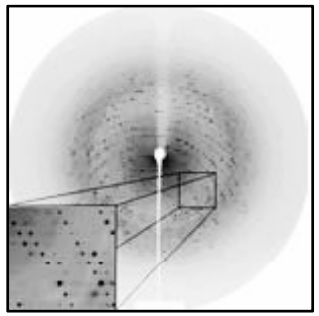
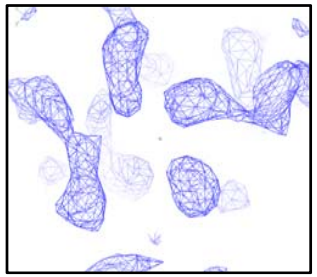
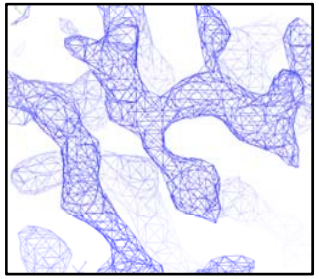
Automated phase improvement and model building with **Parrot** and **Buccaneer**

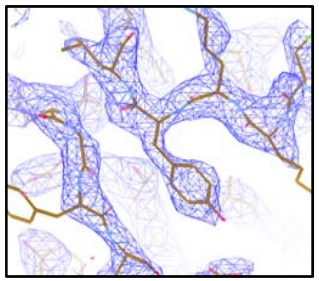
Kevin Cowtan
cowtan@ysbl.york.ac.uk

X-ray structure solution pipeline...









Data collection

Data processing

Experimental phasing

Molecular Replacement

Density Modification

Model building

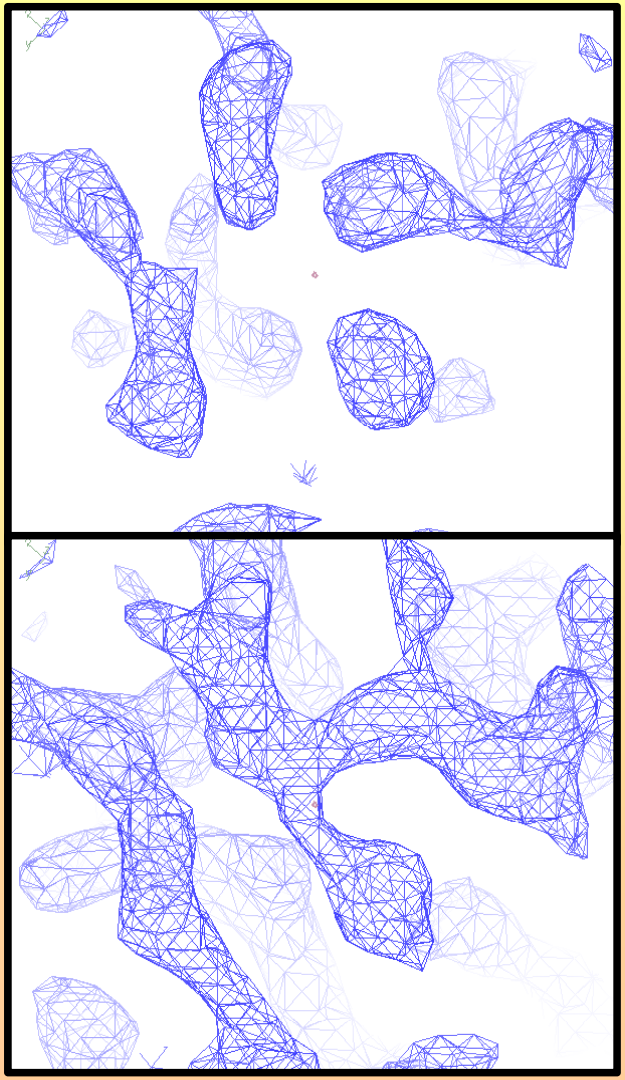
Refinement

Rebuilding Validation

Density Modification

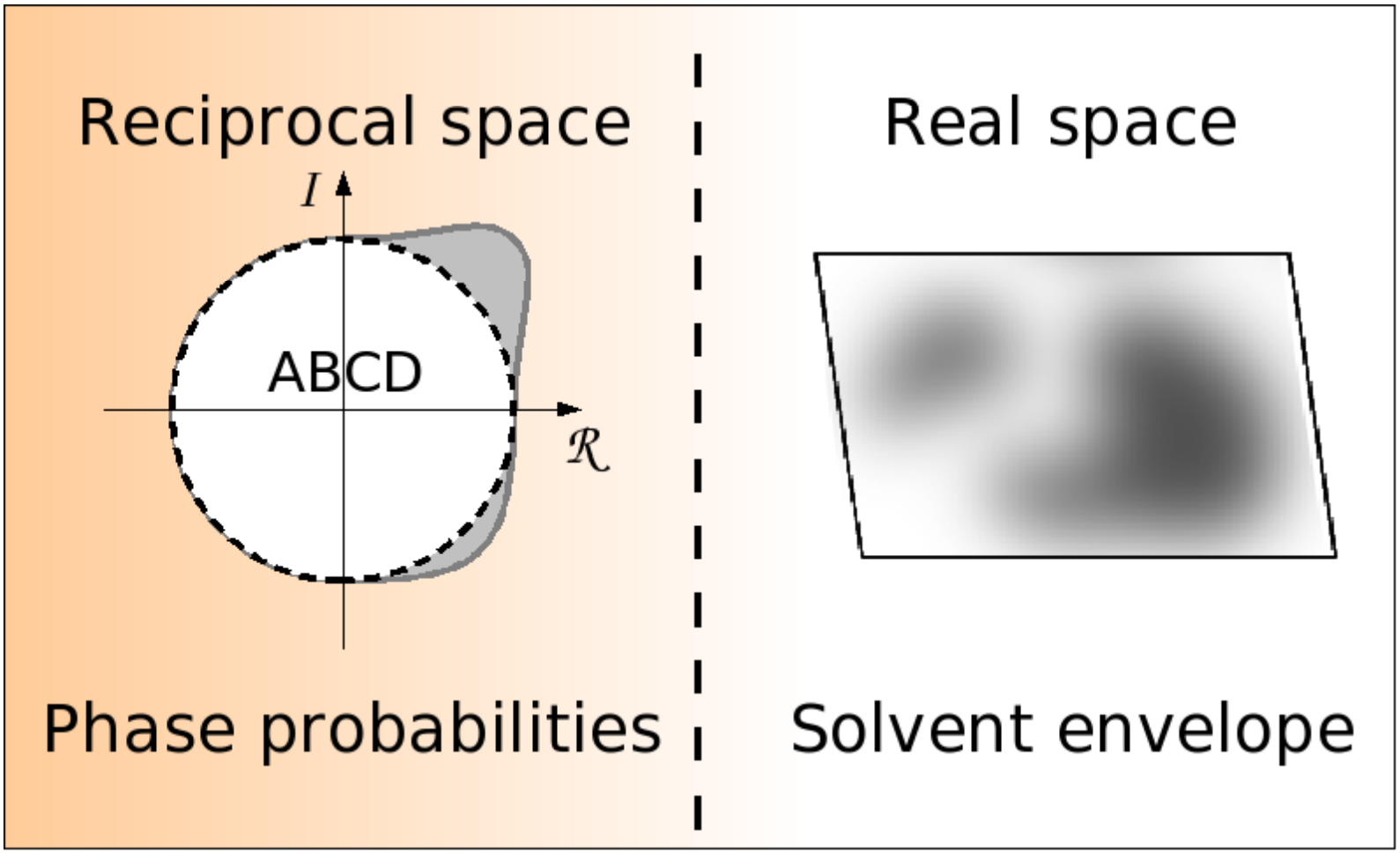
- Traditional density modification: e.g. 'dm', 'solomon', 'parrot', CNS

- Statistical density modification: e.g. 'resolve', 'pirate'



Density modification

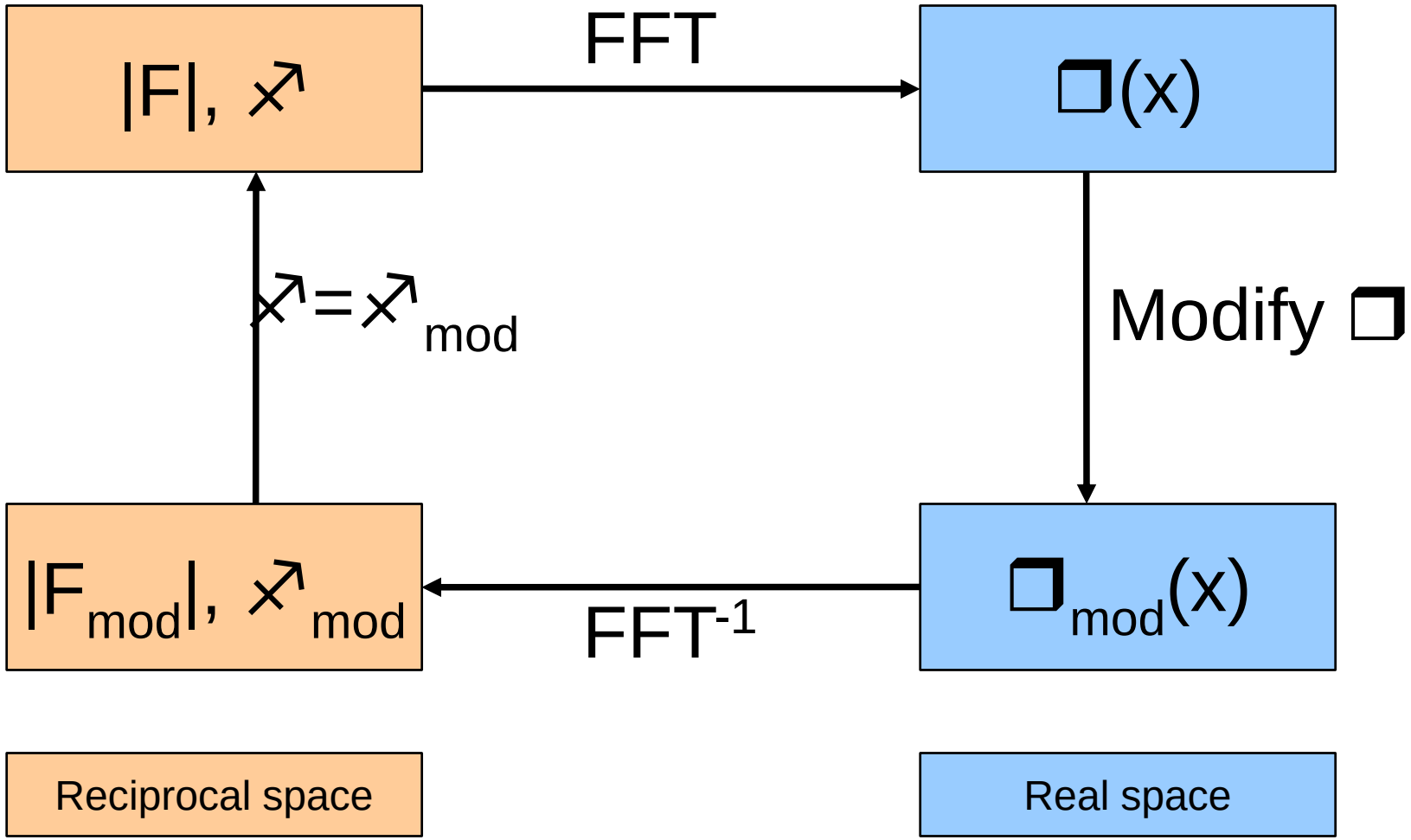
- Density modification is a problem in combining information:



Kevin Cowtan, cowtan@ysbl.york.ac.uk

Density modification

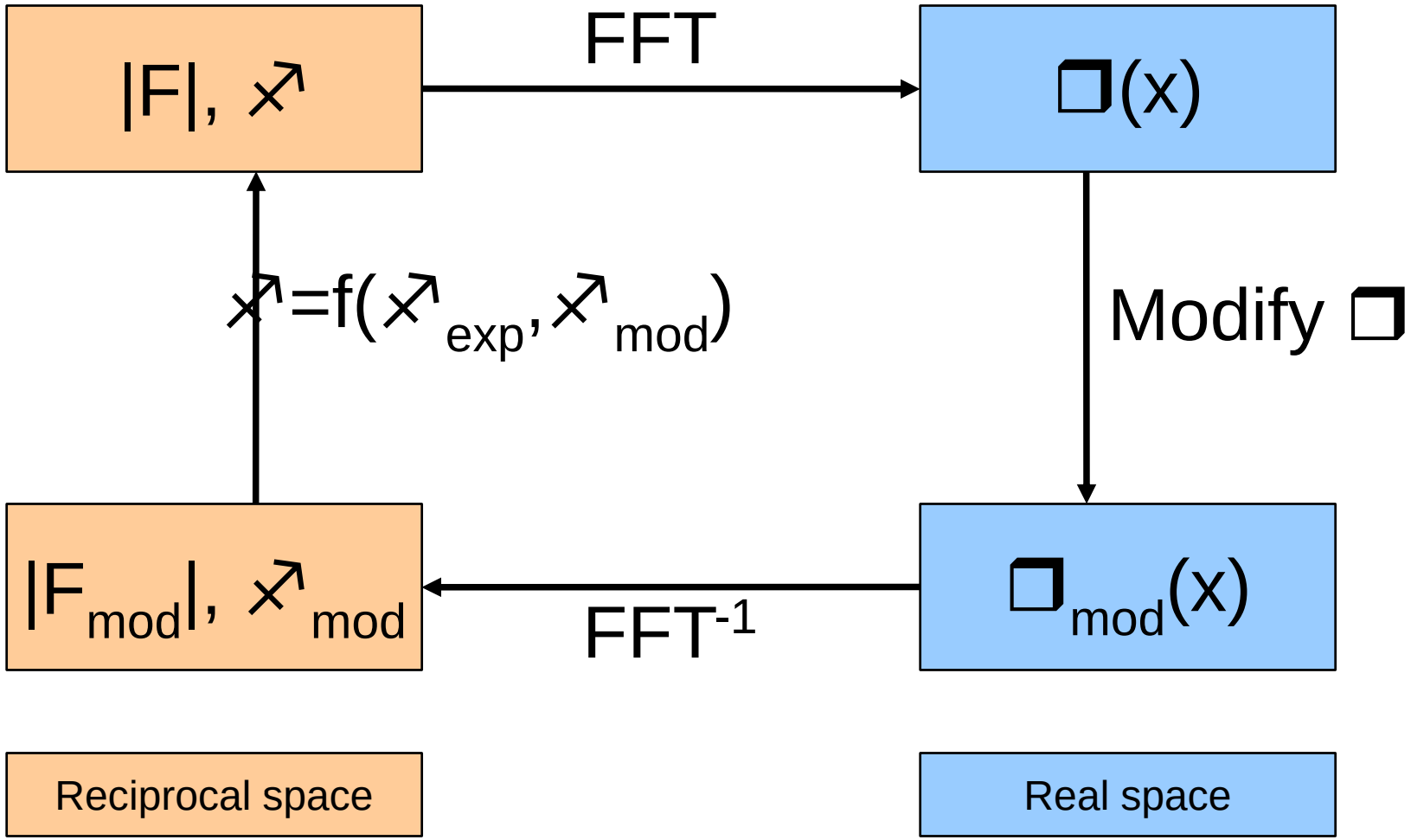
1. Rudimentary calculation:



Kevin Cowtan, cowtan@ysbl.york.ac.uk

Density modification

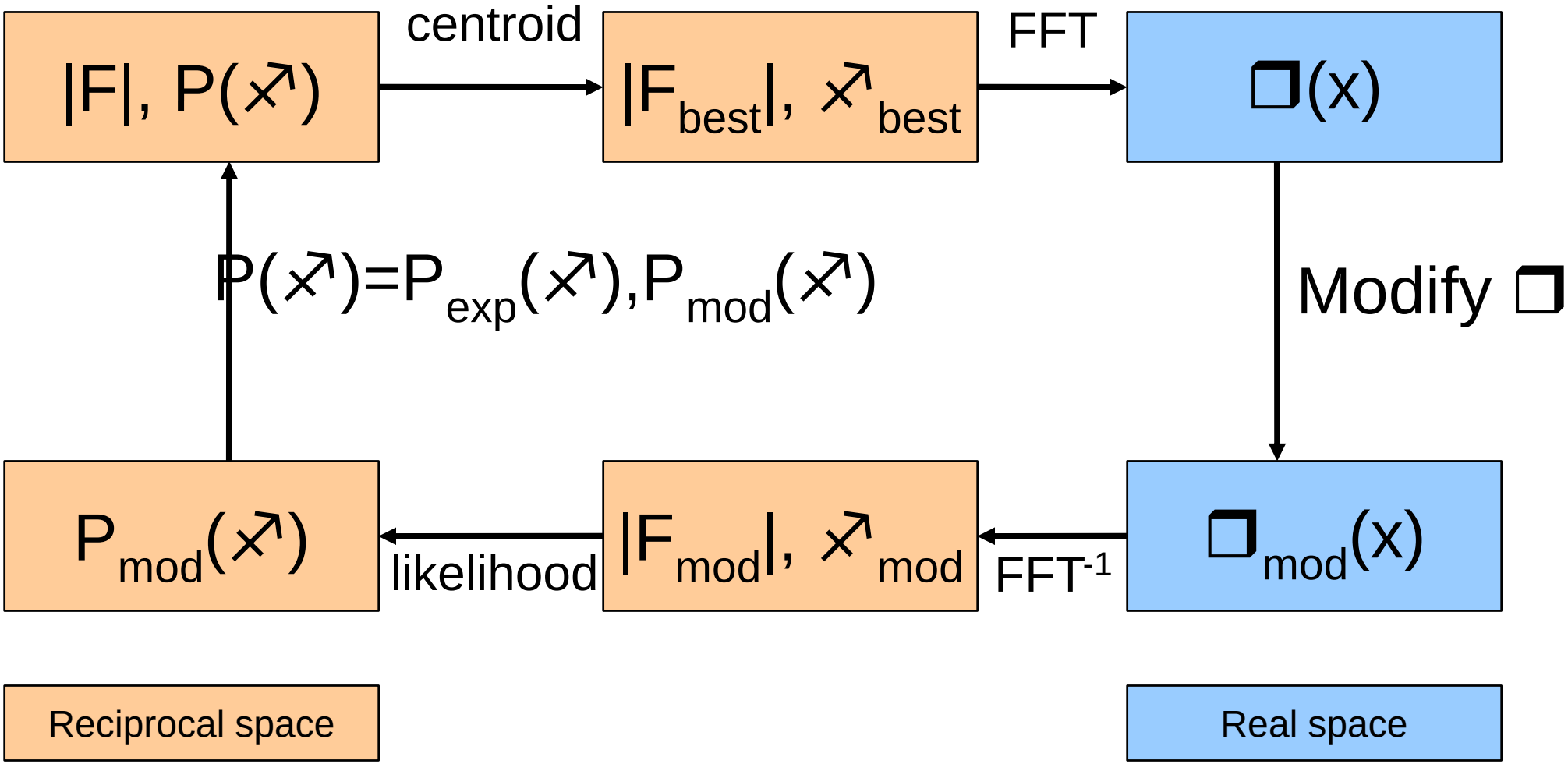
2. Phase weighting:



Kevin Cowtan, cowtan@ysbl.york.ac.uk

Density modification

3. Phase probability distributions:

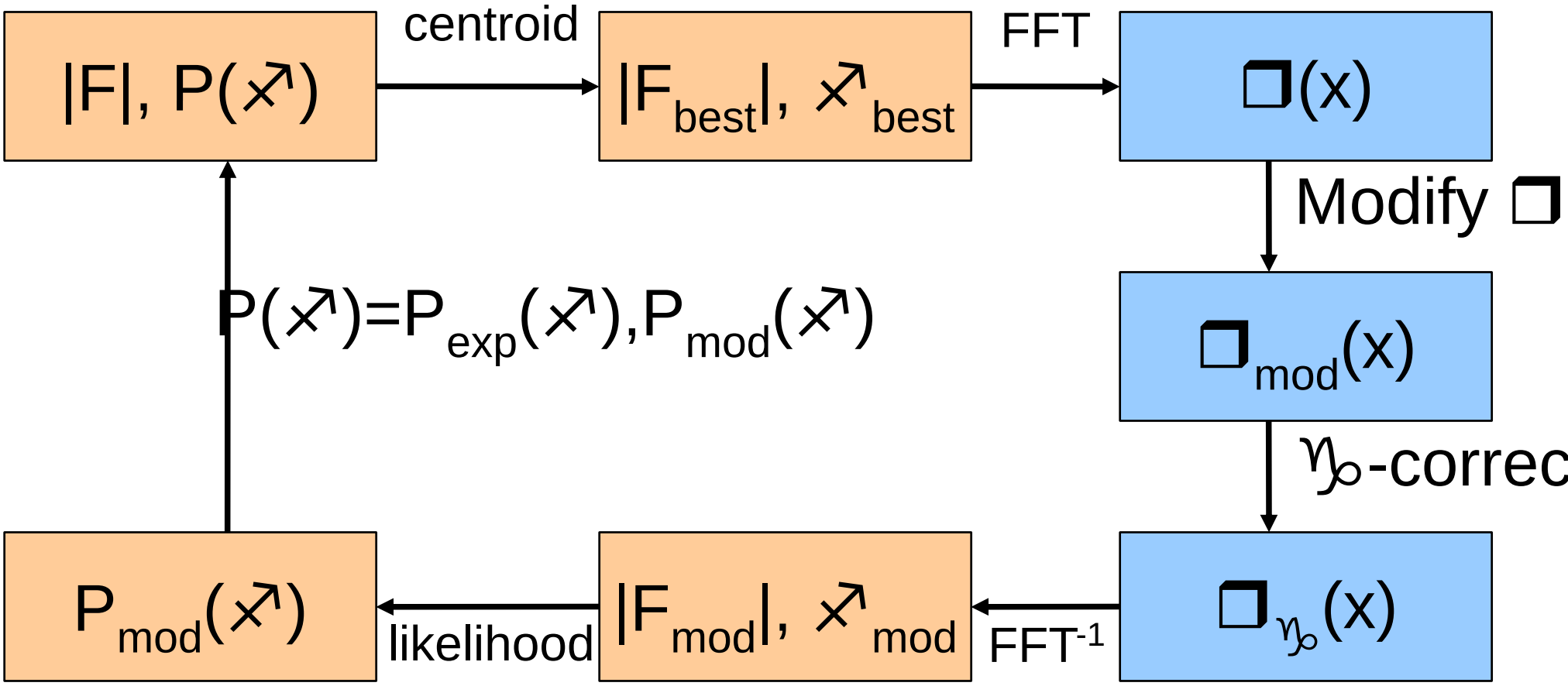


Kevin Cowtan, cowtan@ysbl.york.ac.uk

Density modification

DM, SOLOMON, (CNS)

4. Bias reduction (gamma-correction):



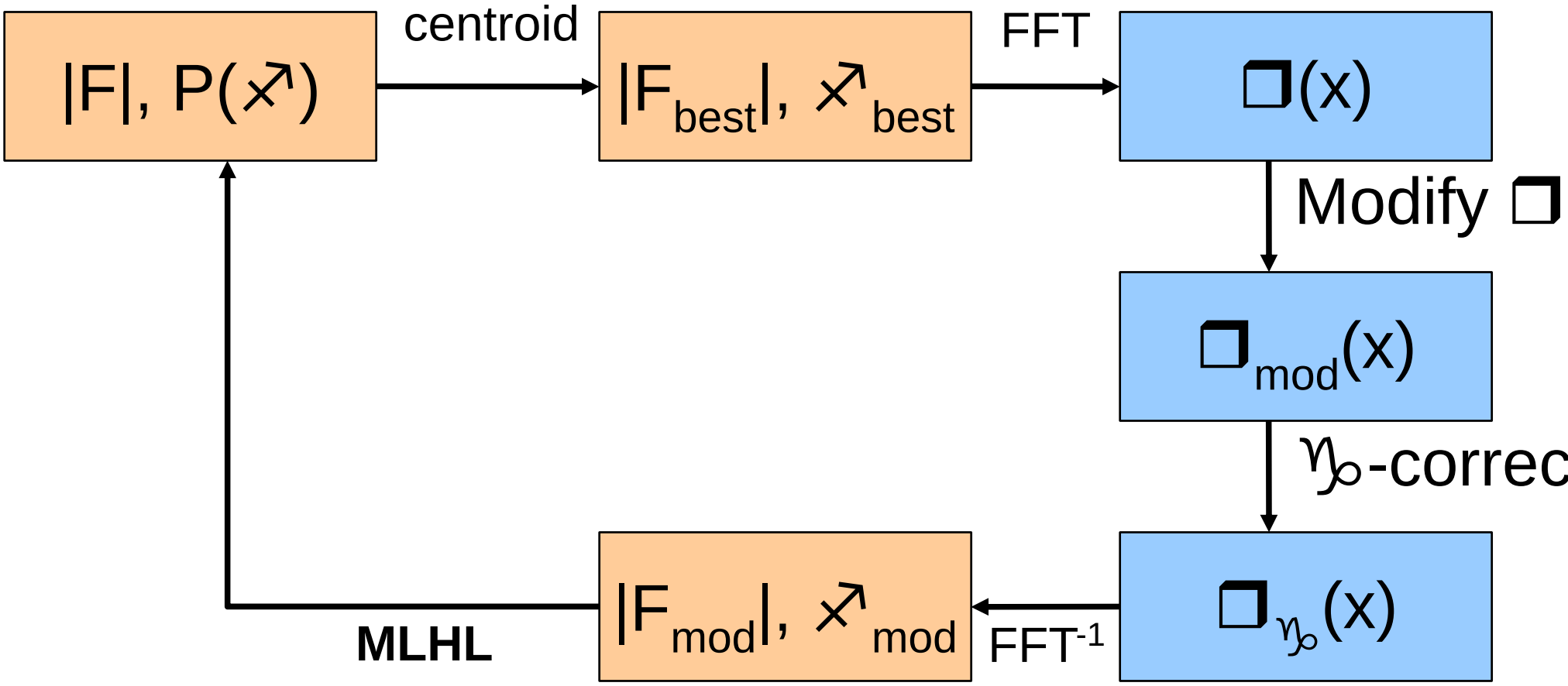
J.P.Abrahams

Kevin Cowtan, cowtan@ysbl.york.ac.uk

Density modification

PARROT

5. Maximum Likelihood H-L:

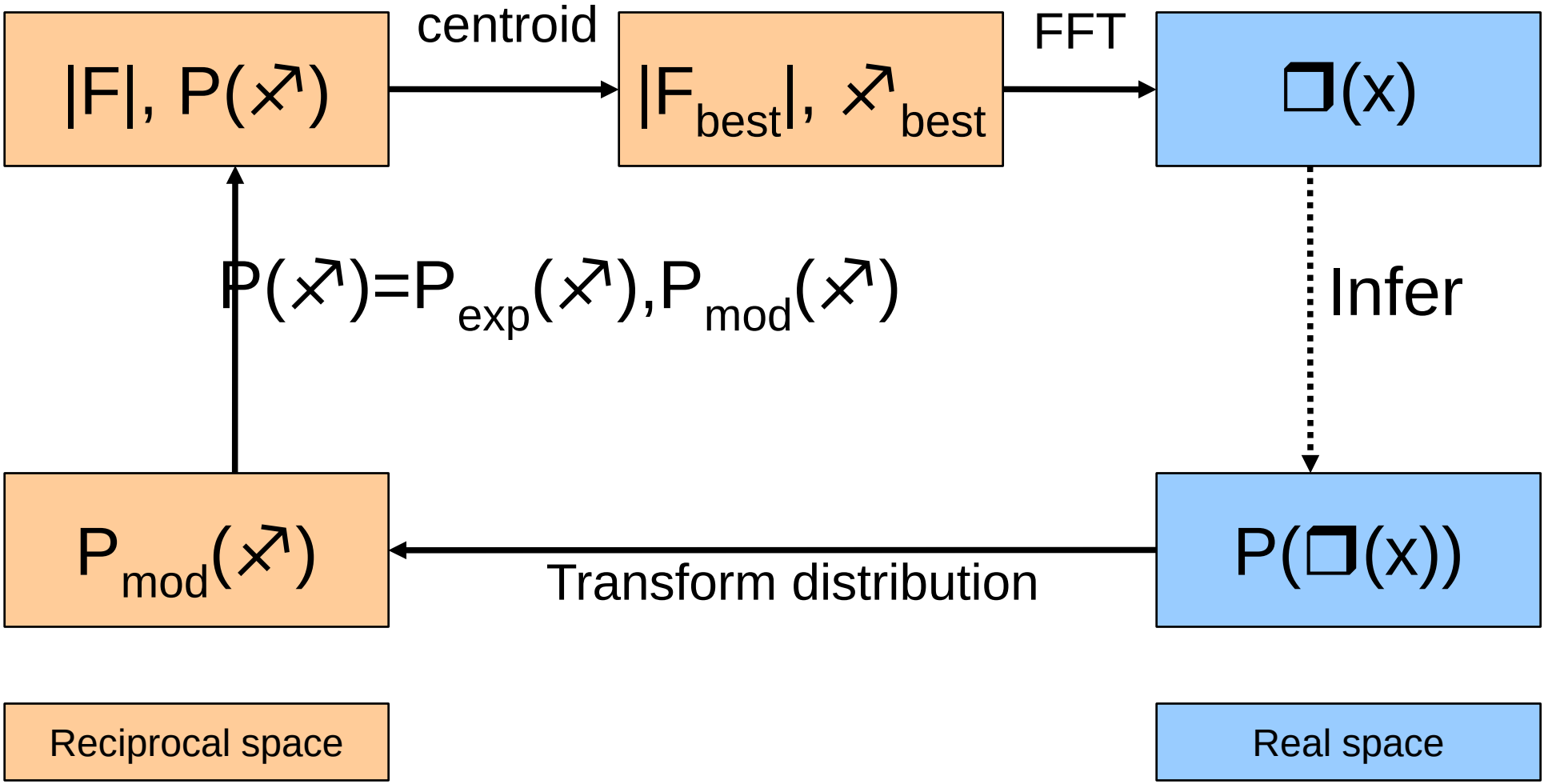


Kevin Cowtan, cowtan@ysbl.york.ac.uk

Density modification

RESOLVE, PIRATE

6. Statistical density modification:

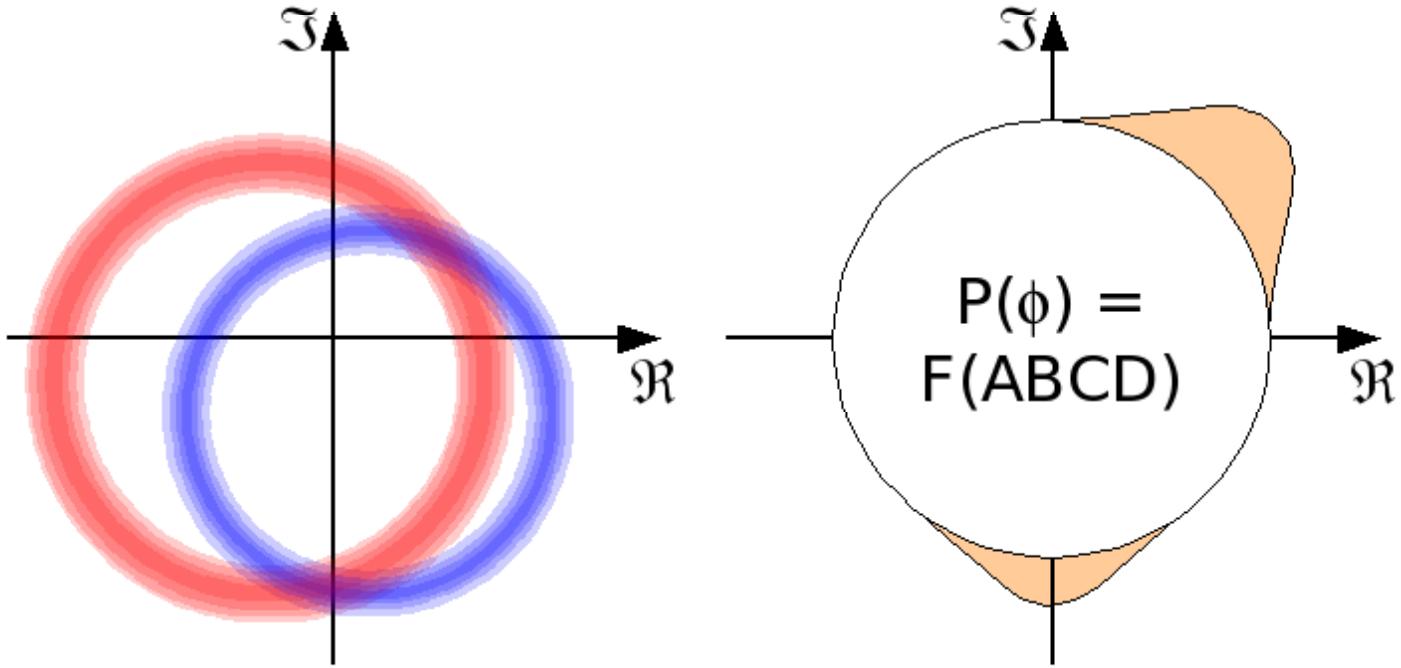


Kevin Cowtan, cowtan@ysbl.york.ac.uk

Density modification

How do we represent phase probabilities?

Henrickson-Lattman coeffs: 4 numbers - A,B,C,D representing a bimodal distribution in phase angle:



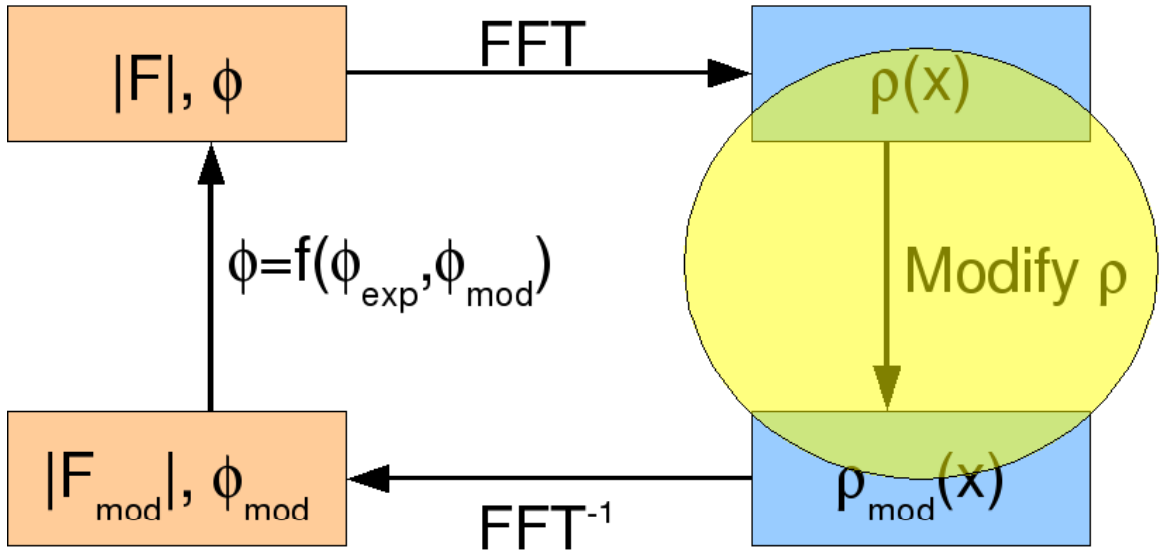
A,B represent a unimodal distribution (equivalent to ↗, FOM)
C,D represent the superimposed biomodality.

Kevin Cowtan, cowtan@ysbl.york.ac.uk

Density modification

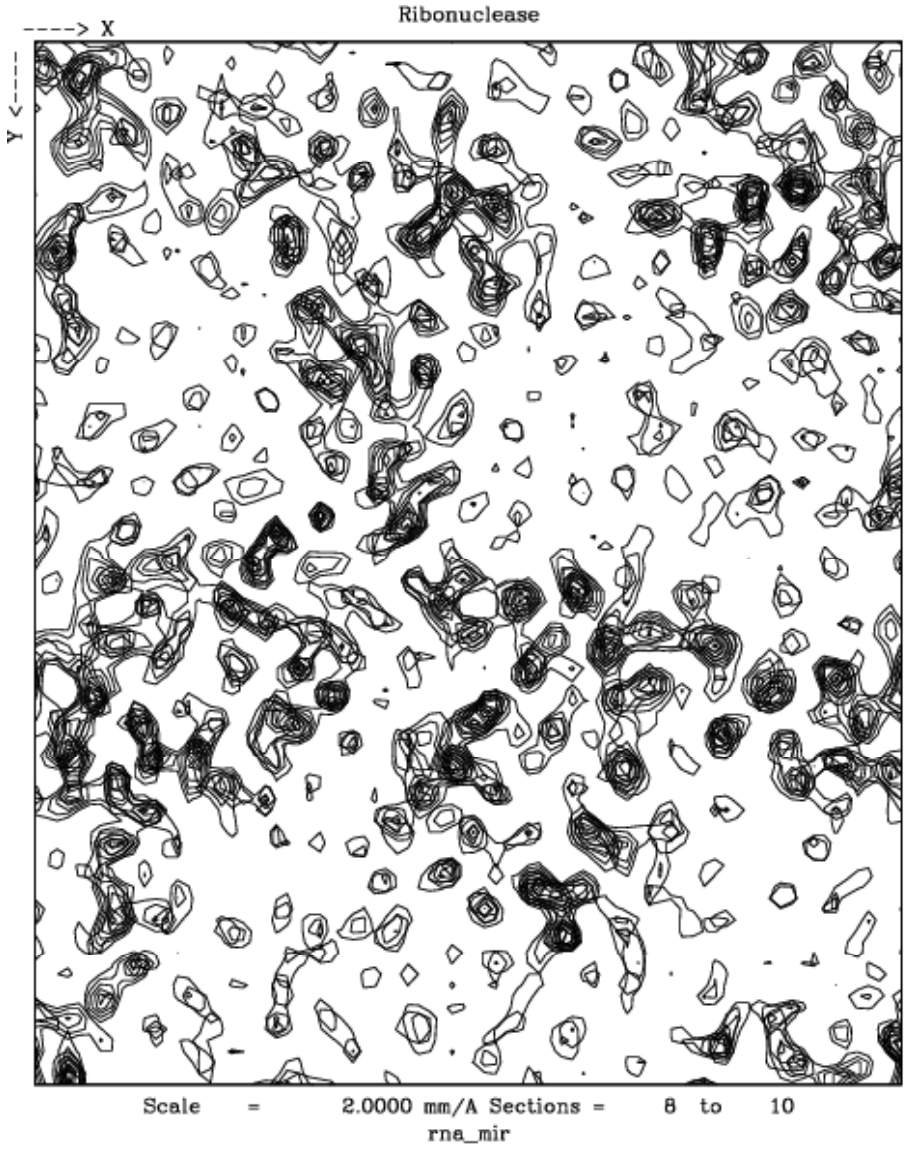
Traditional density modification techniques:

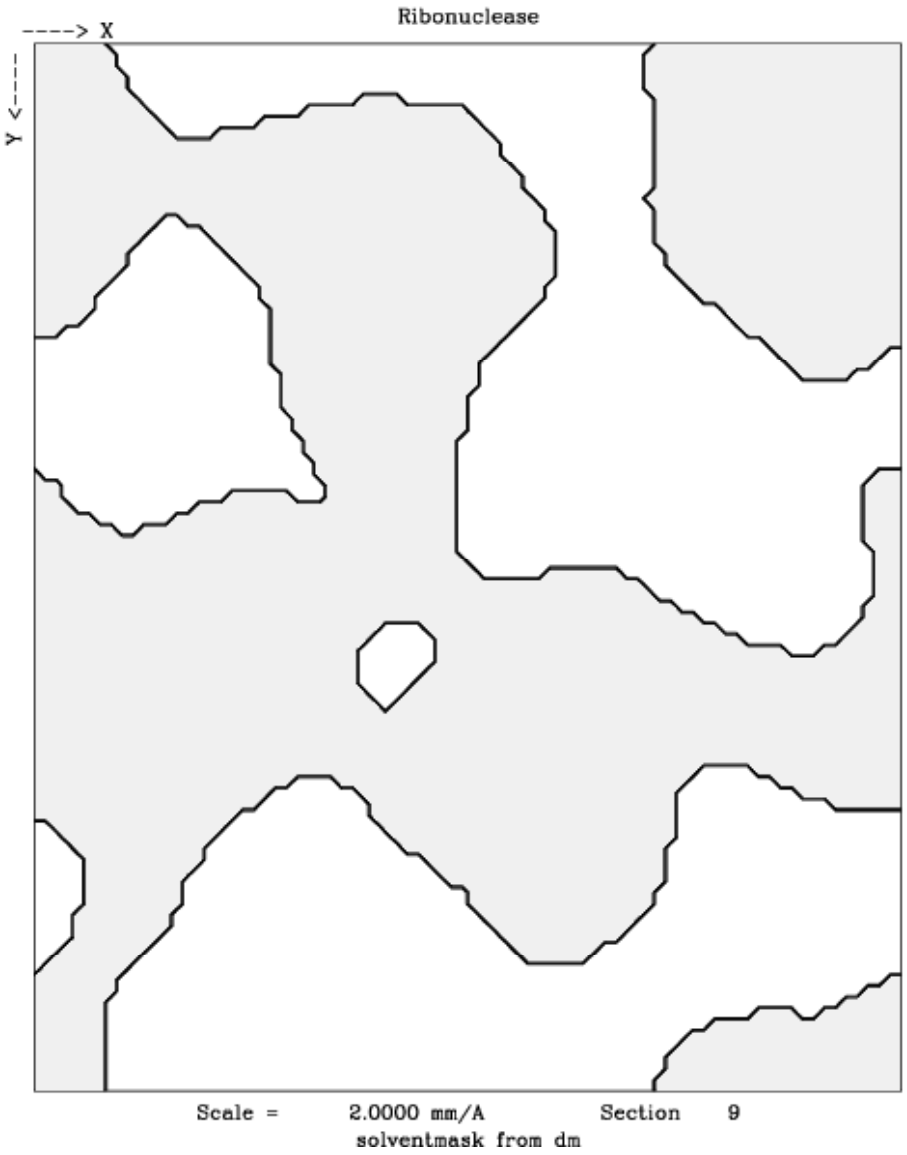
- Solvent flattening
- Histogram matching
- Non-crystallographic symmetry (NCS) averaging



Kevin Cowtan, cowtan@ysbl.york.ac.uk

Solvent flattening





Kevin Cowtan, cowtan@ysbl.york.ac.uk

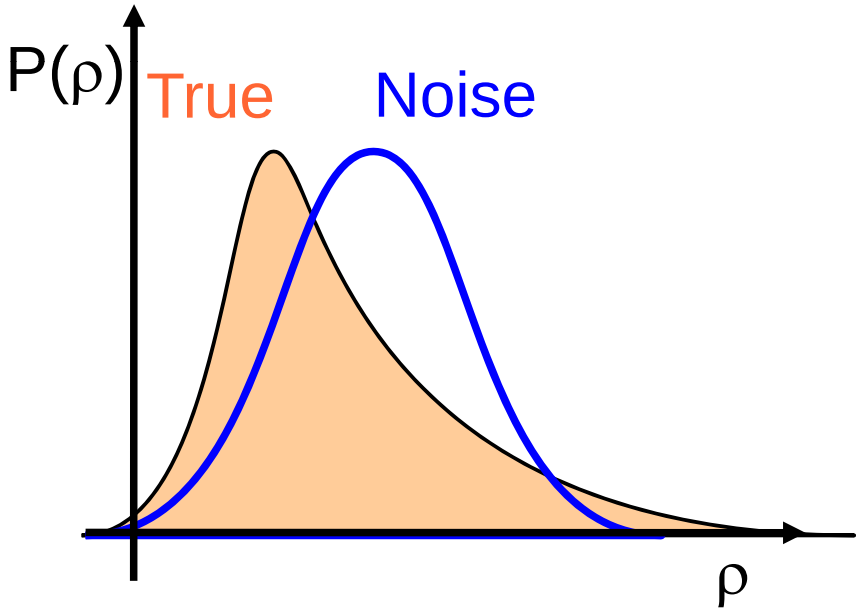
Histogram matching

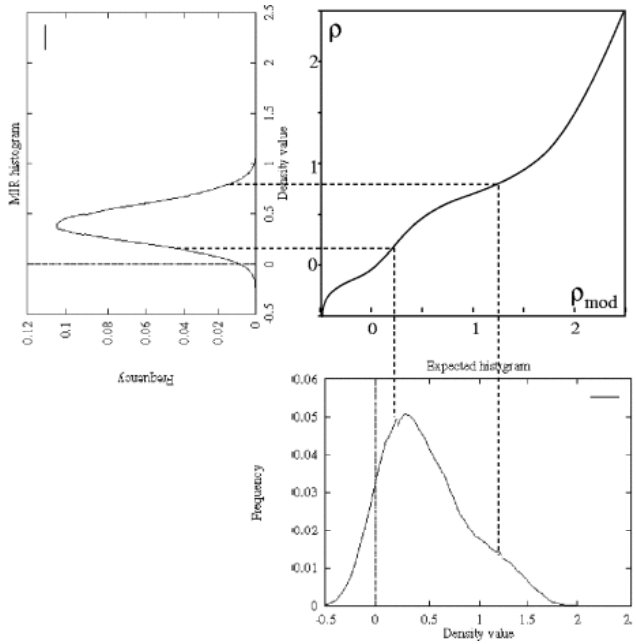
A technique from image processing for modifying the protein region.

- Noise maps have Gaussian histogram.
- Well phased maps have a skewed distribution: sharper peaks and bigger gaps.

Sharpen the protein density by a transform which matches the histogram of a well phased map.

Useful at better than 4A.

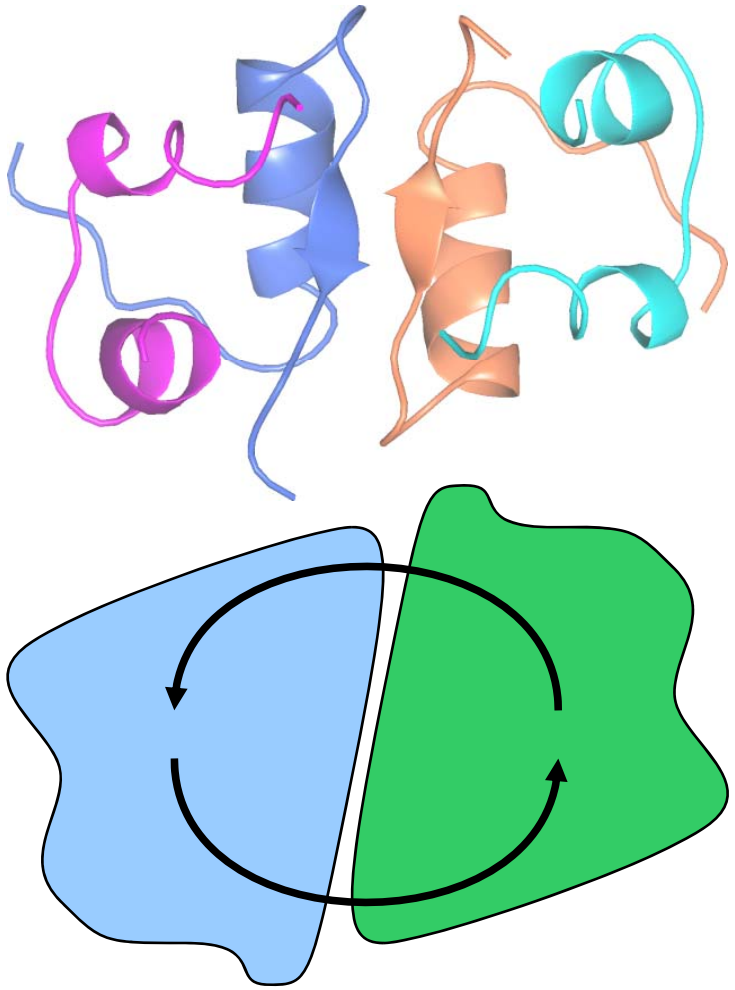




Kevin Cowtan, cowtan@ysbl.york.ac.uk

Non-crystallographic symmetry

- If the molecule has internal symmetry, we can average together related regions.
- In the averaged map, the signal-noise level is improved.
- If a full density modification calculation is performed, powerful phase relationships are formed.
- With 4-fold NCS, can phase from random!

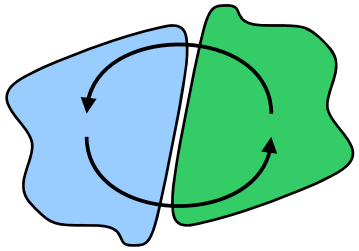


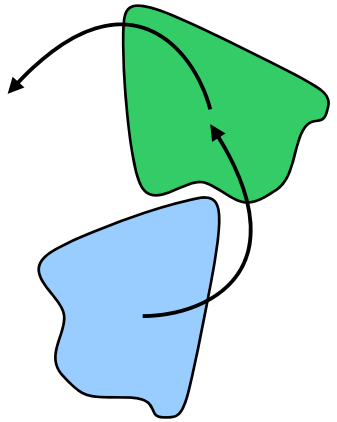
Kevin Cowtan, cowtan@ysbl.york.ac.uk

Non-crystallographic symmetry

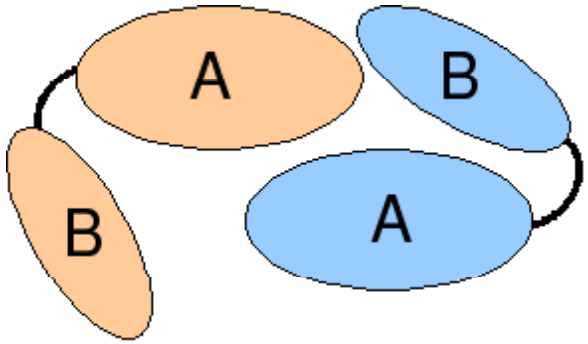
Useful terms:

- Proper and improper NCS: (closed and open)

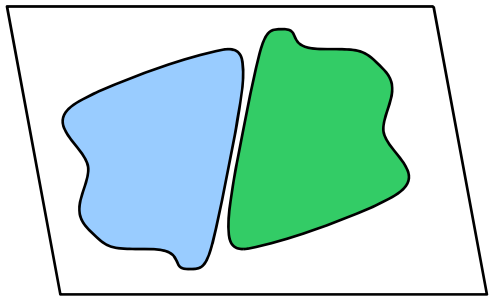


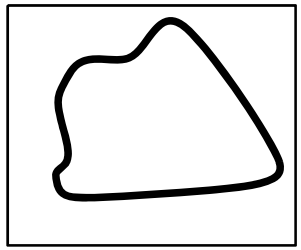


- Multi-domain averaging:



- Multi-crystal averaging:

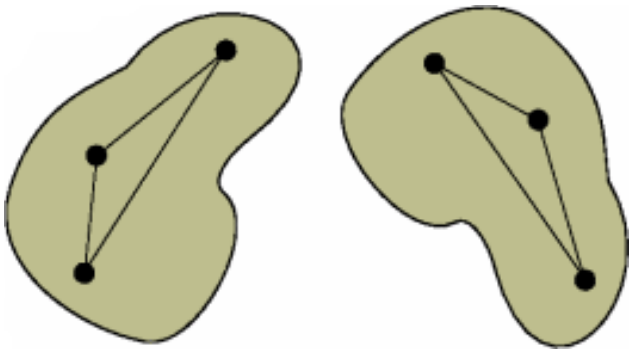




Kevin Cowtan, cowtan@ysbl.york.ac.uk

Non-crystallographic symmetry

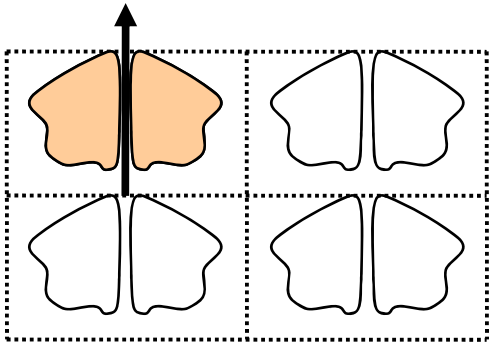
- How do you know if you have NCS?
  - Cell content analysis – how many monomers in ASU?
  - Self-rotation function.
  - Difference Pattersons (pseudo-translation only).
- How do you determine the NCS?
  - **From heavy atoms.**
  - From initial model building.
  - From molecular replacement.
  - *From density MR (hard).*
- Mask determined automatically.



Kevin Cowtan, cowtan@ysbl.york.ac.uk

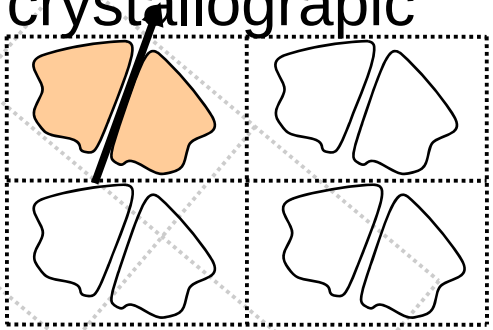
Non-crystallographic symmetry

Crystallographic

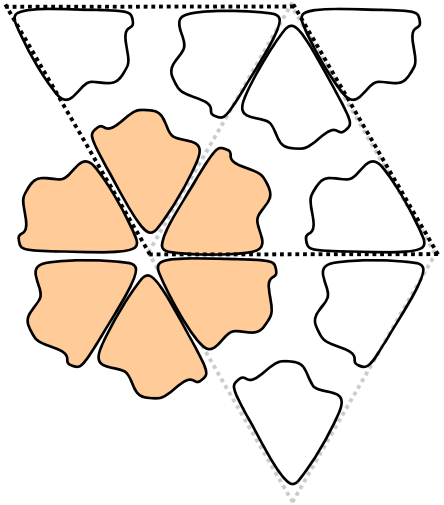


Aligned
2-fold

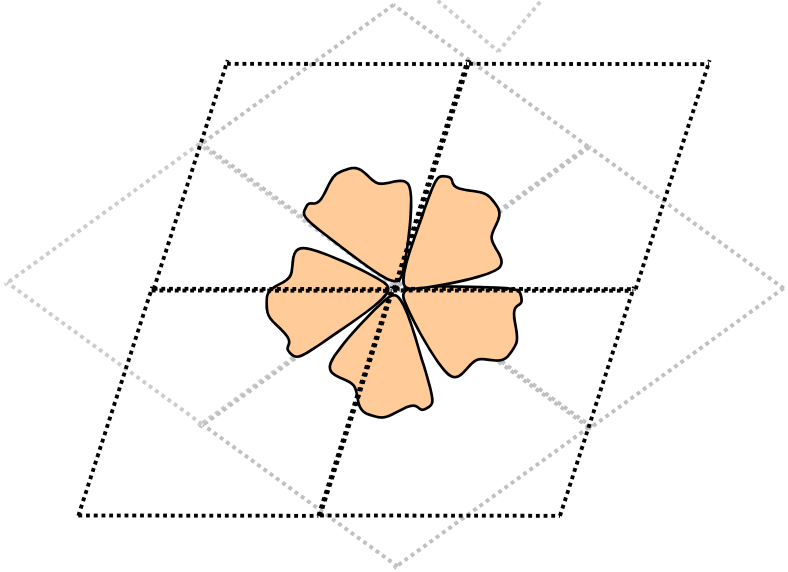
Non-
crystallograpic



Unaligned
2-fold



Aligned
6-fold



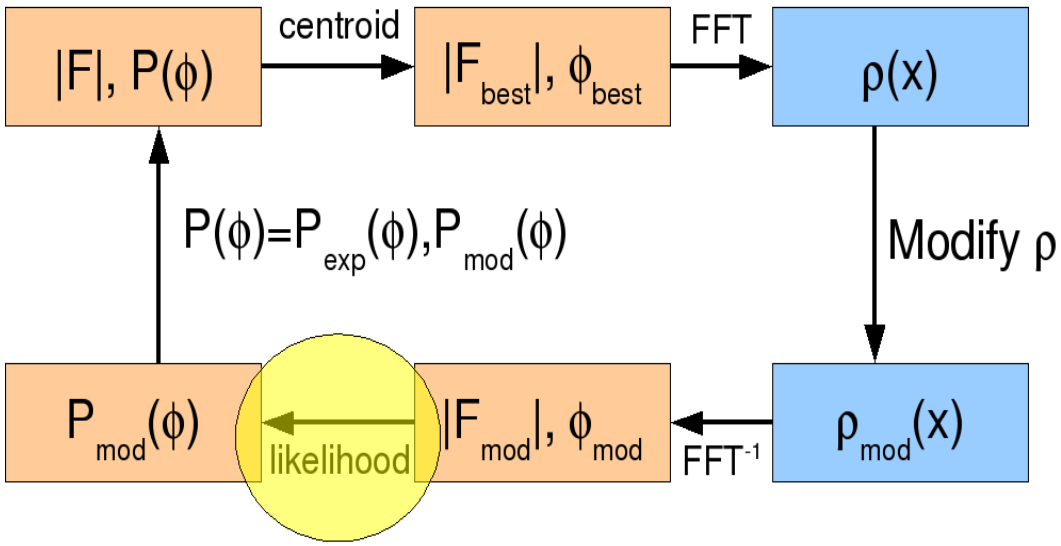
Aligned
5-fold

Kevin Cowtan, cowtan@ysbl.york.ac.uk

Estimating phase probabilities

Problem: How do we go from a single phase estimate to a full phase probability distribution?

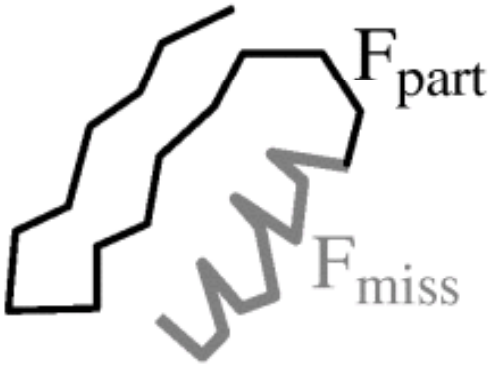
- We need to make an estimate of the error in the estimated phase.
- The errors in the phases are a parameter of the model itself, and may be estimated by likelihood methods.



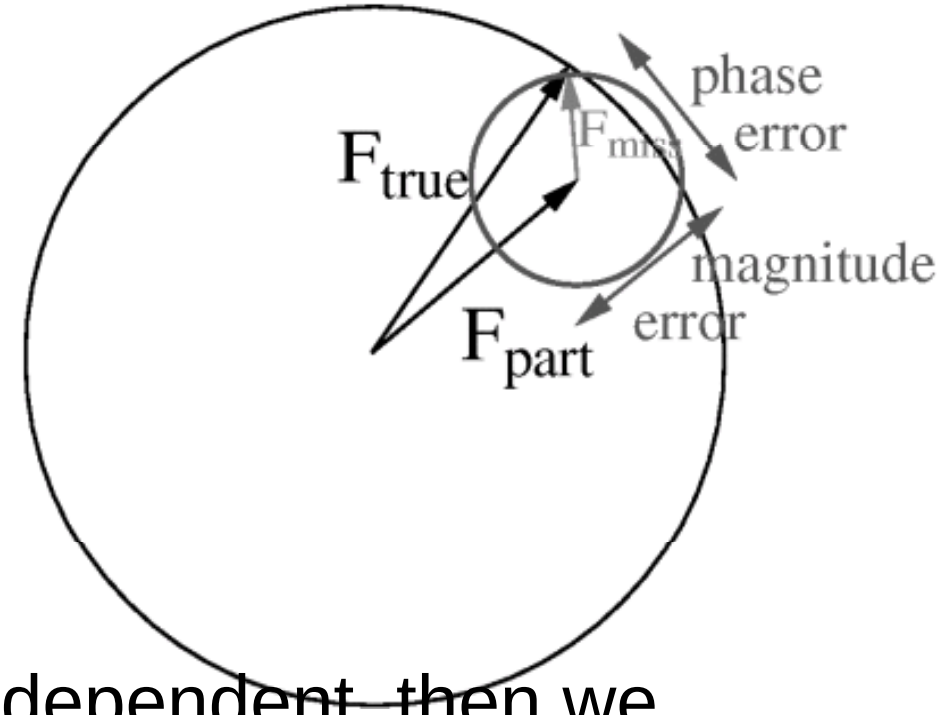
Kevin Cowtan, cowtan@ysbl.york.ac.uk

Estimating phase probabilities

Sim/σ_A weighting:



$F_{true} = F_{part} + F_{miss}$

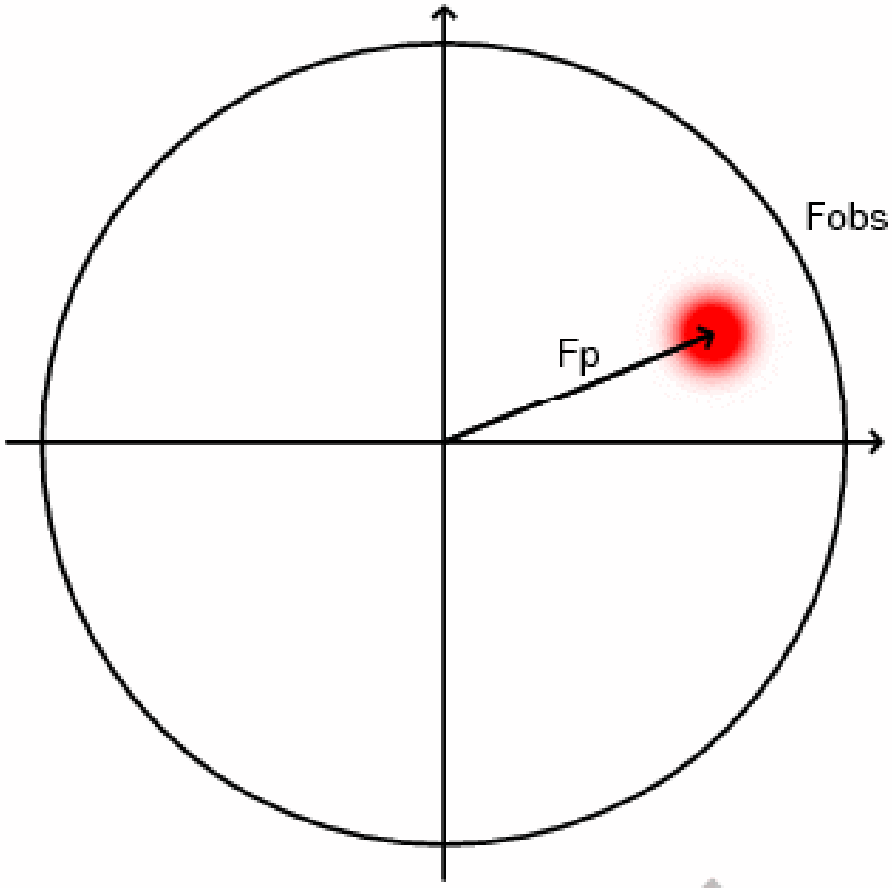


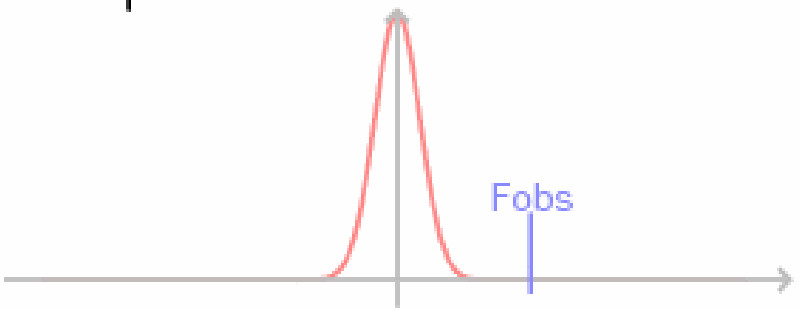
We know $|F_{true}|$, $|F_{part}|$, $\not\angle_{part}$

Assuming $\not\angle_{part}$, $\not\angle_{miss}$ are independent, then we expect the difference in magnitudes between $|F_{true}|$ and $|F_{part}|$, averaged over reflections, to give an indication of the phase error.

Kevin Cowtan, cowtan@ysbl.york.ac.uk

Estimating phase probabilities





Kevin Cowtan, cowtan@ysbl.york.ac.uk

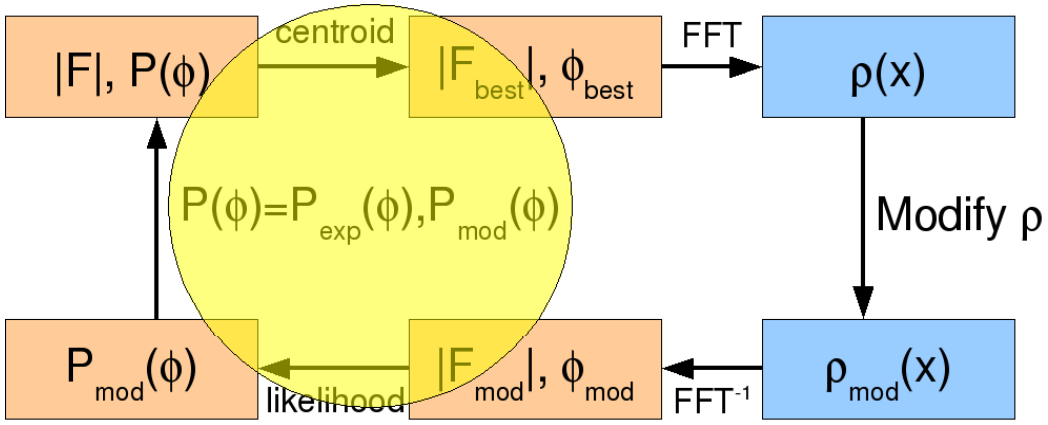
Combining phase probabilities

Once we have an estimate for the error in $\nearrow_{mod}$, we can construct a probability distribution $P_{mod}(\nearrow)$.

The the next cycle can be started with

$$P_{new}(\nearrow) = P_{exp}(\nearrow)P_{mod}(\nearrow)$$

Problem: $P_{exp}(\nearrow)$ and $P_{mod}(\nearrow)$ are not independent. The result is bias, increasing with cycle.



Kevin Cowtan, cowtan@ysbl.york.ac.uk

Bias reduction

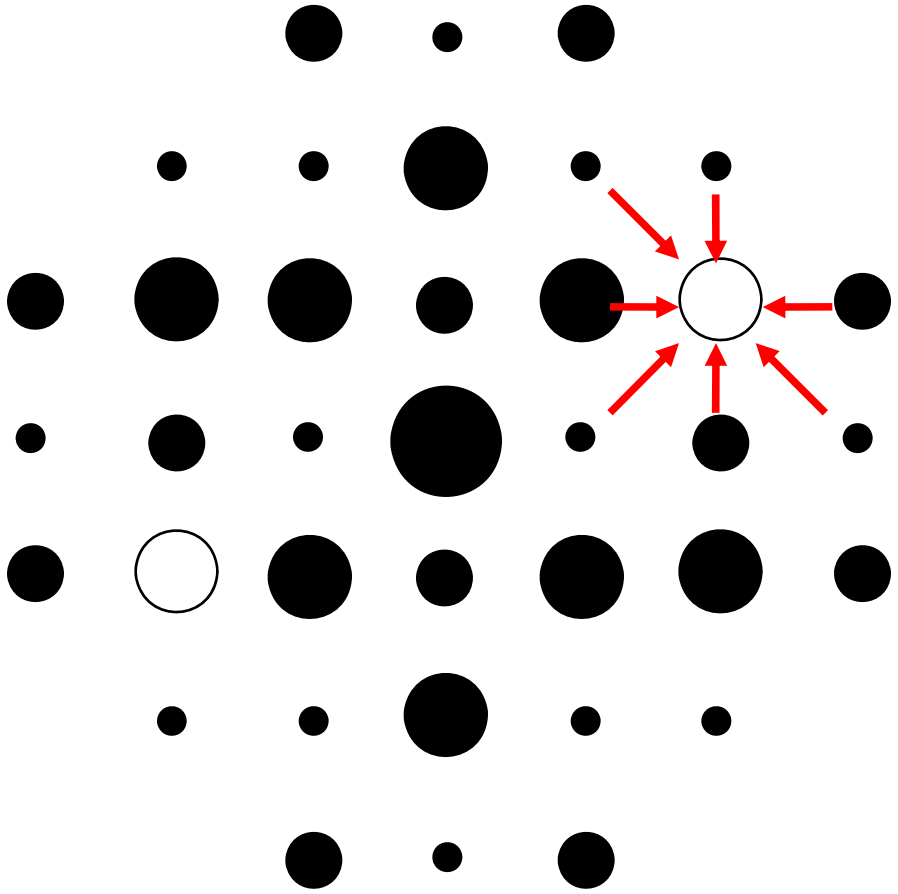
Solution:

Make each reflection only dependent on the other reflections in the diffraction pattern, and not on its own initial value.

Omit one reflection at a time, and use only the modified value of the omitted reflection. (Very slow.)

But can be implemented efficiently:

- Solvent flipping
- The γ-correction



Kevin Cowtan, cowtan@ysbl.york.ac.uk

Density modification in Parrot

Builds on existing ideas:

- DM:
  - Solvent flattening
  - Histogram matching
  - NCS averaging
  - Perturbation gamma
- Solomon:
  - Gamma correction
  - Local variance solvent mask
  - Weighted averaging mask

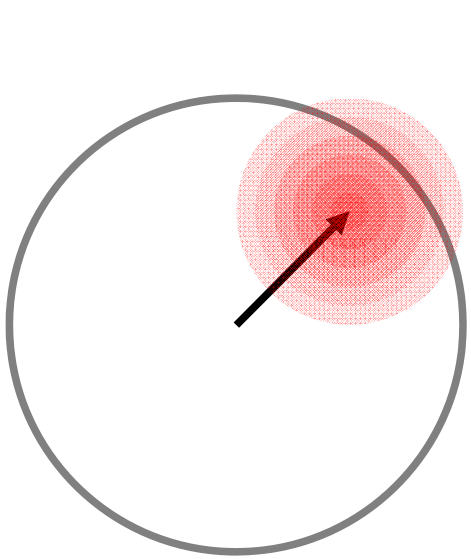
Density modification in Parrot

New developments:

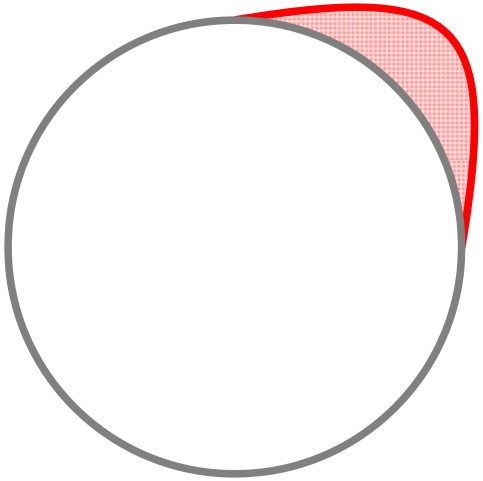
- MLHL phase combination
  - (as used in refinement: *refmac*, *phenix.refine*)
- Anisotropy correction
- Problem-specific density histograms
  - (rather than a standard library)
- Pairwise-weighted NCS averaging...

Estimating phase probabilities

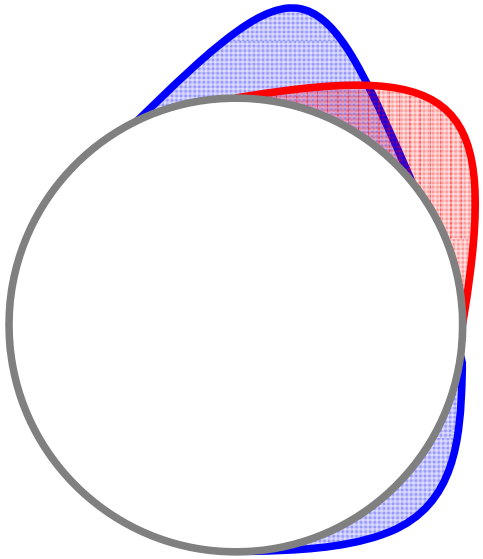
Traditional approach: Rice likelihood function



Estimate the accuracy of the modified F/phase



Turn this into a phase probability distribution

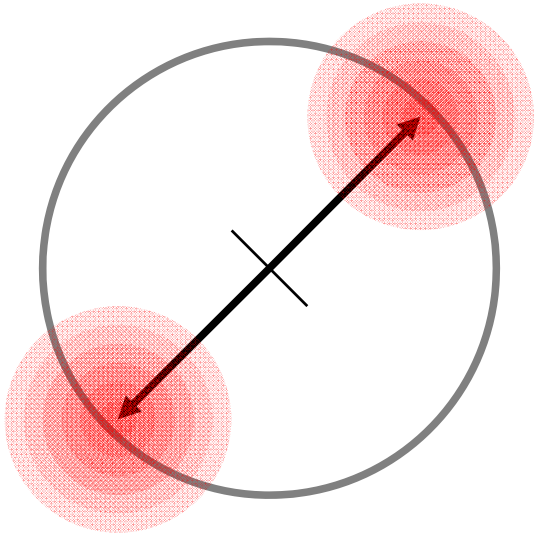


Combine with the experimental phase probability

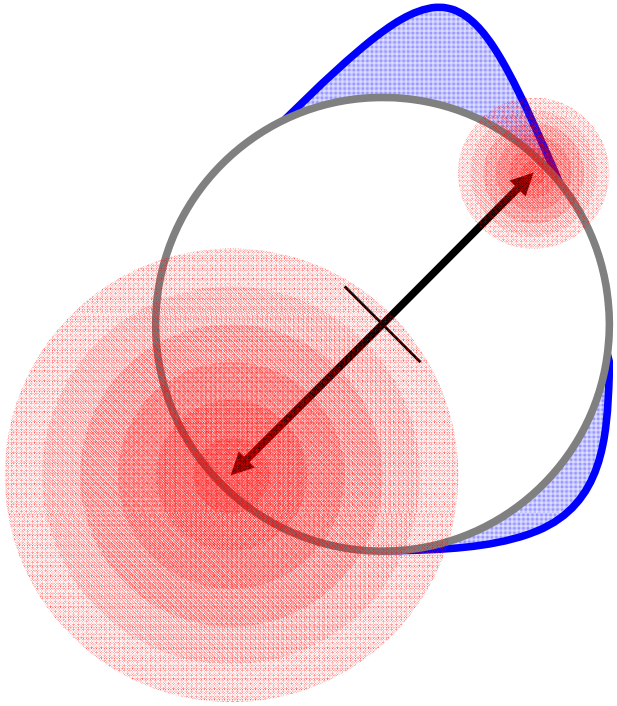
The estimate for the accuracy of the modified F/phase come from the agreement between the modified F and the observed F. **Source of bias.**

Estimating phase probabilities

Problem:



Error estimation does not take into account experimental phase information



The experimental data tells us that the probable error is different in the two cases

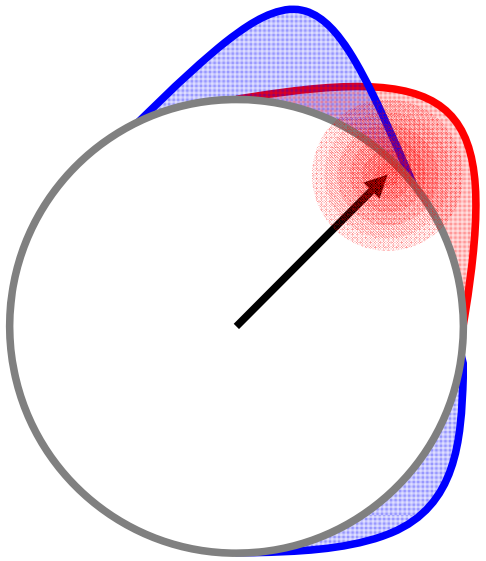
Using the additional information from the phases improves the error model and reduces bias.

Estimating phase probabilities

Solution:

MLHL-type likelihood

target function.



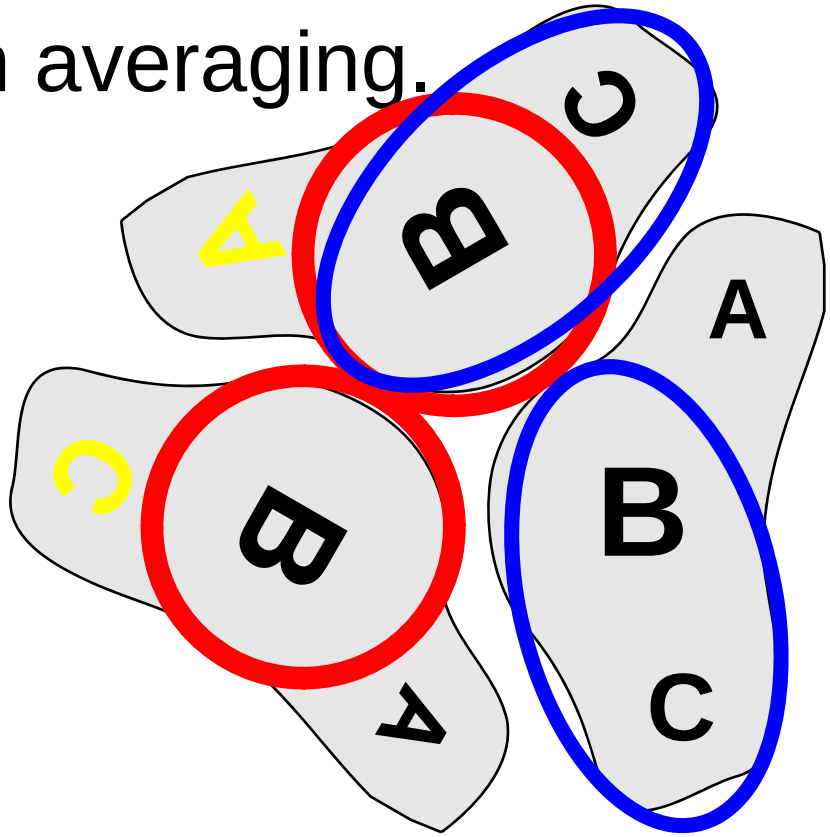
Perform the error estimation and phase combination in a single step, using a likelihood function which incorporates the experimental phase information as a prior.

This is the same MLHL-type like likelihood refinement target used in modern refinement software such as *refmac* or *phenix.refine*.

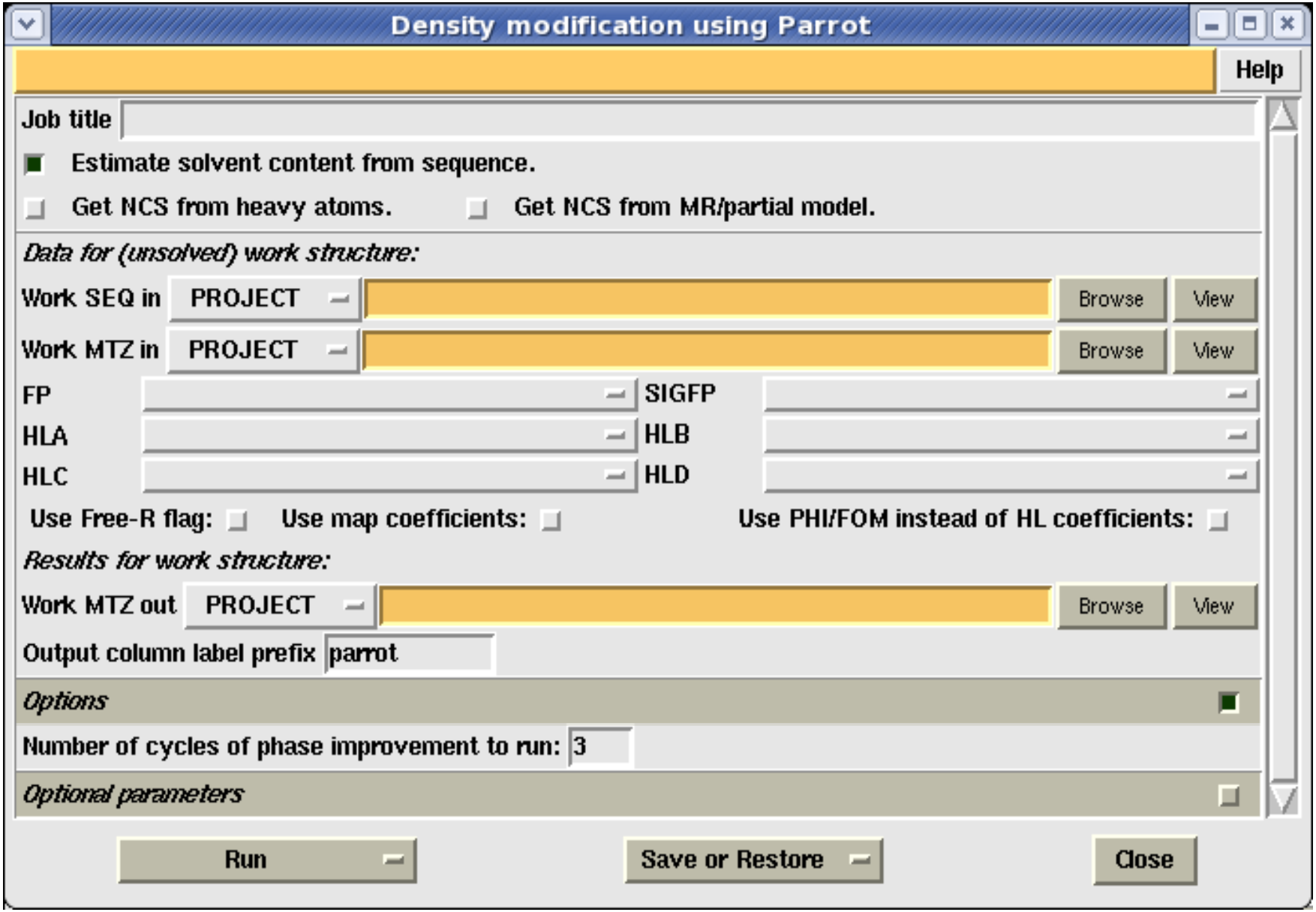
Recent Developments:

Pairwise-weighted NCS averaging:

- Average each pair of NCS related molecules separately with its own mask.
- Generalisation and automation of multi-domain averaging.



Parrot


Density modification using Parrot
Help
Job title
Estimate solvent content from sequence.
Get NCS from heavy atoms.
Get NCS from MR/partial model.
Data for (unsolved) work structure:
Work SEQ in
PROJECT
Browse
View
Work MTZ in
PROJECT
Browse
View
FP
SIGFP
HLA
HLB
HLC
HLD
Use Free-R flag:
Use map coefficients:
Use PHI/FOM instead of HL coefficients:
Results for work structure:
Work MTZ out
PROJECT
Browse
View
Output column label prefix
parrot
Options
Number of cycles of phase improvement to run:
3
Optional parameters
Run
Save or Restore
Close

Parrot

Summary:

A new classical density modification program, employing the latest techniques.

- Fully automated
- Fast
- Better results than DM

Density Modification

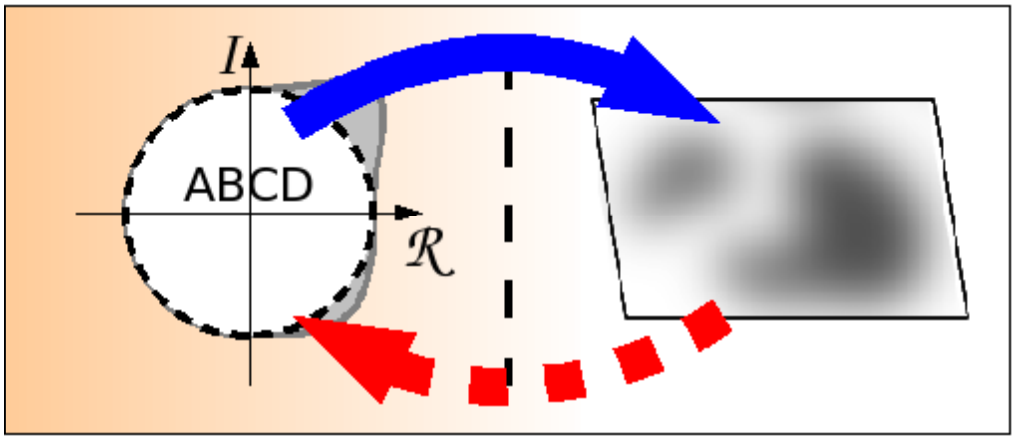
Kevin Cowtan, York.

Statistical density modification:
e.g. Resolve, Pirate

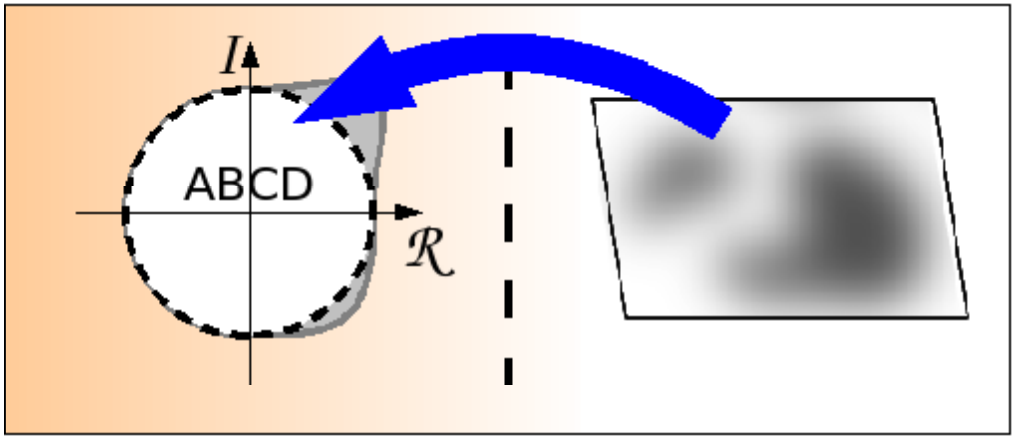
Kevin Cowtan, cowtan@ysbl.york.ac.uk

Density modification

- Traditional density modification:
  *Take the **phases** to the **mask**.*
  Use them to calculate a map.
  But how do we get back to:
  - reciprocal space?
  - probabilities?



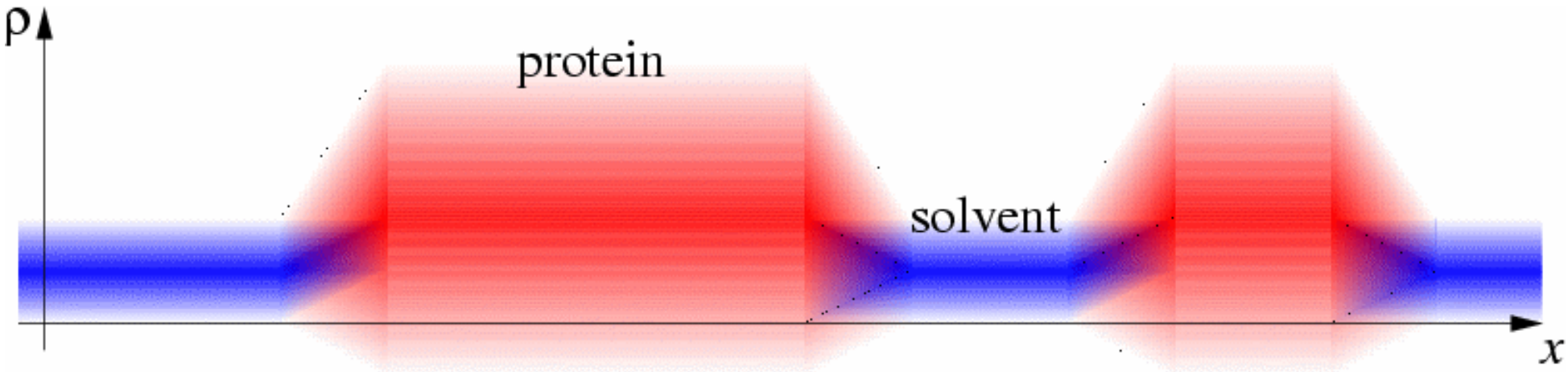
- Statistical density modification:
  *Take the **mask** to the **phases**.*
  - First convert mask to probability.
  - Then transform that probability.



Kevin Cowtan, cowtan@ysbl.york.ac.uk

Statistical density modification

- Form a statistical description of expected map features.

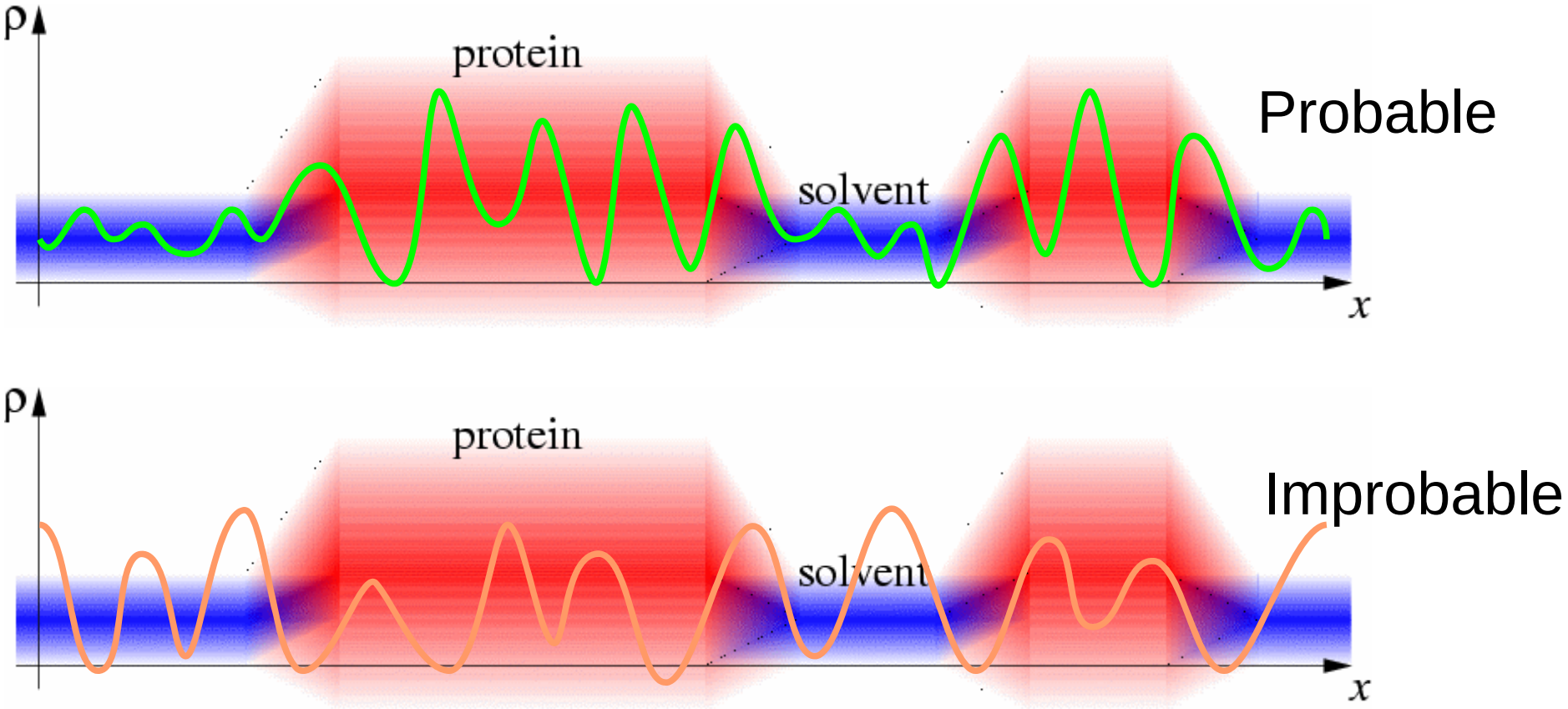


- e.g.
  - Protein has higher mean, and is more peaky (higher variance)
  - Solvent has lower mean, and is flatter (lower variance)

Kevin Cowtan, cowtan@ysbl.york.ac.uk

Statistical density modification

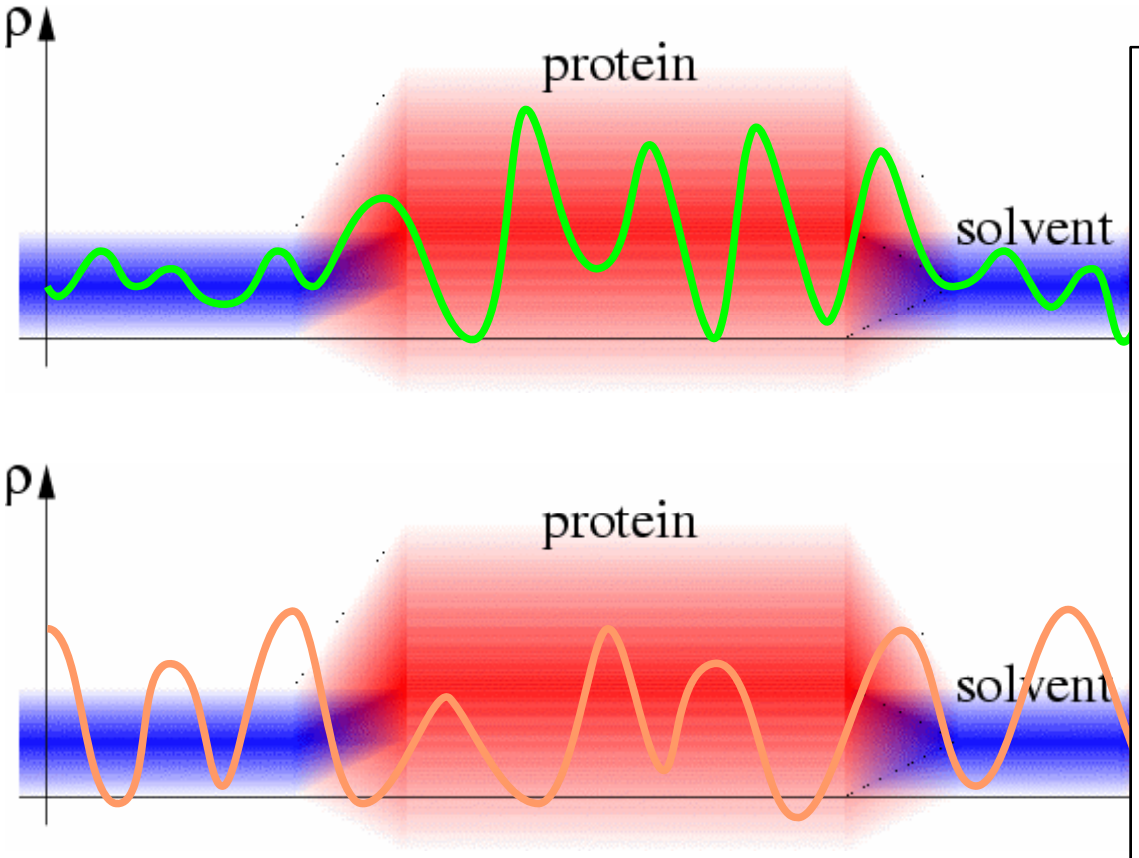
- Probability of a map is determined by how well it fits these distributions:

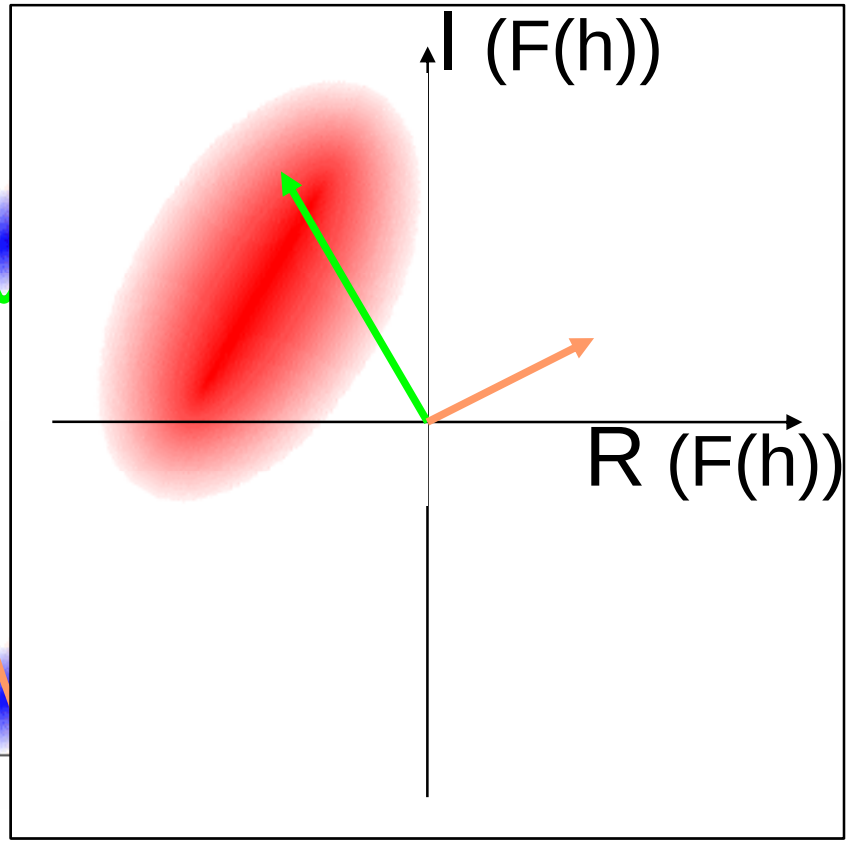


Kevin Cowtan, cowtan@ysbl.york.ac.uk

Statistical density modification

- Probability of each structure factor is given by the probability of the corresponding map.

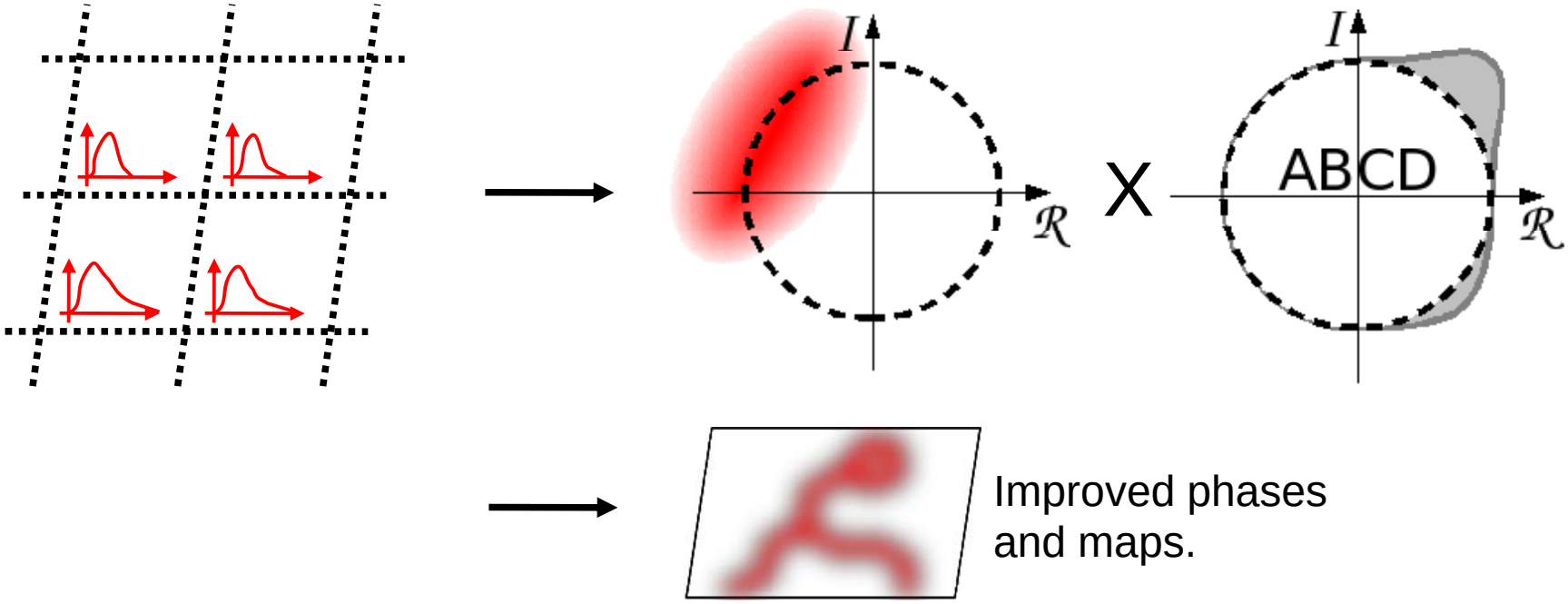




Kevin Cowtan, cowtan@ysbl.york.ac.uk

Statistical density modification

- Obtain per-grid density probability distributions.
- Transform to reciprocal space.
- Combine with experimental phases.
  - Map probability becomes phase probability distribution.



Bricogne (1992) Proc. CCP4 Study Weekend
Bricogne (1997) Methods in Enzymology

Kevin Cowtan, cowtan@ysbl.york.ac.uk

Statistical density modification

Advantages:

- Reduced bias.
- Better phases.

Disadvantages:

- Slow.
- PIRATE in particular works well for some cases and badly for others.

Kevin Cowtan, cowtan@ysbl.york.ac.uk

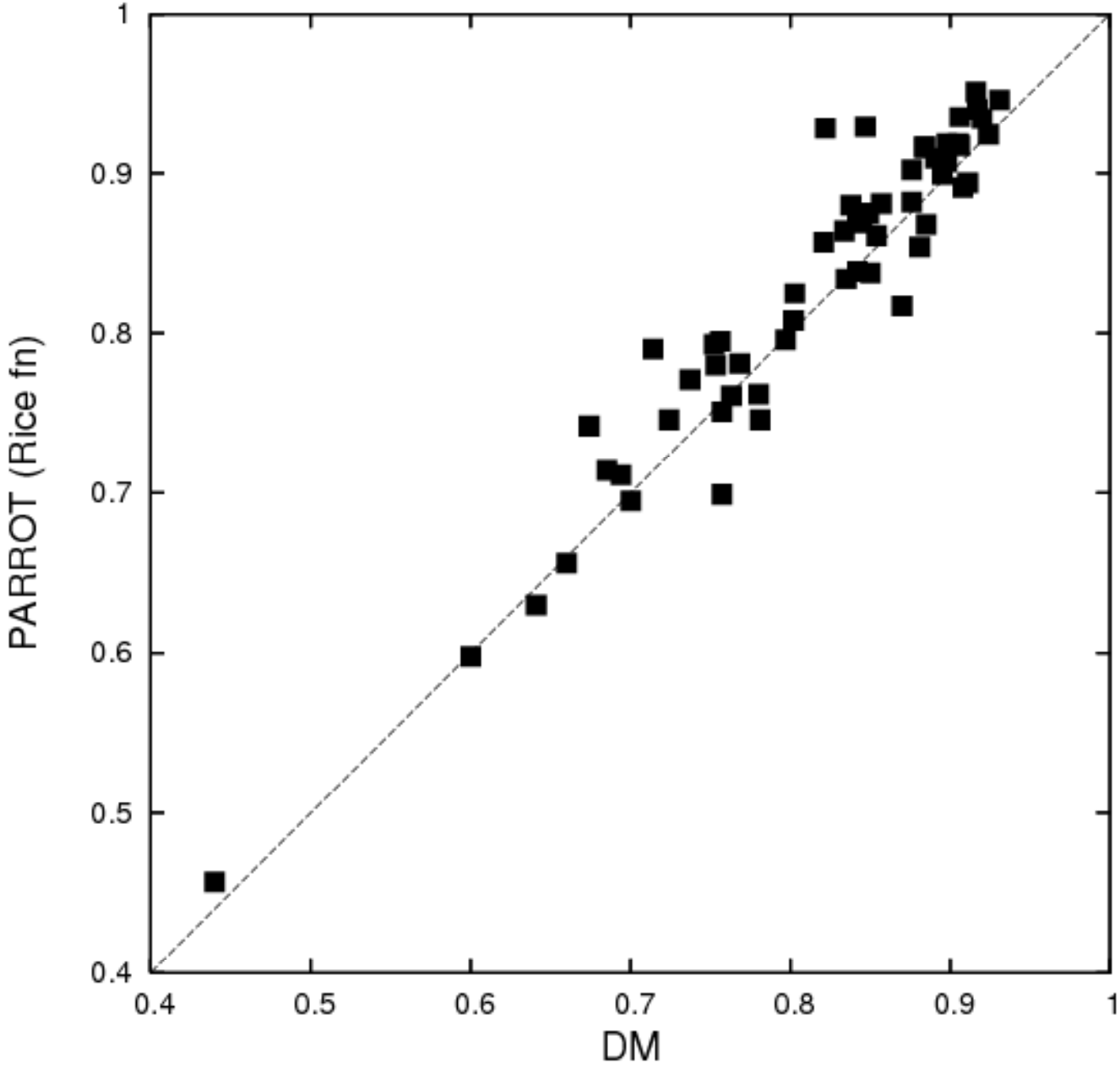
Density Modification

Kevin Cowtan, York.

Some results...

Kevin Cowtan, cowtan@ysbl.york.ac.uk

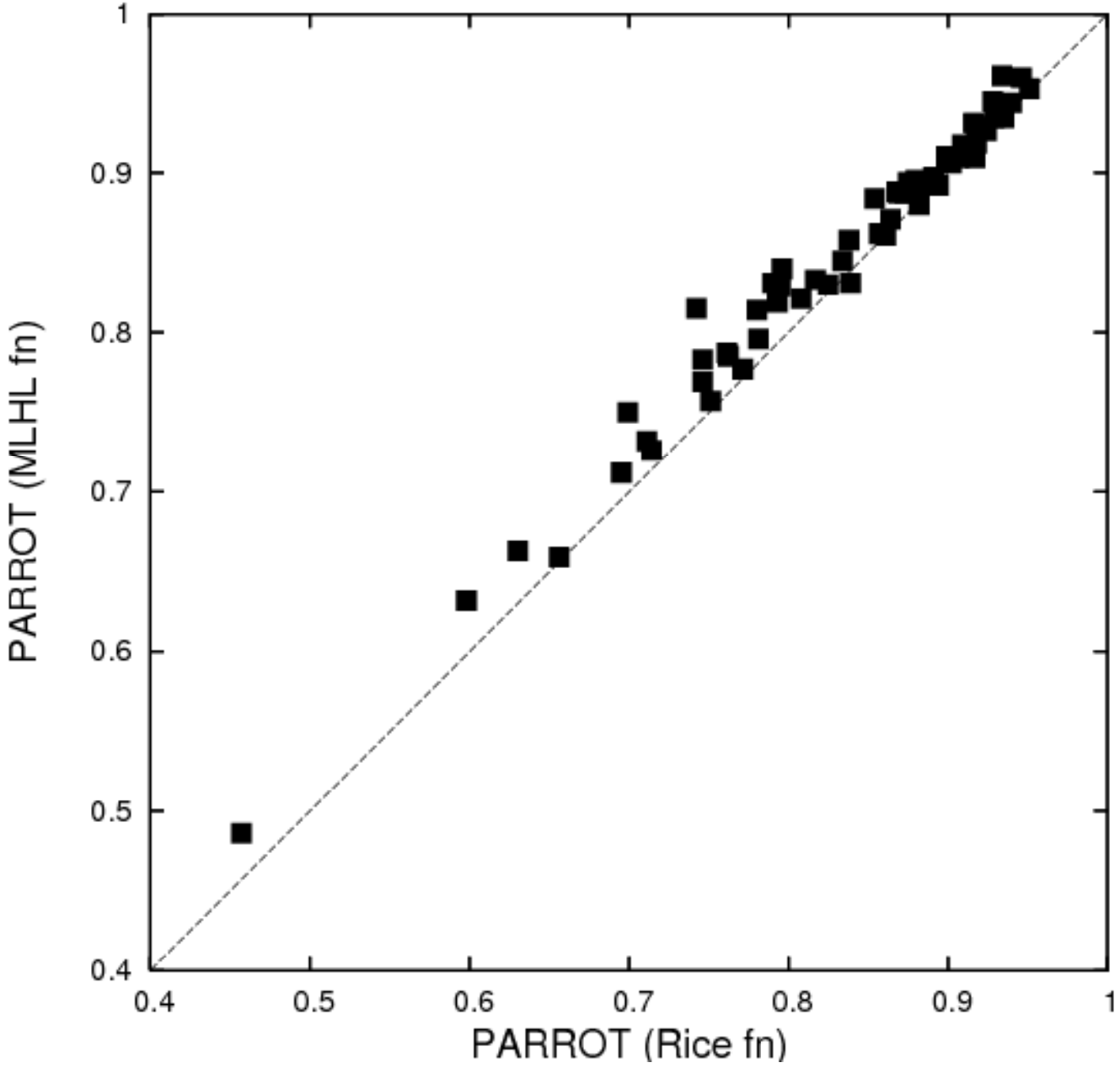
DM vs Parrot



Map correlations

Parrot: No new features enabled.

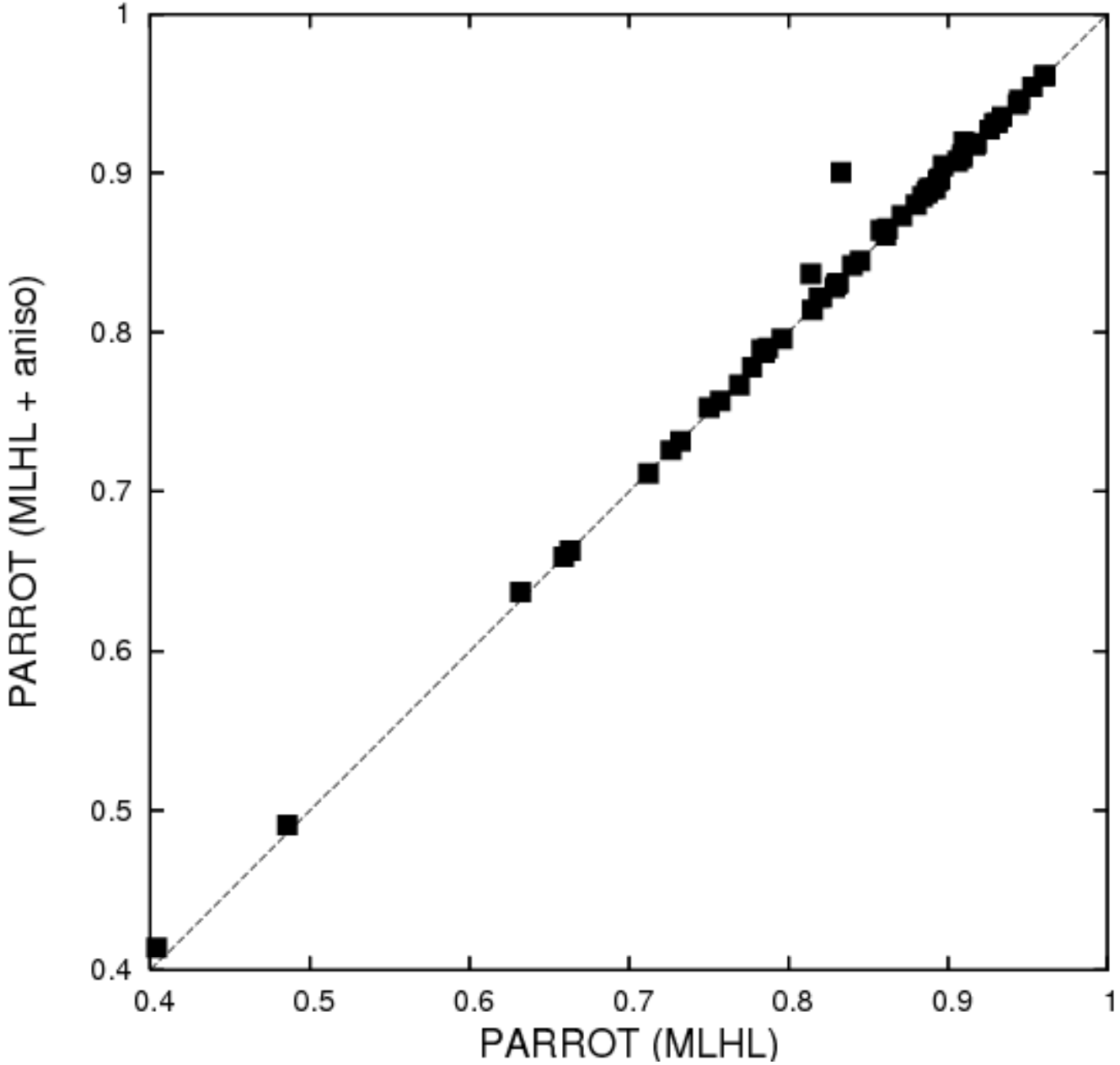
Parrot: Rice vs MLHL



Map correlations

Comparing old and new likelihood functions.

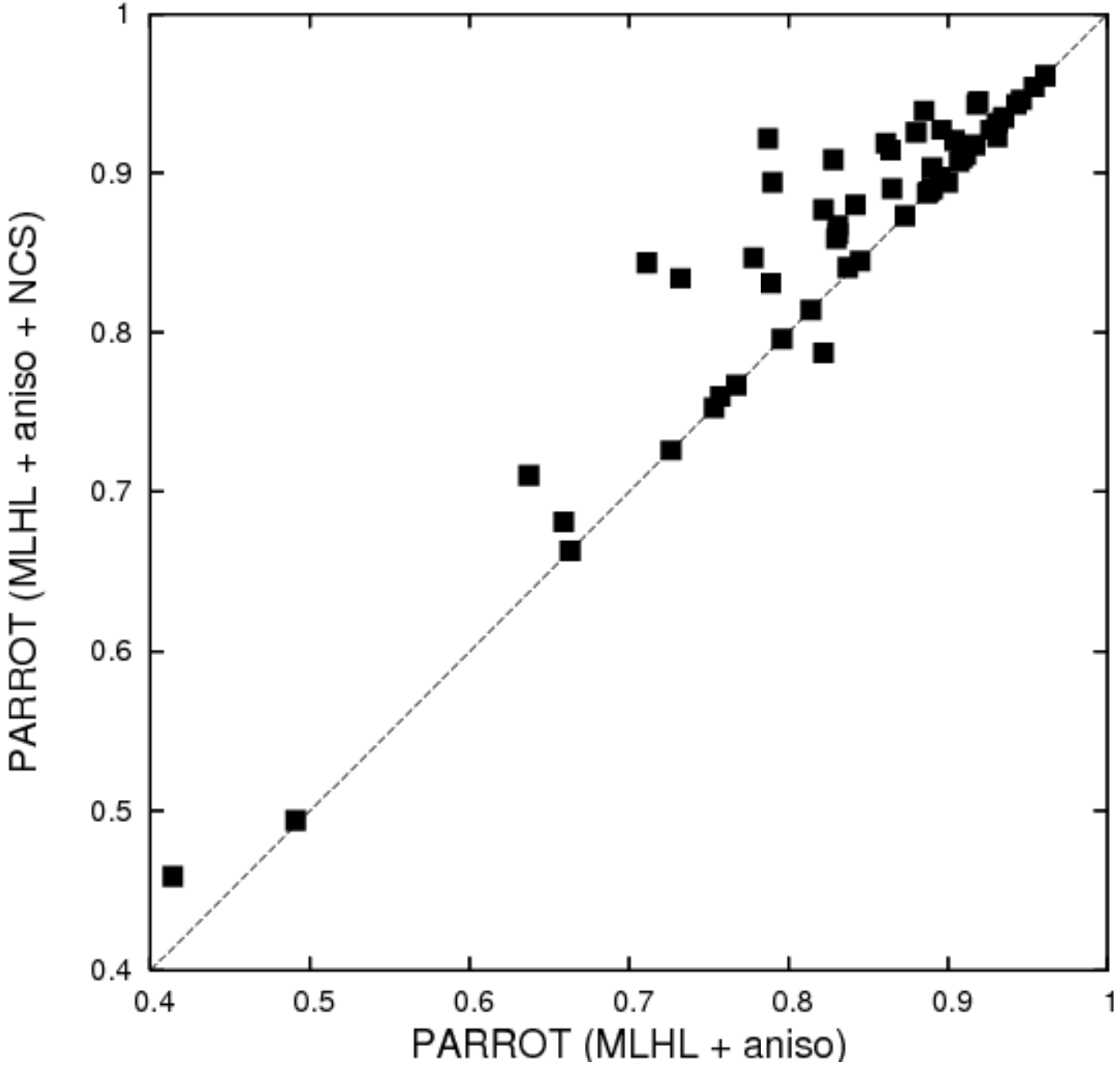
Parrot: Isotropic vs Anisotropic



Map correlations

Comparing with and without anisotropy correction.

Parrot: simple vs NCS averaged



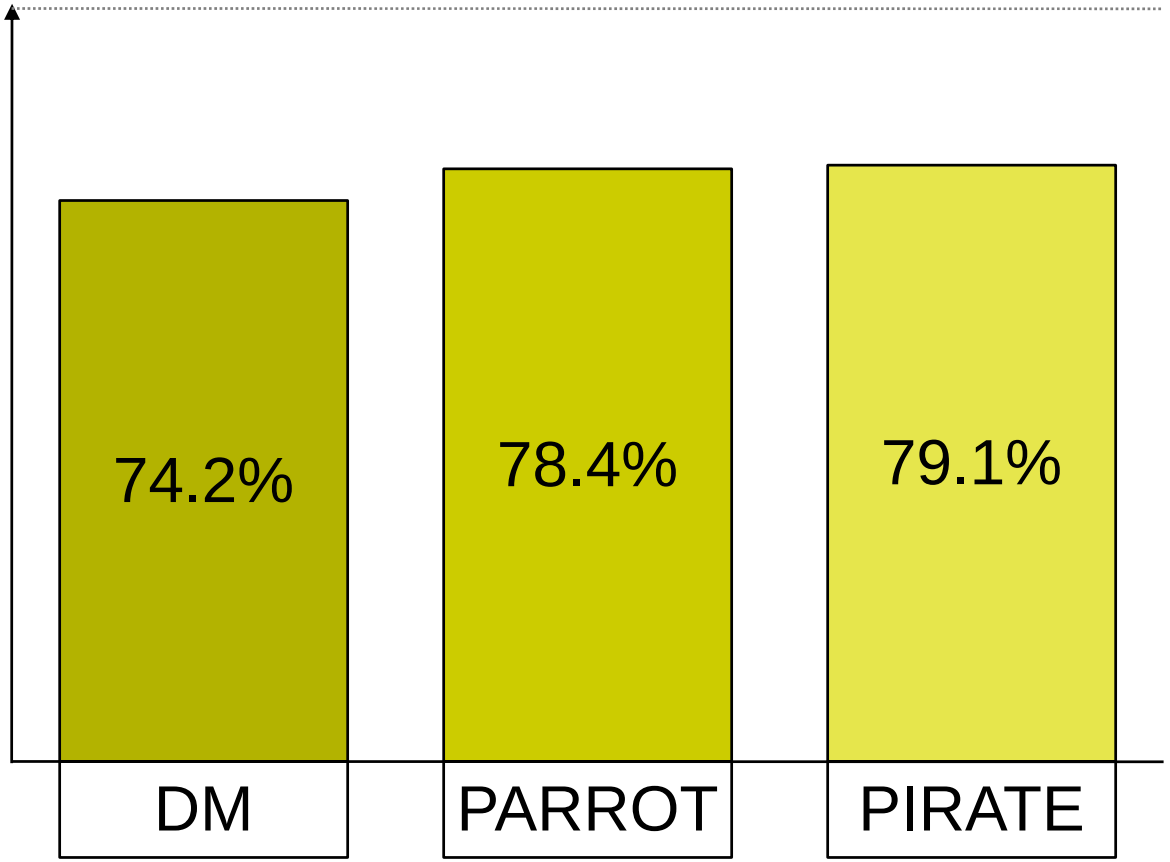
Map correlations

Comparing with and without NCS averaging.

DM vs PARROT vs PIRATE

% residues autobuilt and sequenced

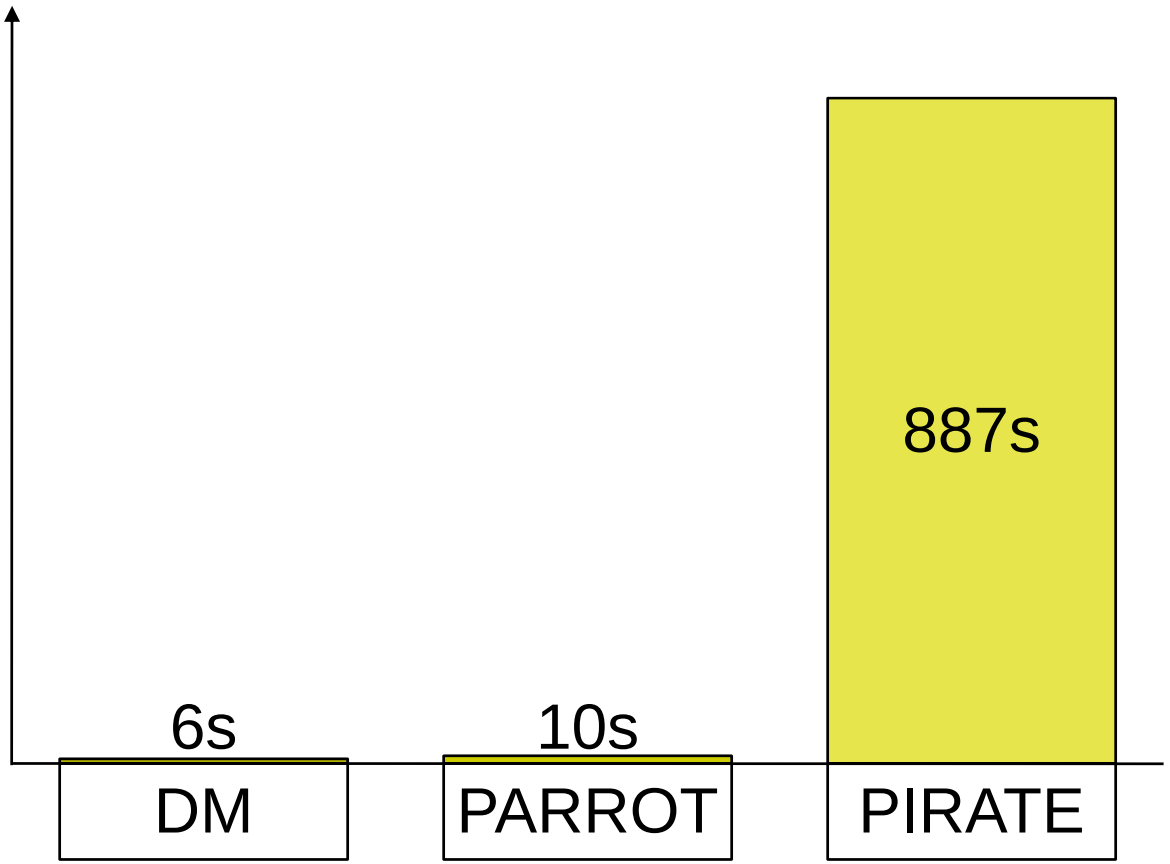
50 JCSG structures, 1.8-3.2A resolution



DM vs PARROT vs PIRATE

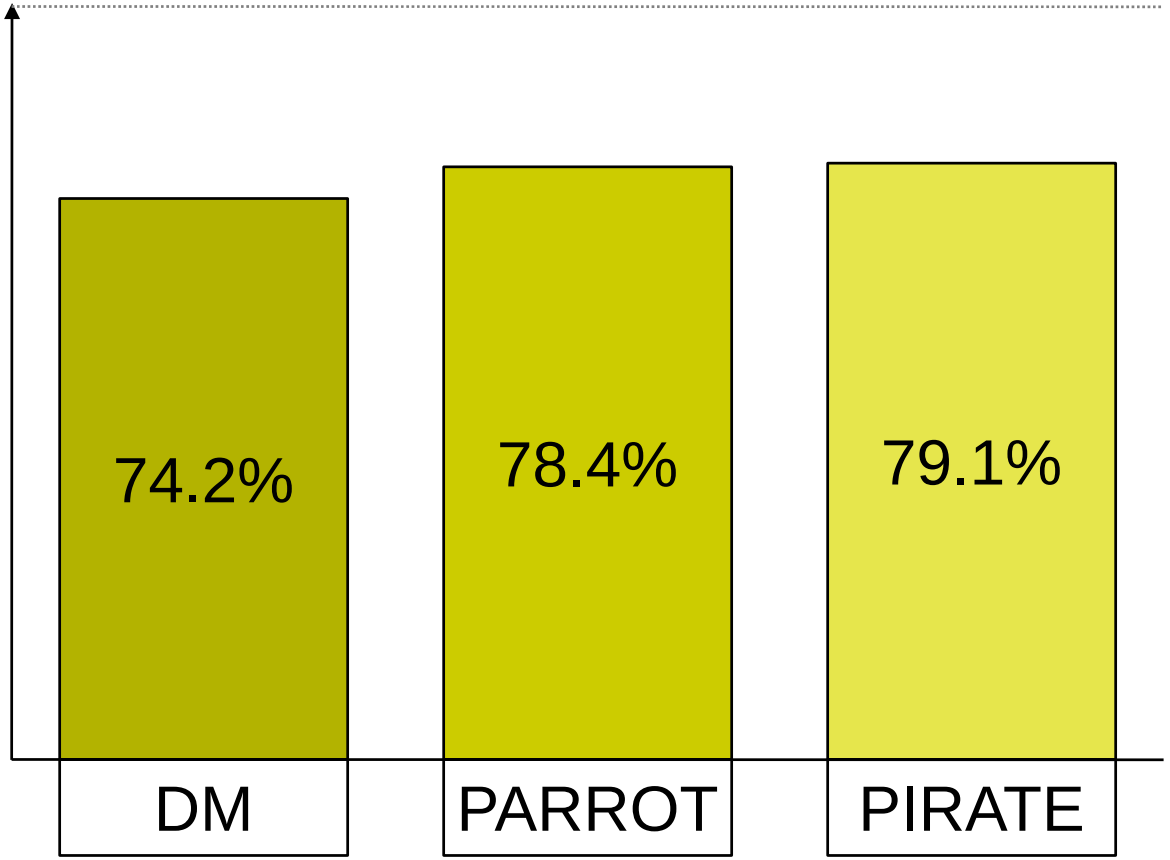
Mean time taken

50 JCSG structures, 1.8-3.2A resolution



DM vs PARROT vs PIRATE

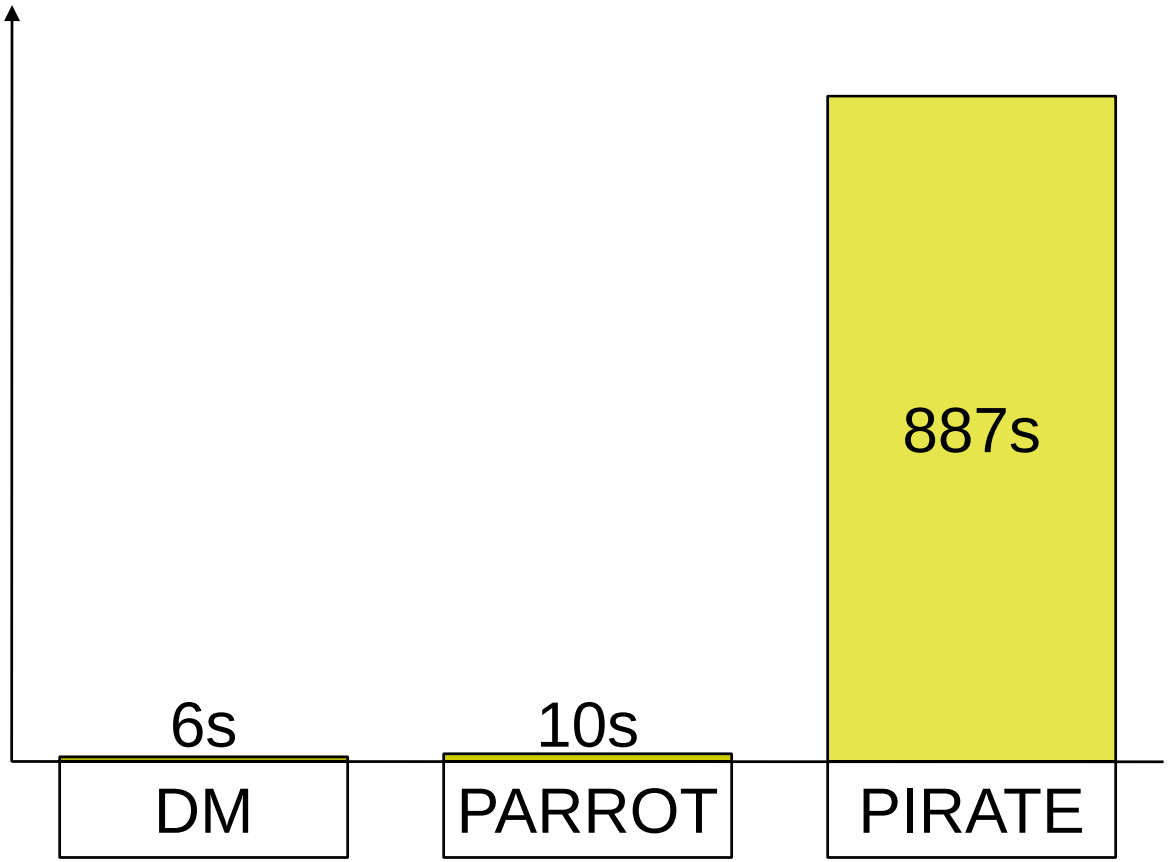
% residues autobuilt and sequenced
50 JCSG structures, 1.8-3.2A resolution



Kevin Cowtan, cowtan@ysbl.york.ac.uk

DM vs PARROT vs PIRATE

Mean time taken
50 JCSG structures, 1.8-3.2A resolution



Kevin Cowtan, cowtan@ysbl.york.ac.uk

Model Building

Model building software:

- Buccaneer