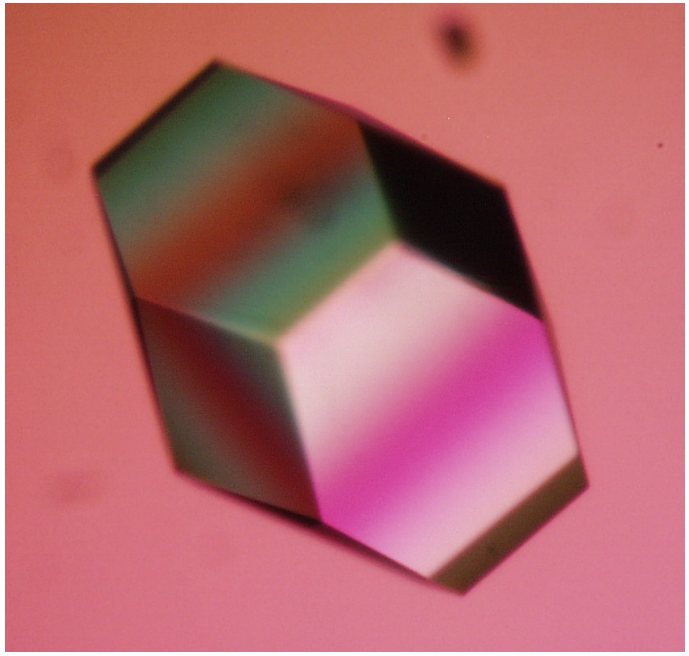
Introduction to experimental phasing

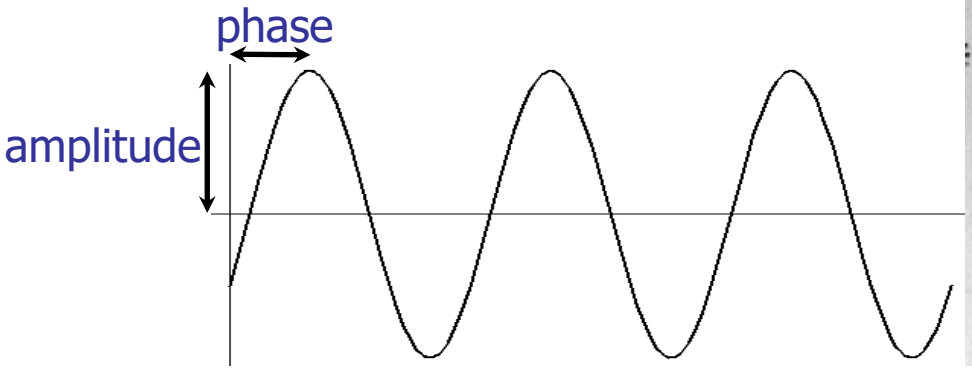

UNIVERSITY OF CAMBRIDGE

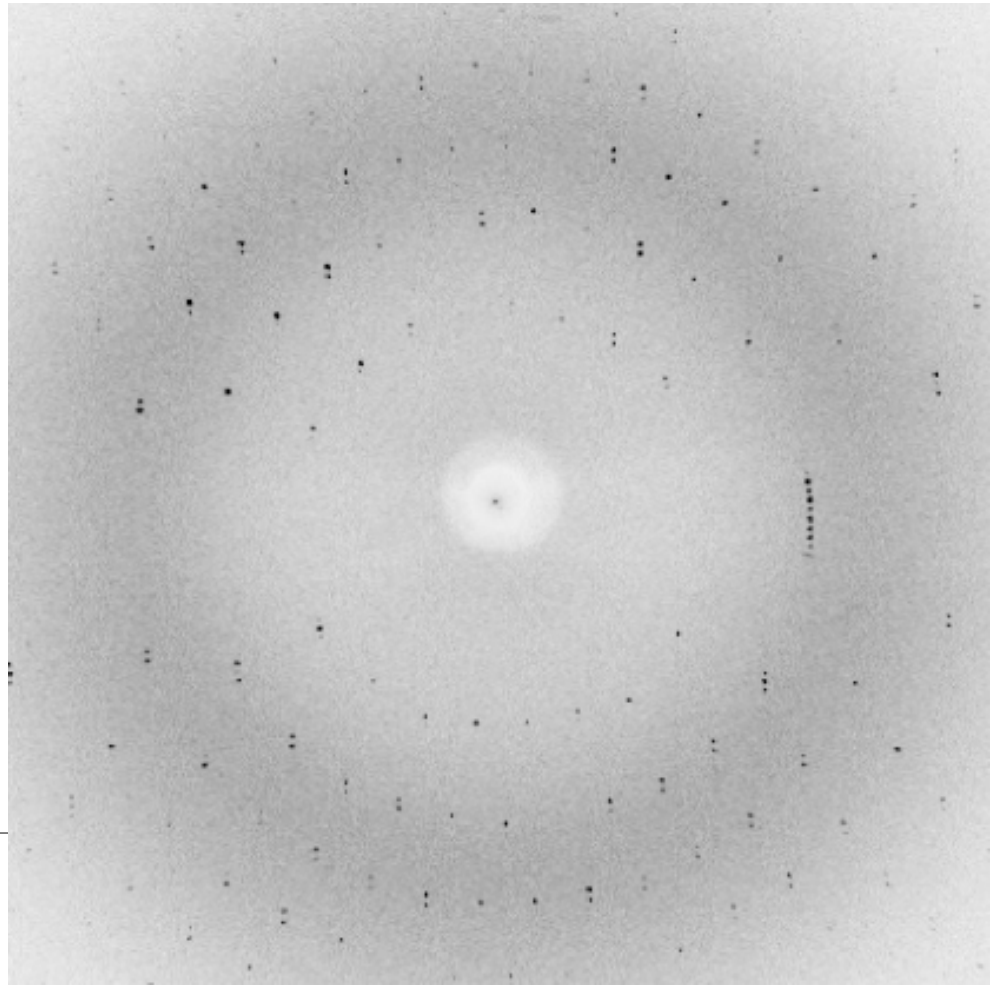
Randy J Read
Department of Haematology

CIMR Cambridge Institute for Medical Research

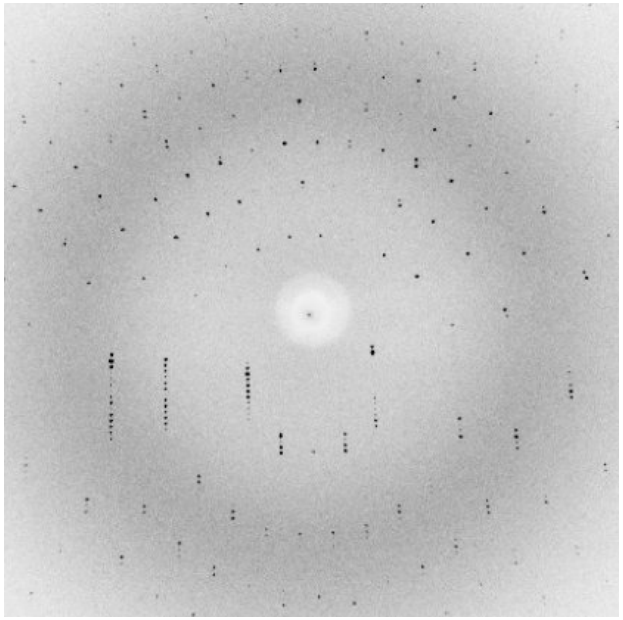
X-ray diffraction

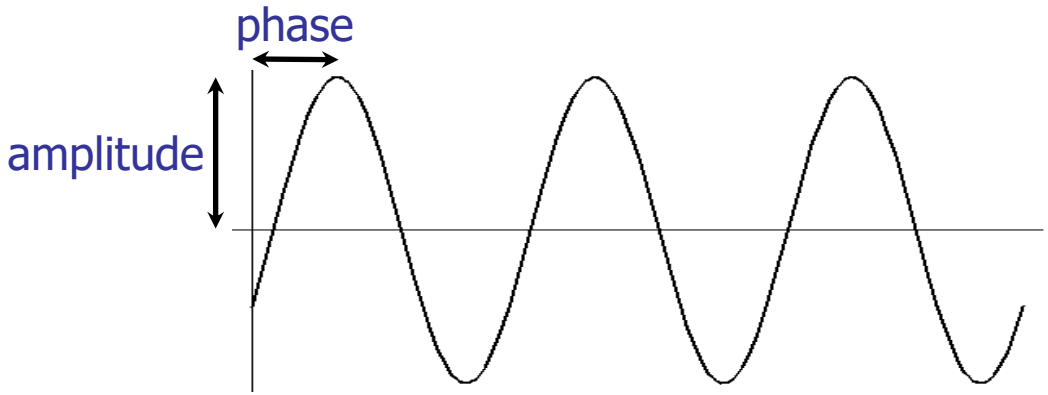






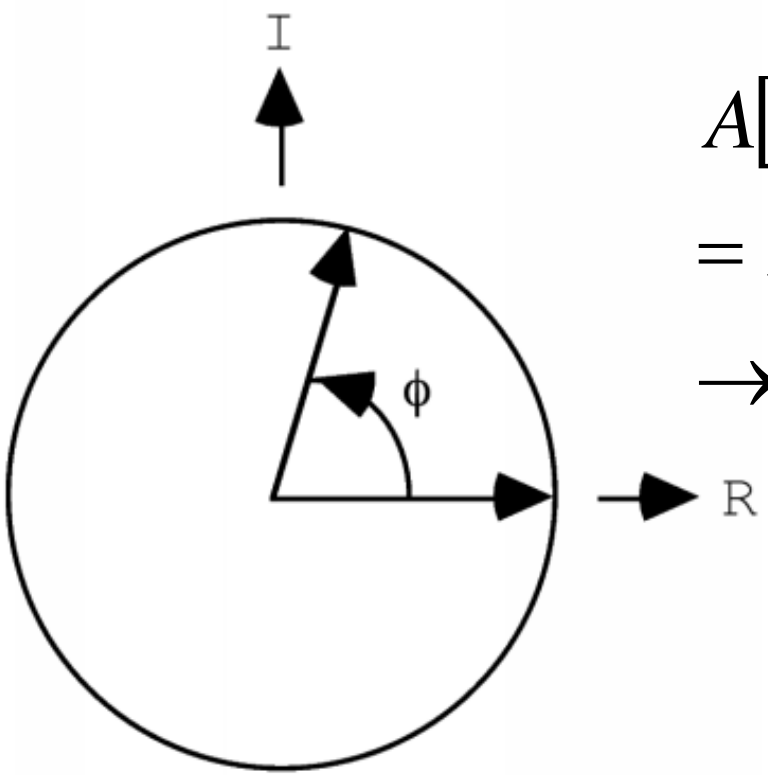
The Phase Problem





Wave as vector (or complex number)

- Wave is x-component of rotating vector
- Initial rotation gives initial phase shift

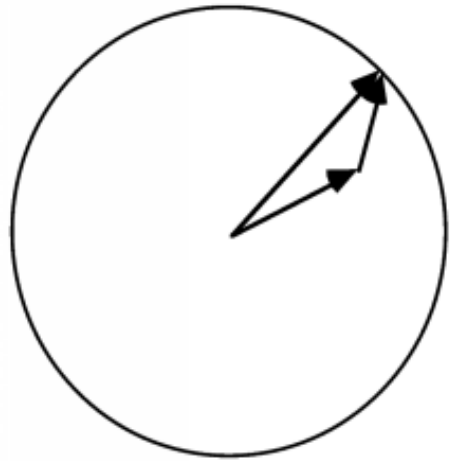


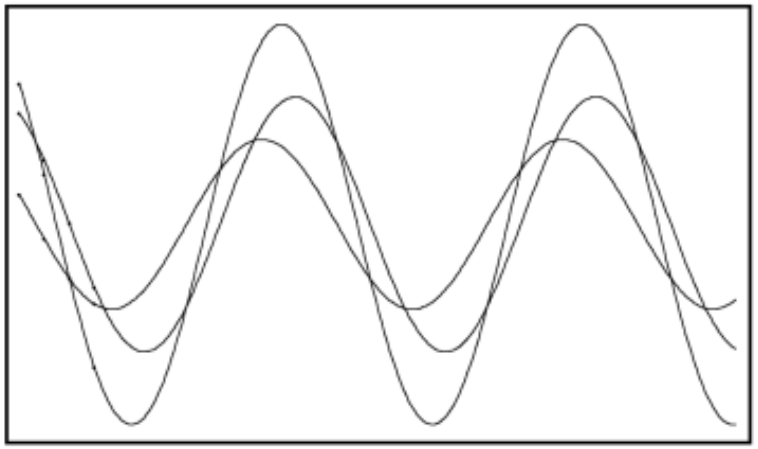
$$A[\cos(\phi + 2\pi\nu t) + i\sin(\phi + 2\pi\nu t)]$$
$$= A\exp[i(\phi + 2\pi\nu t)]$$
$$\rightarrow A\exp(i\phi)$$

Adding waves as vectors

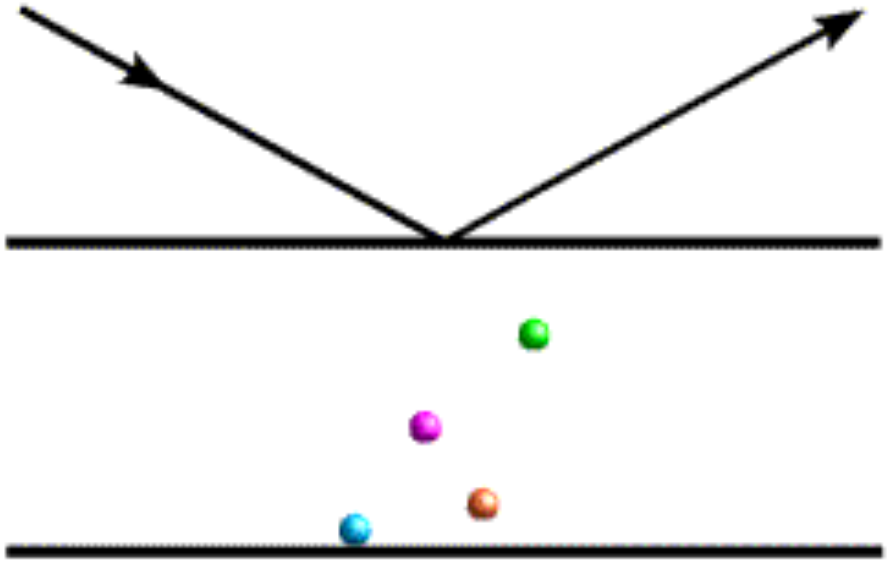
- Adding waves is equivalent to adding vectors
  - easier than trigonometry

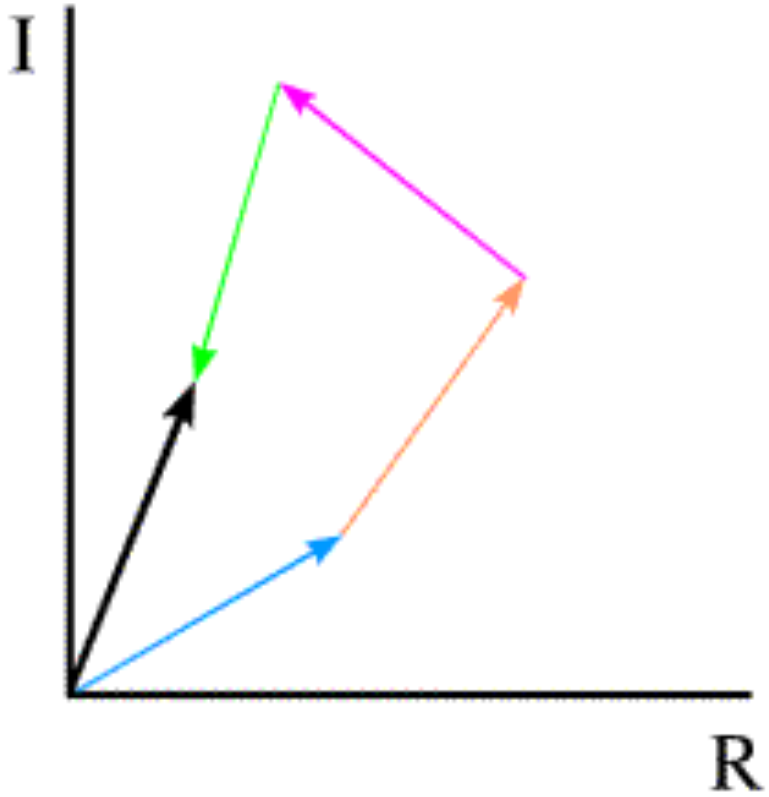
$$A\cos(\alpha+\varphi_1)+B\cos(\alpha+\varphi_2)$$



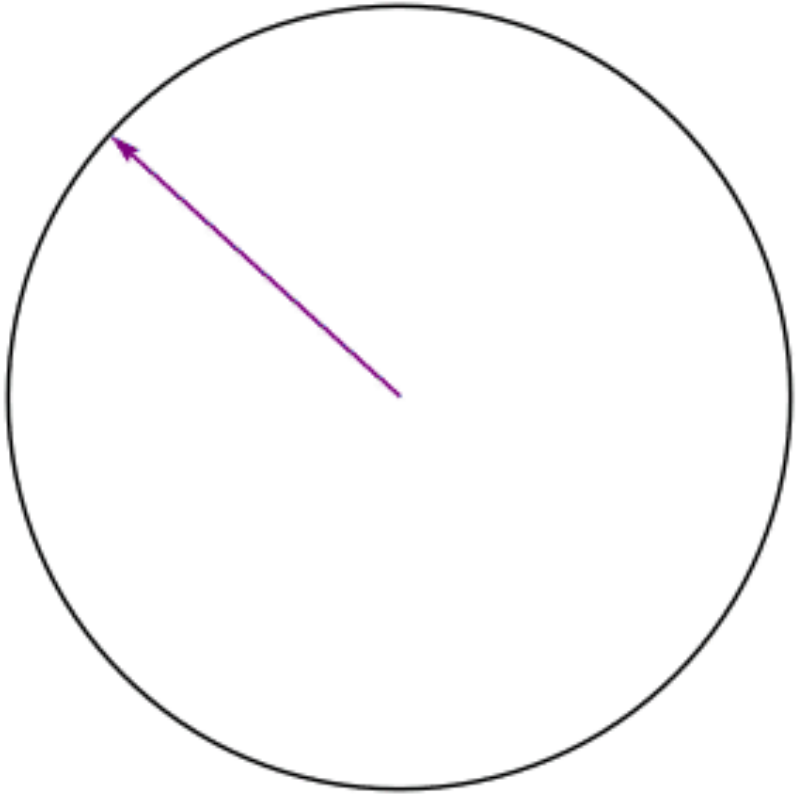


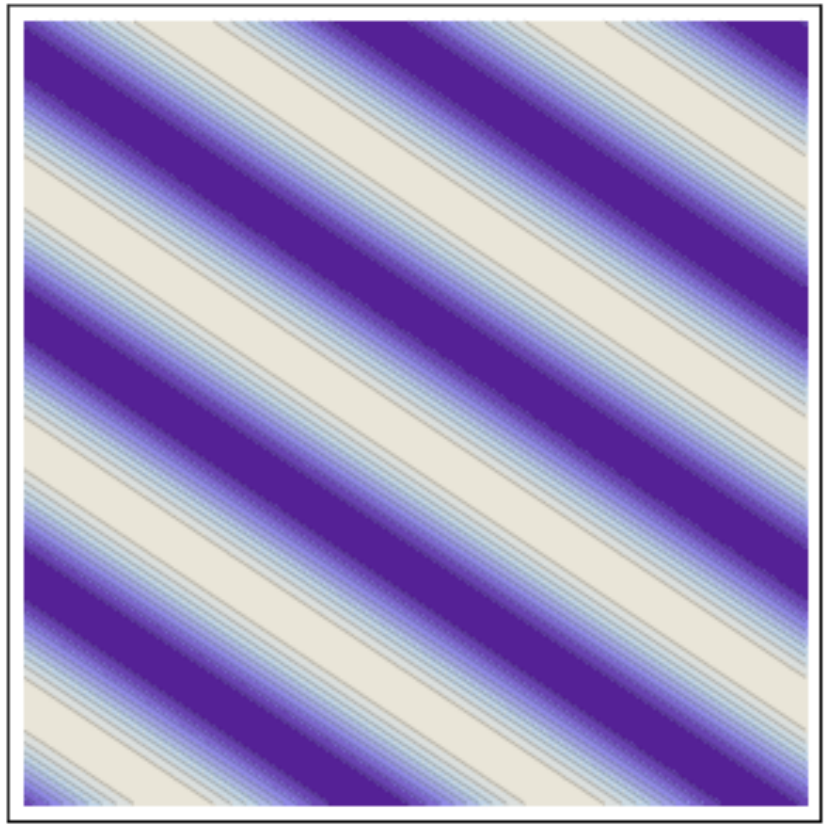
Building up a structure factor



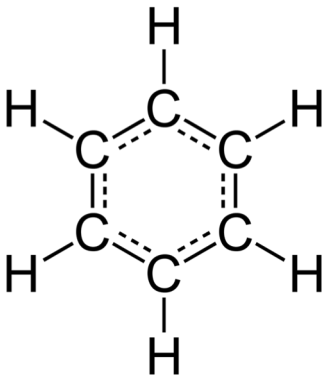


Effect of phase on electron density



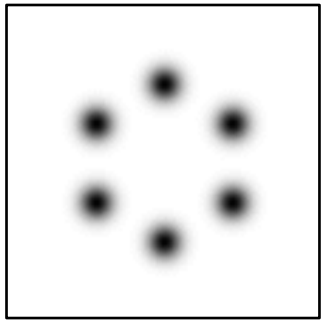


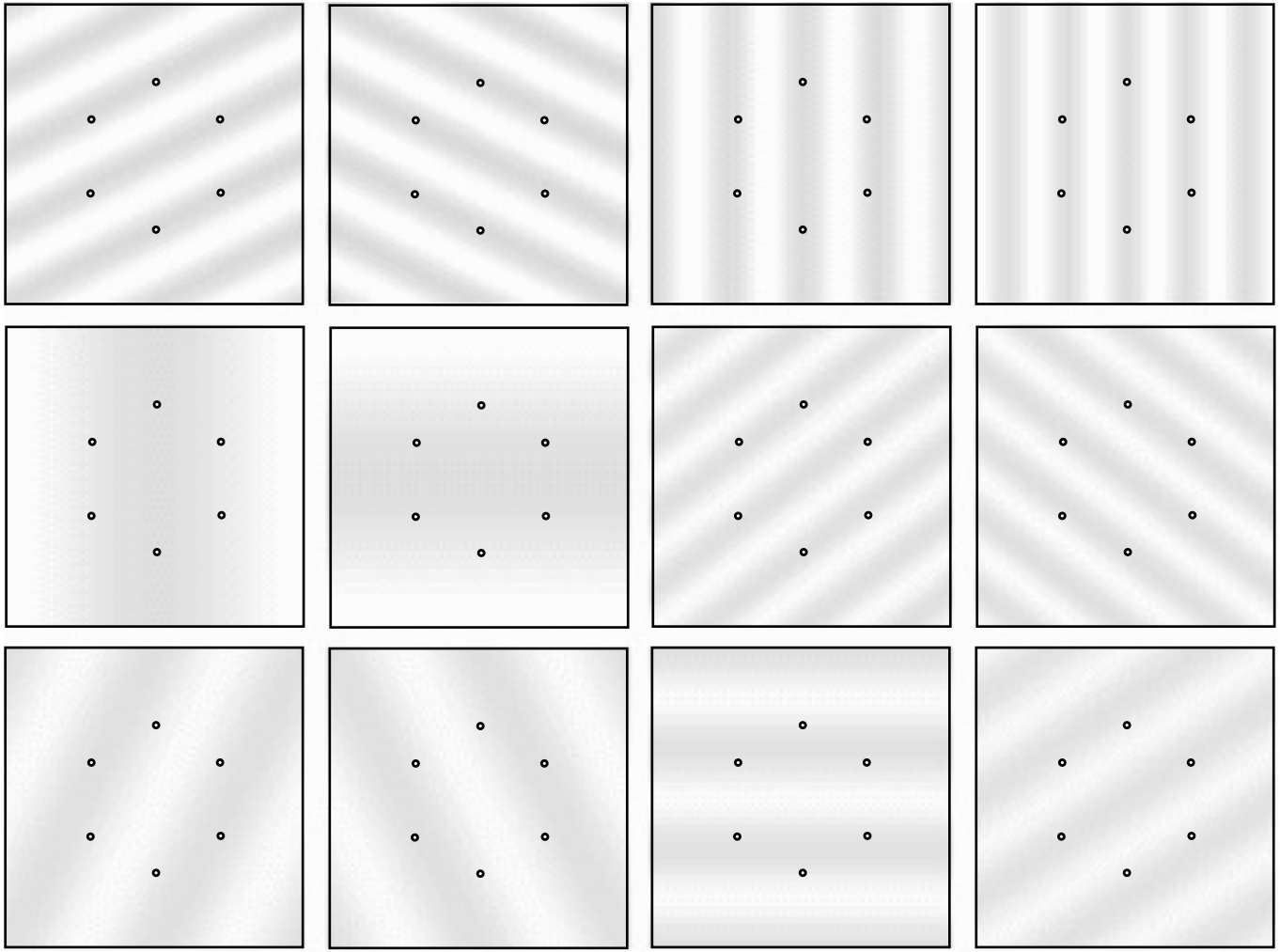
Breaking a picture up into waves



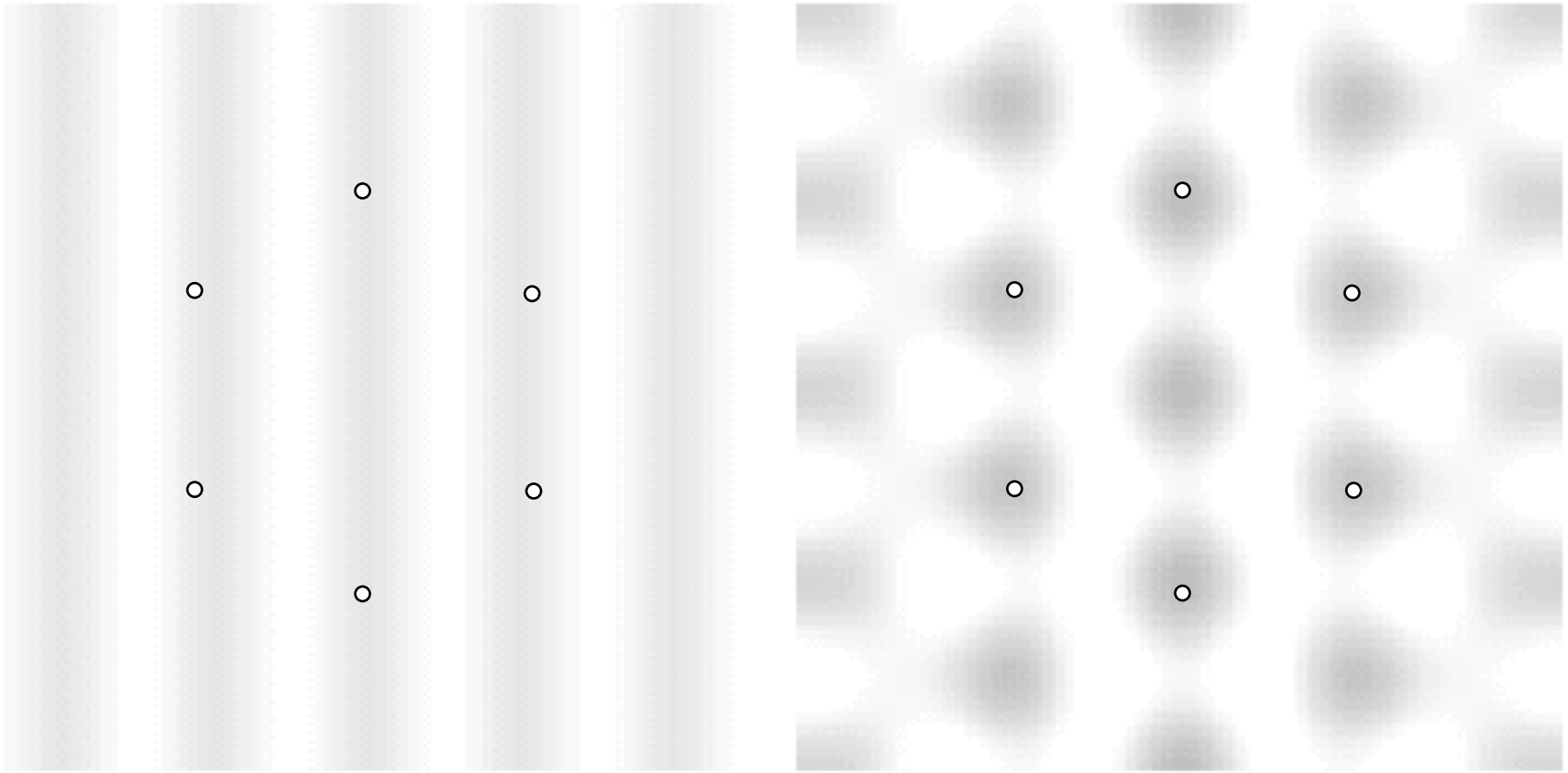
benzene

electrons

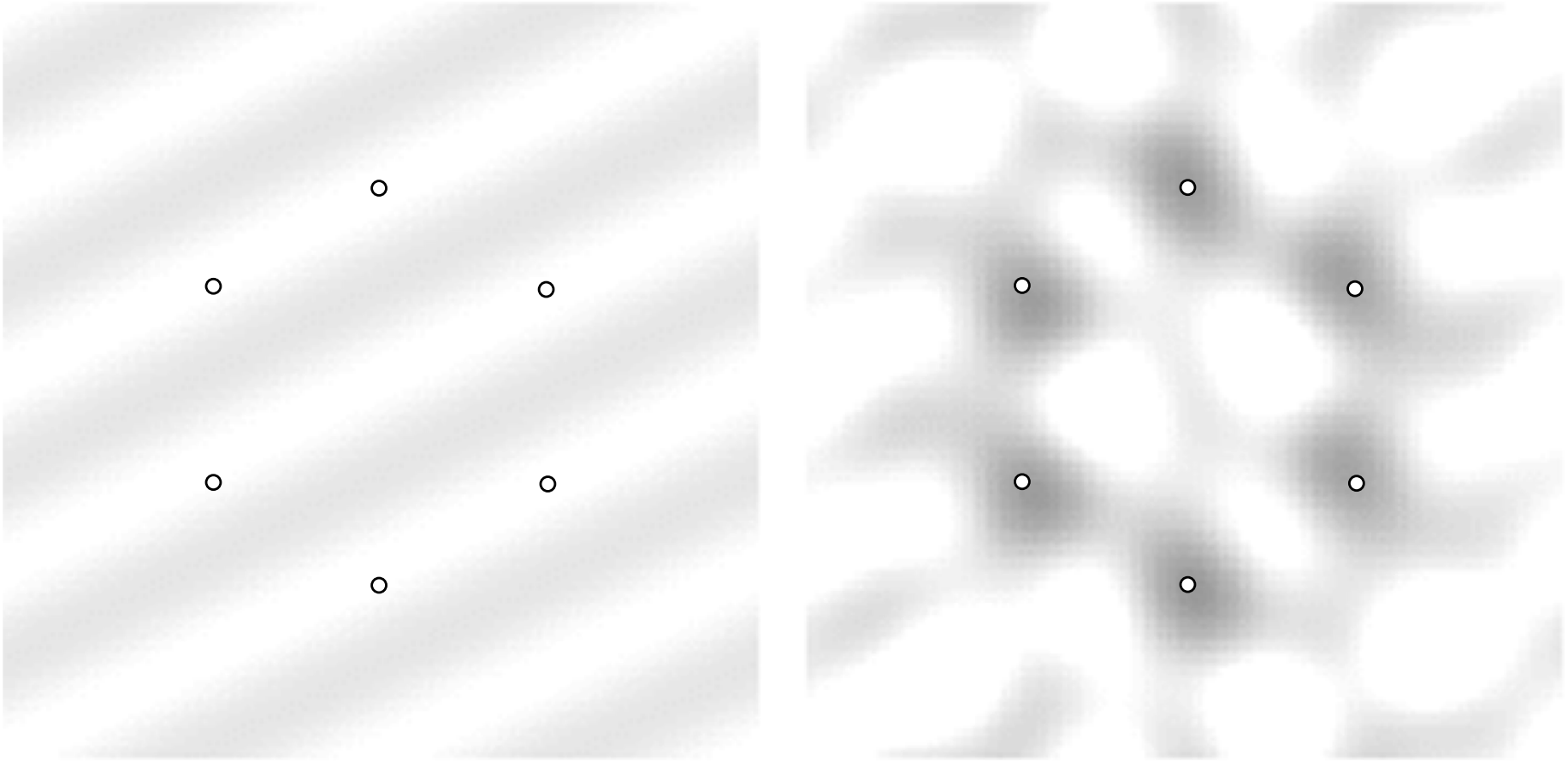




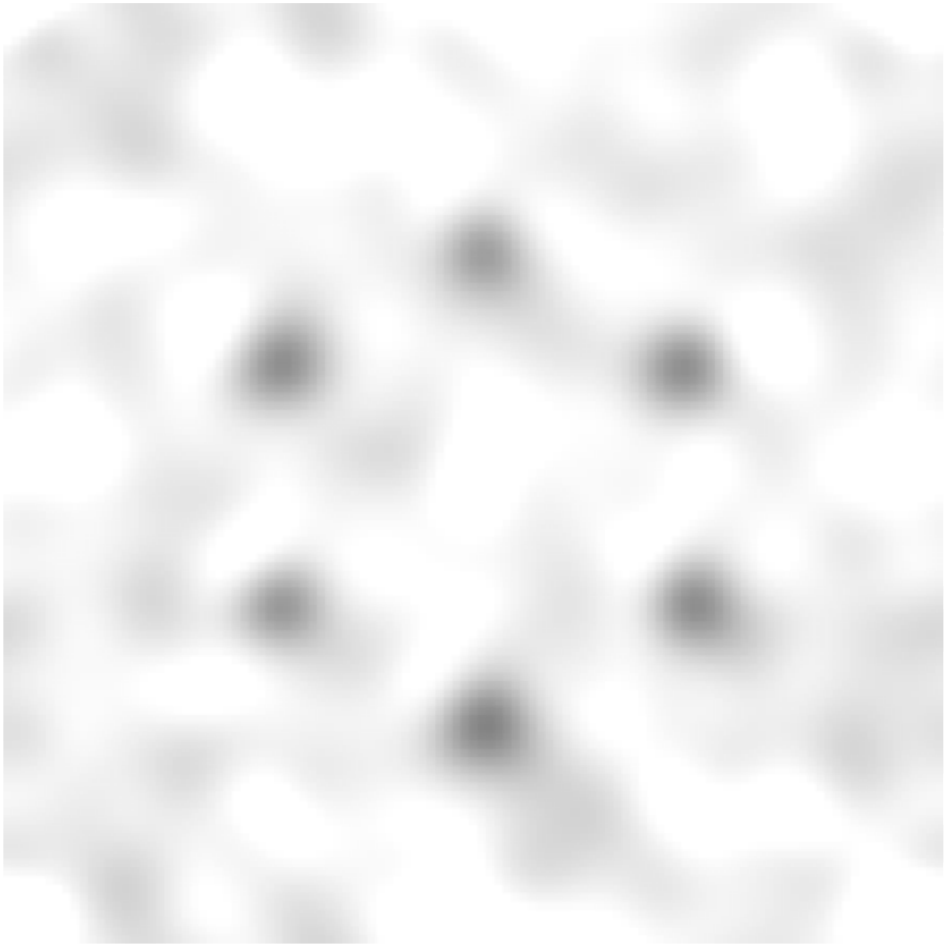
Adding up the top 30 waves



Changing the phase of the top wave



Changing all the phases from correct to random and back

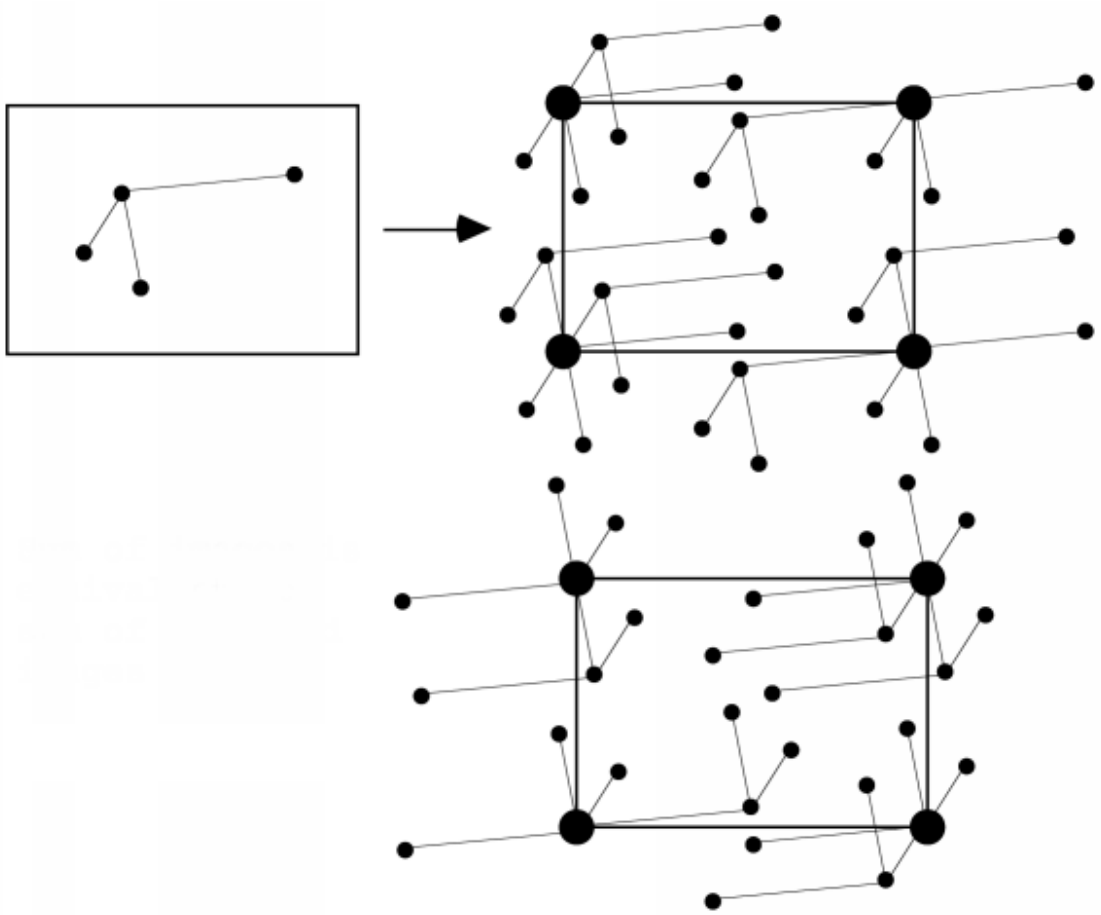


Phasing: *Ab initio* approaches

- Patterson methods
  - Patterson map shows vectors between atoms
    - n^2-n non-origin vectors
- Direct methods
  - based on statistical relationships between phases
    - progressively weaker as number of atoms increases

The Patterson function

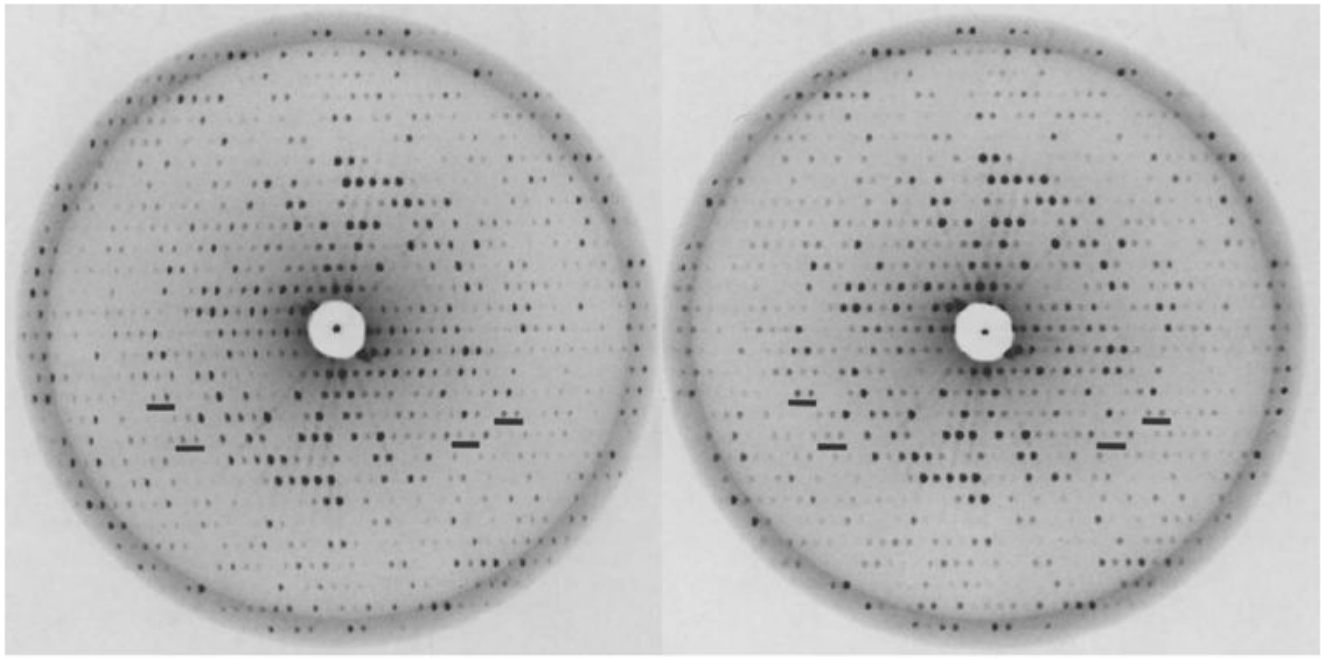
- Map of vectors between atoms
- Superposition of images
- $n^2 - n$ off-origin peaks



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 Friedel 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

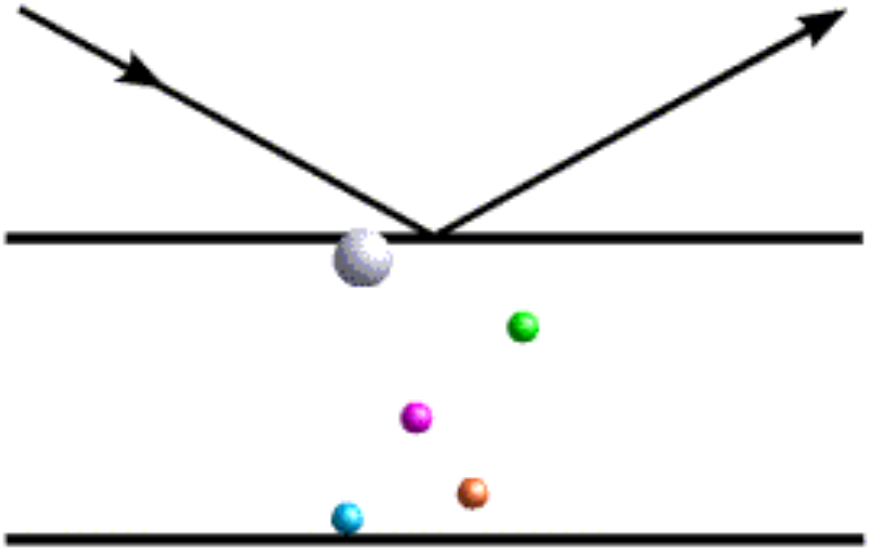
Heavy atom derivative

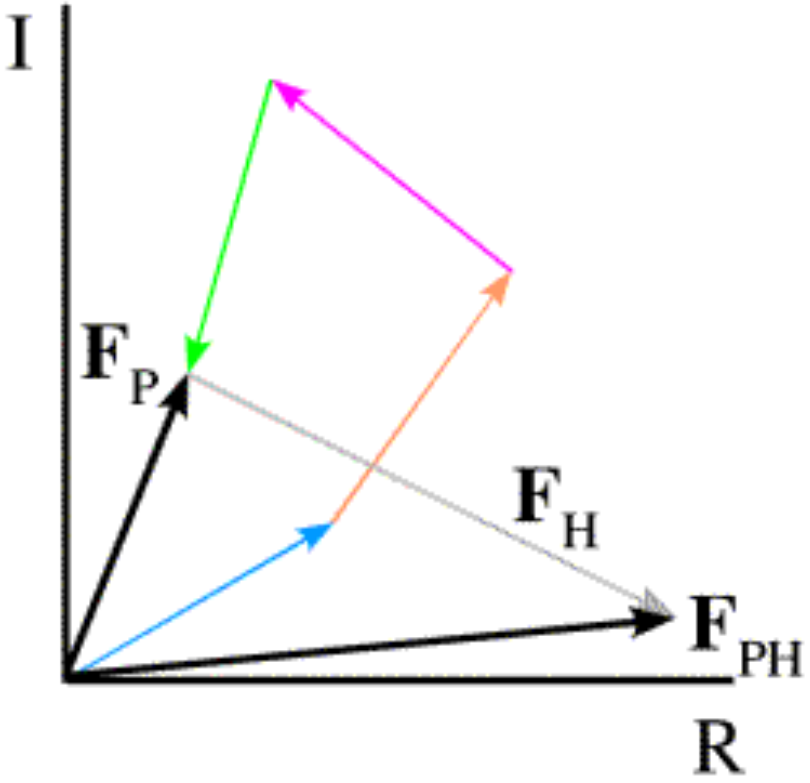
F_{deriv}

from Gale Rhodes, "Crystallography made crystal clear"

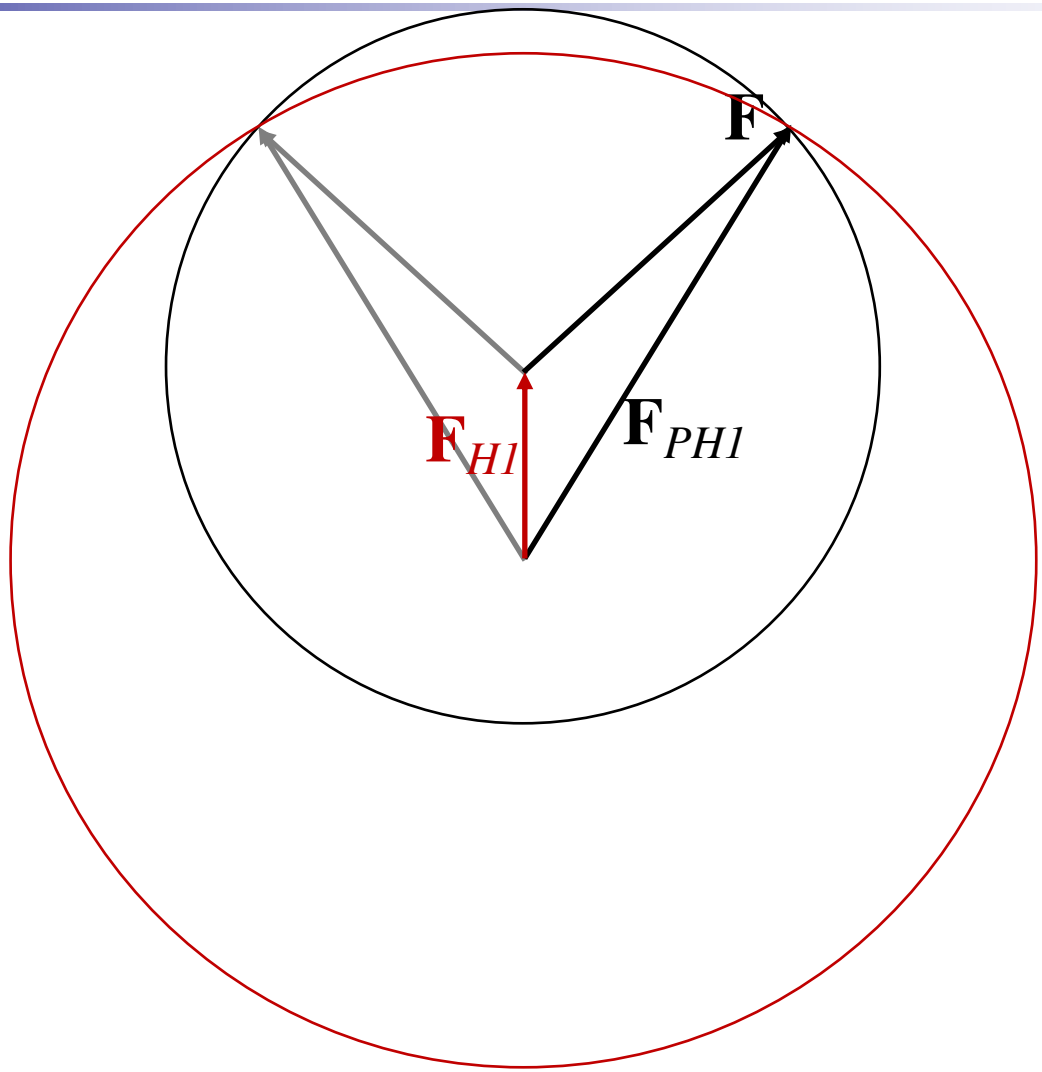
Phase information from perturbed diffraction

- Heavy-atom contribution is vector addition to structure factor





Deducing phase information from isomorphous replacement

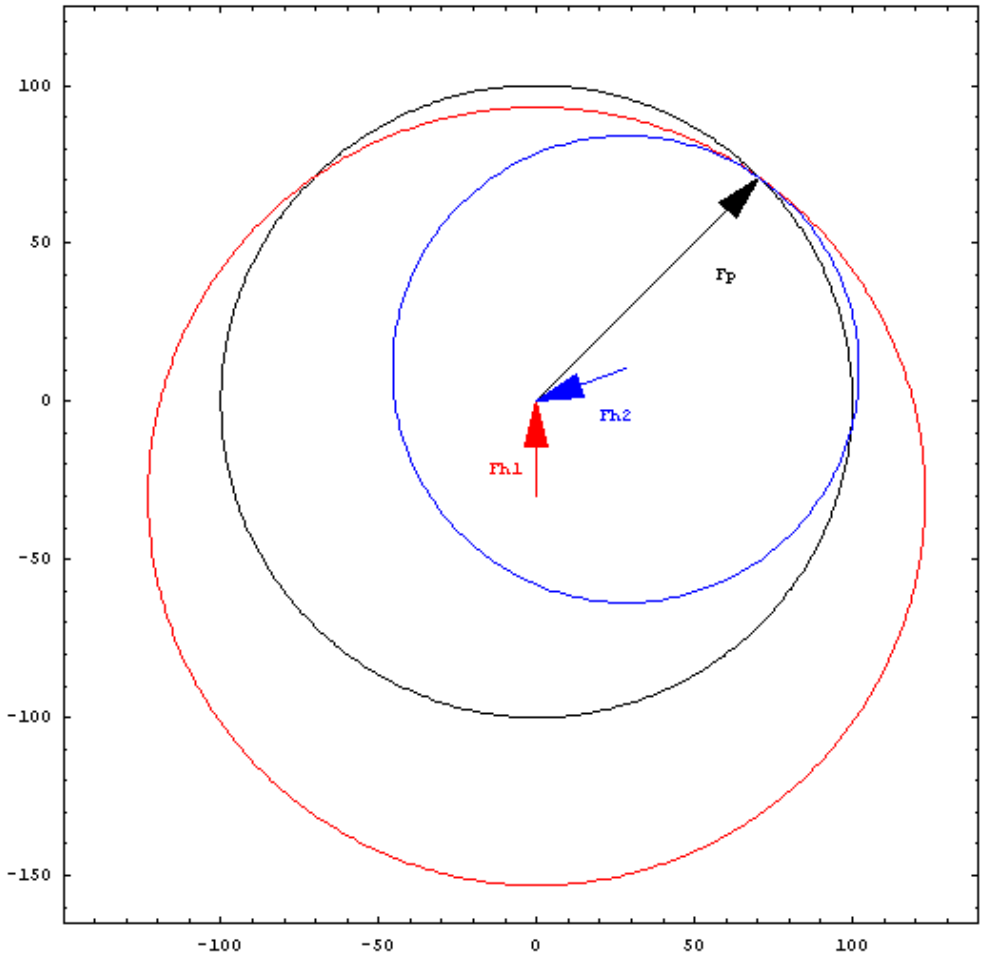


Principle of maximum likelihood

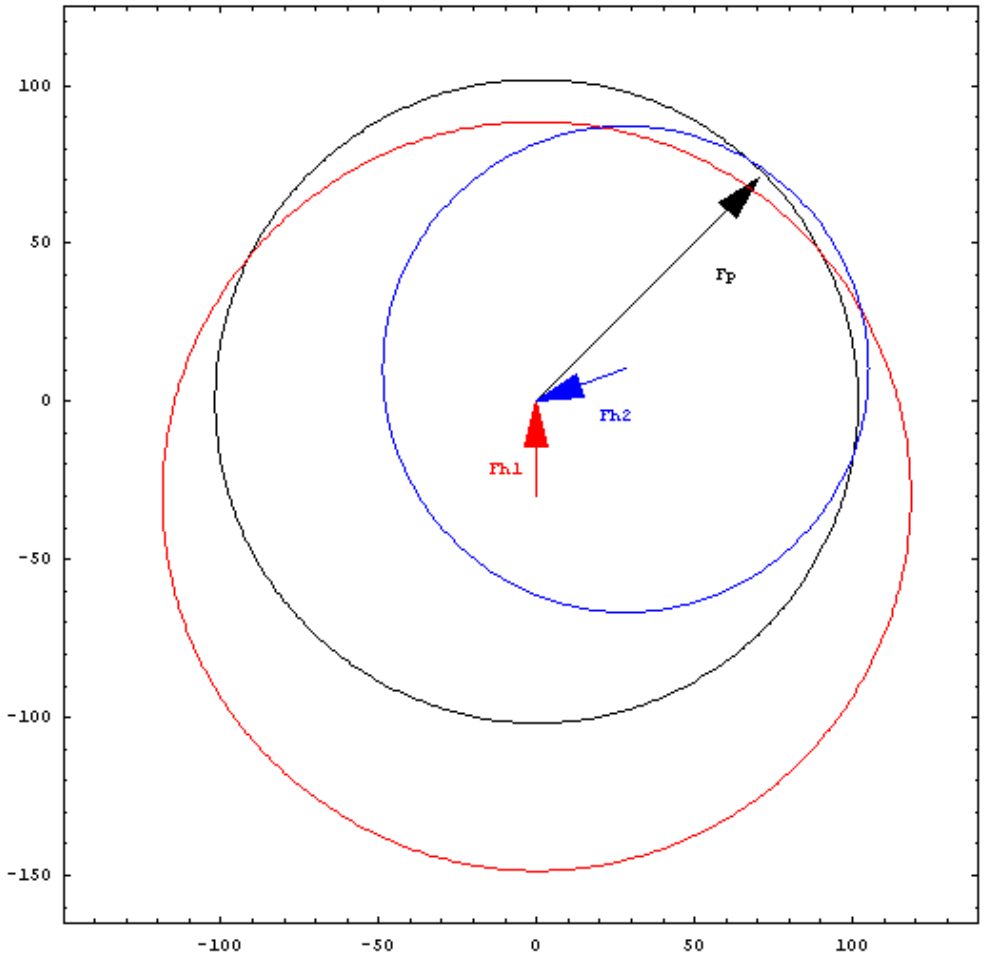
- Best model is most consistent with data
- Measure consistency by probabilities
  - given the model, what is the probability that the data would have been measured?

The Harker construction for multiple isomorphous replacement (MIR)

$$p(|\mathbf{F}_P|,|\mathbf{F}_{PH1}|,|\mathbf{F}_{PH2}|;\mathbf{F},\mathbf{F}_{H1},\mathbf{F}_{H2}) \approx$$
$$p(|\mathbf{F}_P|;\mathbf{F})p(|\mathbf{F}_{PH1}|;\mathbf{F},\mathbf{F}_{H1})p(|\mathbf{F}_{PH2}|;\mathbf{F},\mathbf{F}_{H2})$$



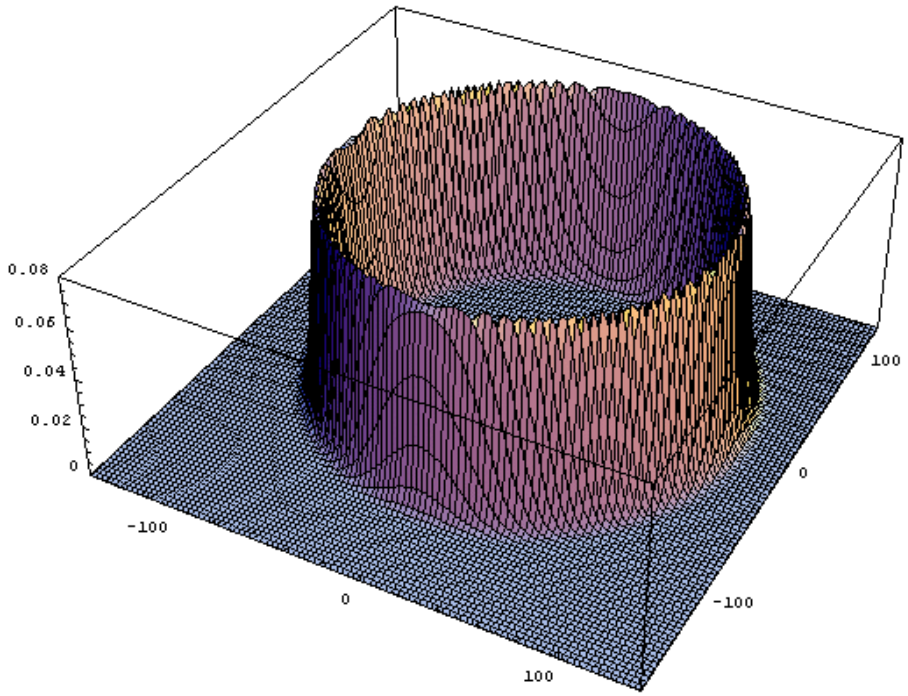
The Harker construction with errors



Probabilistic Harker construction

- Account for measurement errors in native structure factor

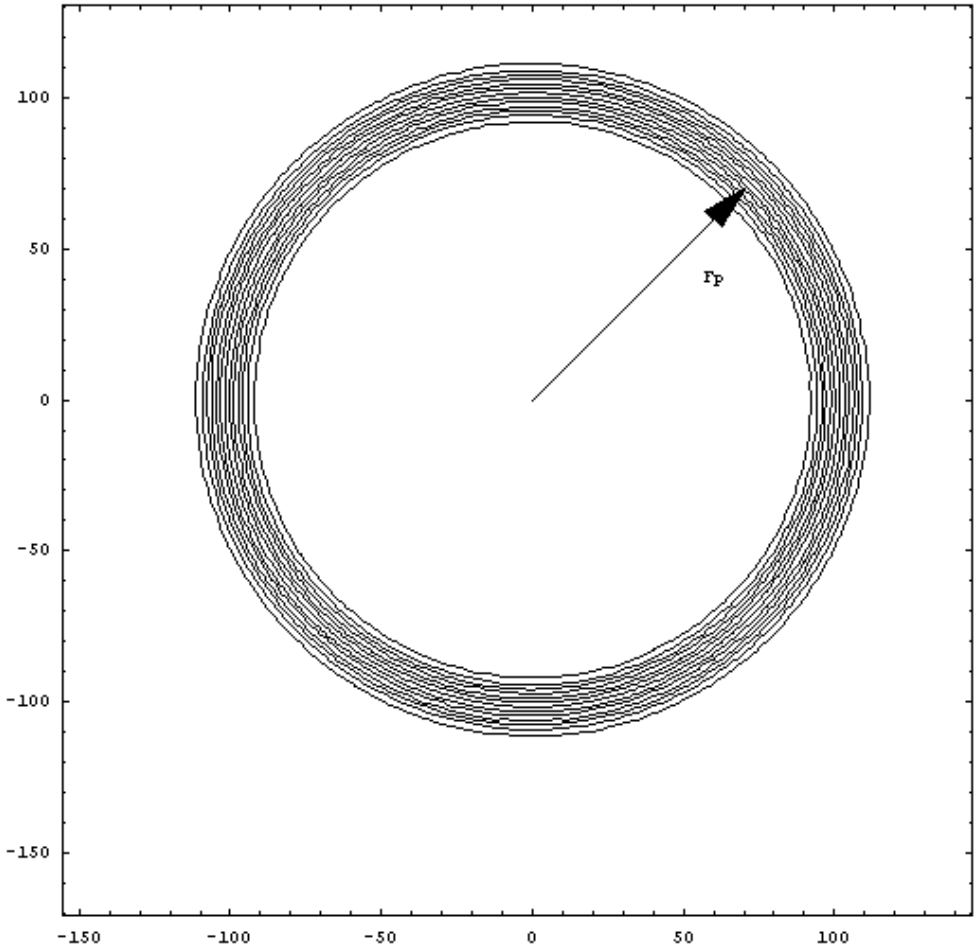
$$p\left(\left|\mathbf{F}_P\right|;\mathbf{F}\right)$$



Probabilistic Harker construction

- Account for measurement errors in native structure factor

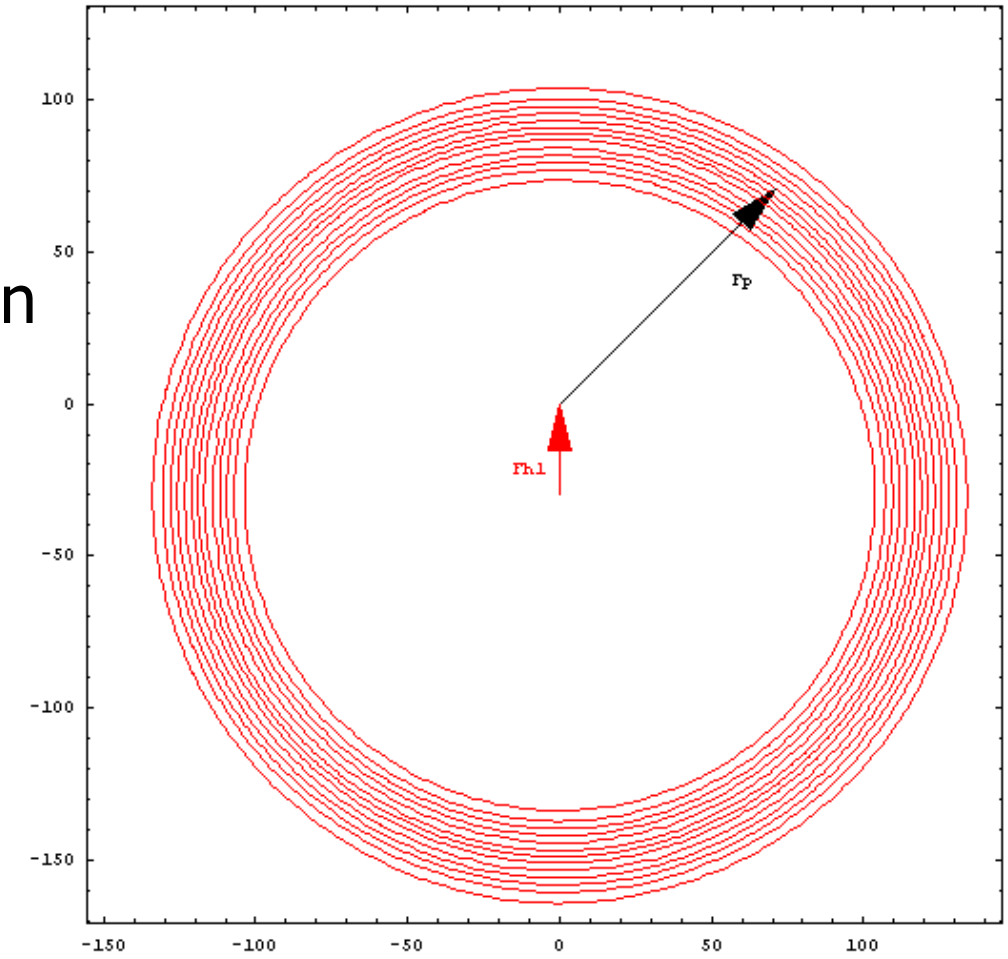
$$p\left(\left|\mathbf{F}_P\right|;\mathbf{F}\right)$$



Probabilistic Harker construction

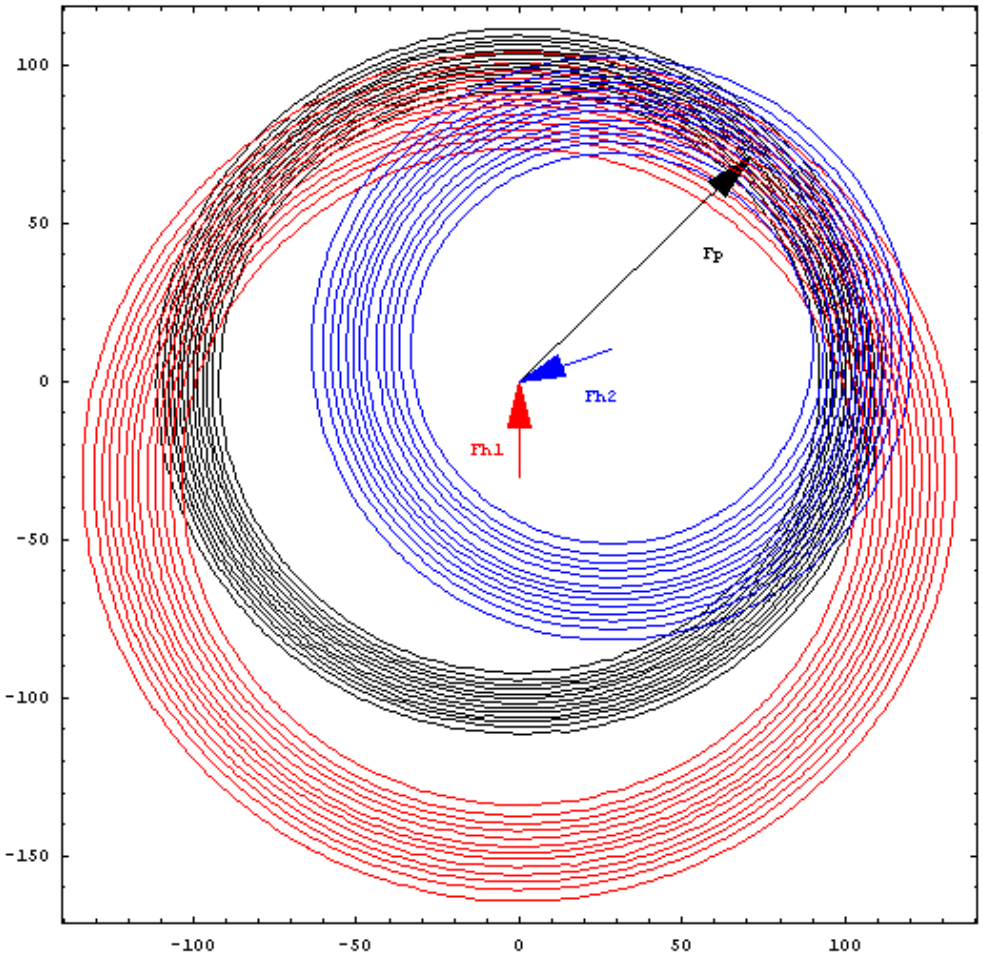
- Account for combined effect of measurement errors, lack of isomorphism and errors in heavy-atom model

$$p\left(\left|\mathbf{F}_{PH1}\right|;\mathbf{F},\mathbf{F}_{H1}\right)$$



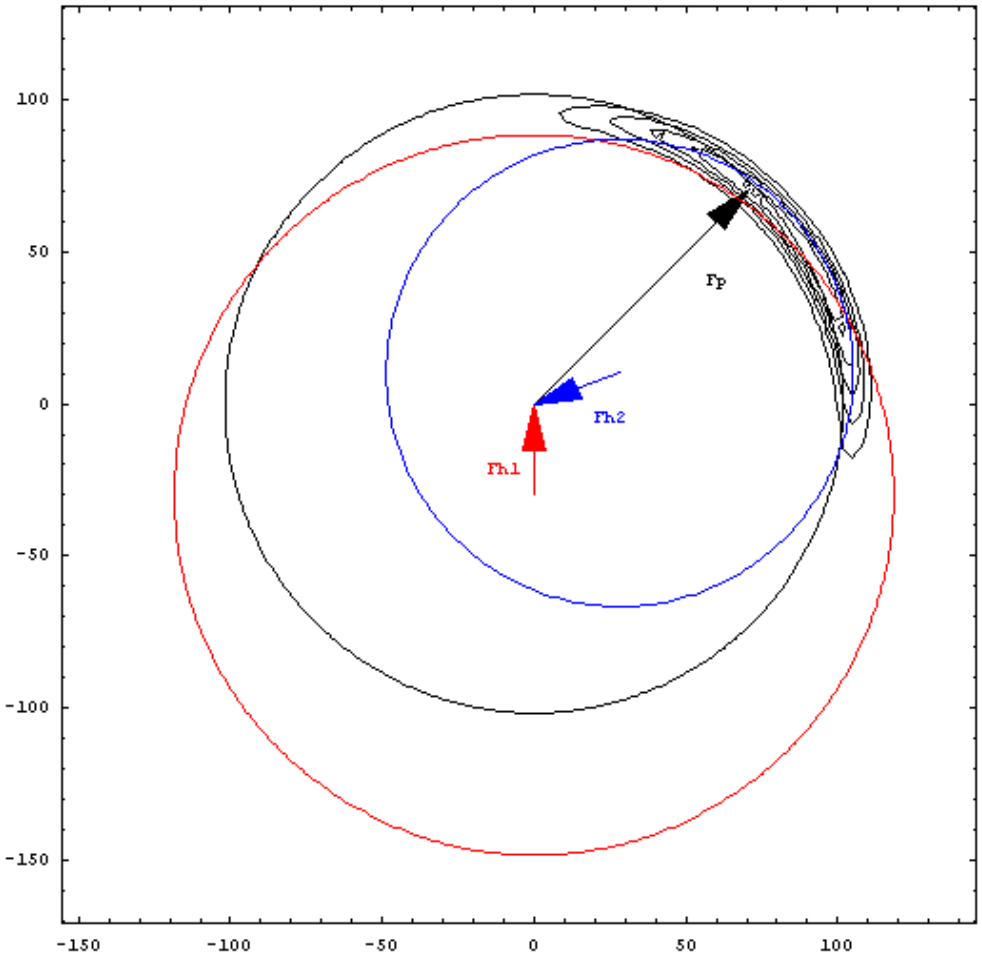
Probabilistic Harker construction

$$p\left(|\mathbf{F}_P|,|\mathbf{F}_{PH1}|,|\mathbf{F}_{PH2}|;\mathbf{F},\mathbf{F}_{H1},\mathbf{F}_{H2}\right)\approx$$
$$p\left(|\mathbf{F}_P|;\mathbf{F}\right)p\left(|\mathbf{F}_{PH1}|;\mathbf{F},\mathbf{F}_{H1}\right)p\left(|\mathbf{F}_{PH2}|;\mathbf{F},\mathbf{F}_{H2}\right)$$



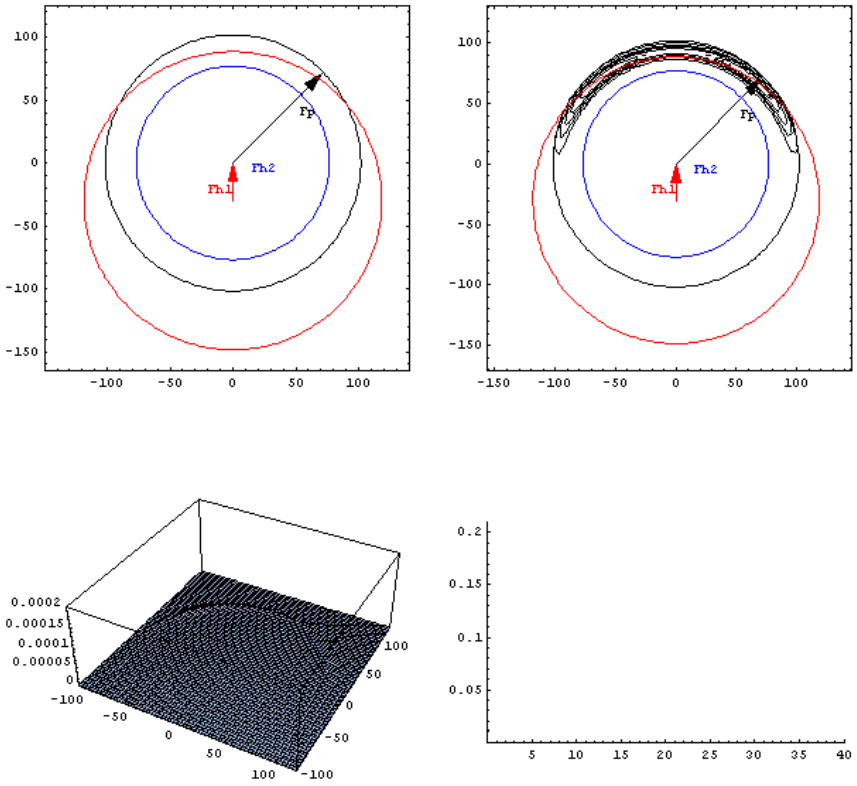
Probabilistic Harker construction

$$p\left(|\mathbf{F}_P|,|\mathbf{F}_{PH1}|,|\mathbf{F}_{PH2}|;\mathbf{F}_{H1},\mathbf{F}_{H2}\right)=$$
$$\int_{\mathbf{F}} p(\mathbf{F})\,p\left(|\mathbf{F}_P|,|\mathbf{F}_{PH1}|,|\mathbf{F}_{PH2}|;\mathbf{F},\mathbf{F}_{H1},\mathbf{F}_{H2}\right)d\mathbf{F}$$



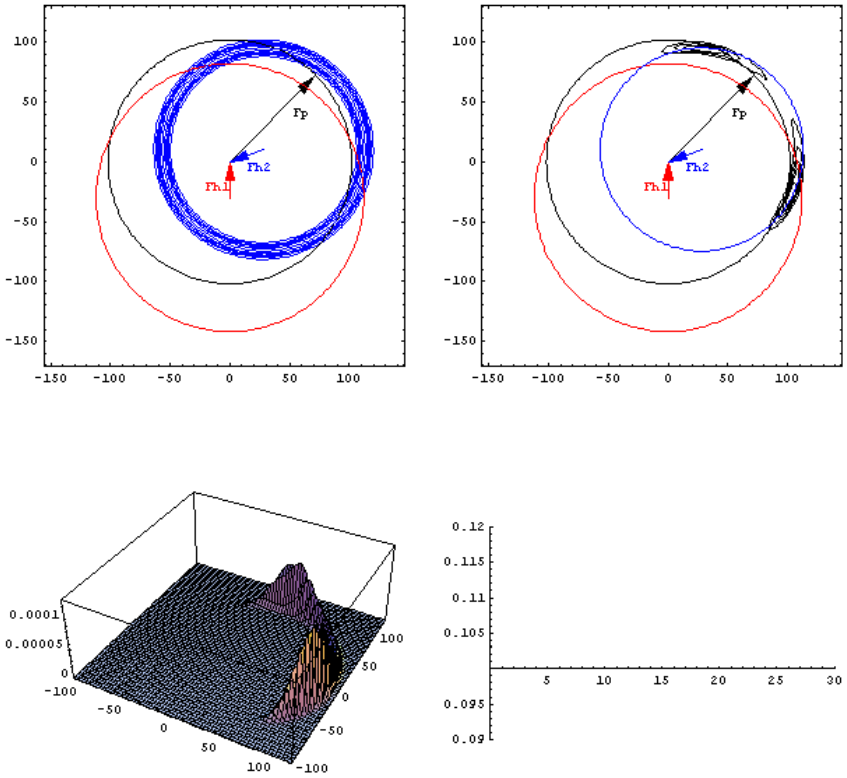
Likelihood as function of heavy atom model

- See animation at http://www-structmed.cimr.cam.ac.uk/Course/Adv_phasing/Expphasing.html#sub2a



Likelihood as function of lack-of-isomorphism error

- See animation at http://www-structmed.cimr.cam.ac.uk/Course/Adv_phasing/Expphasing.html#sub2b



“Best phase” and figure of merit

- Experimental phasing determines a probability distribution for the phase, not an exact value
- Blow & Crick asked how to make the best map
  - map that minimises RMS error from true map
  - for each term, use the probability-weighted average of the (complex) structure factor
    - center of mass of circle weighted by phase probability
    - average is inside circle: measured amplitude is downweighted by figure of merit (= expected cosine of phase error)
    - “best phase” is not necessarily the most probable phase

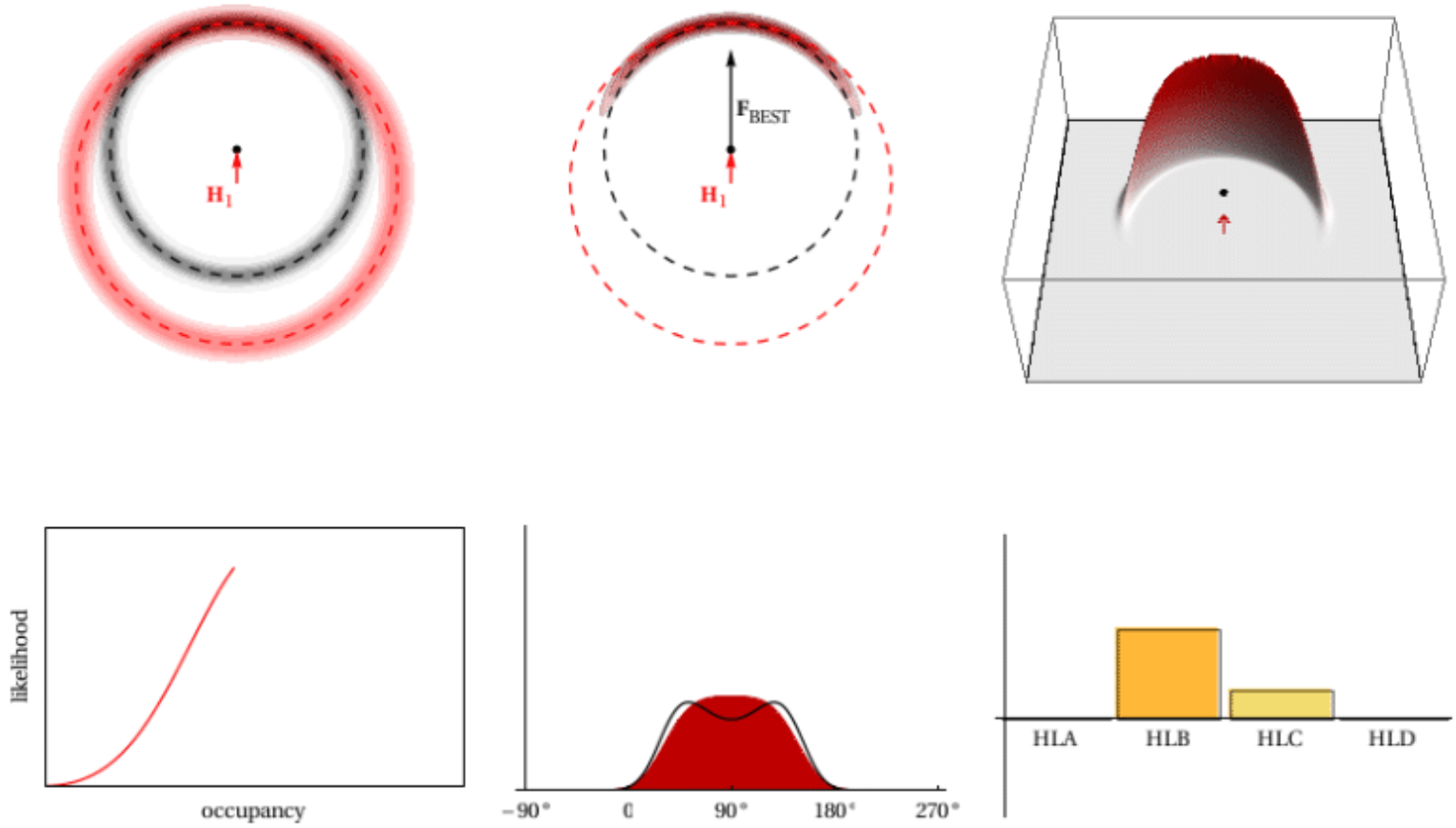
Hendrickson-Lattman coefficients

- MIR phase probabilities have one or two peaks
  - two crossing points for any pair of circles
  - one peak can be larger
- Phase probability distribution can be approximated with Hendrickson-Lattman (HL) coefficients

$$p_{HL}(\varphi) = N_{HL} \times \exp(A_{HL}\cos\varphi + B_{HL}\sin\varphi + C_{HL}\cos 2\varphi + D_{HL}\sin 2\varphi)$$

- Independent phase information can be combined by adding HL coefficients

Best (centroid) map for SIR phasing



animation produced by Airlie McCoy

Making heavy-atom derivatives

- Ordered binding depends on chemical properties of heavy-atom reagents
  - bonding (e.g. Hg with sulfhydryls)
  - hydrophobic interactions
    - *e.g.* xenon
  - ionic interactions
    - exchange with native metal sites
    - halides
  - may be pH dependent, some metals react with precipitants

Lack of isomorphism

- Heavy-atom reagents can perturb crystal
  - change crystal packing
  - change protein conformation
- Commonly accompanied by change in cell dimensions
  - absolute size of cell change determines extent of non-isomorphism
    - not fractional or percentage change!
  - but agreement of data as a function of resolution is best measure

Dealing with poor isomorphism

- Consider solving derivative alone by SAD or MAD
  - heavy atoms tend to be strong anomalous scatterers
  - final refinement with best data set, usually native
- Poorer isomorphism = stronger multi-crystal averaging!
  - understood in terms of sampling of molecular transform

SAD phasing

- Currently the top choice for structures that are not solved by molecular replacement
  - one data set
  - no problem with isomorphism (except possibly data merged from multiple crystals)
- Tomorrow's lecture!