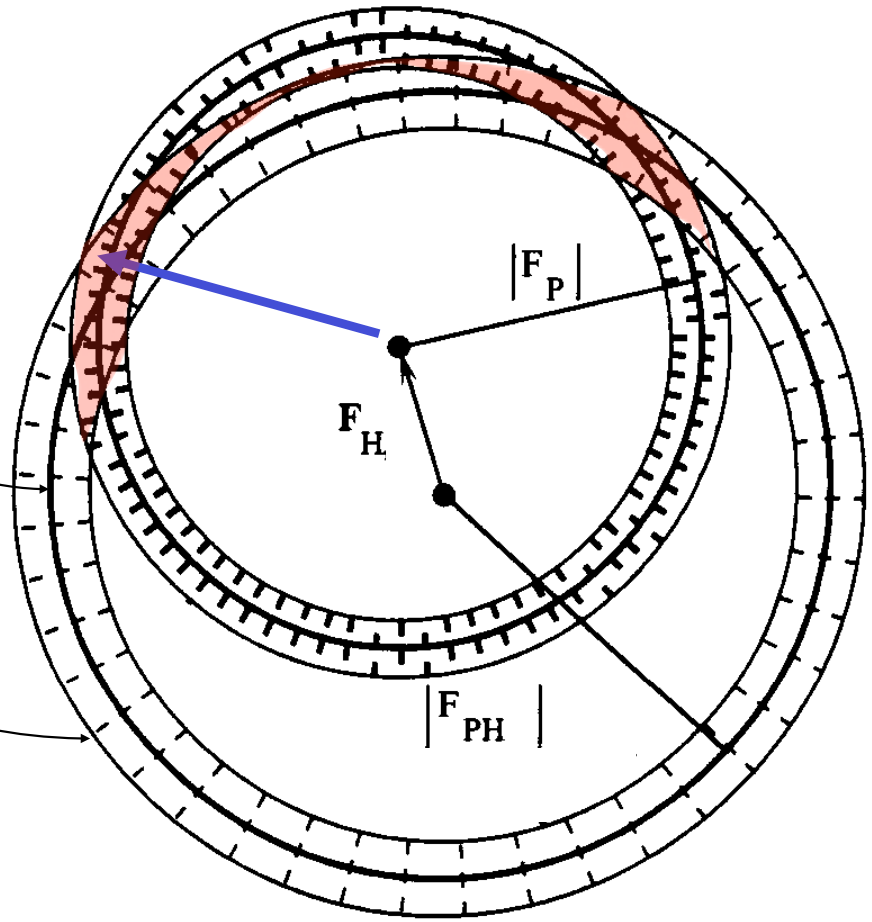
Protein Structure Determination
Xray Lecture 7, 8

Anomalous Dispersion
Density modification
Refinement

Phase probability distribution

Radii are F_p and $\mathbf{k}*F_{ph}$

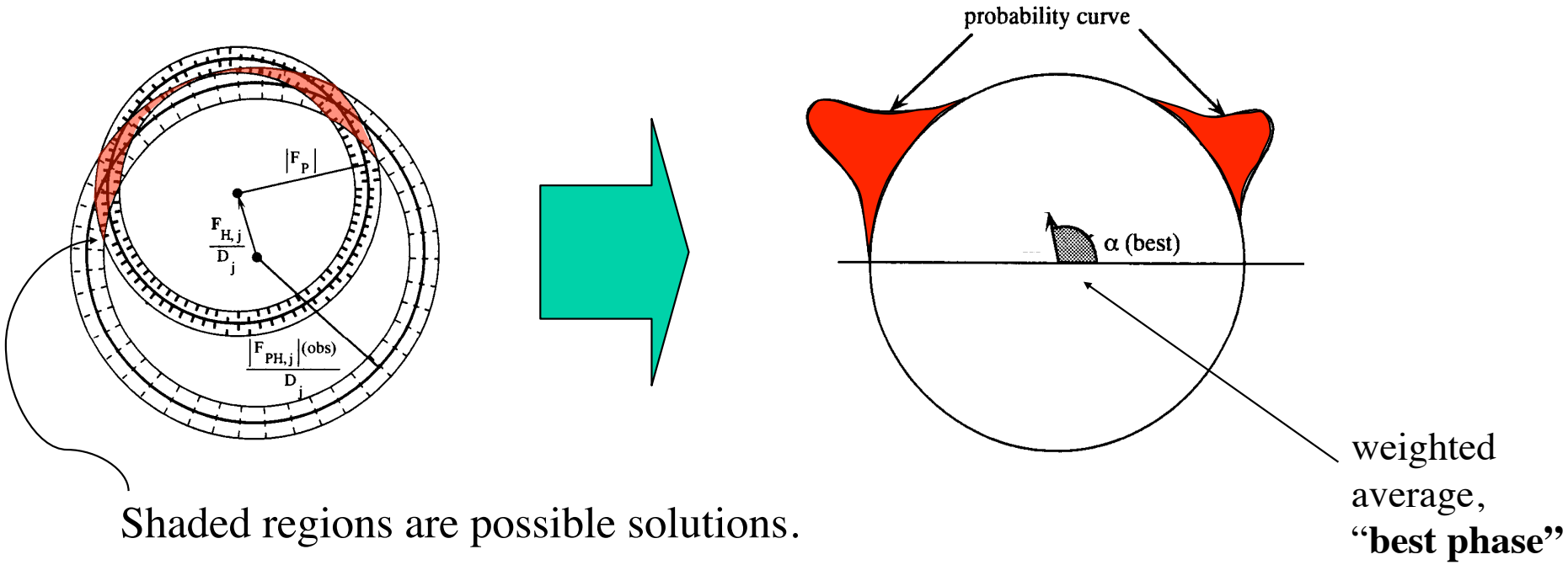
Width are σ_p and $\mathbf{k}*\sigma_{ph}$



The red area are the places in Argand space where both F_P and F_{PH}-F_H can be

Most probable versus best phase

The degree of overlap in the amplitude rings is a measure of phase probability.

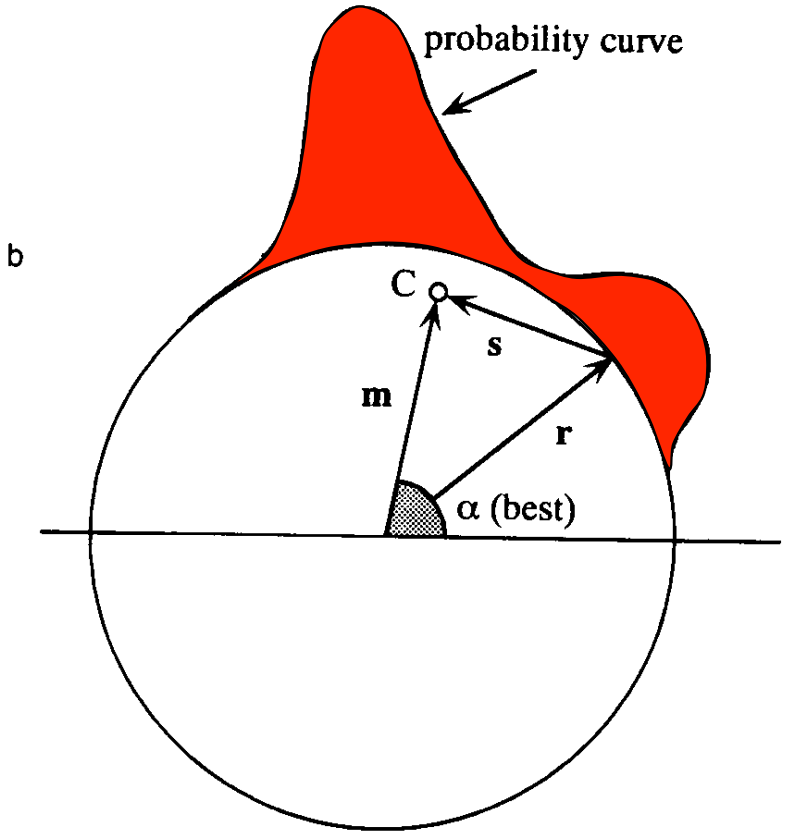


Shaded regions are possible solutions.

The "best phase" for calculating density is the probability-weighted average phase. It may have low probability, even zero!

Figure of merit

Figure of merit "m" is a measure of how good the phases are.



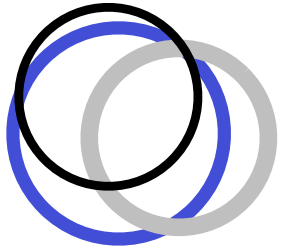
C is the "center of mass" of a ring of phase probabilities (probability is the "mass"). Assume the radius of the ring is 1. If the probabilities are sharply distributed, **m≈1**. If they are distributed widely, **m** is smaller.

$$\mathbf{F_{best}(hkl) = m\,F(hkl)\,e^{-i\alpha_{best}}}$$

Reverse FT using figure of merit, *m*, and best phase.

$$\rho(\mathrm{r}) = \sum_h F_{best}(h) e^{-2\pi h \bullet r} = \sum_h m \left| F(h) \right| e^{i\alpha_{best}} e^{-2\pi h \bullet r}$$

In class exercise: Figure of Merit



F_P=5.00 σ=0.5

F_{PH1}=5.50 σ=0.8 F_{H1}=2.23 α_{H1}=-63.4°

F_{PH2}=4.50 σ=0.9 F_{H2}=0.50 α_{H2}=-164°

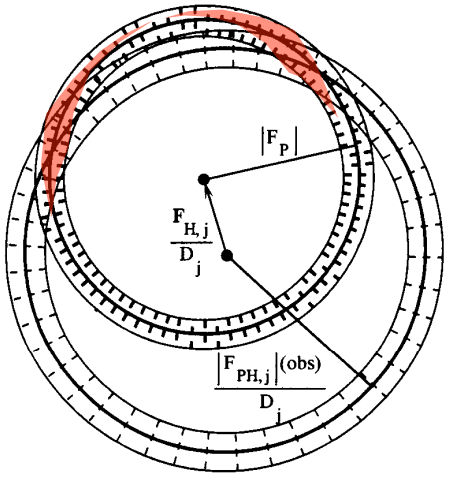
(1) Draw three circles separated by vectors F_{H1} and F_{H2}.

(2) Draw circular "error bars" of width 2σ.

(3) Draw circle plot of F_p phase probabilities.

(4) By eye, estimate the centroid *c* of probability.

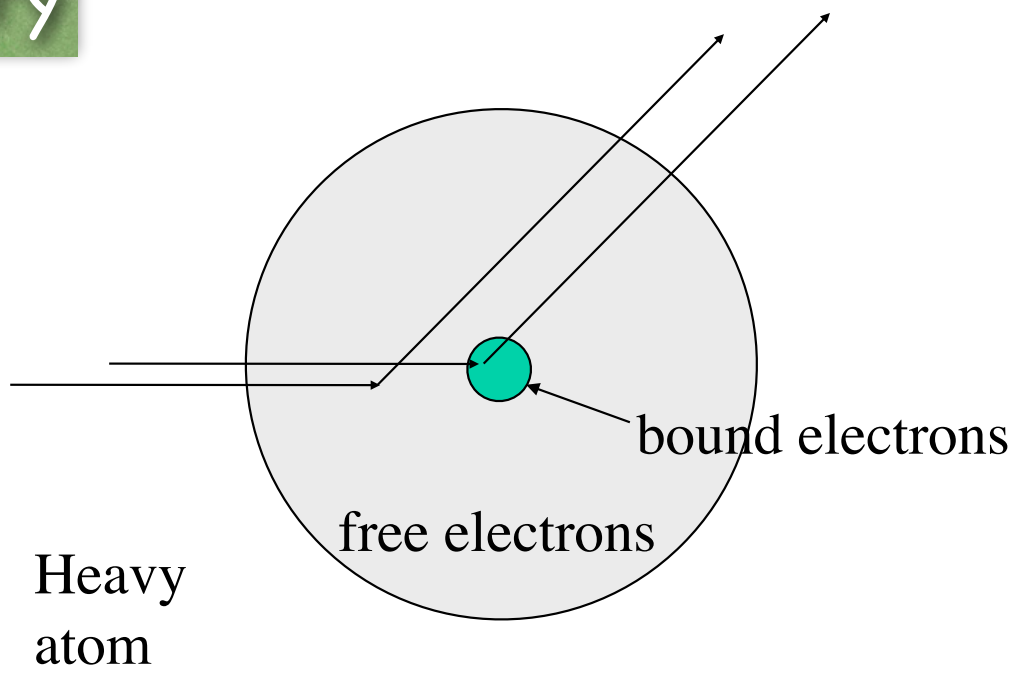
(5) What is the Figure of Merit, *m*?



Anomalous dispersion

Theory

real

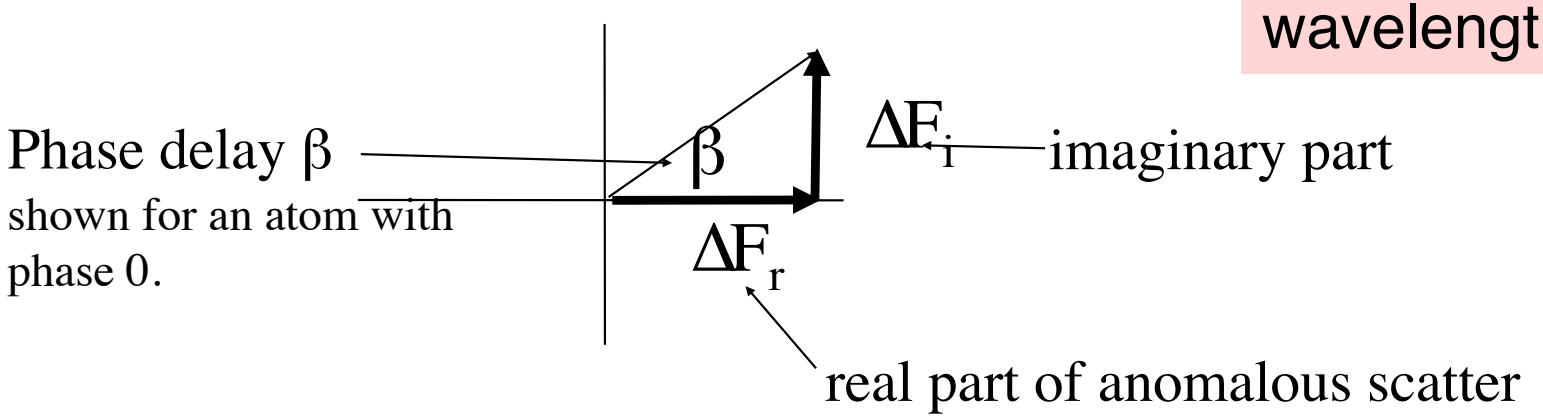


Innermost, bound electrons scatter with a phase shift, relative to free electrons.

Think of it as delayed scatter.

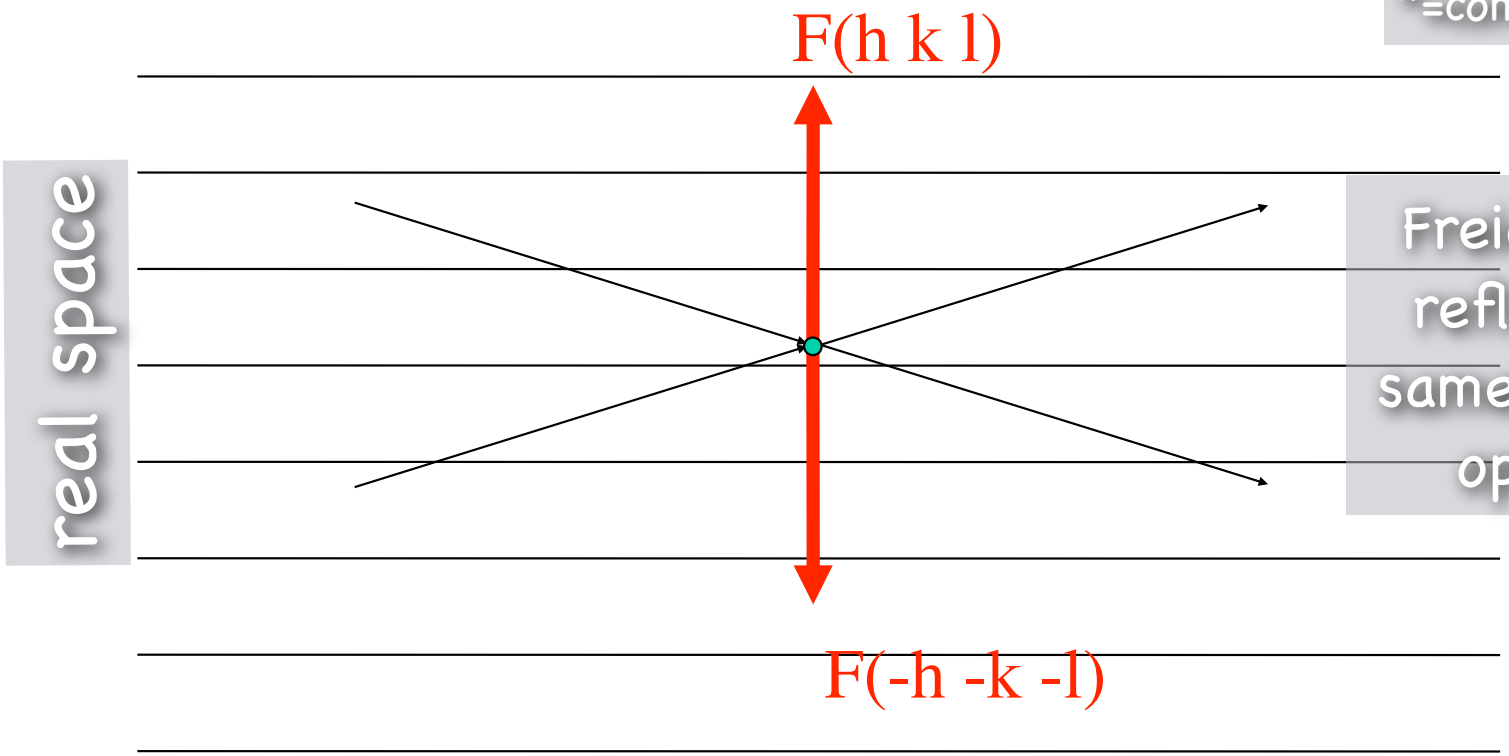
Only some heavy atoms are anomalous at some wavelengths.

Argand

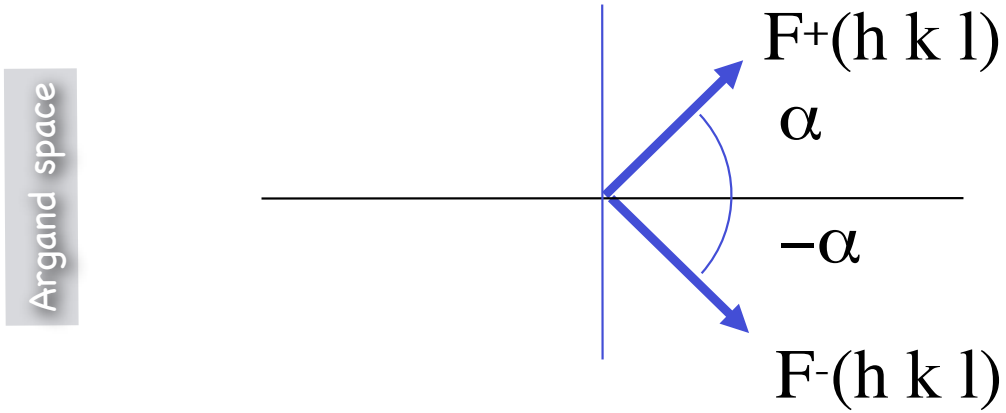


Friedel's Law: F(hkl) = F*(-h-k-l)

*=complex conjugate



Freidel reflection reflect from the same Bragg planes, opposite side.



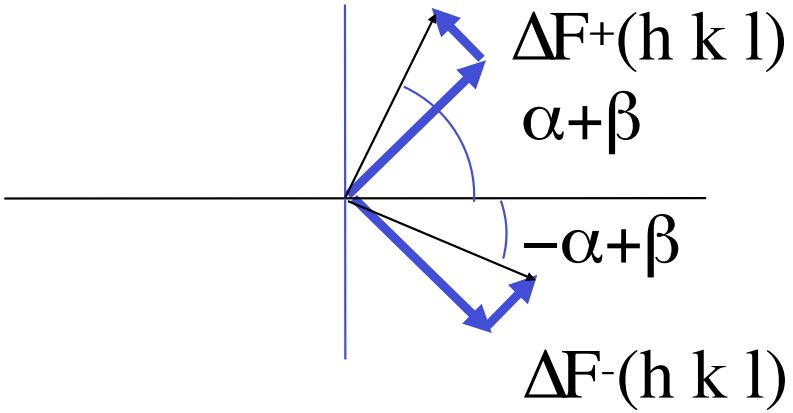
Friedel's Law

…made to be broken…

Friedels have same amplitude, opposite sign.

Anomalous dispersion breaks Friedel's Law

We call one Friedel the "plus" and the other the "minus" depending on the sign of h.



+ and - contributions for just a heavy atom with anomalous scattering.

Phase delay is always a positive angle whether it's the + or the - reflection. So Freidel's Law is now broken.

Argand

Multiwavelength Anomalous Dispersion

Taking advantage of the breaking of Friedel's Law to get three dataset from one crystal.

Selenomethionine (sMet) is methionine with the Sulfur replaced by a Selenium.

Selenium scatters anomalously at 0.98Å, and normally at 1.54Å.

In Multiwavelength Anomalous Dispersion (**MAD**) we collect data on one crystal, two wavelengths.

Result is 3 data sets from one crystal

subscript key: wavelength ($\lambda 1$ or $\lambda 2$) and Friedel (+ or -)

Datasets: $F_{\lambda 1}$ $F^{+}_{\lambda 2}$ $F^{-}_{\lambda 2}$

MAD vector math

λ_1 is a wavelength where Se does not absorb, normal diffraction.

λ_2 is a wavelength where Se does absorb, anomalous ΔF is added.

$F^{-}\lambda_1(h\ k\ l)$

$F^{+}\lambda_1(h\ k\ l)$

} Friedel mates collected at non-anomalous wavelength (equivalent)

$F^{-}\lambda_2(h\ k\ l)$

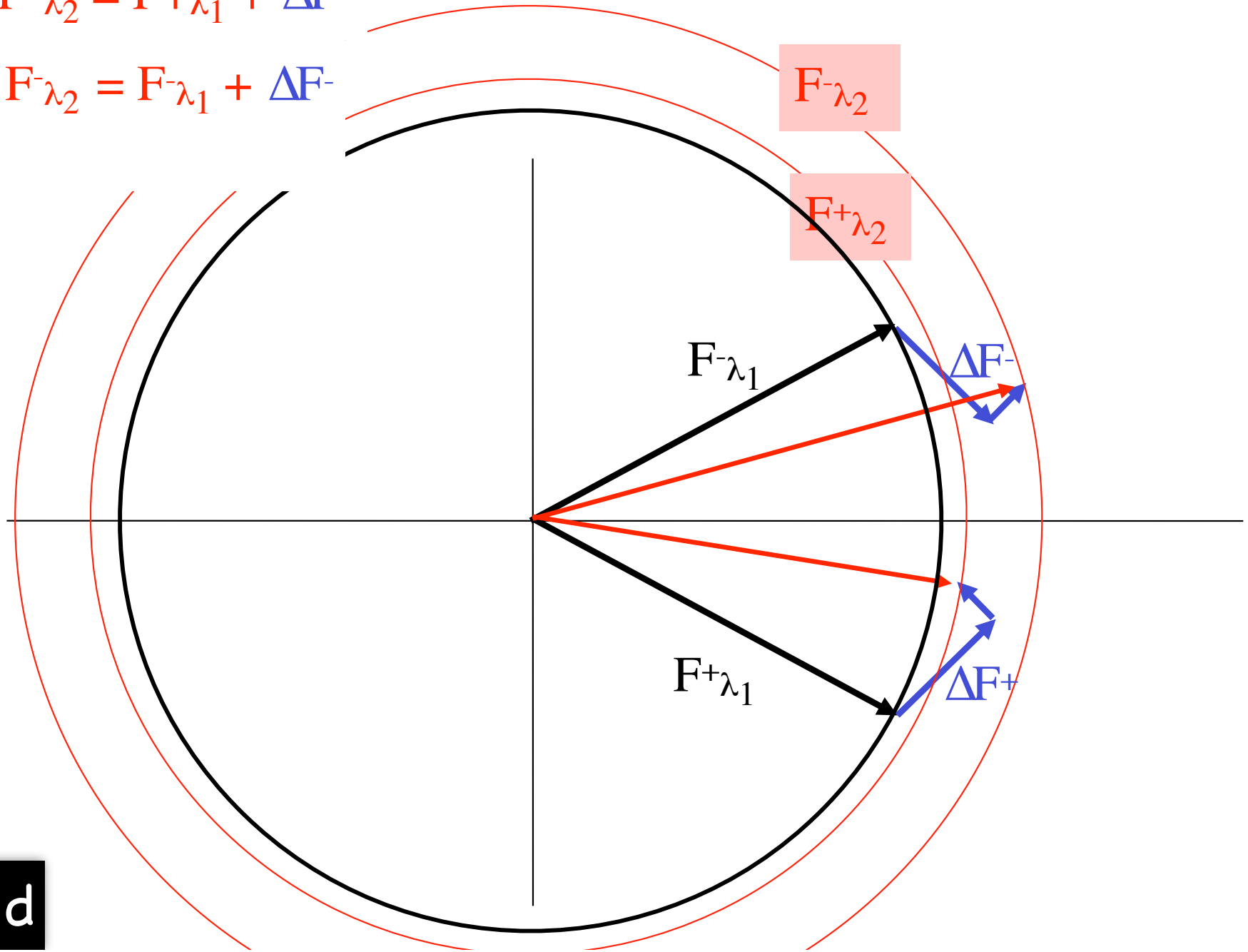
$F^{+}\lambda_2(h\ k\ l)$

Friedel mates collected anomalous wavelength (not equivalent)

After adding anomalous, vector addition is not the mirror image

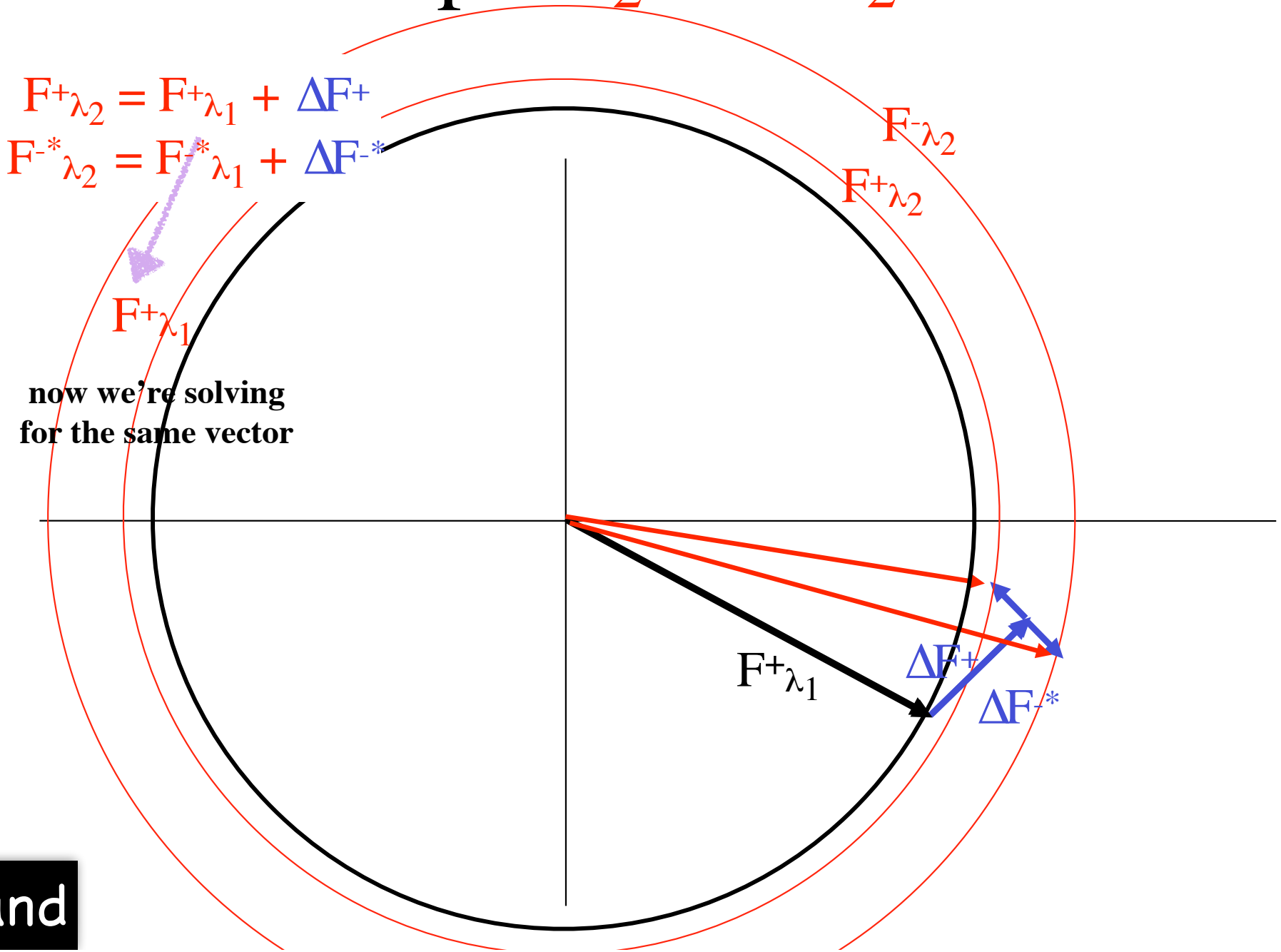
$F^{+}\lambda_2 = F+\lambda_1 + \Delta F^{+}$

$F^{-}\lambda_2 = F^{-}\lambda_1 + \Delta F^{-}$



Argand

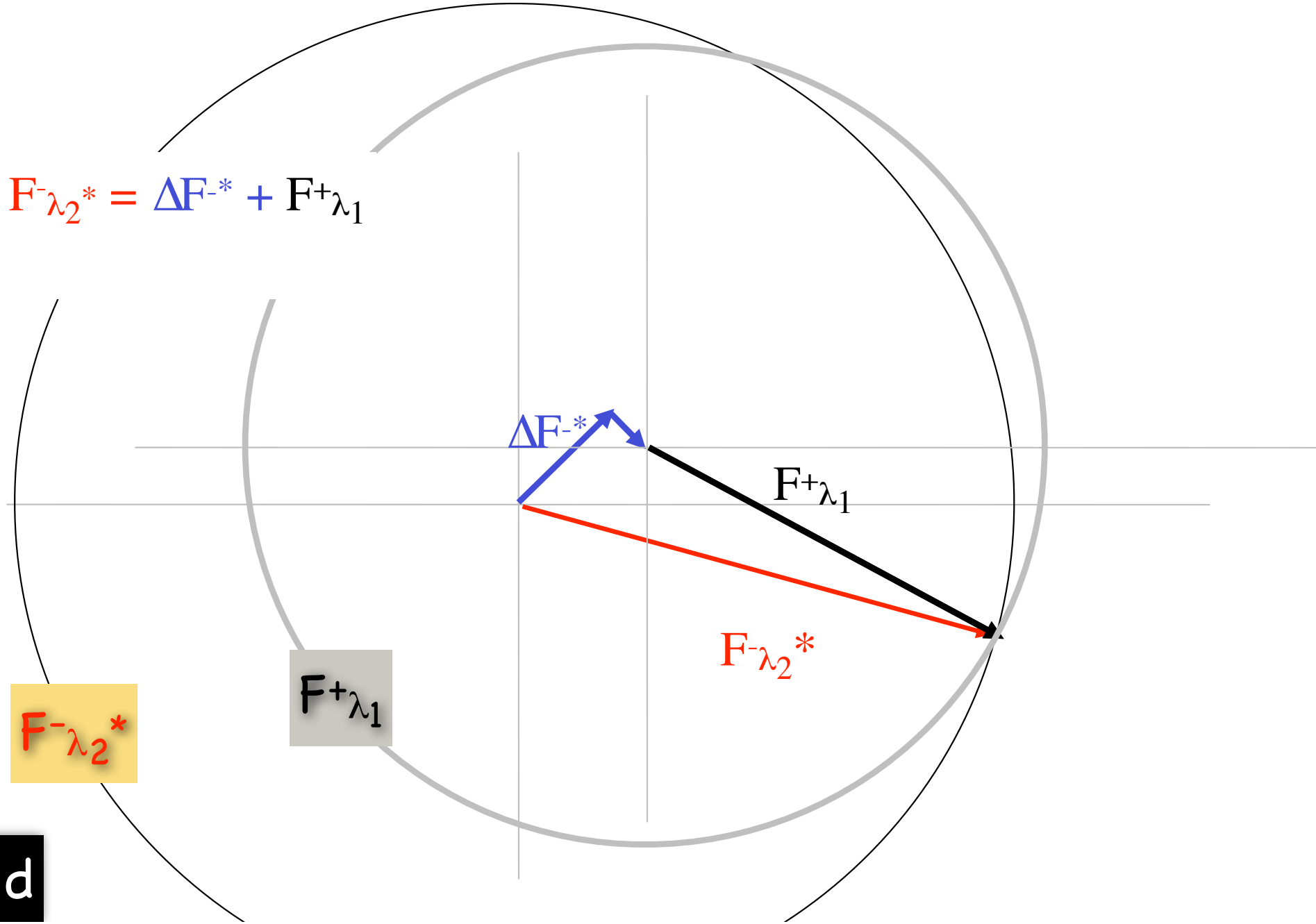
Flip $F^{-}\lambda_2$ to $F^{-}\lambda_2{}^{*}$



Argand

switch order of addition (offset circles)

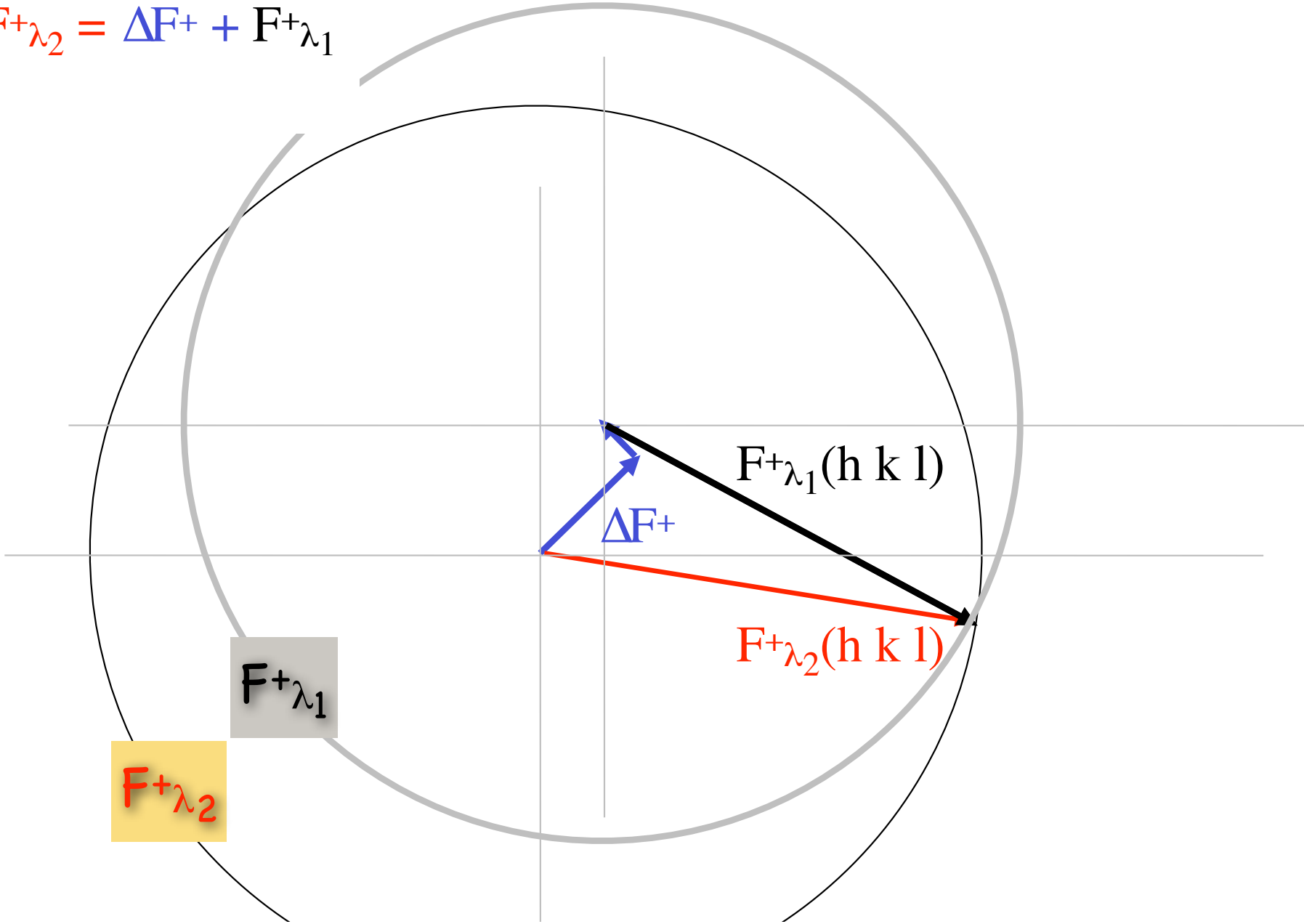
$F^{-}\lambda_2^{*} = \Delta F^{-*} + F^{+}\lambda_1$



Argand

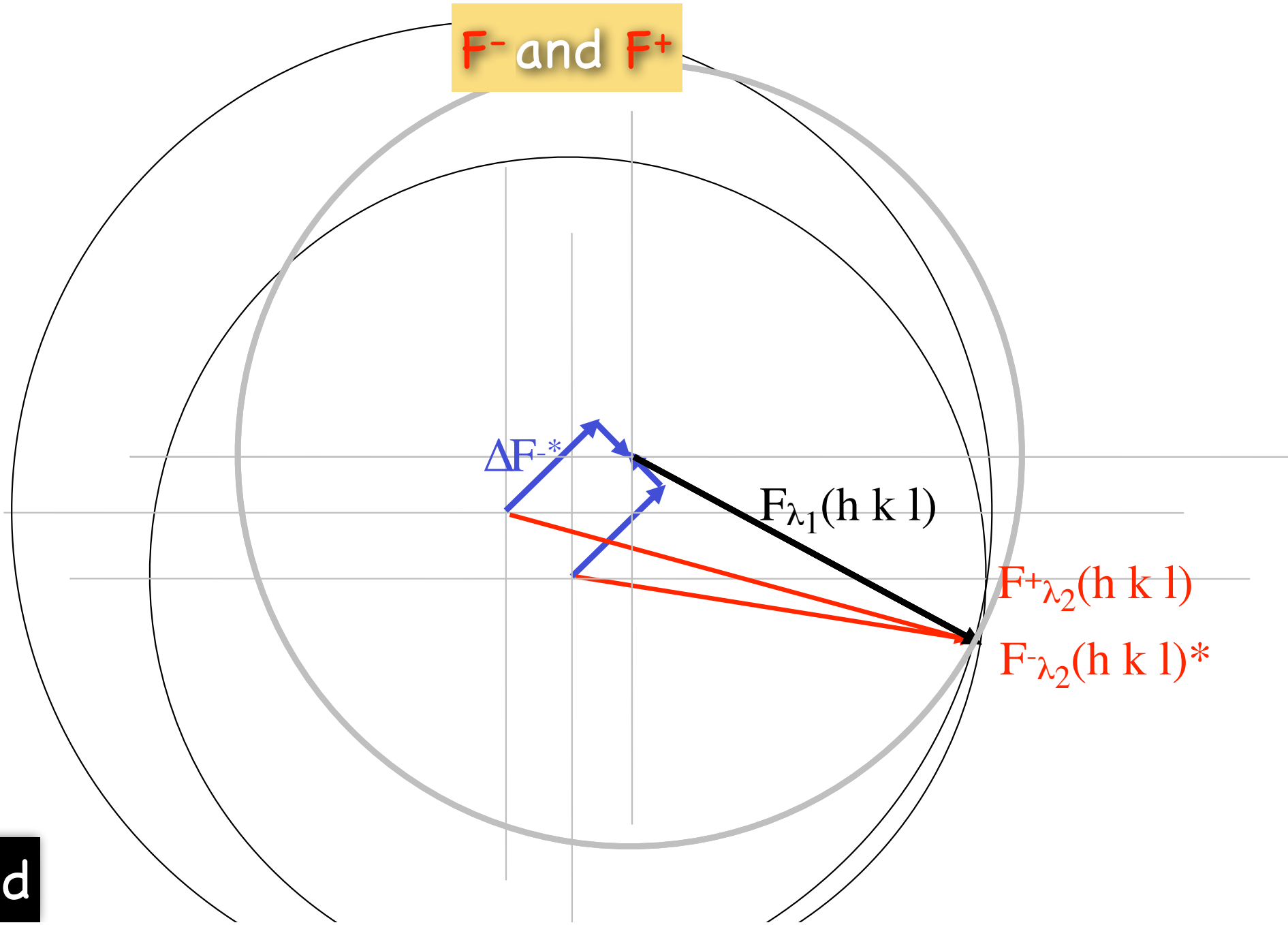
switch order of addition (offset circles)

$F^+_{\lambda_2} = \Delta F^+ + F^+_{\lambda_1}$



Argand

complete Harker diagram for MAD



Argand

Exercise 5: Solve a MAD Harker

Use compass, ruler, graph, colored pencils. Do the vector addition.

measured intensities

$F_{\lambda_1} = 12.0$ $\alpha_{\lambda_1} = ????$

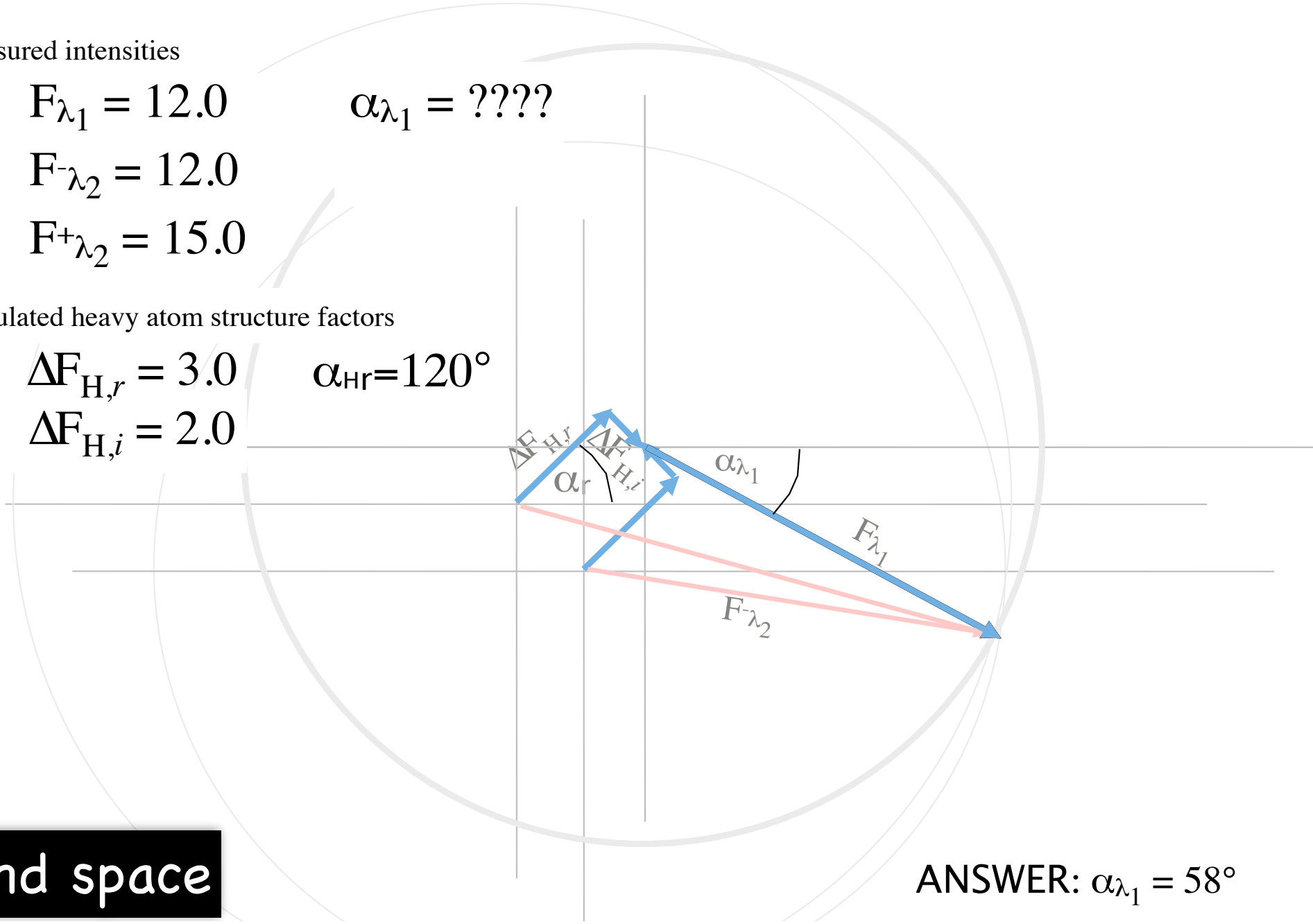
$F^{-}_{\lambda_2} = 12.0$

$F^{+}_{\lambda_2} = 15.0$

calculated heavy atom structure factors

$\Delta F_{H,r} = 3.0$ $\alpha_{Hr}=120°$

$\Delta F_{H,i} = 2.0$



Argand space

ANSWER: $\alpha_{\lambda_1} = 58°$

Solving MAD using math

$$F^{-}_{\lambda_2}(h\ k\ l)^* = F_{\lambda_1}(h\ k\ l) + \Delta F^{-}(h\ k\ l)$$

$$= |F^{-}_{\lambda_2}(h\ k\ l)|e^{-i\alpha_{2}^{-}} = |F_{\lambda_1}(h\ k\ l)|e^{i\alpha_1} + \Delta F^{-}(h\ k\ l)e^{i(\alpha_h-\beta_h)}$$

real part $|F^{-}_{\lambda_2}(h\ k\ l)|\cos(\alpha_{2}^{-}) = |F_{\lambda_1}(h\ k\ l)|\cos(\alpha_1) + |\Delta F^{-}(h\ k\ l)|\cos(\alpha_h-\beta_h)$

imaginary part $-|F^{-}_{\lambda_2}(h\ k\ l)|\sin(\alpha_{2}^{-}) = |F_{\lambda_1}(h\ k\ l)|\sin(\alpha_1) + |\Delta F^{-}(h\ k\ l)|\sin(\alpha_h-\beta_h)$

$$F^{+}_{\lambda_2}(h\ k\ l) = F_{\lambda_1}(h\ k\ l) + \Delta F^{+}(h\ k\ l)$$

$$= |F^{+}_{\lambda_2}(h\ k\ l)|e^{i\alpha_{2}^{+}} = |F_{\lambda_1}(h\ k\ l)|e^{i\alpha_1} + \Delta F^{+}(h\ k\ l)e^{i(\alpha_h+\beta_h)}$$

real part $|F^{-}_{\lambda_2}(h\ k\ l)|\cos(\alpha_{2}^{+}) = |F_{\lambda_1}(h\ k\ l)|\cos(\alpha_1) + |\Delta F^{-}(h\ k\ l)|\cos(\alpha_h+\beta_h)$

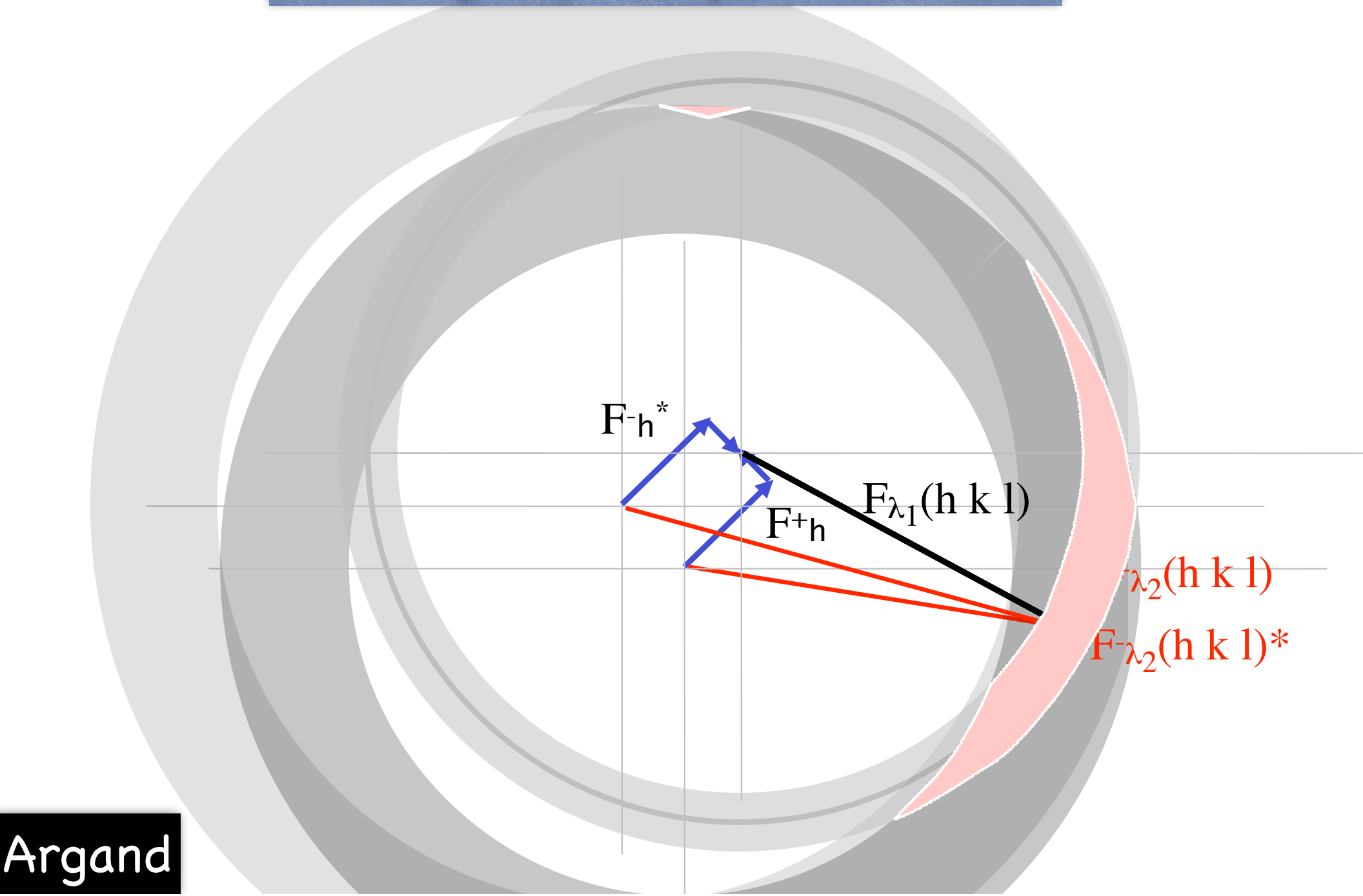
imaginary part $|F^{-}_{\lambda_2}(h\ k\ l)|\sin(\alpha_{2}^{+}) = |F_{\lambda_1}(h\ k\ l)|\sin(\alpha_1) + |\Delta F^{-}(h\ k\ l)|\sin(\alpha_h+\beta_h)$

Four equations , Three unknowns.

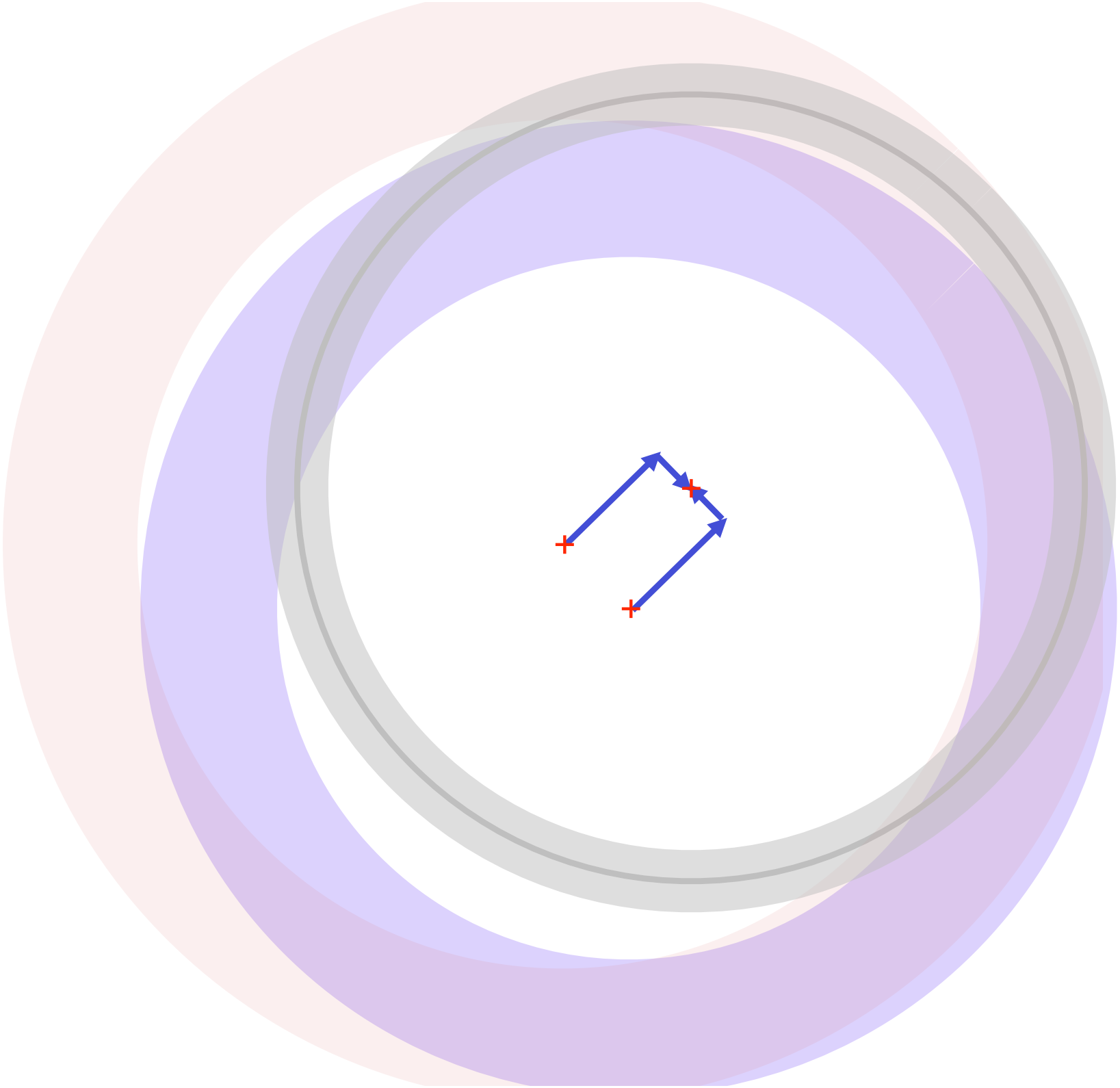
Knowns: all three amplitudes $|F|$, phase and amplitude of real part α_h, phase and amplitude of imaginary part β_h

Unknowns: phases α_1 , α_{2}^{-}, α_{2}^{+}

Phase probability distribution

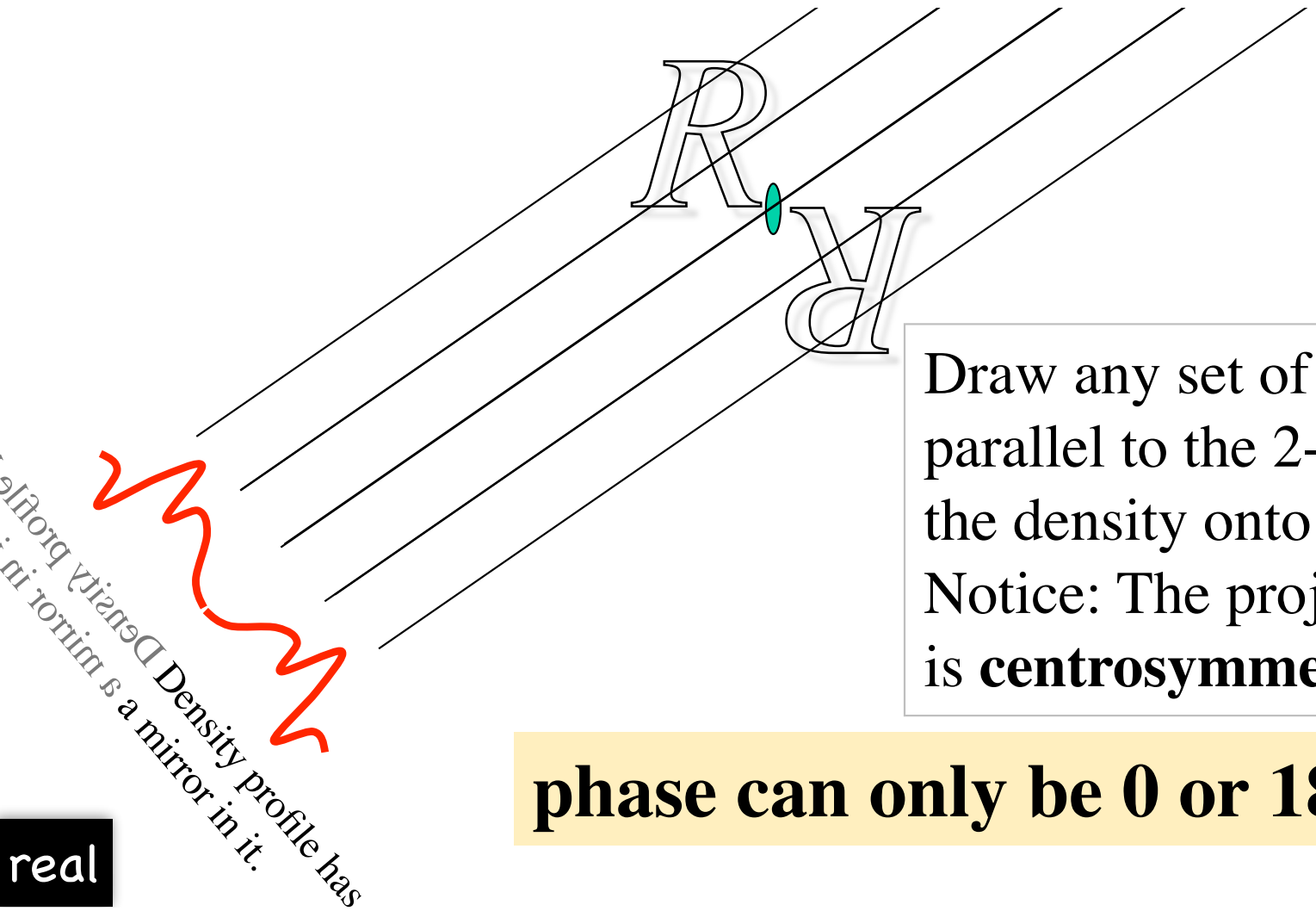


Argand



F's that are both Syms AND Friedels are **centric reflections**

Example: F(0 k l) and F(0 -k -l) when **a** is a 2-fold are centric



Draw any set of Bragg planes parallel to the 2-fold. Project the density onto a line.
Notice: The projected density is **centrosymmetric**.

phase can only be 0 or 180°.

Why Centric reflections are so useful

- All **centric** reflections have real phase = 0° or 180°

Therefore for **centric reflections**:

$|F_{ph}| = |F_p| \pm |F_h|$

....is *exact**. Not approximate.
Plus, no phase error for centric reflections.

± = + if the phase of Fp and Fh are both 0, or both 180,
± = – otherwise.

If $|F_{ph}| < |F_p|$ and the phase of F_h is -180°, what is the phase of F_p?
180°? or -180°?
What if If $|F_{ph}| > |F_p|$?

Summary: Heavy atom phasing methods

SIR = single isomorphous replacement, without anomalous.

Fourier transform uses: Figure-of-merit weighted amplitudes, alpha-best phases and centric reflections.

MIR = multiple isomorphous replacement, without anomalous. Same Fourier terms, but Figure-of-merit is generally better than for SIR.

Summary: Heavy atom phasing methods

SAD = single wavelength anomalous dispersion. Phases from F+ and F- from one crystal.

Fourier transform uses: Figure-of-merit weighted amplitudes, alpha-best phases and centric reflections.

MAD = multi-wavelength anomalous dispersion. Phases from three datasets from one crystal at 2 wavelengths (or more). F+, F- at anomalous wavelength, and F at non-anomalous wavelength.

Let's calculate an electron density map

A unit wave. Applies a phase shift of $2\pi h \bullet r$. A function of reciprocal lattice position **h** and fractional coordinates r.

$$\rho(\mathrm{r}) = \sum_h F_{best}(h) e^{-2\pi h \bullet r} = \sum_h m \left| F_o(h) \right| e^{i\alpha_{best}} e^{-2\pi h \bullet r}$$

density, a function of real space position r, in fractional coordinates

a structure factor (wave), a function of reciprocal lattice position **h**.

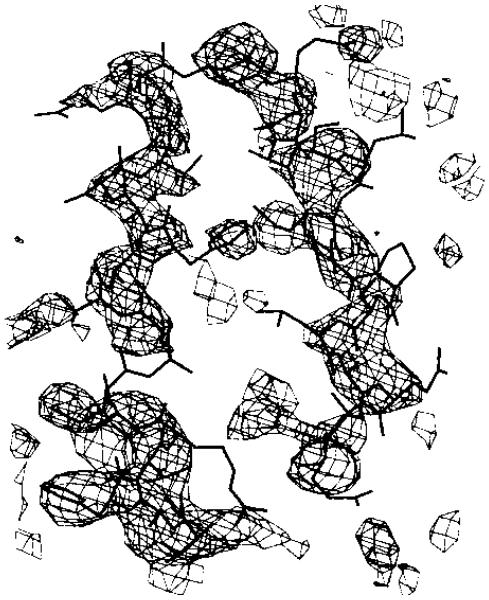
F_{best} is F_o times the figure of merit m. The phase is α_{best}.

Is the initial map good enough?

(1) The map is calculated using α_{best}.

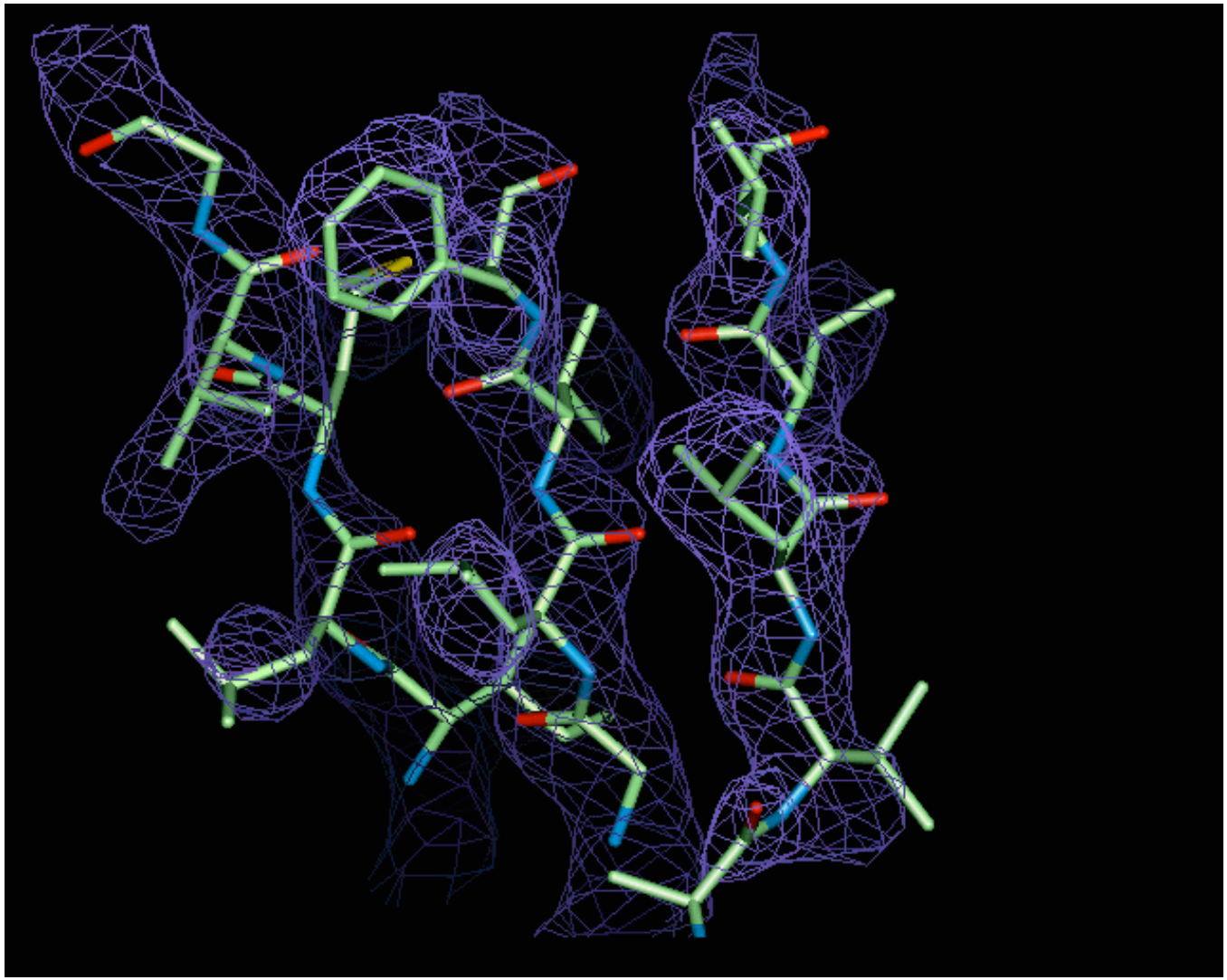
(2) The map is contoured and displayed using {InsightII, MIDAS, XtalView, FRODO, O, ...}

(3) A “trace” is attempted.



real

Model building



e- density cages (1 σ contours) displayed using InsightII

Information used to build the first model:

Density

$$\rho(\mathrm{r}) = \sum_{h} F_{best}(h) e^{-2\pi h \bullet r} = \sum_{h} m \left| F(h) \right| e^{i\alpha_{best}} e^{-2\pi h \bullet r}$$

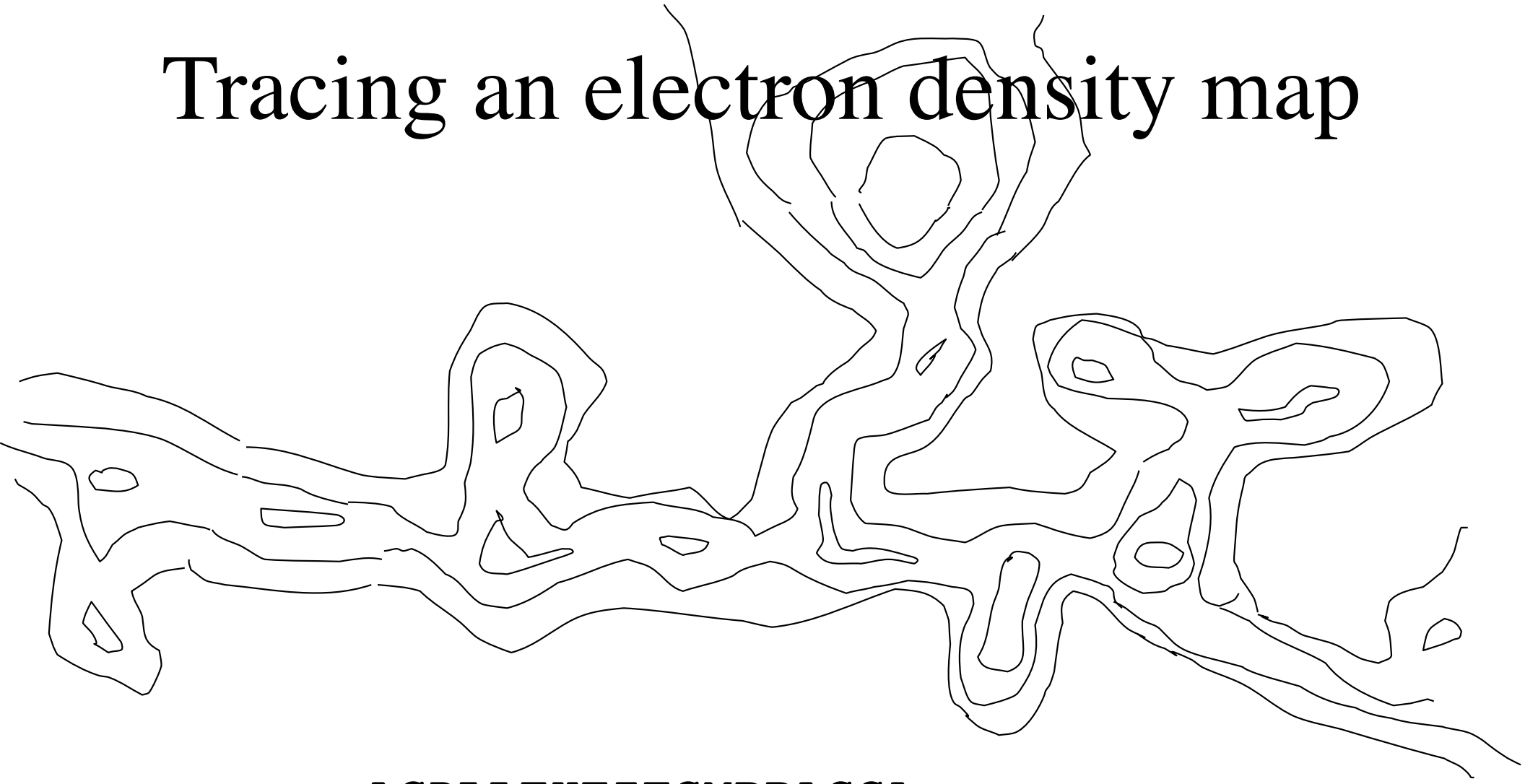
Sequence and Stereochemistry

Models are built initially by identifying characteristic sidechains (by their shape) then tracing forward and backward along the backbone density until all amino acids are in place.

Alpha-carbons can be placed by hand, and numbered, then an automated program will add the other atoms (MaxSprout).

Class exercise:

Tracing an electron density map



sequence: `AGDLLEHEIFGMPPAGGA`

Can you locate the density above in the sequence?

real

R-factor: How good is the model?

Calculate F_{calc}'s based on the model.

Compute R-factor

$$R = \frac{\sum_h \left| |F_{obs}(h)| - |F_{calc}(h)| \right|}{\sum_h |F_{obs}(h)|}$$

Depending on the space group, an R-factor of ~55% would be attainable by scaled random data.

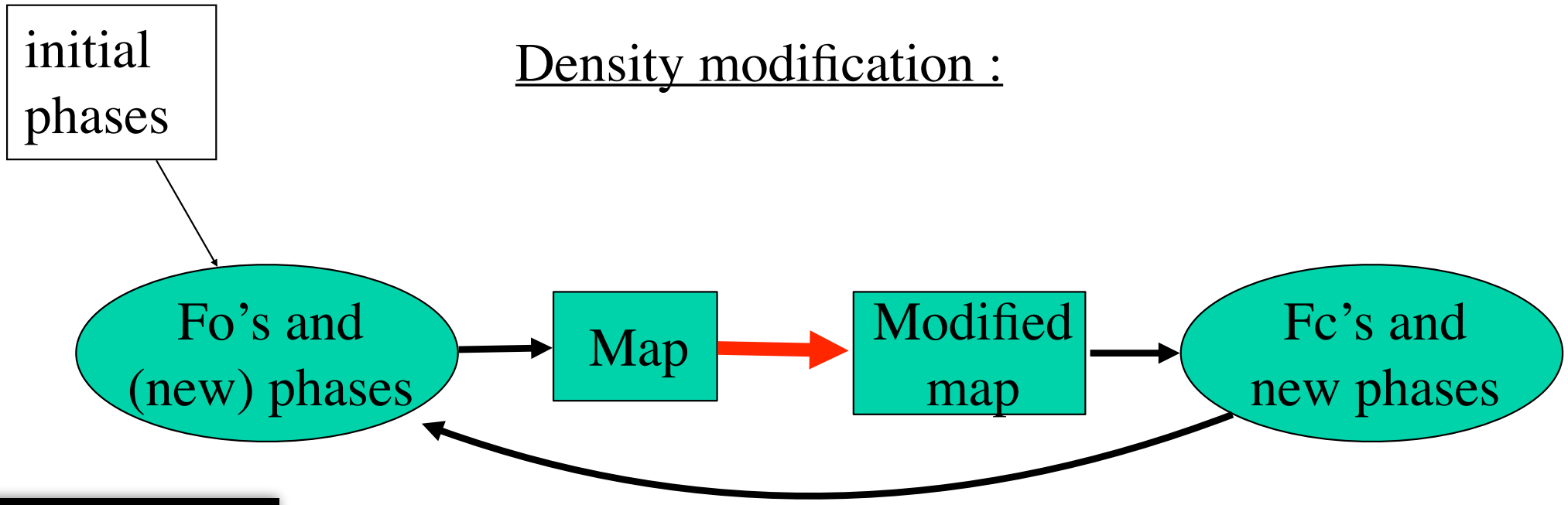
The R-factor must be < ~50%.

Note: It is possible to get a high R-factor for a **correct model**. What kind of mistake would do this?

Argand)

What can you do if the phases are not good enough?

1. Collect more heavy atom derivative data
2. Try *density modification* techniques.



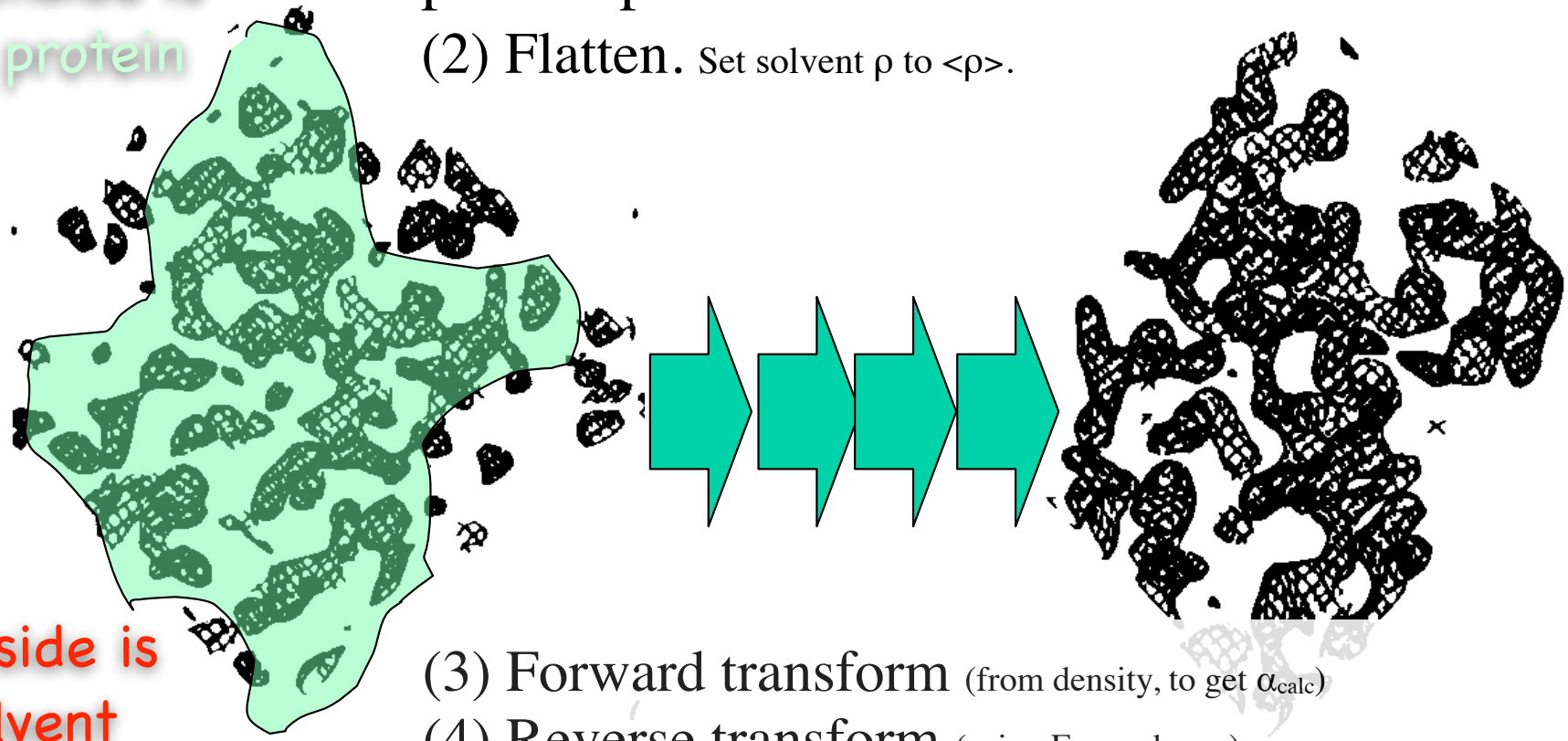
real/reciprocal

Density modification techniques

1. Solvent Flattening

(1) Draw envelope around protein part

(2) Flatten. Set solvent ρ to <ρ>.



(3) Forward transform (from density, to get α_{calc})

(4) Reverse transform (using F_{obs} and α_{calc})

real

2. Skeletonization

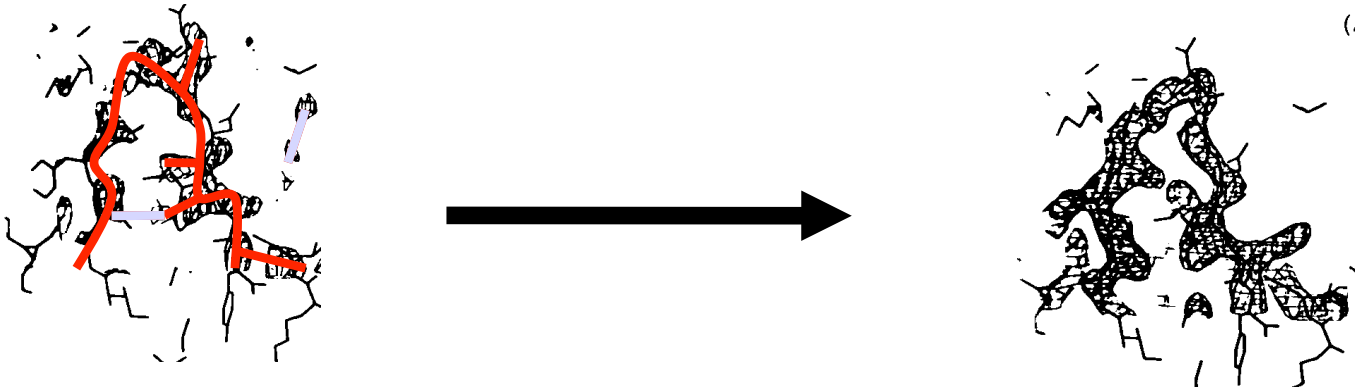
Make the density look like a chain

(1) Calculate map.

(2) Skeletonize it (draw ridge lines)

(3) Prune skeleton so that it is “protein-like”

(4) Back transform the skeleton to get new phases.



Protein-like means: (a) no cycles, (b) no islands

3. Non-crystallographic symmetry

Average NCS-equivalent density

NCS transformation is: $\underline{M}r_1 + v = r_2$

(1) To find $\underline{M}$, use Self Rotation function.

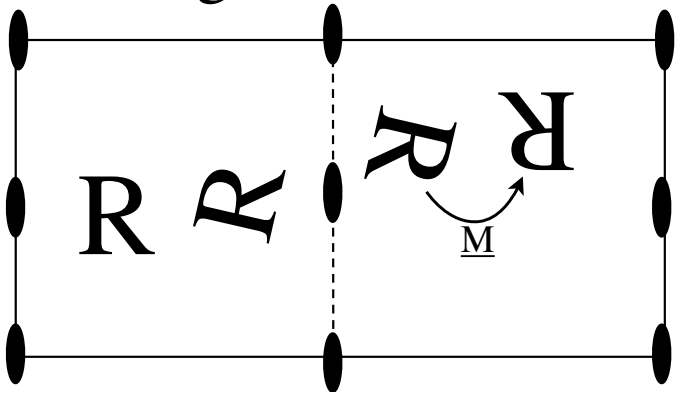
(2) To find v, use Patterson Correlation function.

(3) Calculate map (ρ) using F_{obs} and α_{calc}.

(4) Set $\rho'(r_1) = \rho'(r_2) = <\rho(r_1) + \rho(r_2)>$

(4) Back transform to get new α_{calc}.

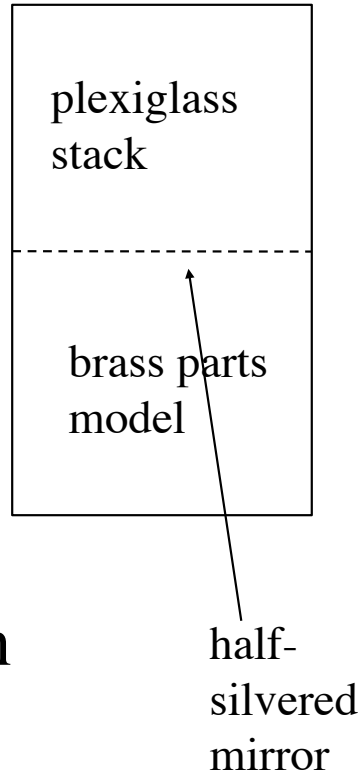
repeat 3-5 till converged



real space

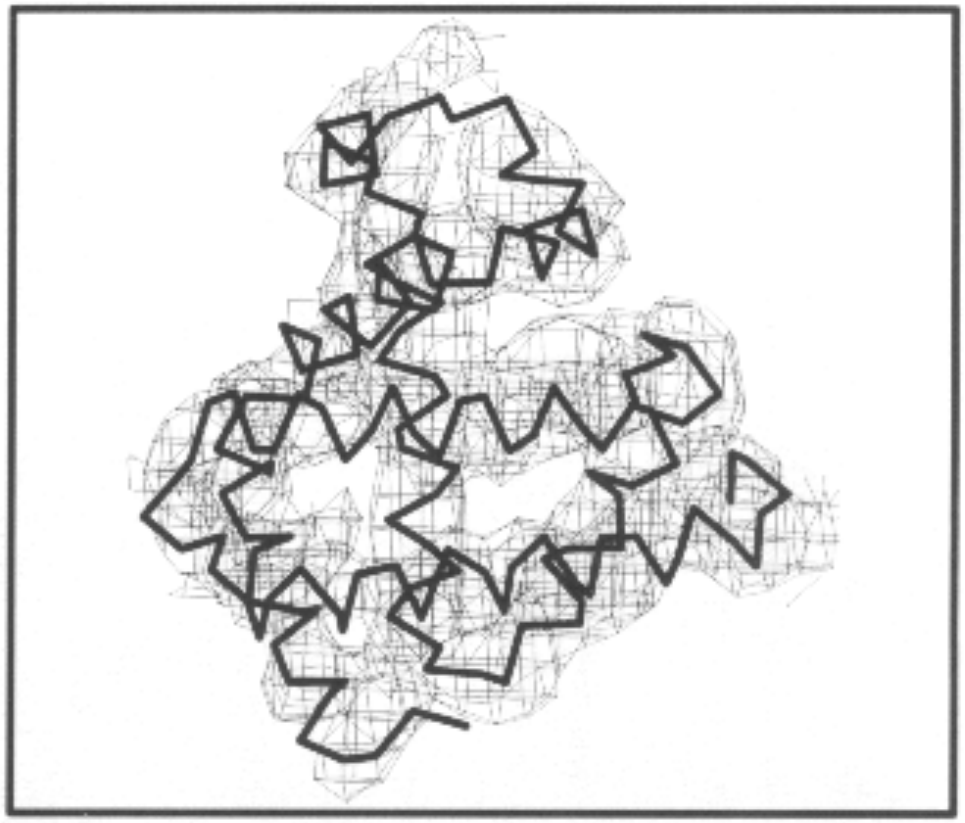
A survey of electron density maps

The early (low-resolution) days:



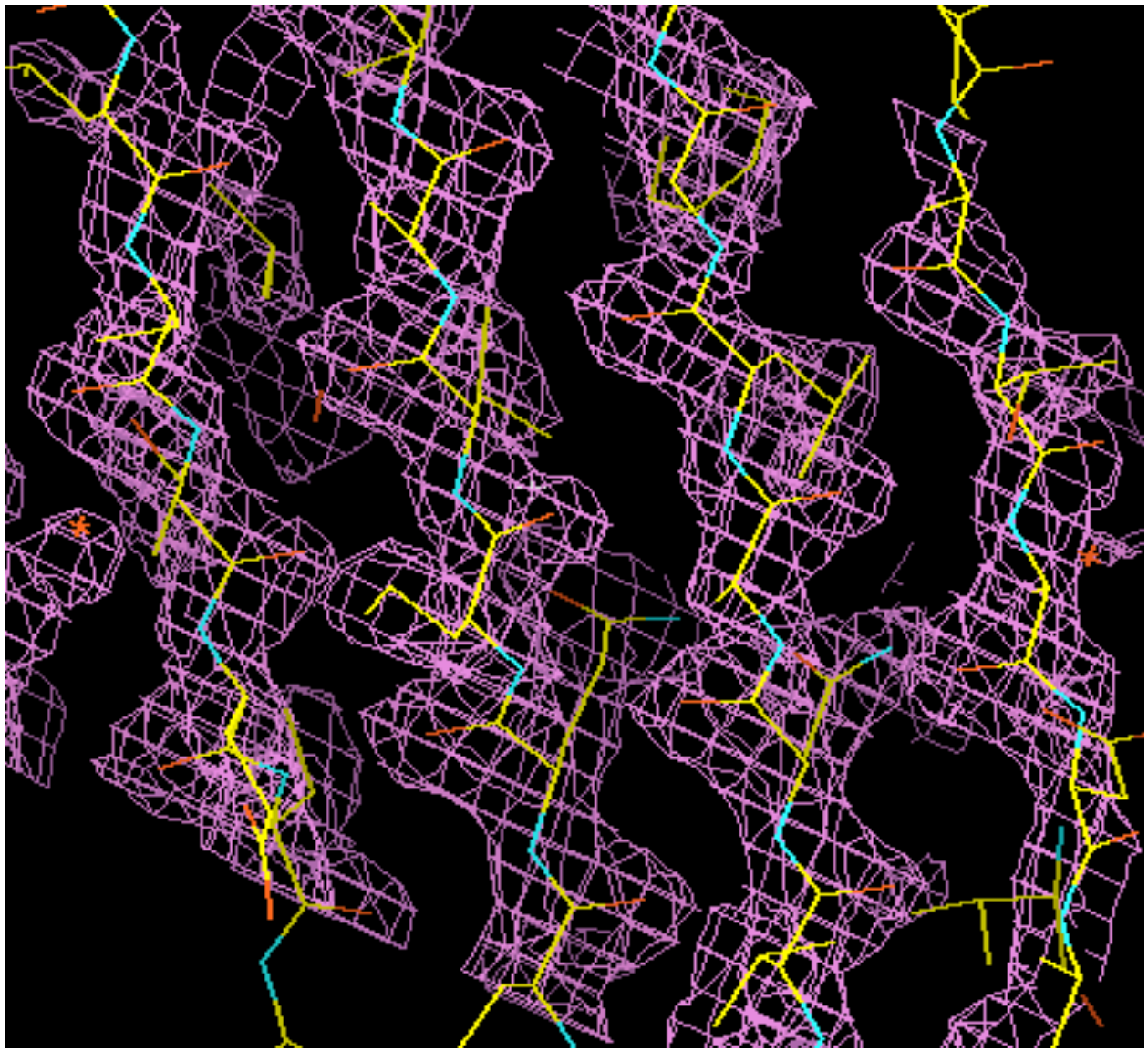
Before computers, maps were contoured on stacked pieces of plexiglass. A “Richards box” was used to build the model.

Low-resolution



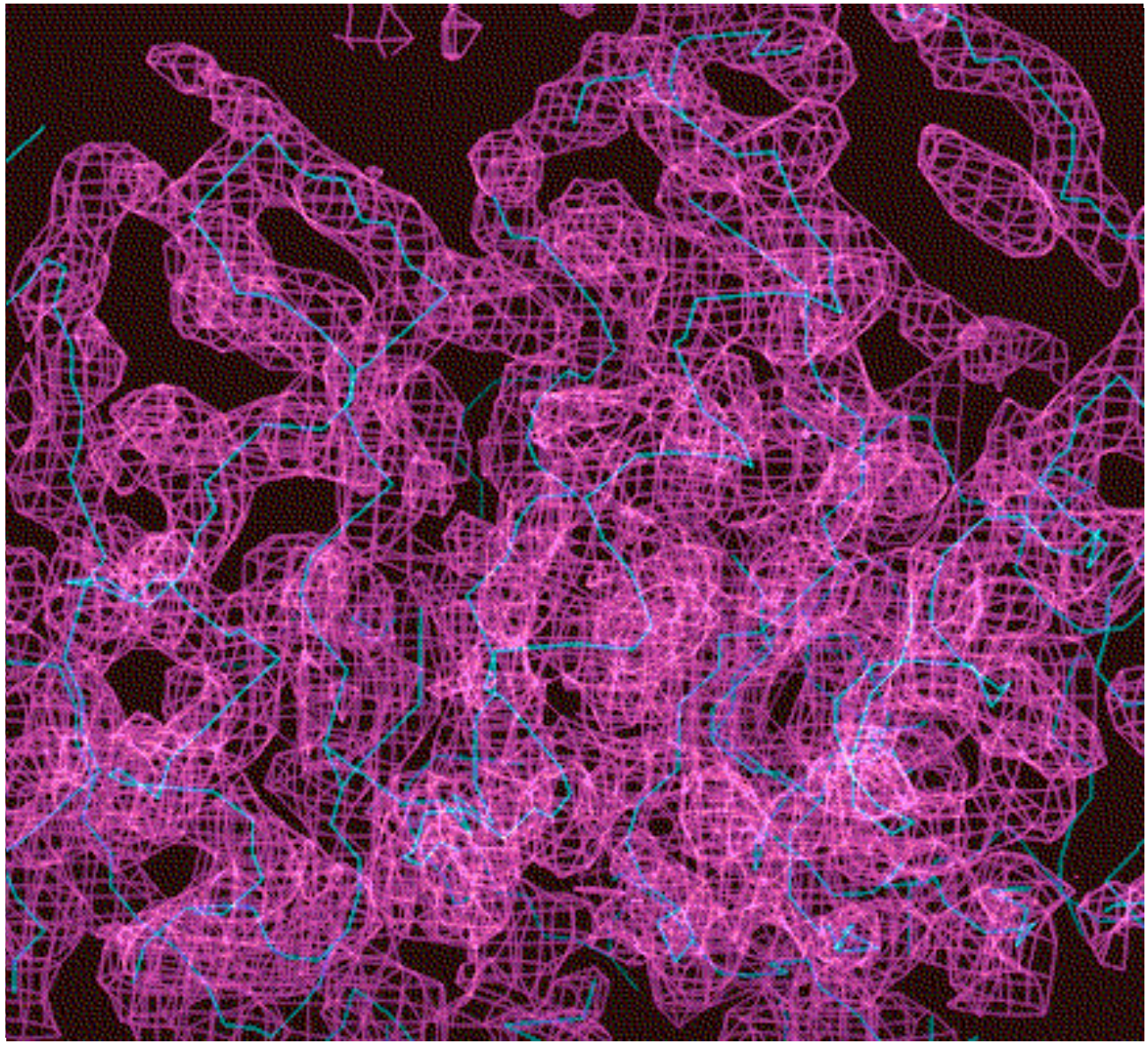
At 4-6Å resolution, alpha helices look like **sausages**.

Medium resolution

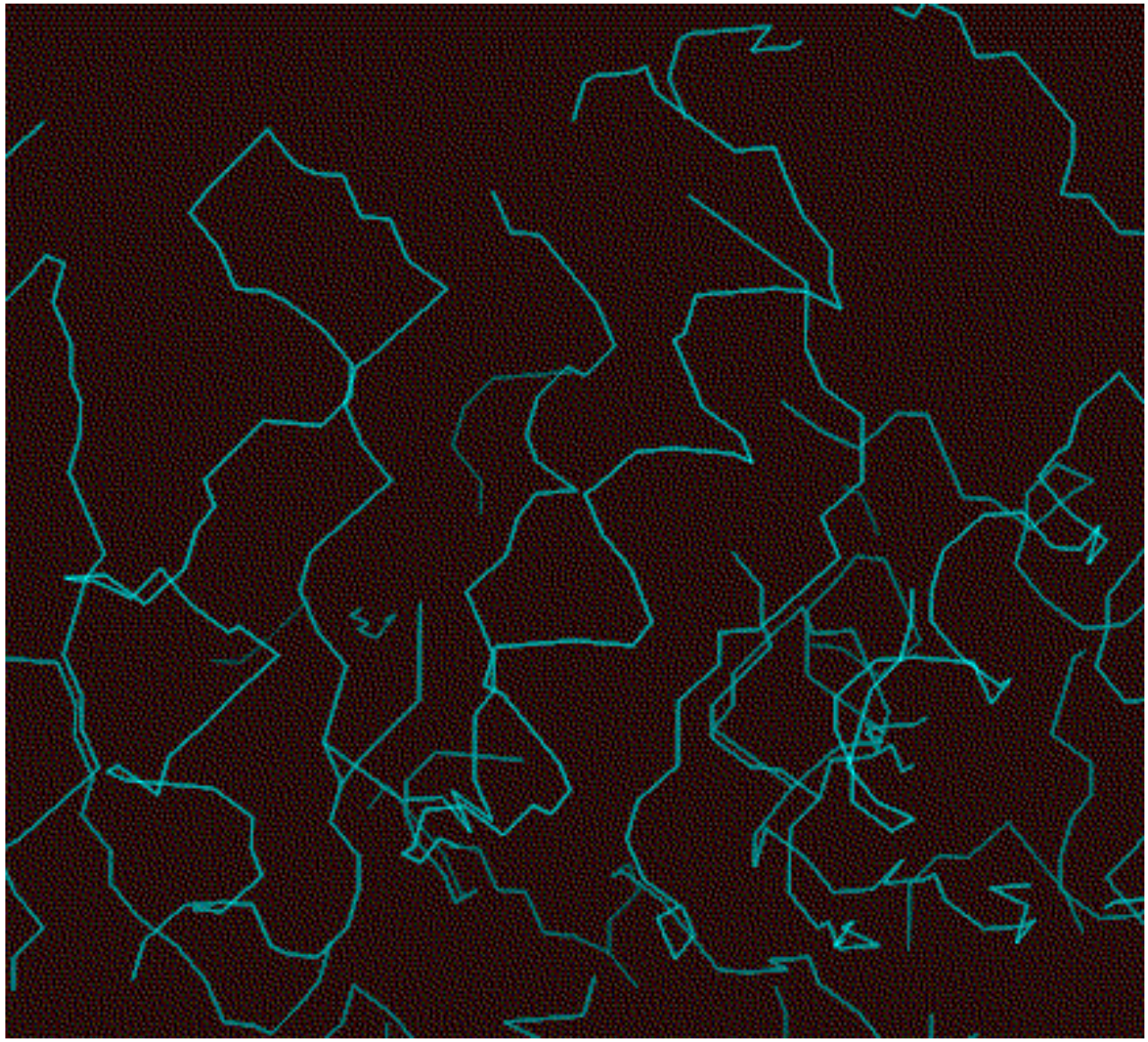


~3Å data is good enough to see the backbone with space in between.

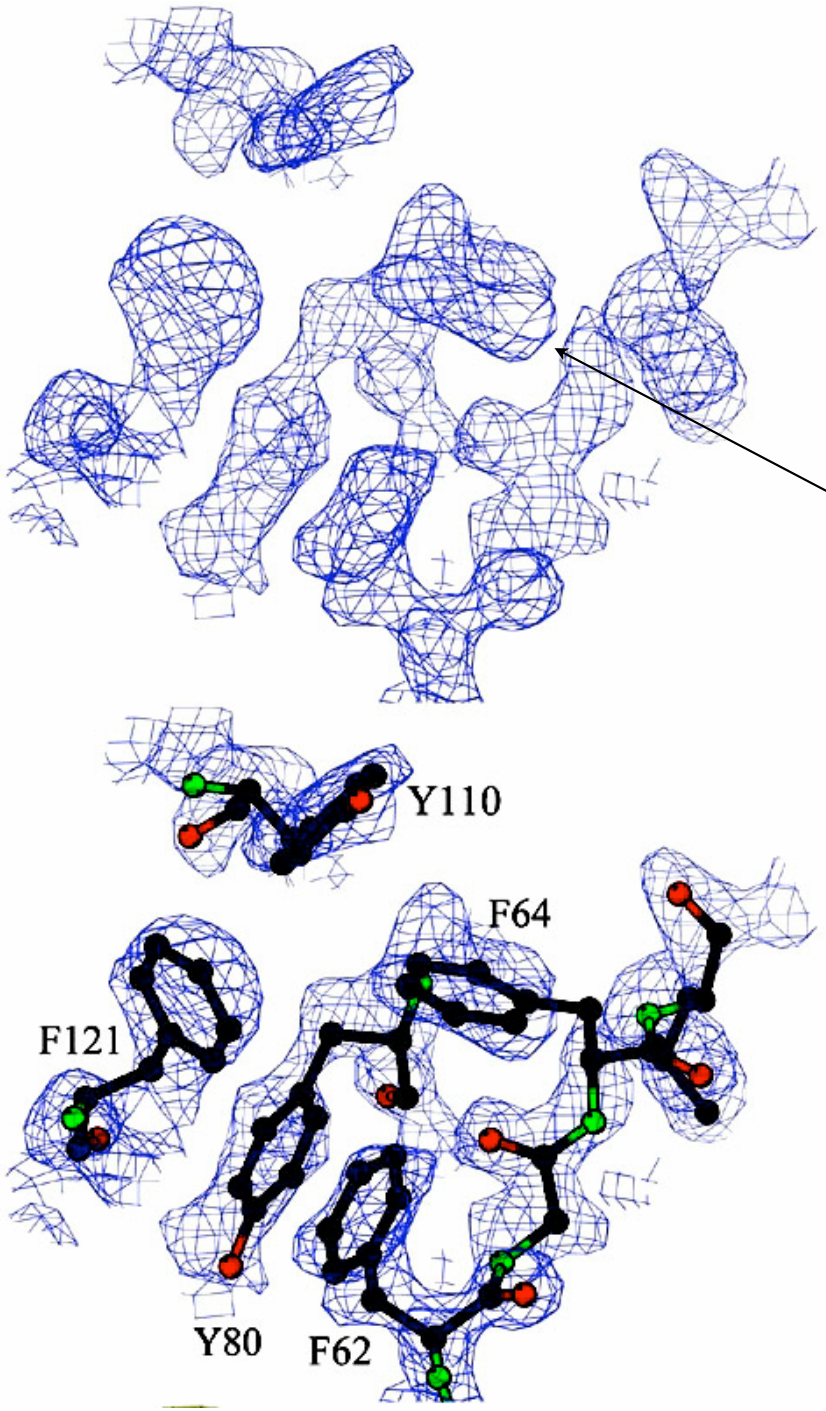
Automatic density tracing



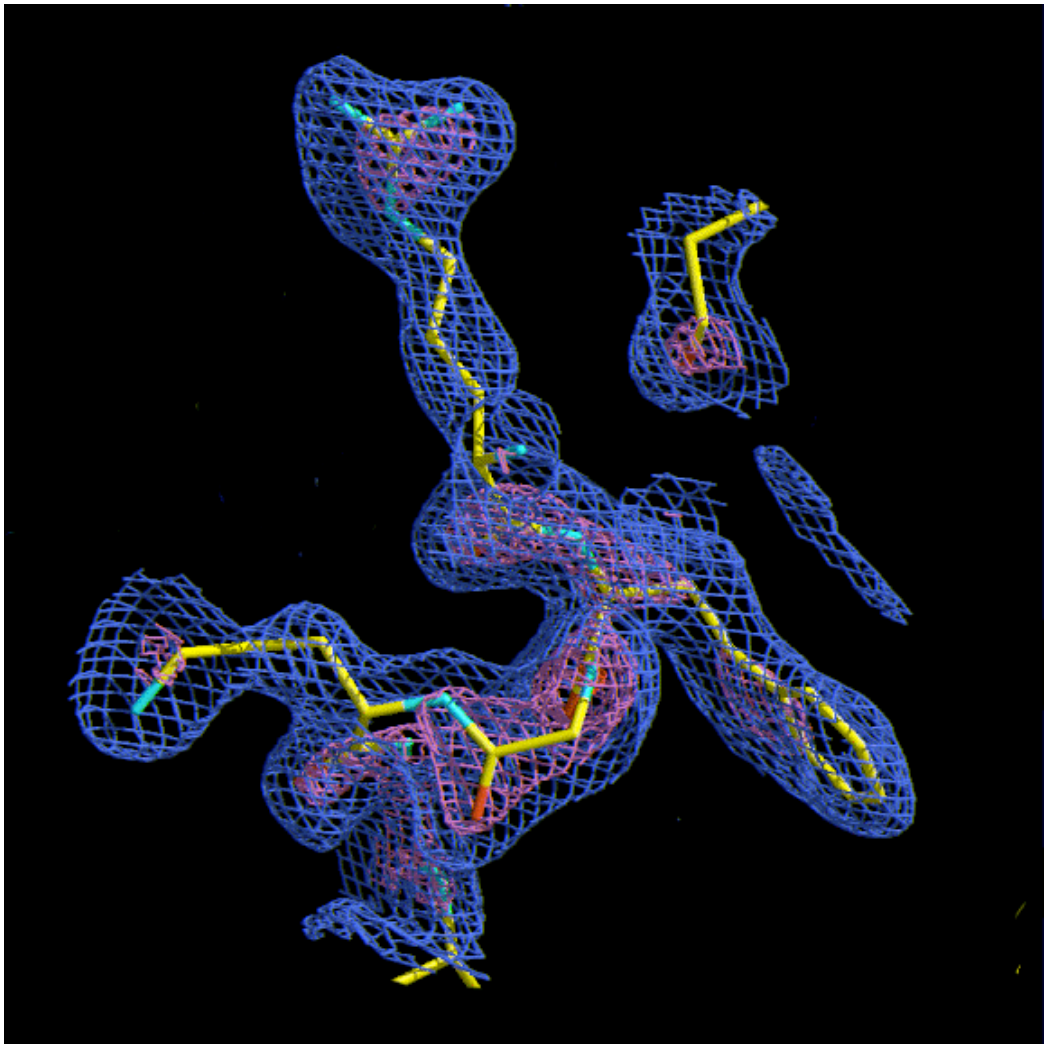
The program BONES traces the density automatically, if the phases are good.



BONES models need to be manually connected and sidechains attached.
MaxSPROUT converts a fully connected trace to an all-atom model.

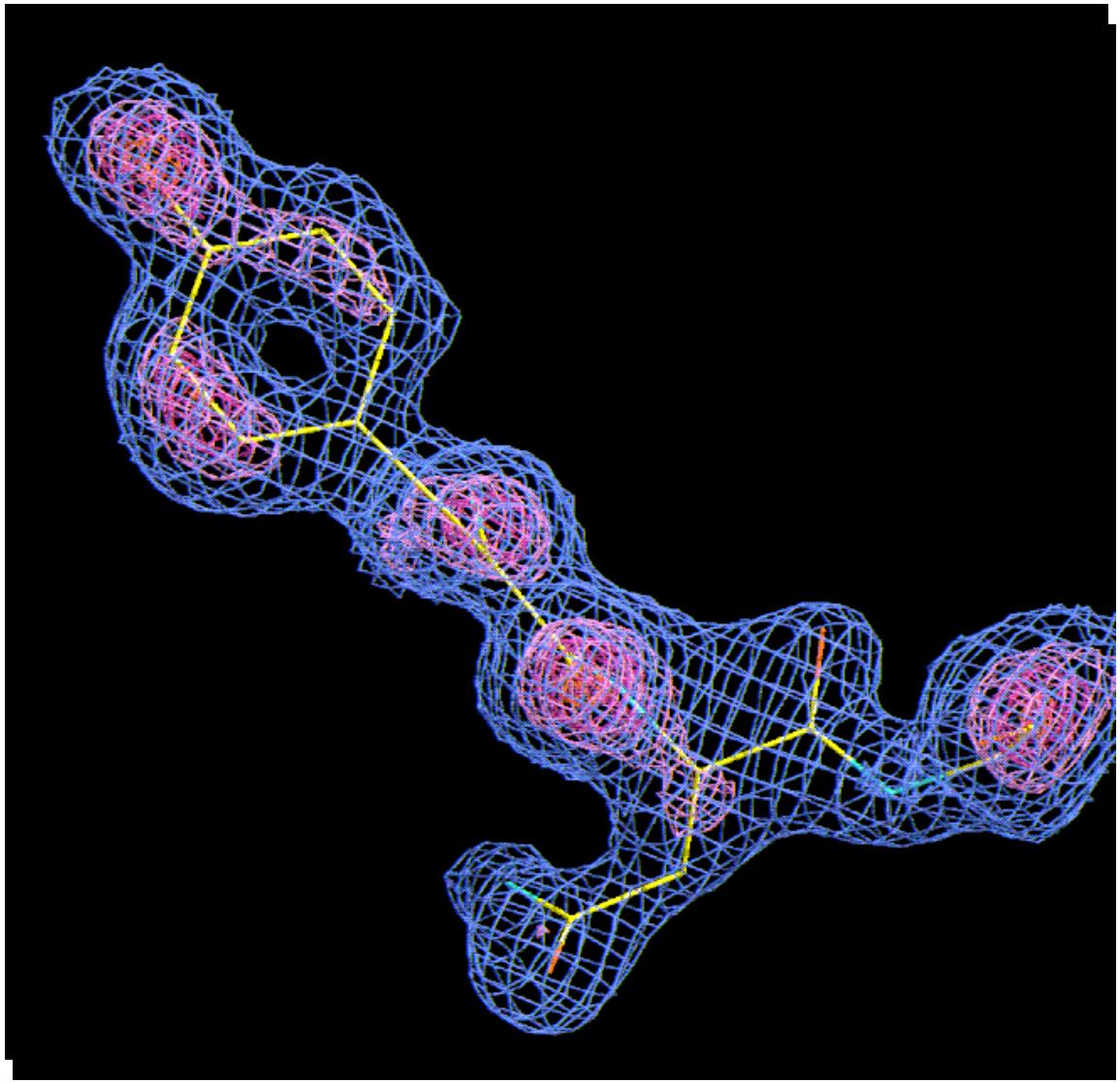


discontinuous density can be resolved by model building and refinement.



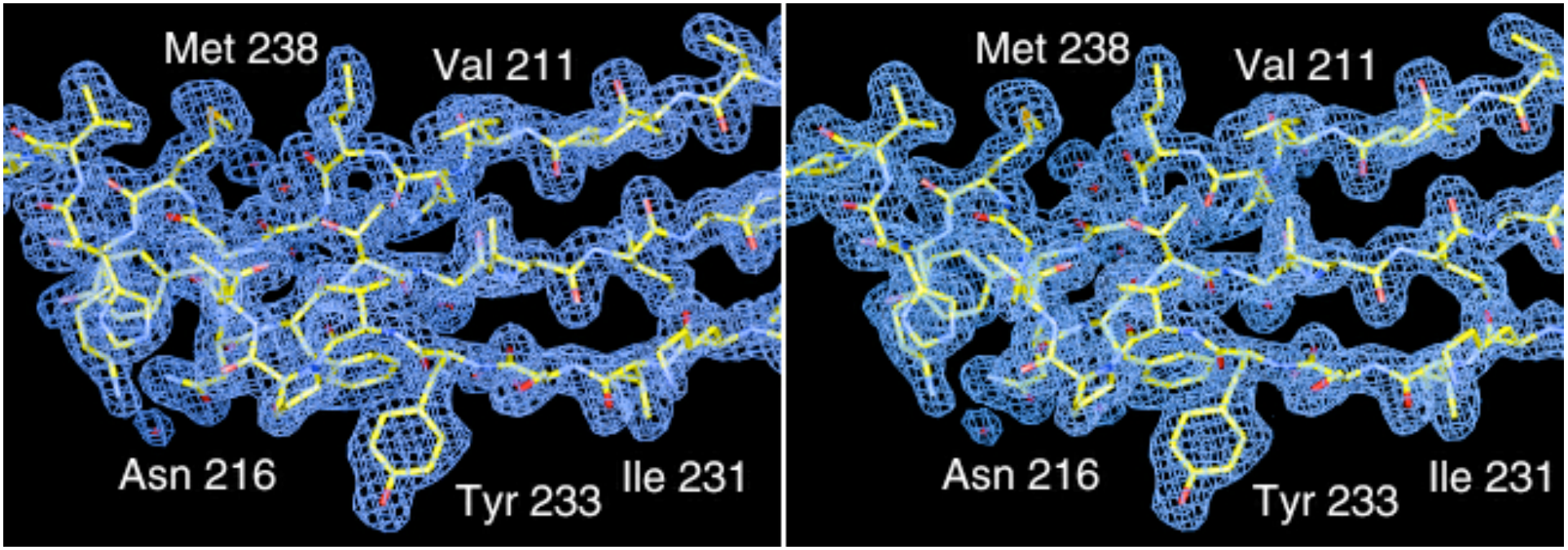
Contouring at two density cutoffs sometimes helps

Holes in rings are a good thing



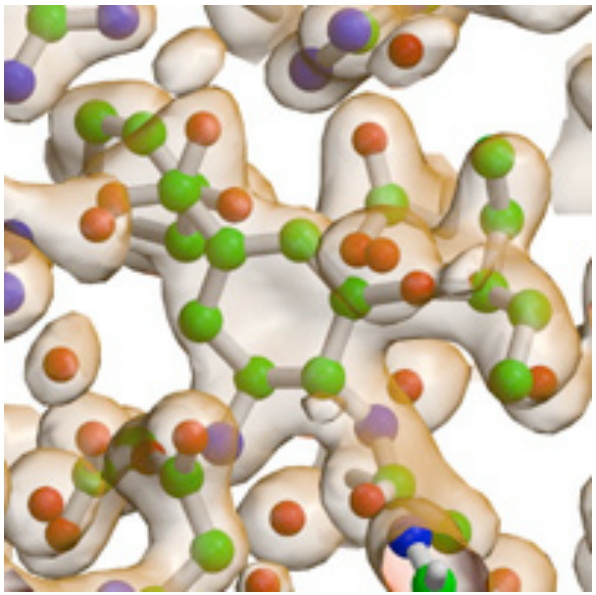
Seeing a hole in a tyrosine or phenylalanine ring is universally accepted as proof of good phases. You need at least 2Å data.

It is very useful to see in stereo



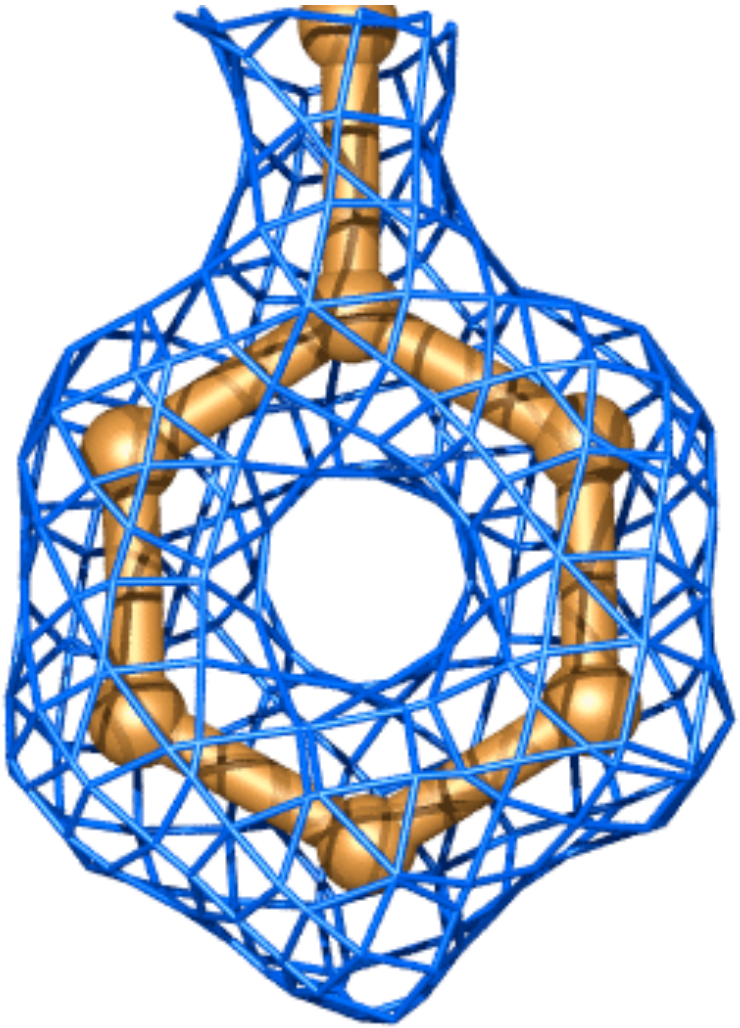
Relax your eyes. Look into the distance.

Fancy rendering

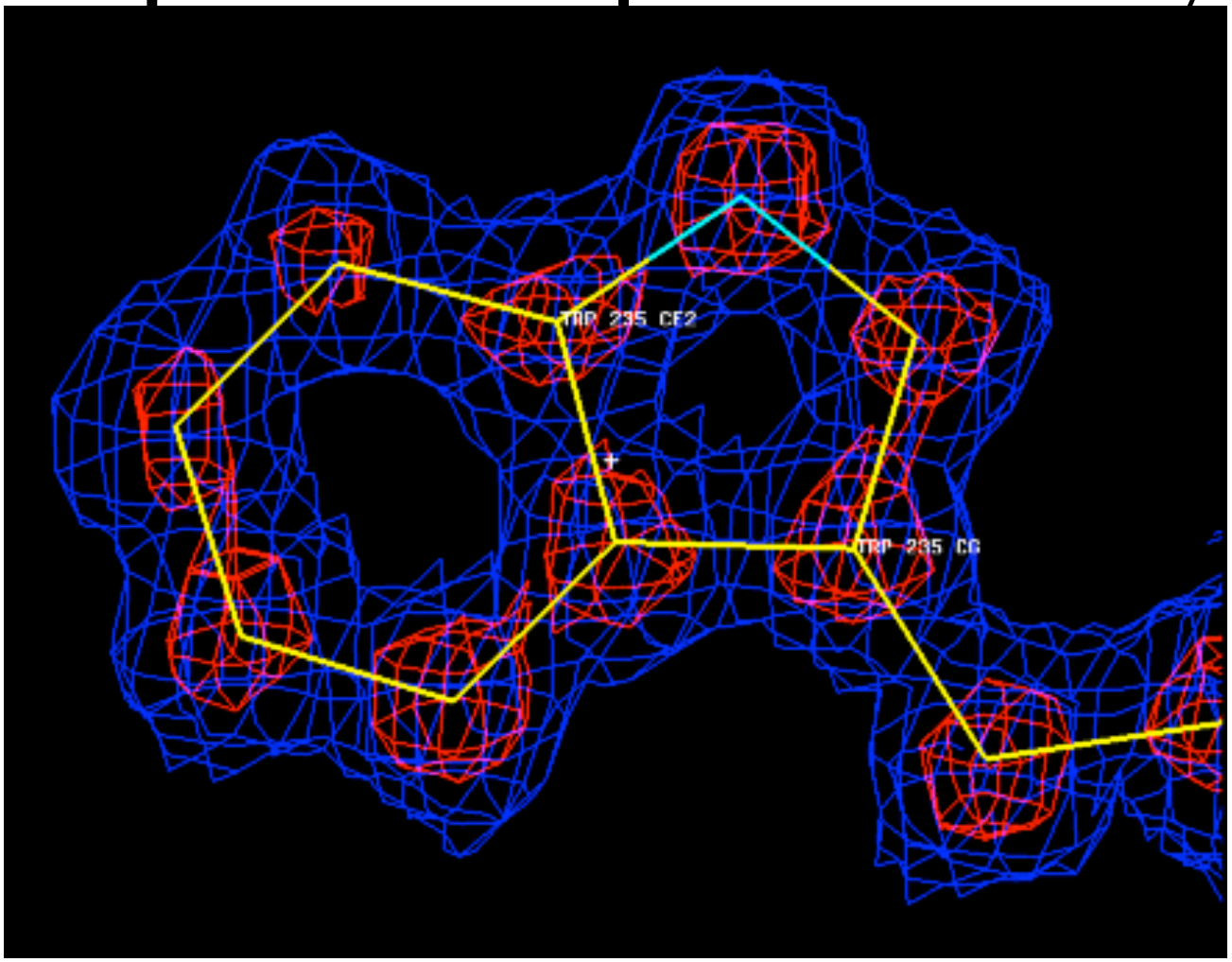


"CONSCRIPT: A program for generating electron density isosurfaces for presentation in protein crystallography." M. C. Lawrence, P. D. Bourke

Great map: holes in rings

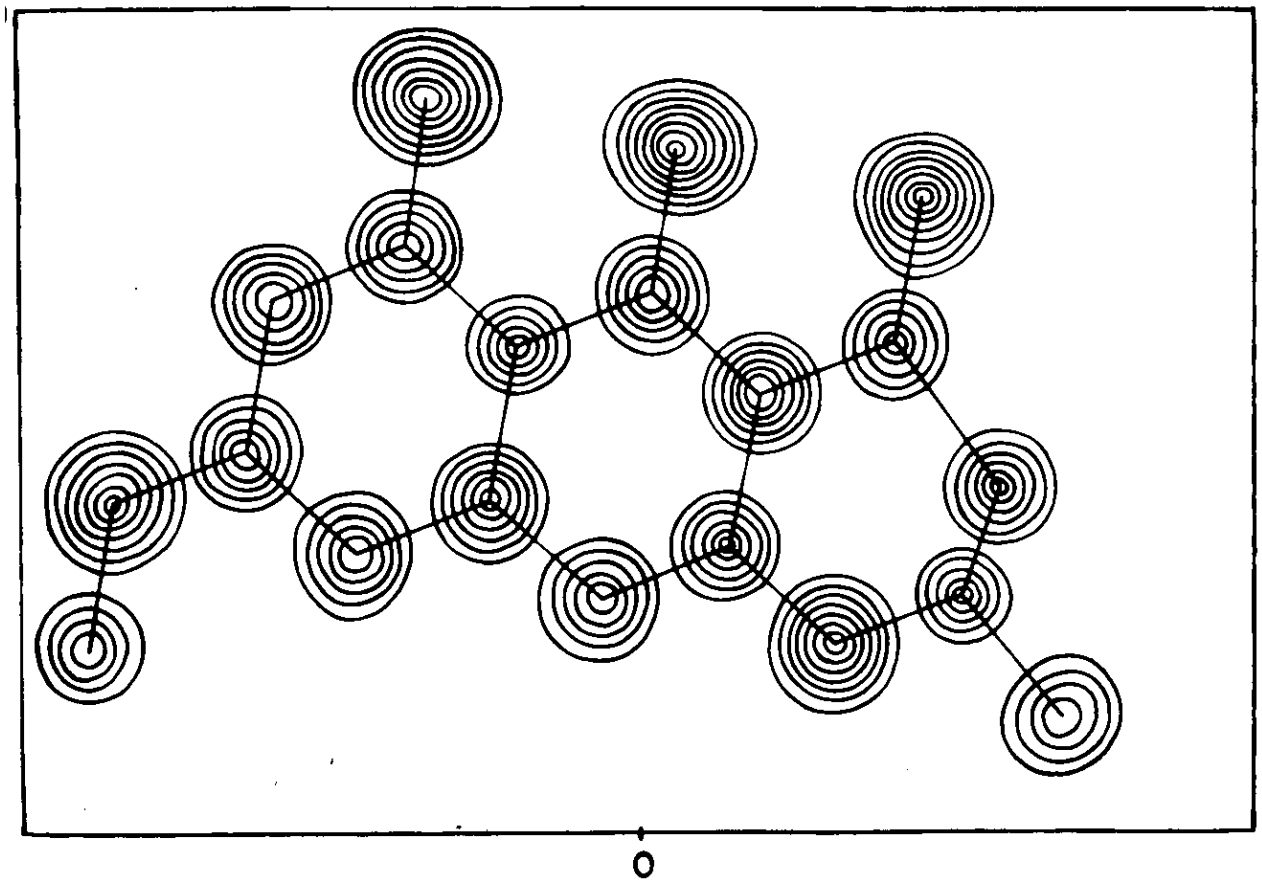


Superior map: Atomicity



Rarely is the data this good. 2 holes in Trp. All atoms separated.

Only small molecule structures look this good



Atoms are separated down to several contours. Proteins are never this well-ordered. But this is what the density really looks like.

Refinement

- Calculate the *gradient* of the R-factor with respect to each atomic position.
- Move each atom is down-hill in the gradient.
- Manage stereochemistry.

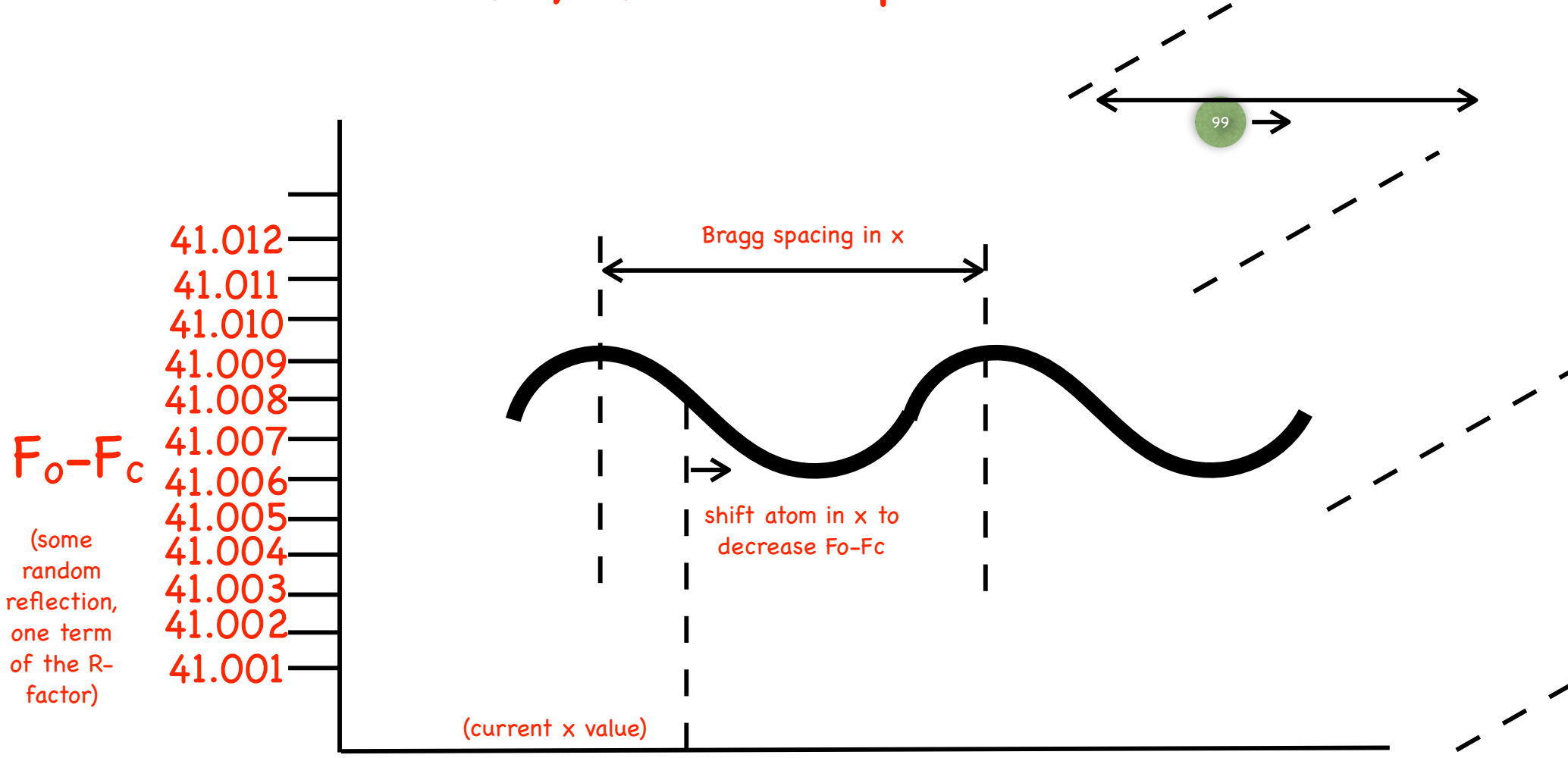
$$R = \frac{\sum_h \left| |F_{obs}(h)| - |F_{calc}(h)| \right|}{\sum_h |F_{obs}(h)|}$$

$$\frac{dR}{dv_i}$$

coordinates of atom *i*

refinement

Deriv. of R-factor w/respect to any atom coordinate (x,y,z) is a simple cosine function.

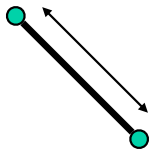


the x coordinate of atom 99

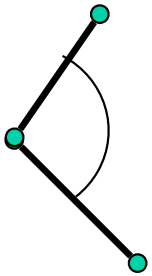
dR/dx_{99} = sum of $d|F_o-F_c|/dx_{99}$ over all $|F_o-F_c|$.

What is stereochemistry?

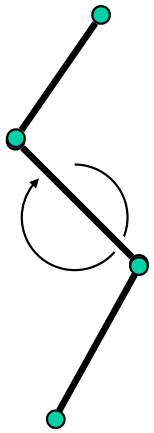
- Bond lengths
- Angles
- Chirality
- Planar groups
- Restraints are used to maintain “ideal” geometry.



bond lengths



bond angles



torsion angles

hybrid F's

- F_c's are calculated from the atomic coordinates
- A new electron density map calculated from the F_c's *would only reproduce the model*. (of course!)
- Instead we use the *observed amplitudes* $|F_o|$, and the *model phases*, α_c.

Hybrid back transform: $$\rho(\dot{r}) = \sum_h \left|F_{obs}(h)\right| e^{-i\left(2\pi\left(h \bullet \dot{r}\right) + \alpha_{calc}(h)\right)}$$

.

Hybrid maps show places where the current model is wrong and needs to be changed.

Difference map: F_o-F_c amplitudes

The F_o "native" map $\rho(F_o)$ differs from the Fc map $\rho(F_c)$ in places where the *model is wrong*. So we take the difference. In the difference map:

Missing atoms? ⇒ $\rho(F_o\text{-}F_c) > 0.0$

Wrongly placed atoms? ⇒ $\rho(F_o\text{-}F_c) < 0.0$

Correctly modeled atoms? ⇒ $\rho(F_o\text{-}F_c) = 0.0$

Q: Subtracting densities (real space) is the same as subtracting amplitudes (reciprocal space) and transforming. T or F?

2Fo-Fc

The F_o *map* plus the *difference map is*

F_o where the differences are zero (the atoms are *correct*)

Less than F_o where the model has wrong atoms.

Greater than F_o where the model is missing atoms.

$F_o + (F_o - F_c) = 2F_o - F_c$

free R-factor

The free R-factor is the test set residual, calculated the same as the R-factor, but on the "test set".

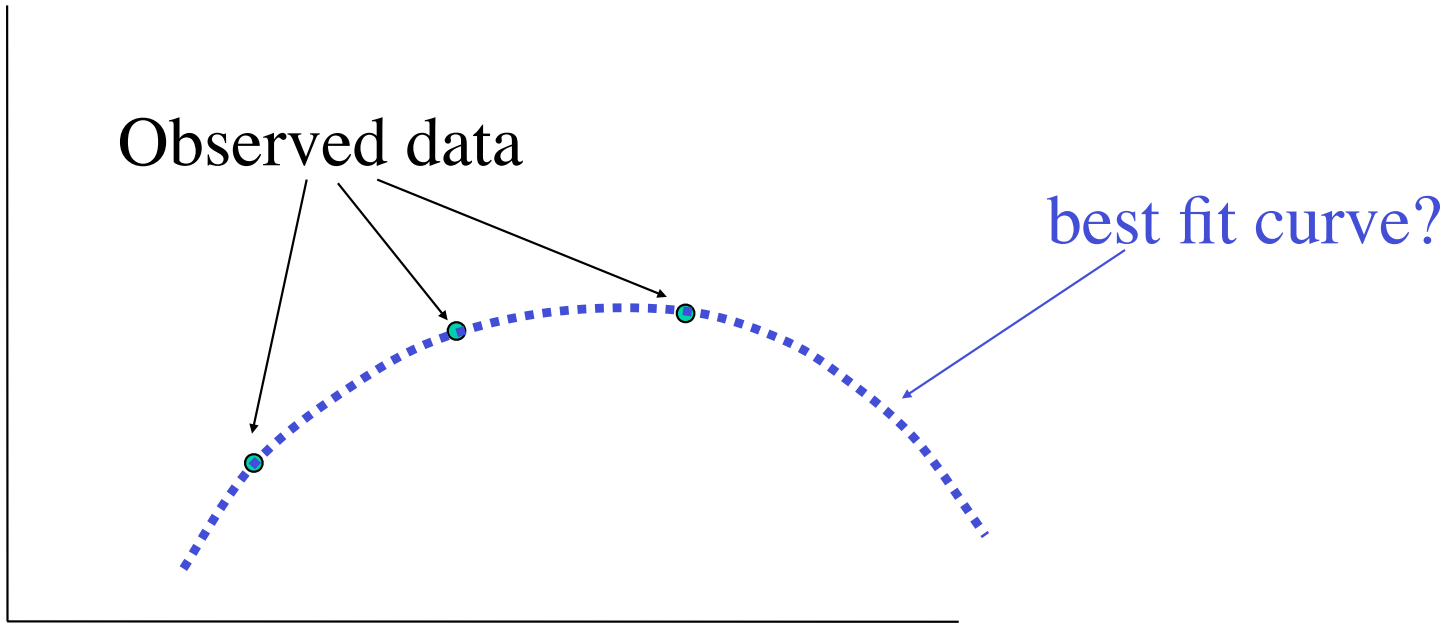
Free R-factor asks "how well does your model predict the data it hasn't seen?"

$$R_{free} = \frac{\sum_{h \in T} \left| |F_{obs}(h)| - k|F_{calc}(h)| \right|}{\sum_{h \in T} |F_{obs}(h)|}$$

Note: the only difference is which *hkl* are used to calculate.

Why cross-validate?

Because if you have more parameters than data, you can over-fit the data.

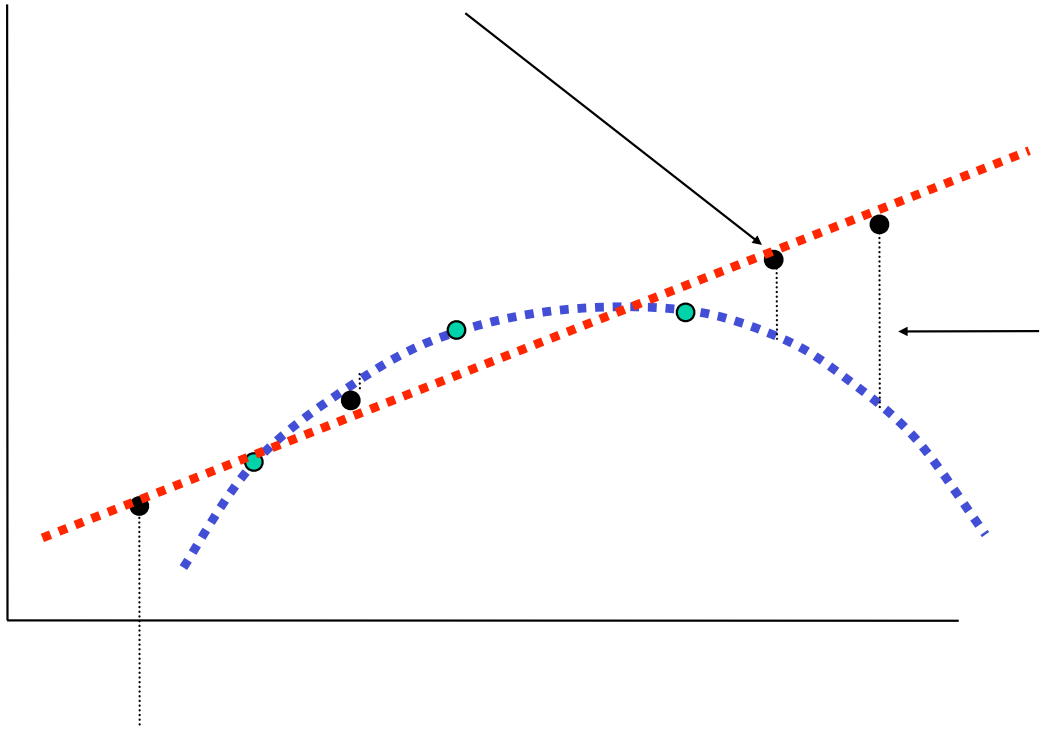


R-factor = 0.000??

Fitting and overfitting

Fit is correct if *additional data*, not used in fitting the curve, fall on the curve.

Low residual in the "**test set**" *justifies* the fit.



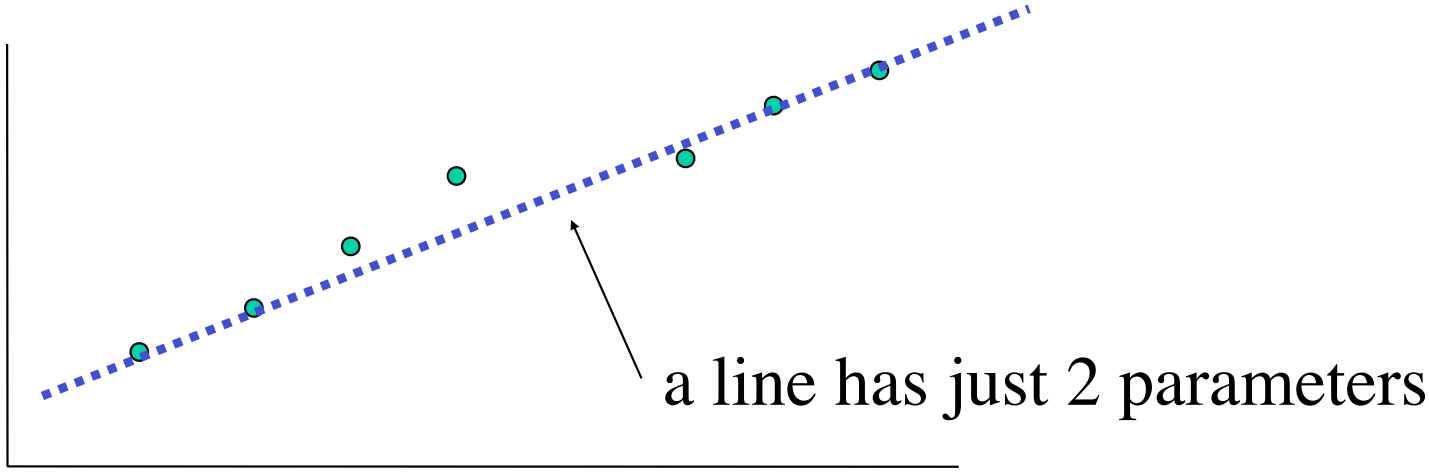
residual is not zero now.

cross-validation

=Measuring the residual on data (the "test set") that were not used to create the model.

The residual on test data is likely to be small if

$$\frac{data}{parameters}$$ is large.

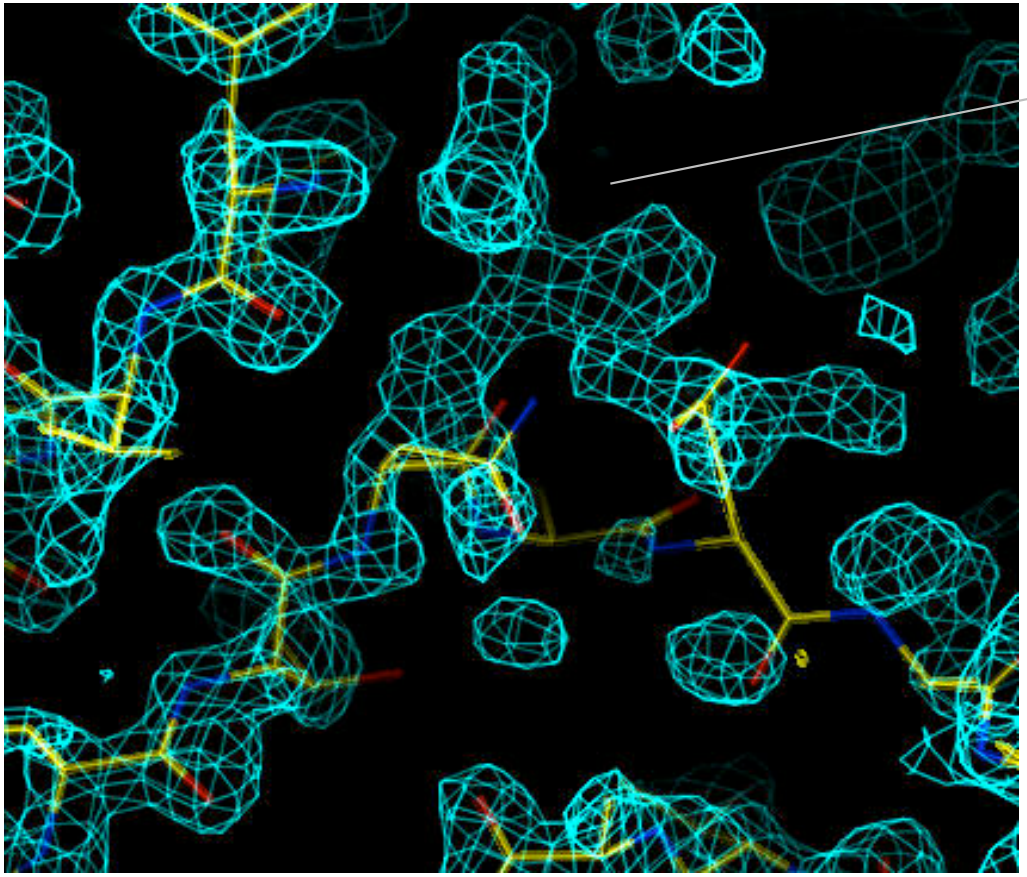


Phase bias

The model *biases* the phases.

The effect of phase bias is local.

To correct phase bias, we must **omit** the erroneous part of the model and calculate a map, called an "OMIT MAP".



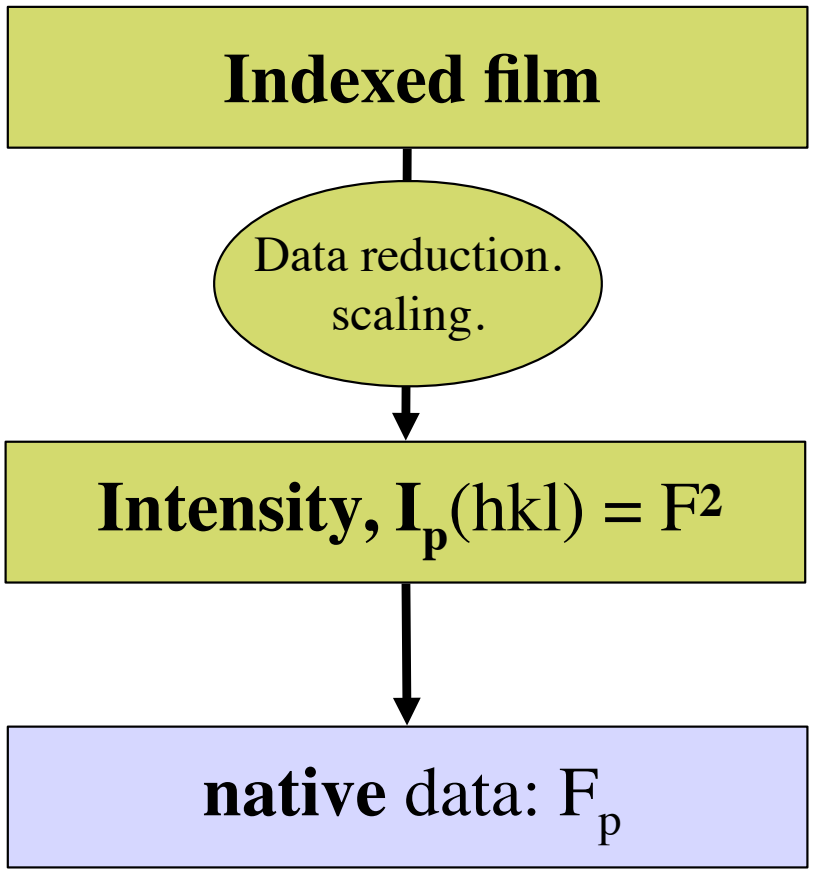
The true density s showing up where the wrong model was omitted.

$2F_o\text{-}F_c = F_o + (F_o - F_c)$
= map + difference map

...is used for an omit map.

review

From crystal to data



I is always a relative term

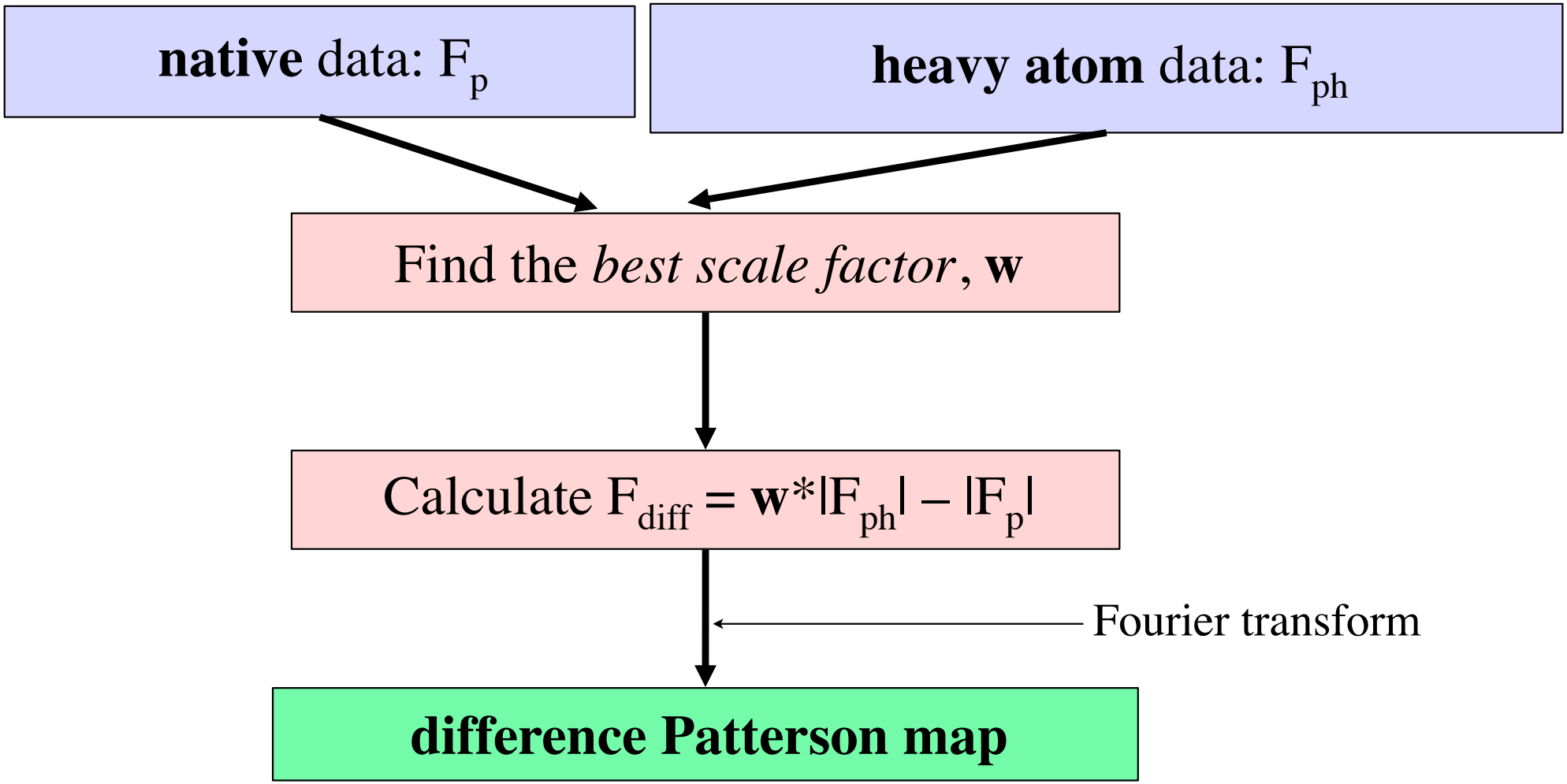
Bigger crystal, higher I

Better crystal, higher I

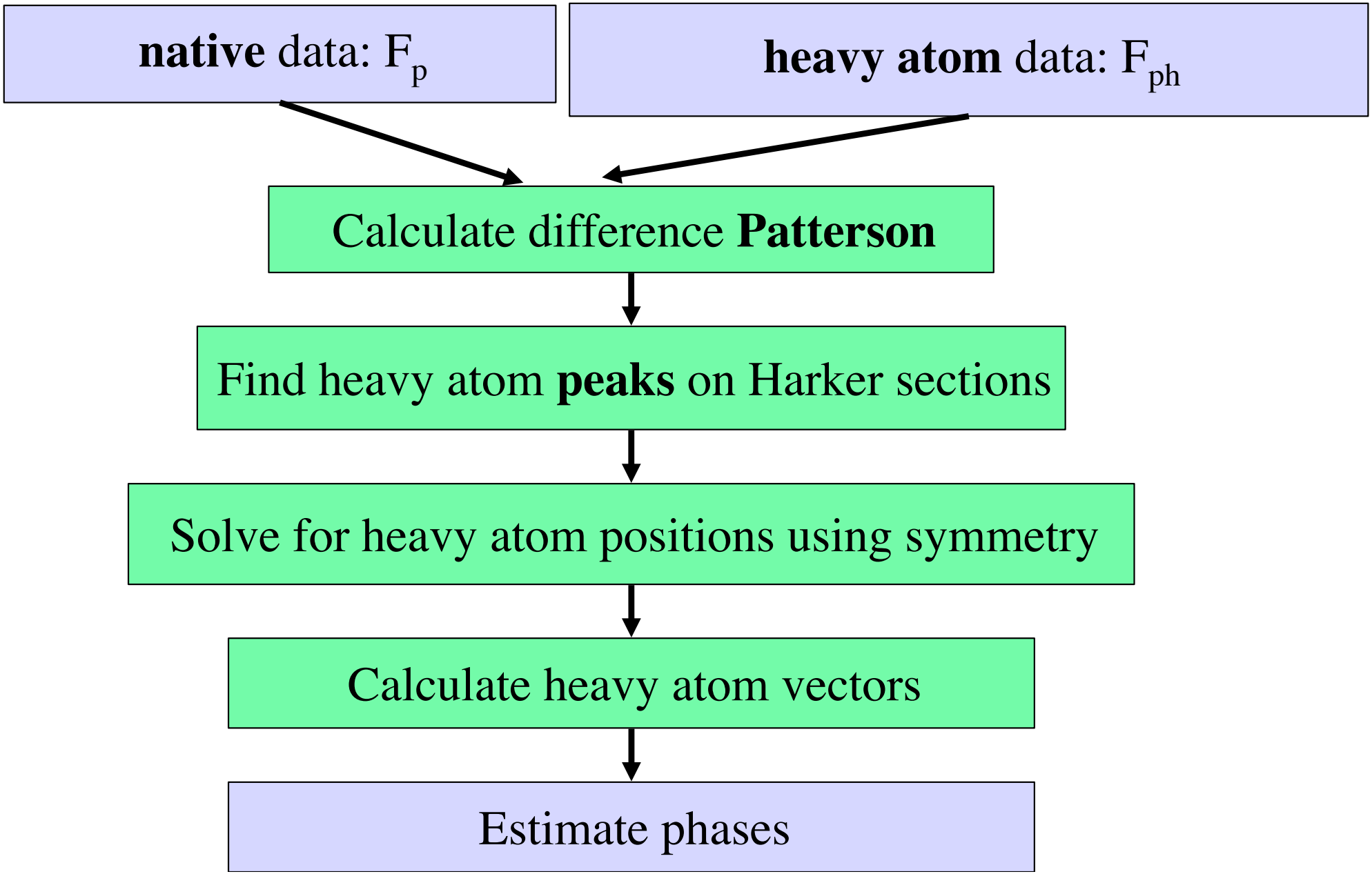
Longer exposure, higher I

More intense Xrays, higher I

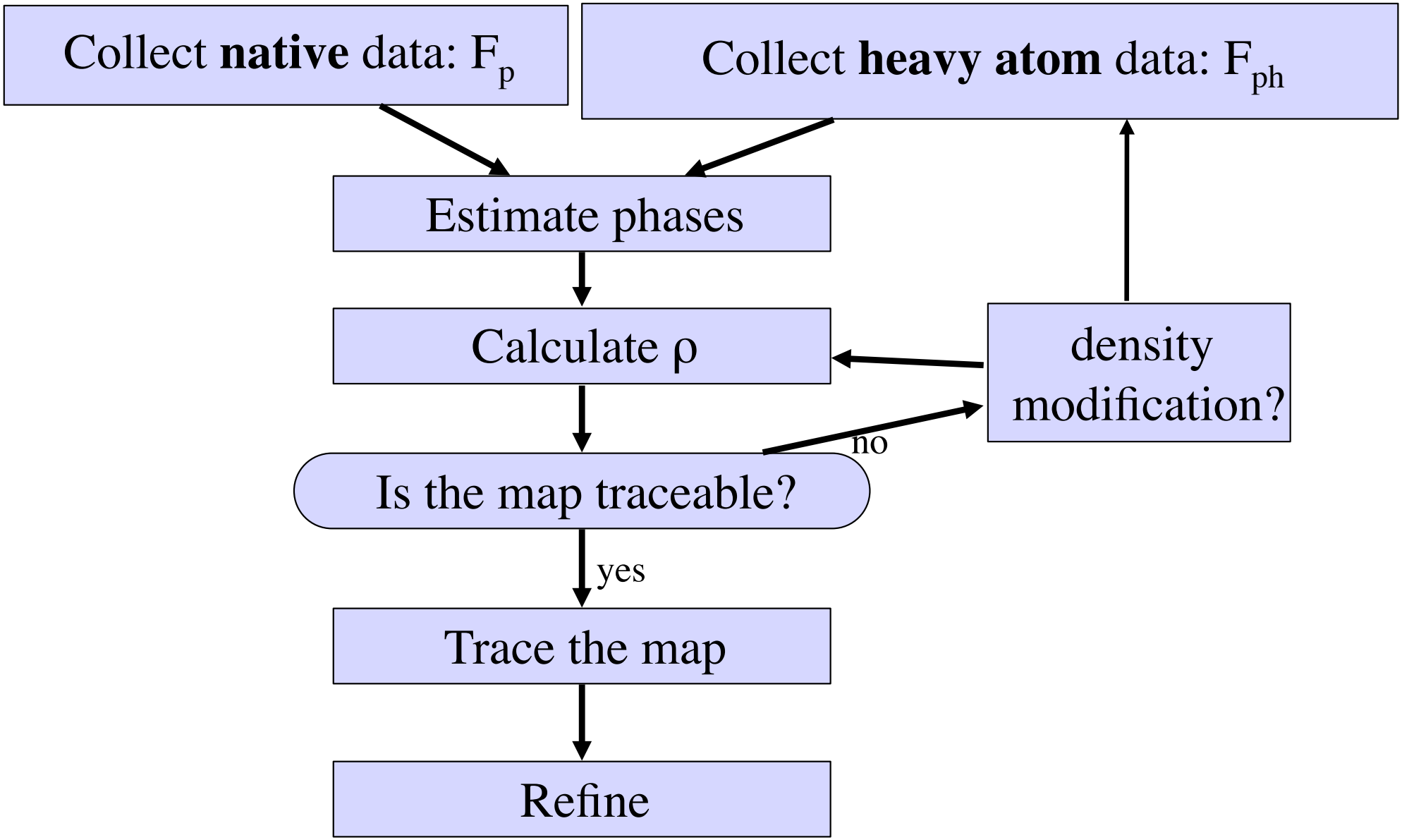
From data to Patterson map



From data to phases



From data to model



Exercise 6 -- Install software

- ~~Install Phenix~~
- Install Coot ~~(...if it doesn't come with Phenix)~~
- ~~On Monday bring computer, start up Phenix and follow along using the files linked to the course web page.~~