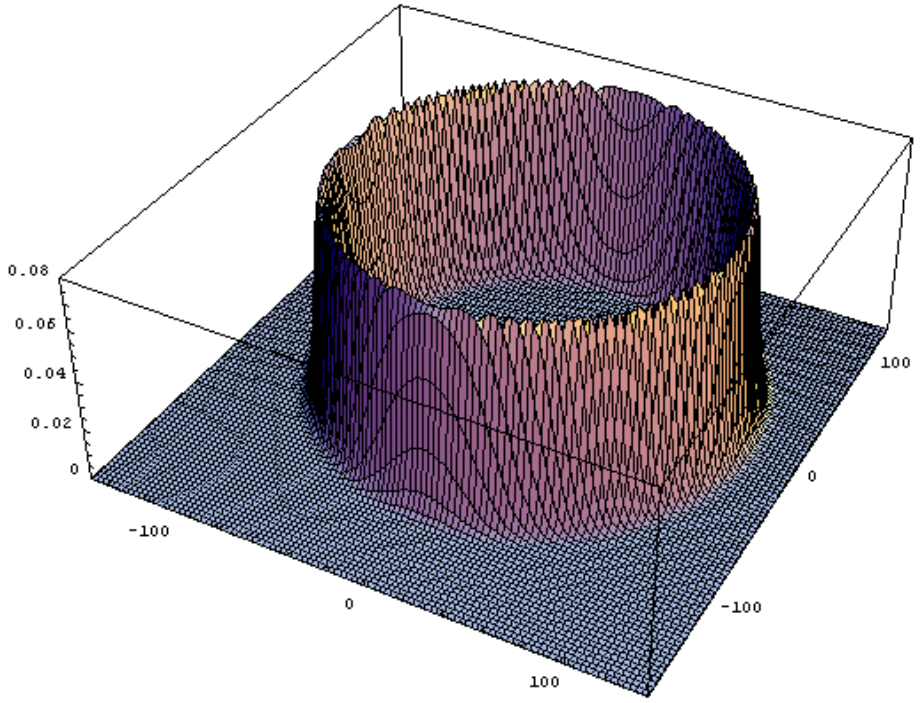
- Heavy-atom reagents can perturb crystal
  - change crystal packing
  - change protein conformation
- Commonly accompanied by change in cell dimensions

The Harker construction with errors



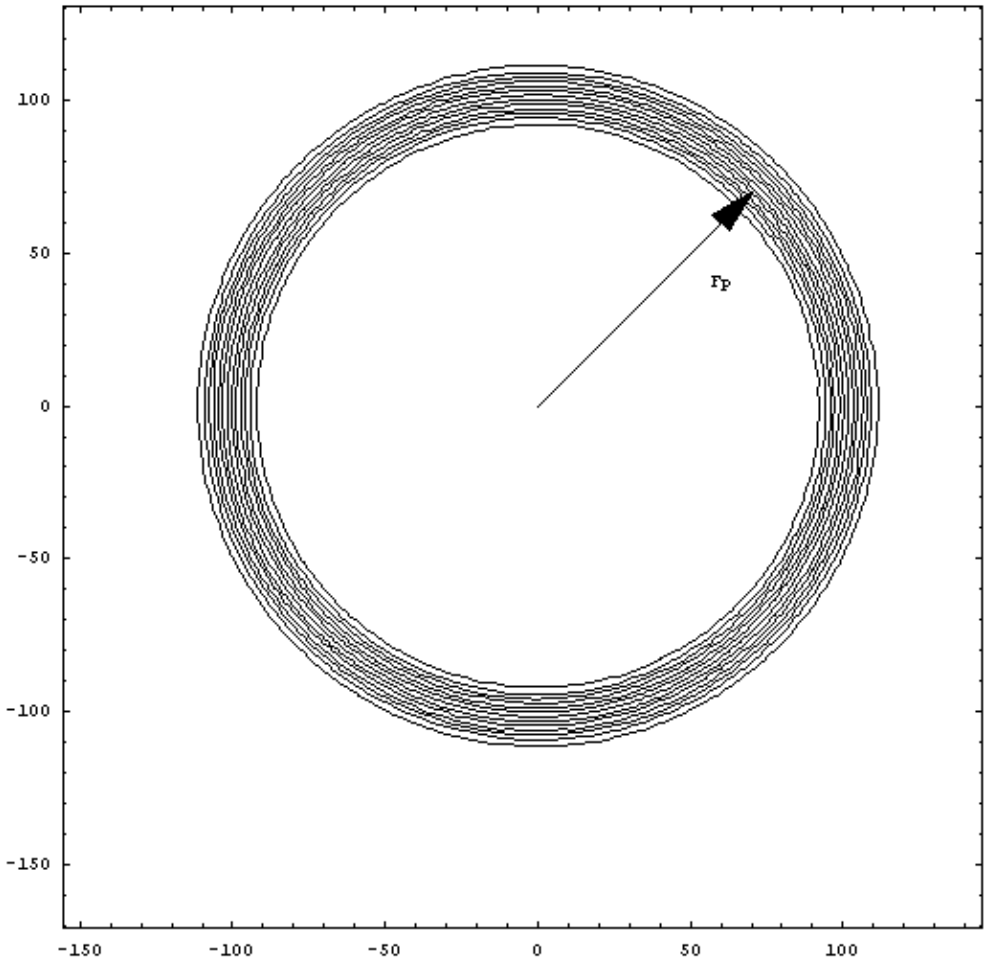
Probabilistic Harker construction

- Account for measurement errors in native structure factor



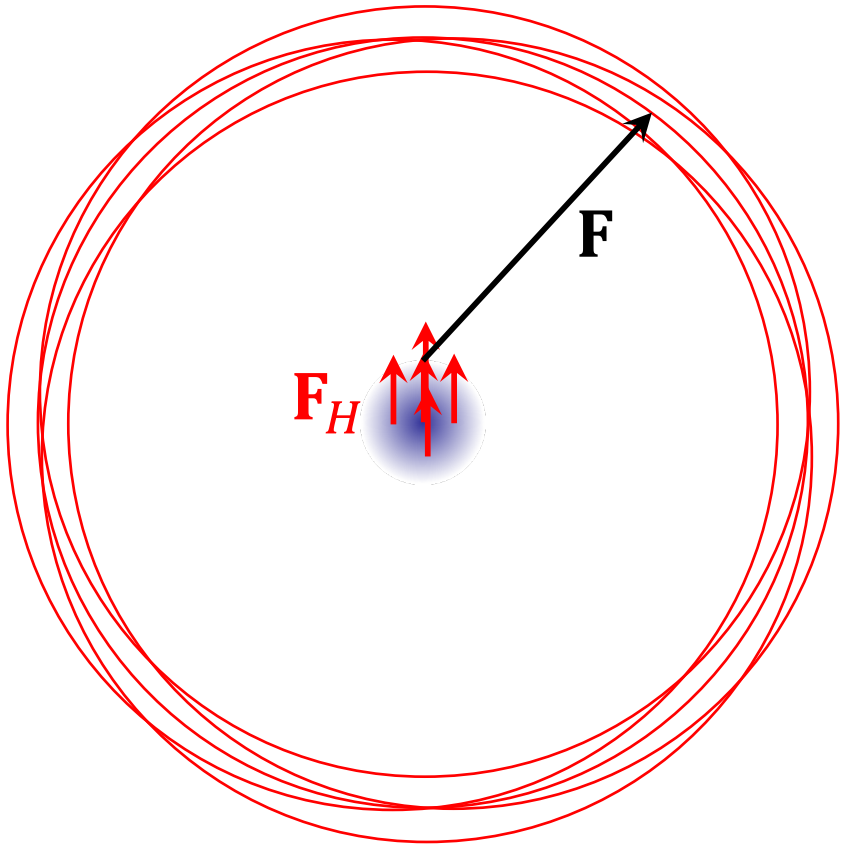
Probabilistic Harker construction

- Account for measurement errors in native structure factor



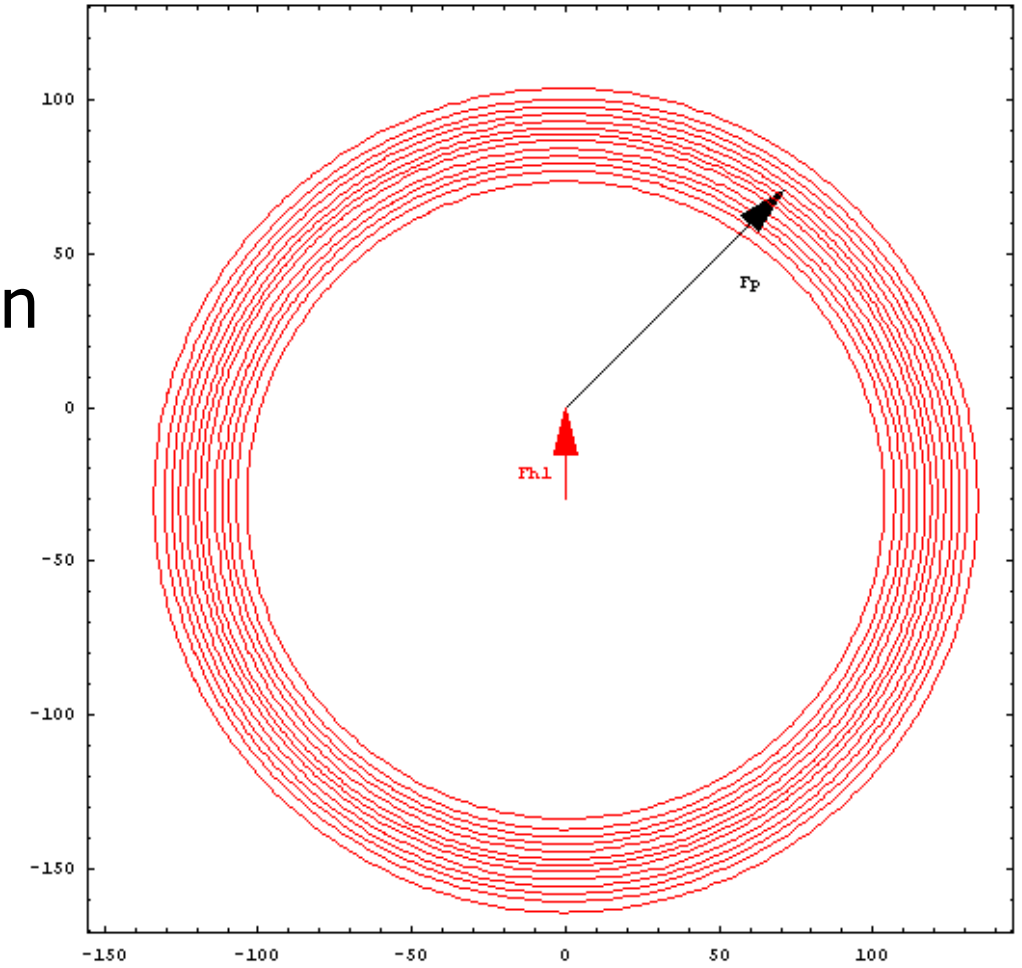
Probabilistic Harker construction

- Account for errors in heavy-atom model

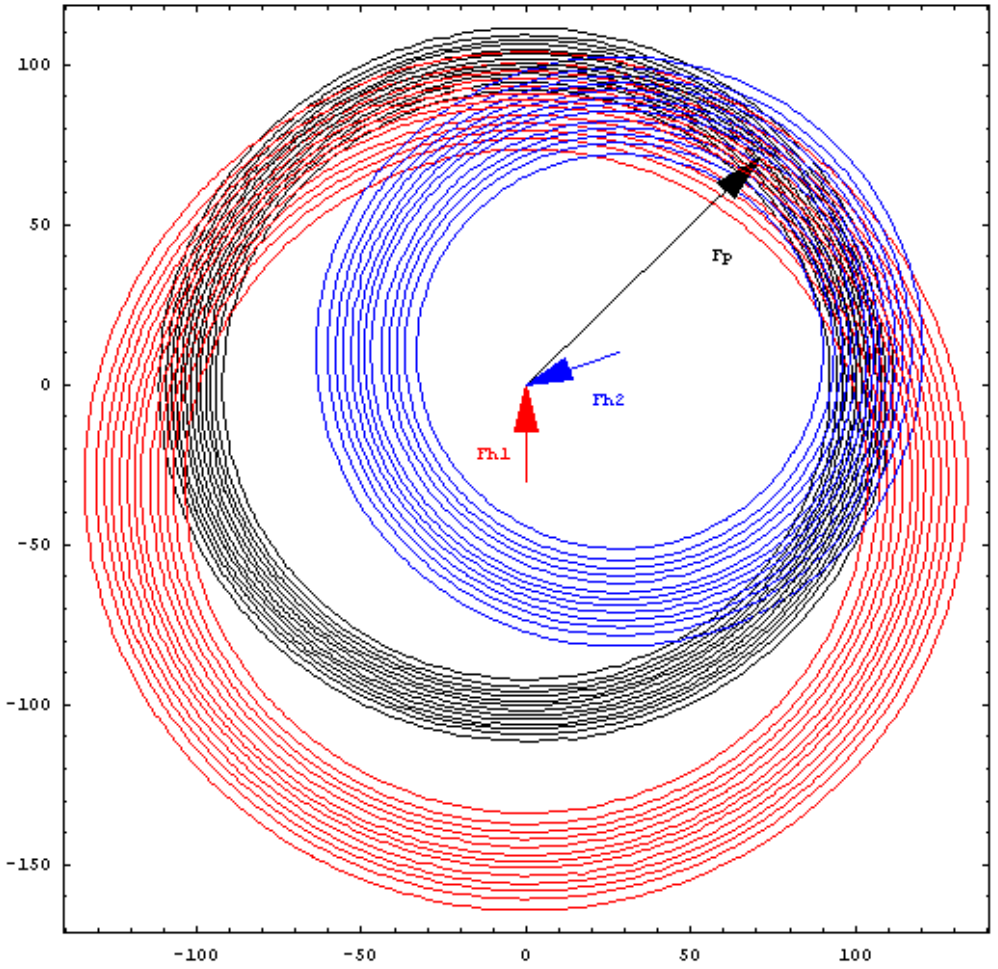


Probabilistic Harker construction

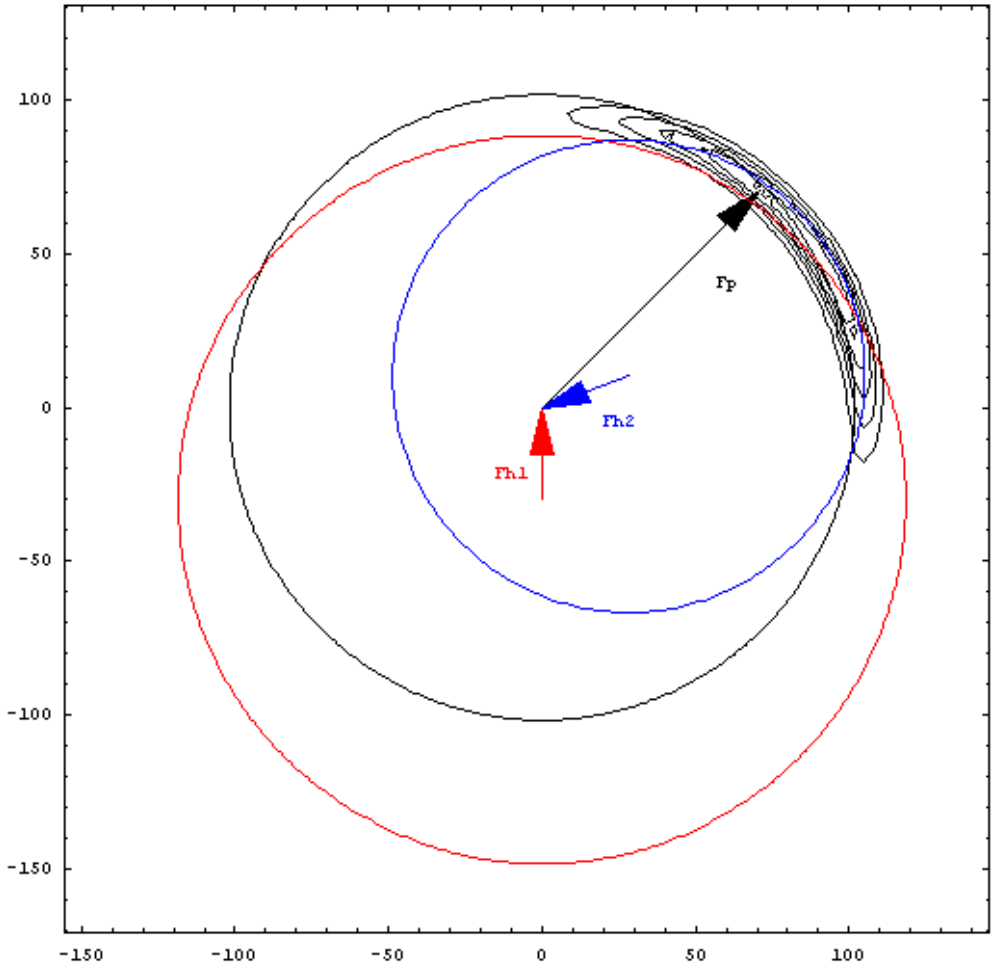
- Account for combined effect of measurement errors, lack of isomorphism and errors in heavy-atom model



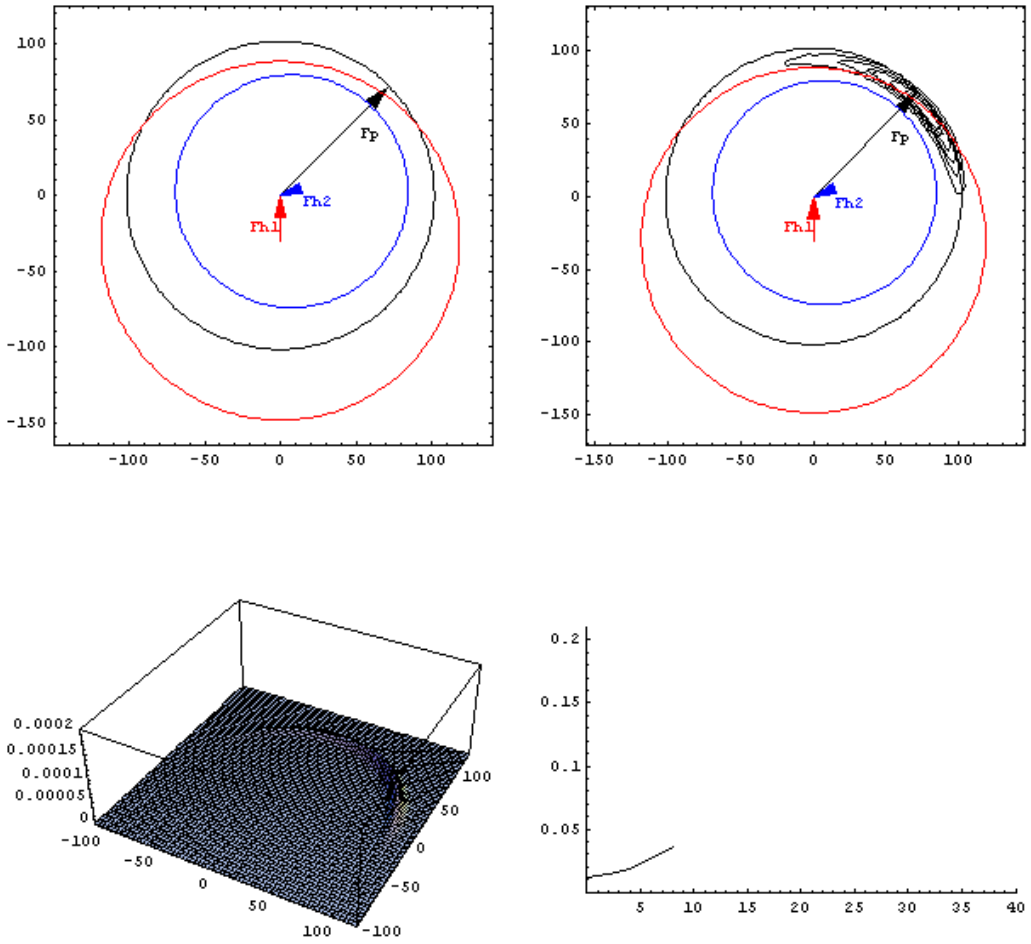
Probabilistic Harker construction



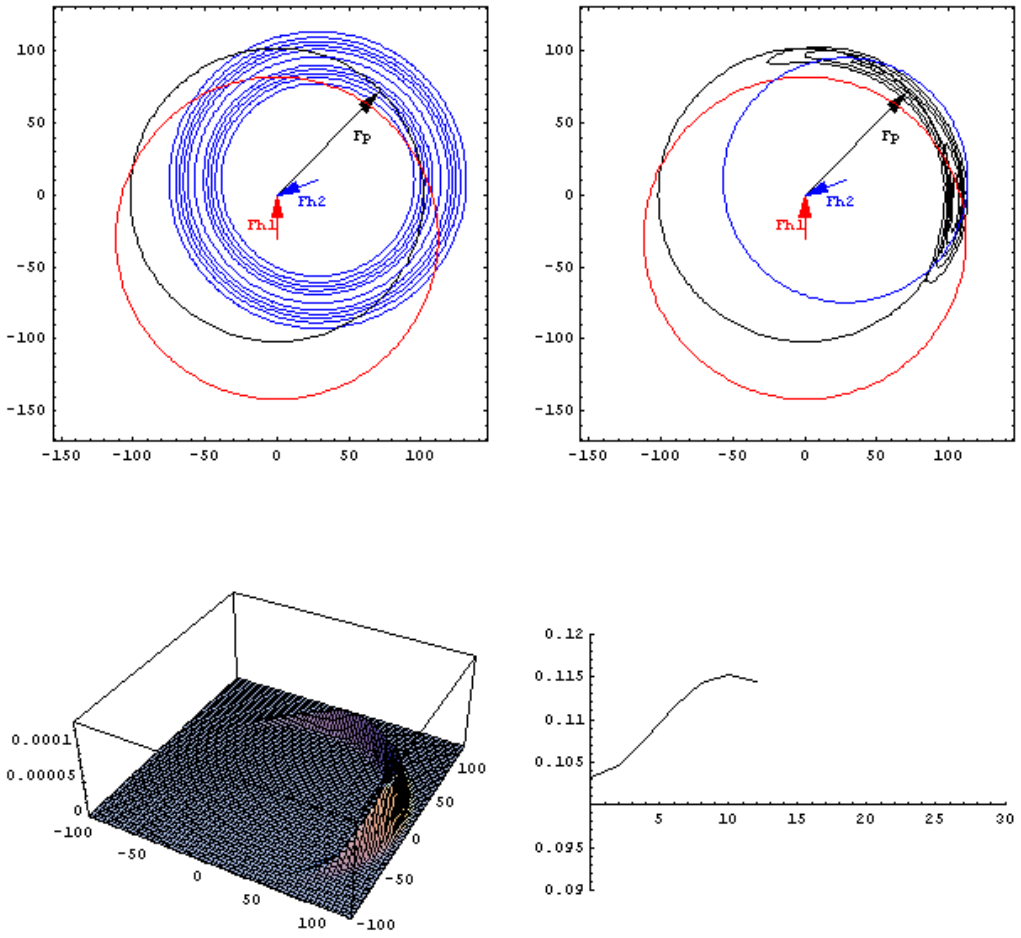
Probabilistic Harker construction



Likelihood as function of heavy atom model



Likelihood as function of lack-of-isomorphism error

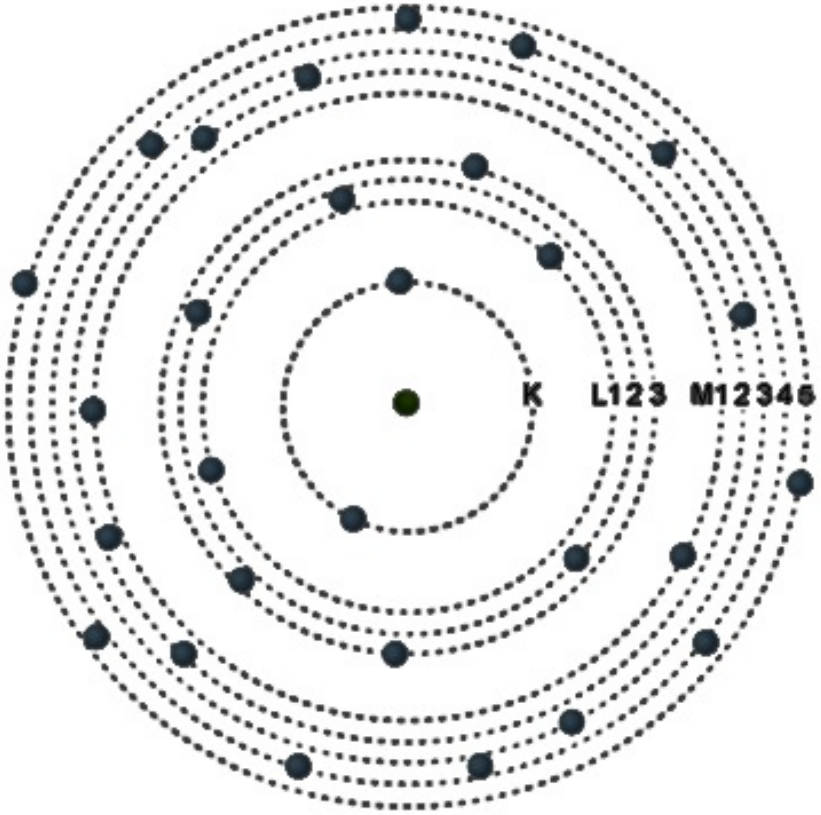


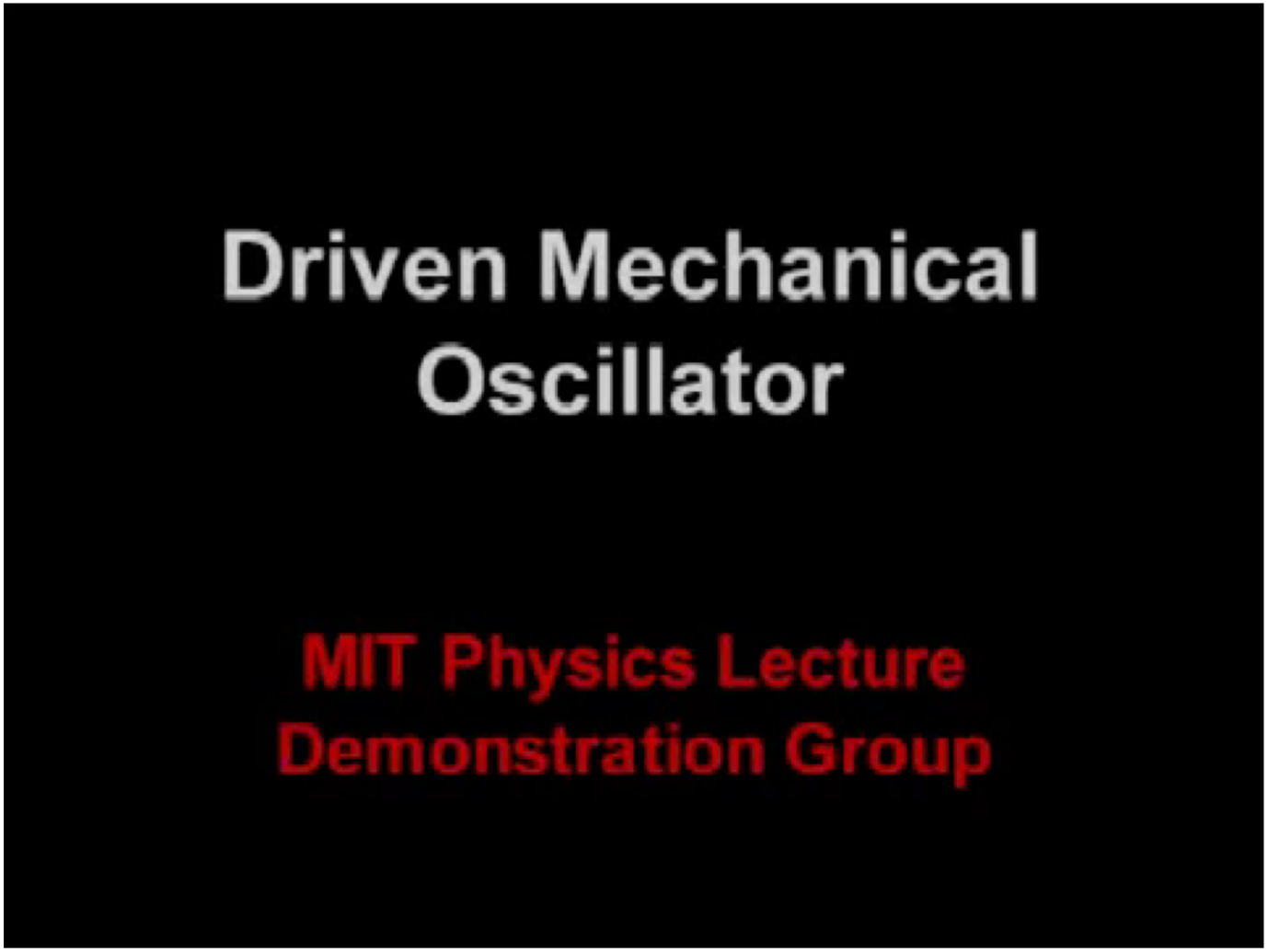
SAD: single-wavelength anomalous diffraction

- Most popular way to solve structures by experimental phasing (over 70% and rising)
- Can be done with intrinsic S and CuKα X-rays
- SAD phasing theory is very good
- Easy to automate
- Can be very fast
  - Can be done from single dataset
    - no lack of isomorphism
- May need multiple crystals
  - and careful data processing

Anomalous scattering

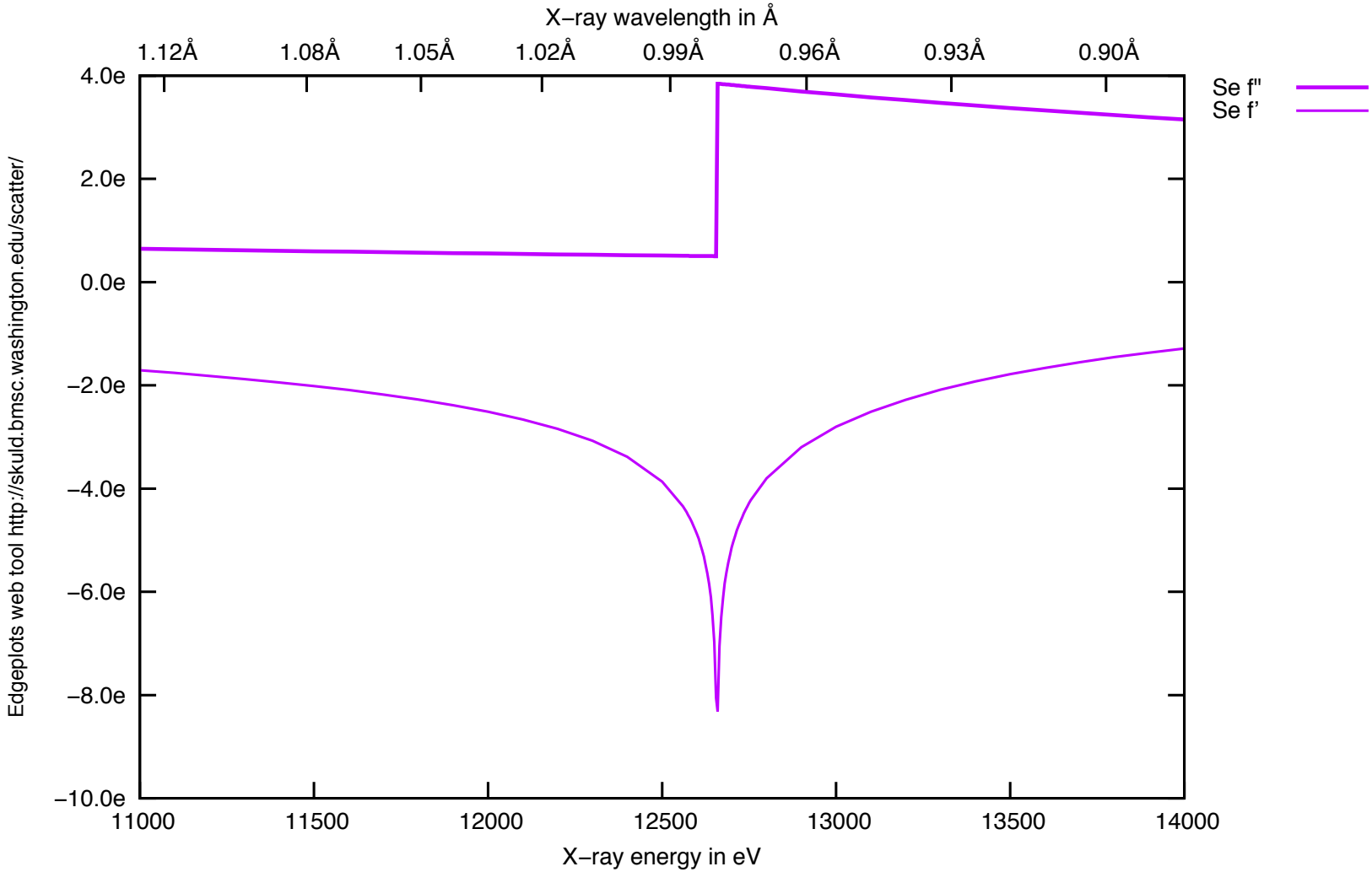
- Anomalous scattering is due to the electrons being tightly bound (particularly in K & L shells)
- In classical terms, the electrons scatter as though they have resonant frequencies





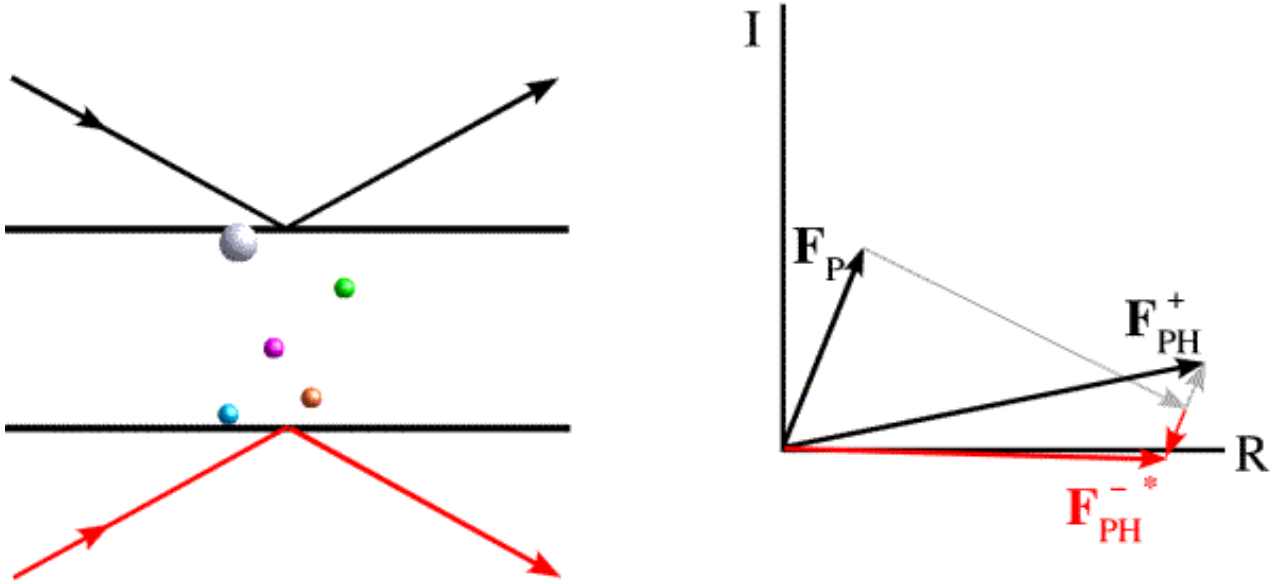
https://www.youtube.com/watch?v=aZNnwQ8HJHU

Anomalous scattering as a function of wavelength: $f = f_0 + f' + i\,f''$



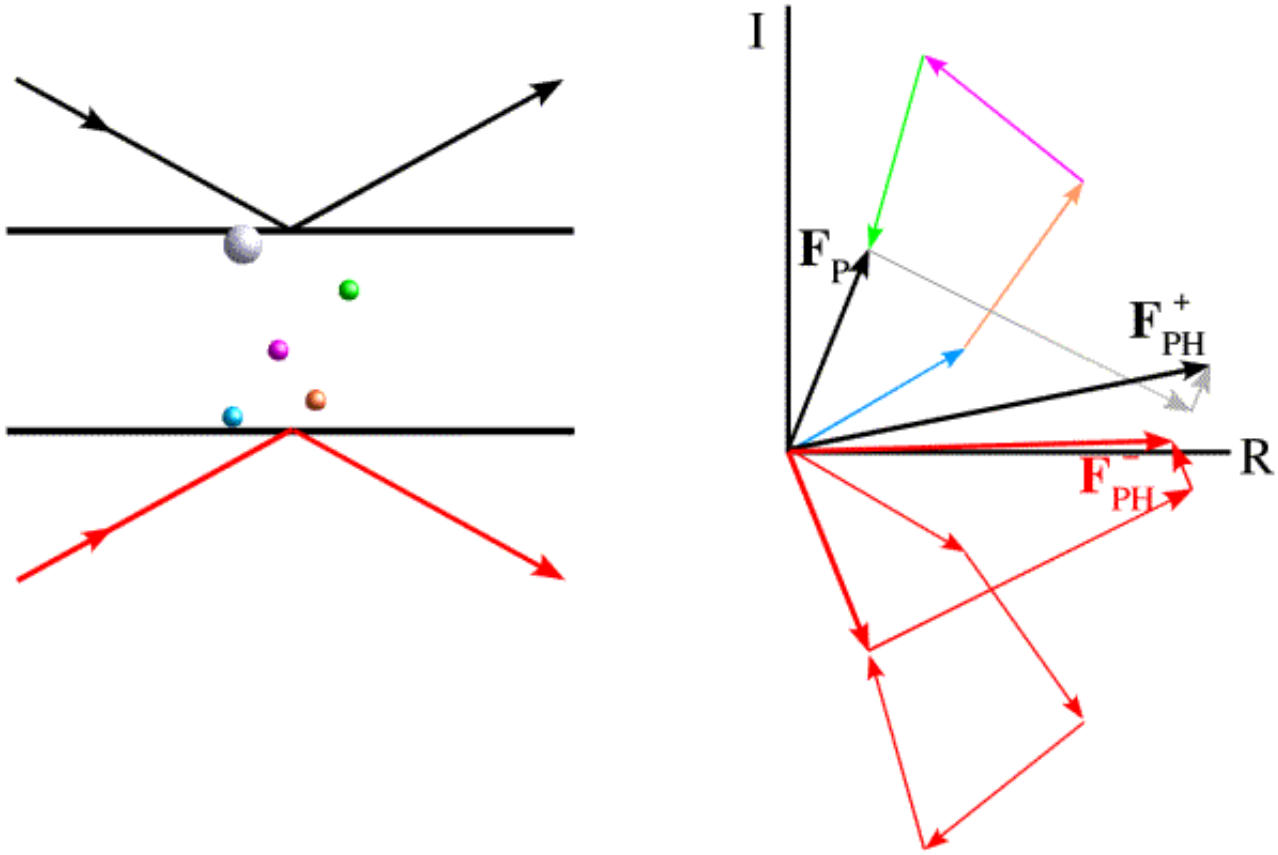
Diffraction with anomalous scatterers

- SAD: single-wavelength anomalous diffraction

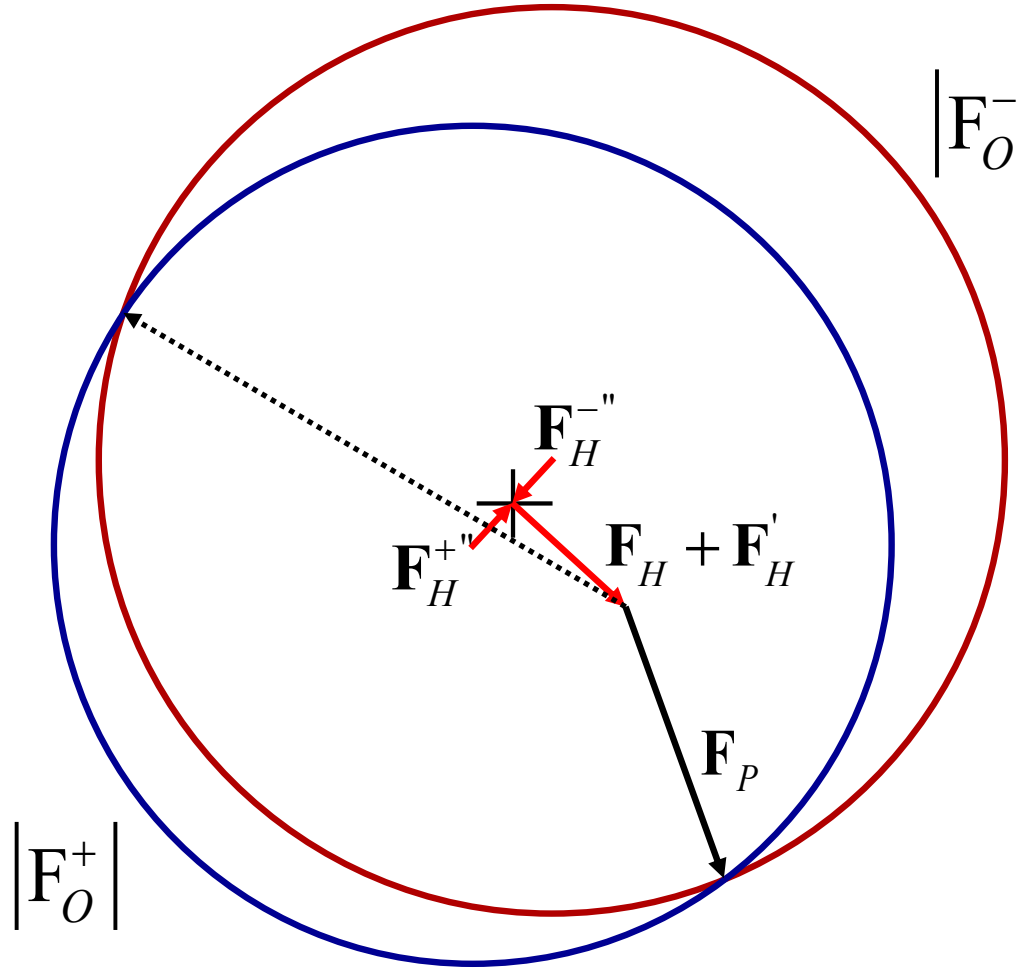


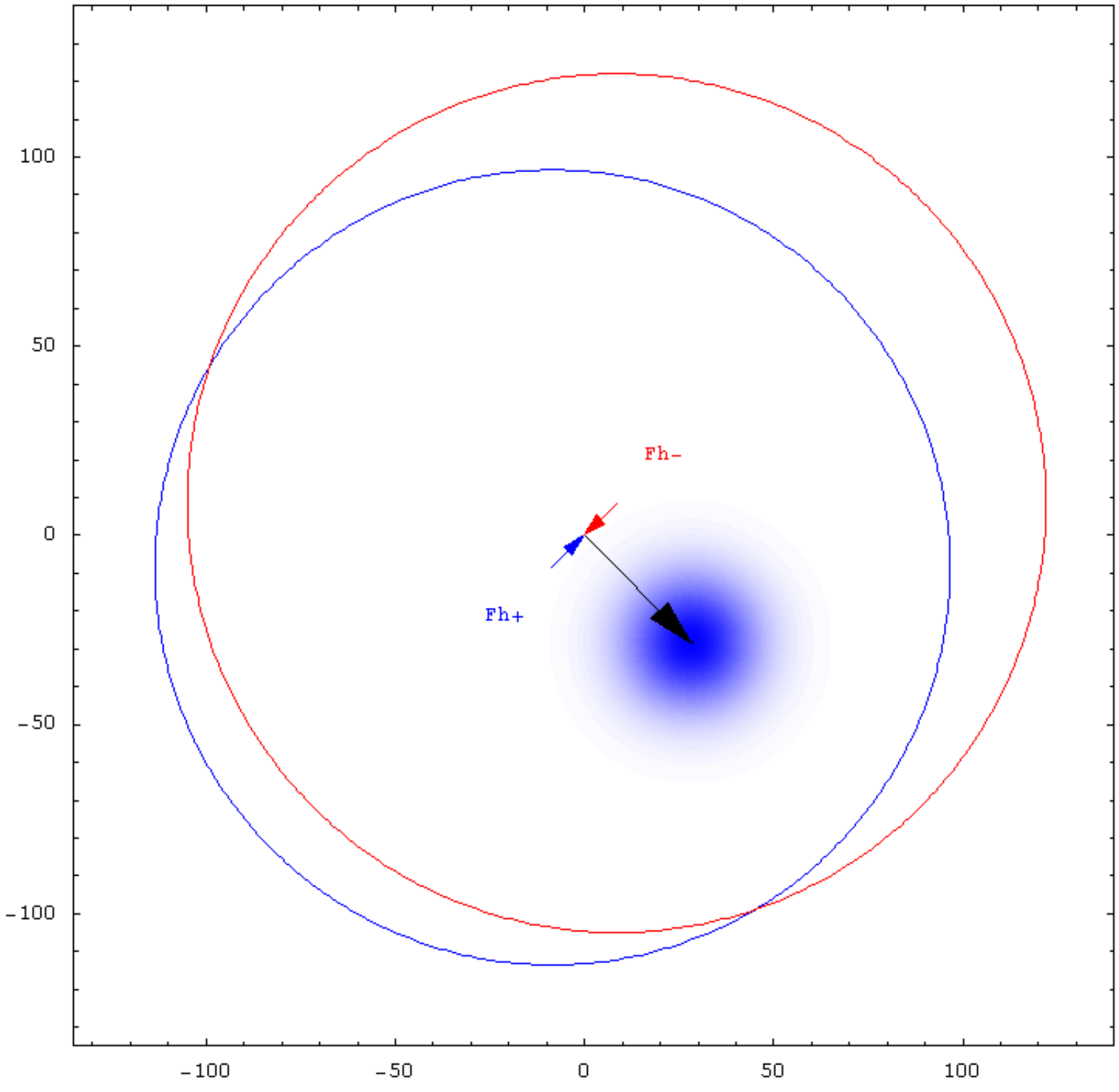
Diffraction with anomalous scatterers

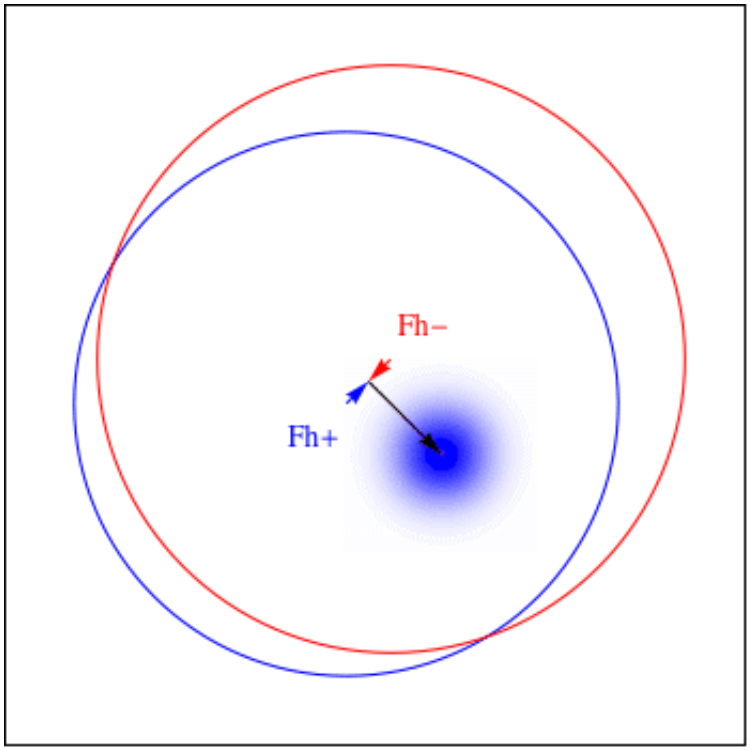
- SAD: single-wavelength anomalous diffraction

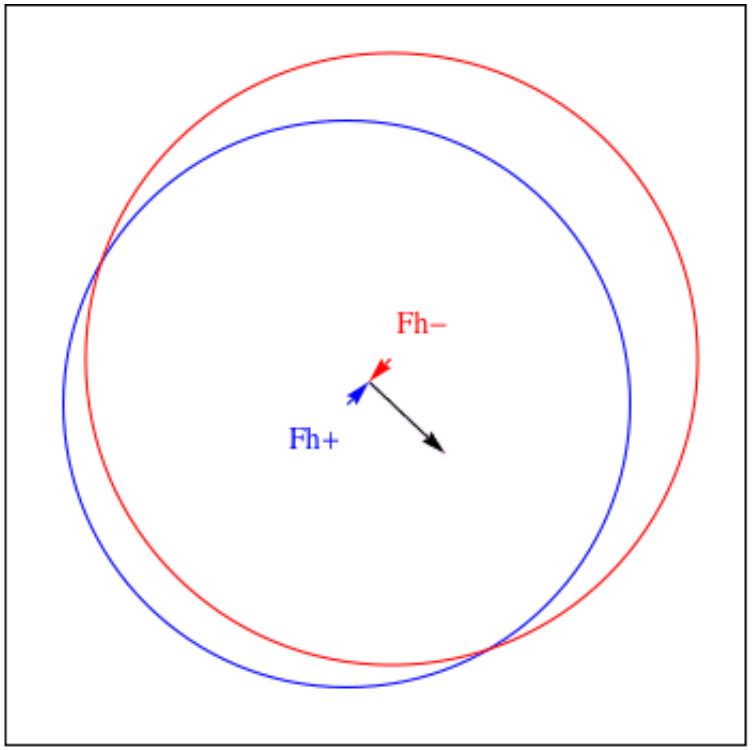


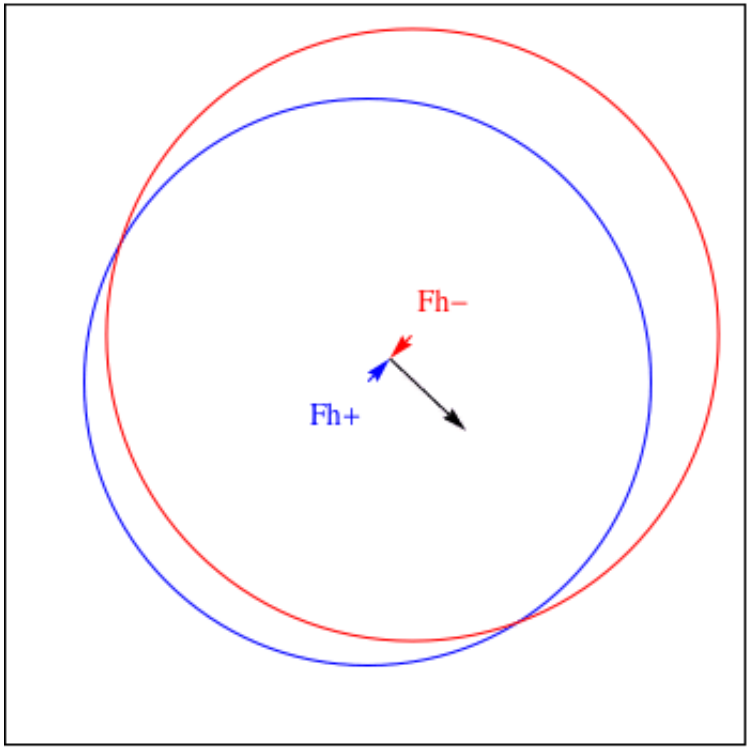
Harker construction for SAD phasing

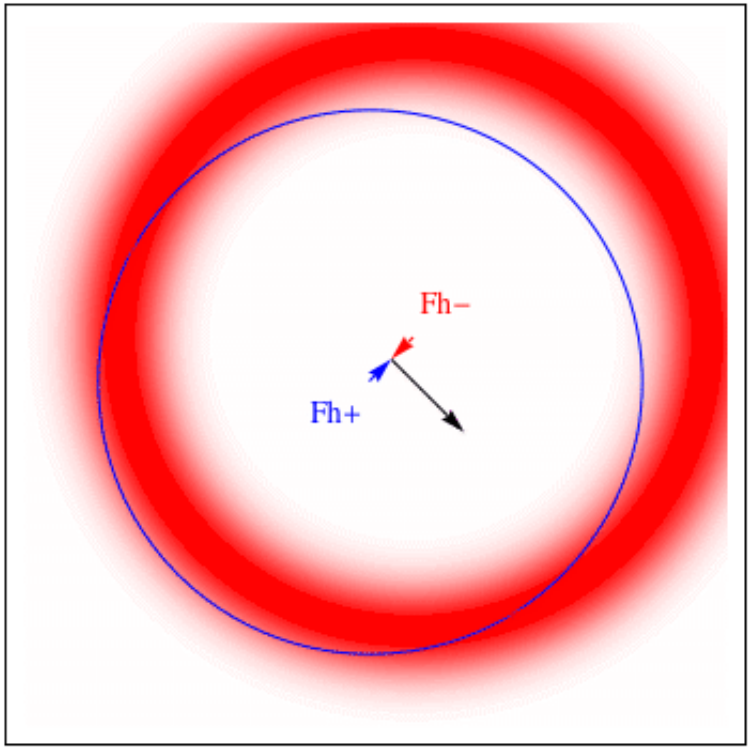



Fh-
Fh+
100
50
0
-50
-100
-100
-50
0
50
100


Fh−
Fh+

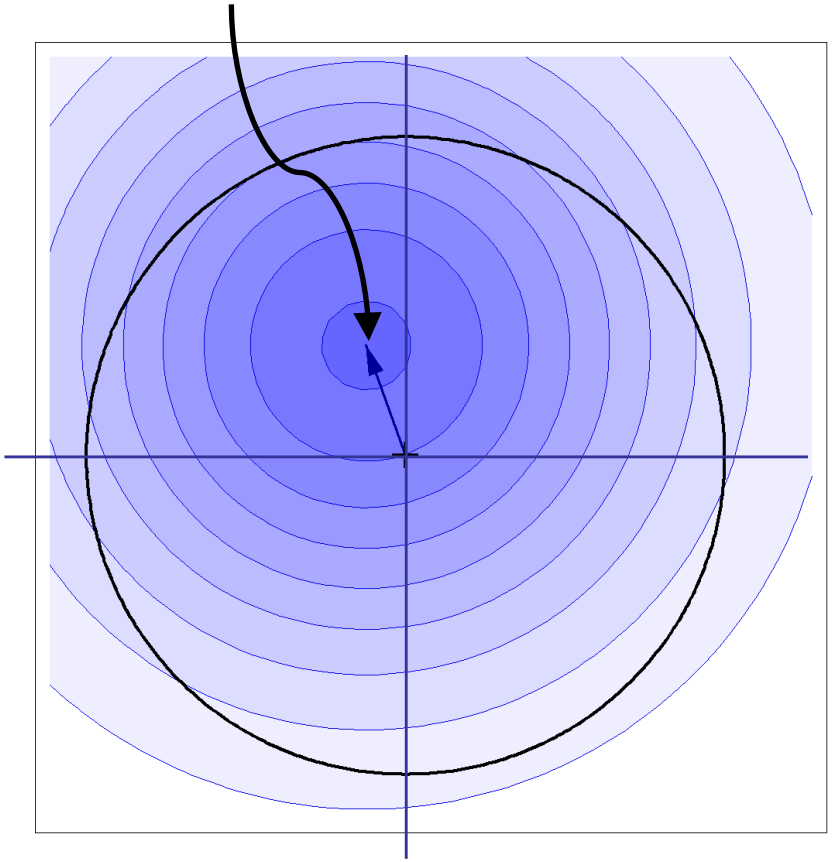

Fh−
Fh+


Fh−
Fh+


Fh−
Fh+

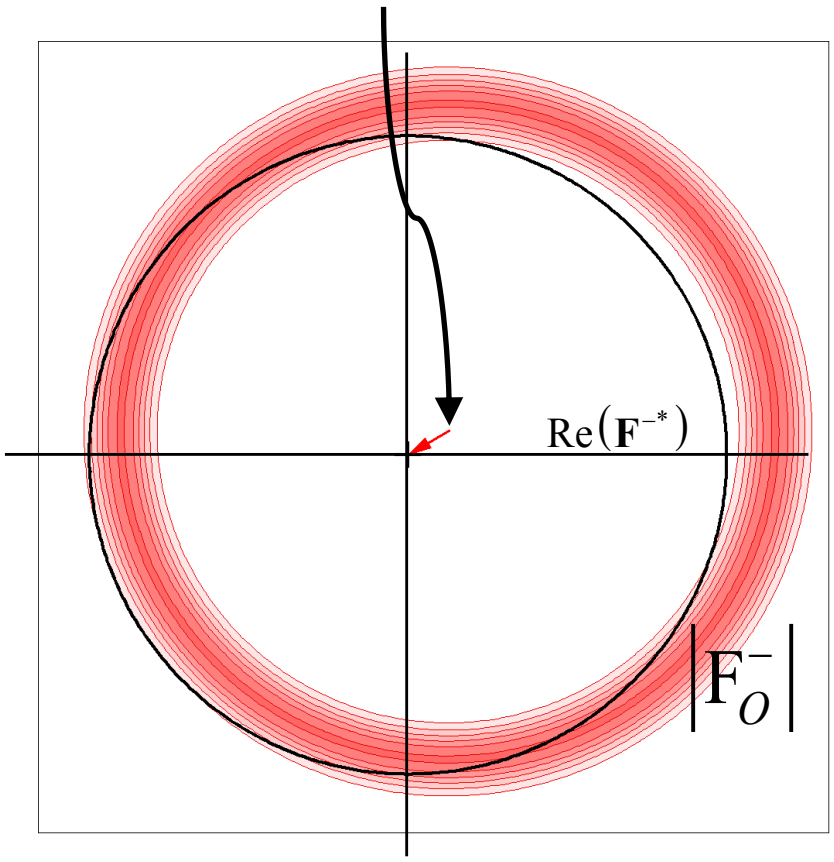
Intuitive understanding of SAD phasing

Expected value of $\mathbf{F}^{-*}$ $(\mathbf{H}^{-*})$



$P\left(\mathrm{F}_O^-, \alpha_O^-; \mathbf{H}^{-*}\right)$

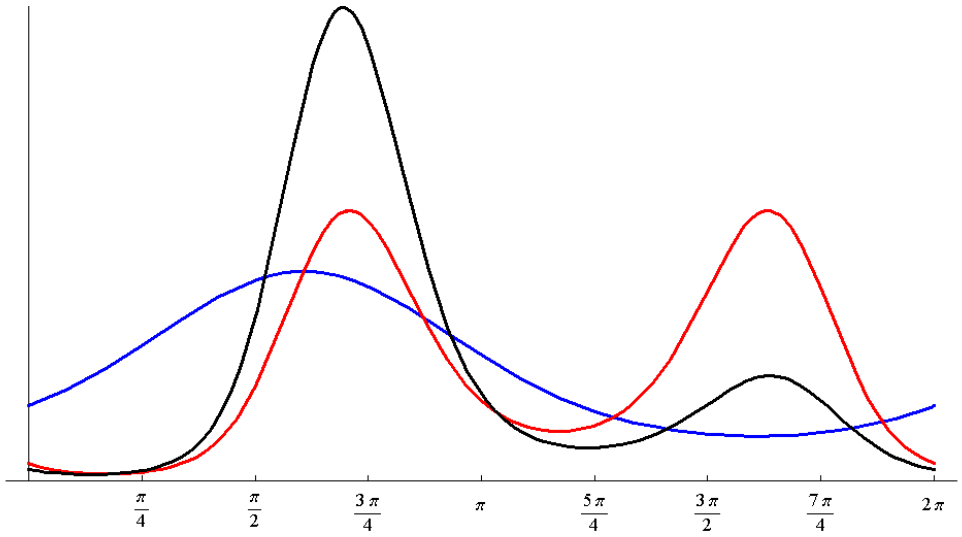
Expected difference between $\mathbf{F}^{+}$ and $\mathbf{F}^{-*}$



$P\left(\mathrm{F}_O^+; \mathrm{F}_O^-, \mathbf{H}^+, \mathbf{H}^{-*}\right)$

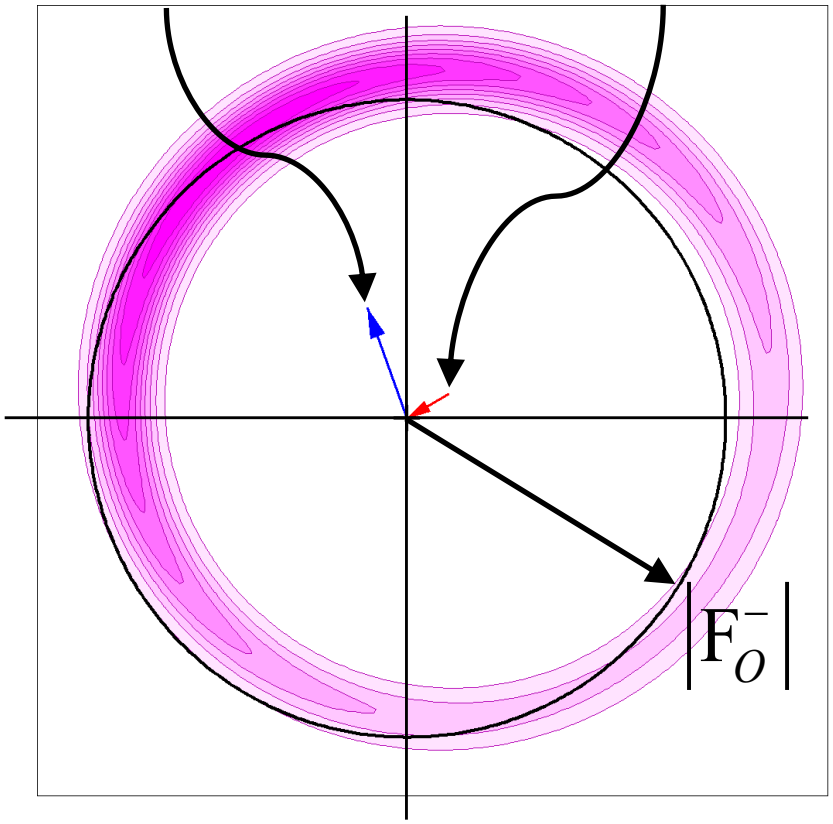
Intuitive understanding of SAD phasing

Total likelihood is integral of the product of the two distributions under the black circle



Expected value of $\mathbf{F}^{-*}$ $(\mathbf{H}^{-*})$

Expected difference between $\mathbf{F}^{+}$ and $\mathbf{F}^{-*}$



SAD log-likelihood gradient (LLG) map

- Compute derivative of log-likelihood with respect to heavy atom structure factor
- Fourier transform gives map of where likelihood target would like to see changes in anomalous scatterer model
- Very sensitive to minor sites
  - picks up sites identified as water molecules in refined structures determined by halide soaks
- Used to improve substructure determination in phenix.hyss (Tom Terwilliger, Gabór Bunkóczi)
- http://www.phaser.cimr.cam.ac.uk/index.php/Tutorials
  - tutorial with data for lysozyme iodide soak

Combining MR and SAD

- CuKα data to 1.9Å on hen egg-white lysozyme
  - can't find sulfurs with HySS or SHELXD
- Solve by MR with goat alpha-lactalbumin (40% identical)
- Use MR model as "substructure" for SAD
  - look for S atoms in LLG map (finds all 10 S, 5-9 Cl^-)
  - phases automatically combine MR and SAD
- Automated fitting with density-modified map
  - tutorial with these data is available

Availability of *Phaser* SAD phasing

- Available in CCP4 and Phenix
  - part of ccp4i Phaser-EP pipeline in CCP4
  - standalone SAD phasing in ccp4i2 and CCP4cloud
  - part of AutoSol pipeline in Phenix

Background information

- "*Phaser* crystallographic software", McCoy, Grosse-Kunstleve, Adams, Winn, Storoni & Read (2007), *J. Appl. Cryst.* **40**, 658-674.
  - plus papers cited here
- "Liking likelihood", Airlie J. McCoy (2004), *Acta Cryst. D***60**, 2169-2183.

- http://www.phaser.cimr.cam.ac.uk/index.php
- http://www.phaser.cimr.cam.ac.uk/index.php/Tutorials
- http://www-structmed.cimr.cam.ac.uk/Course

Acknowledgements

- Phaser: Airlie McCoy, Gábor Bunkóczi, Rob Oeffner, Kaushik Hatti, Massimo Sammito, Duncan Stockwell
- Phenix: Ralf Grosse-Kunstleve, Nigel Moriarty, Paul Adams, Tom Terwilliger