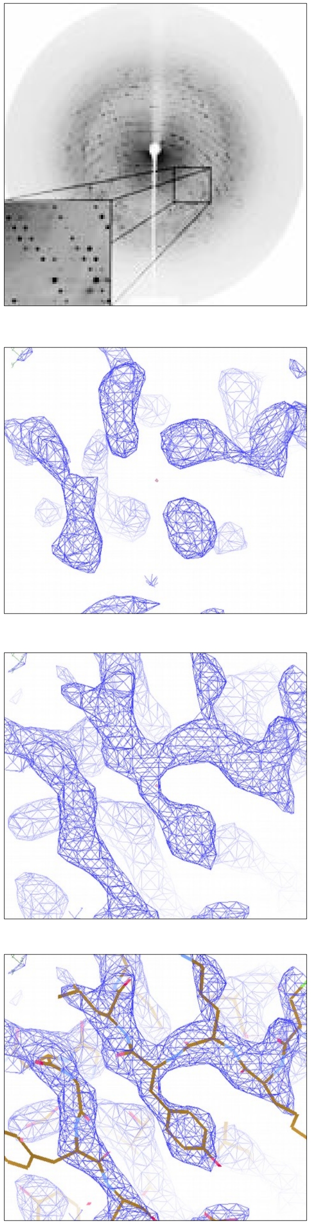
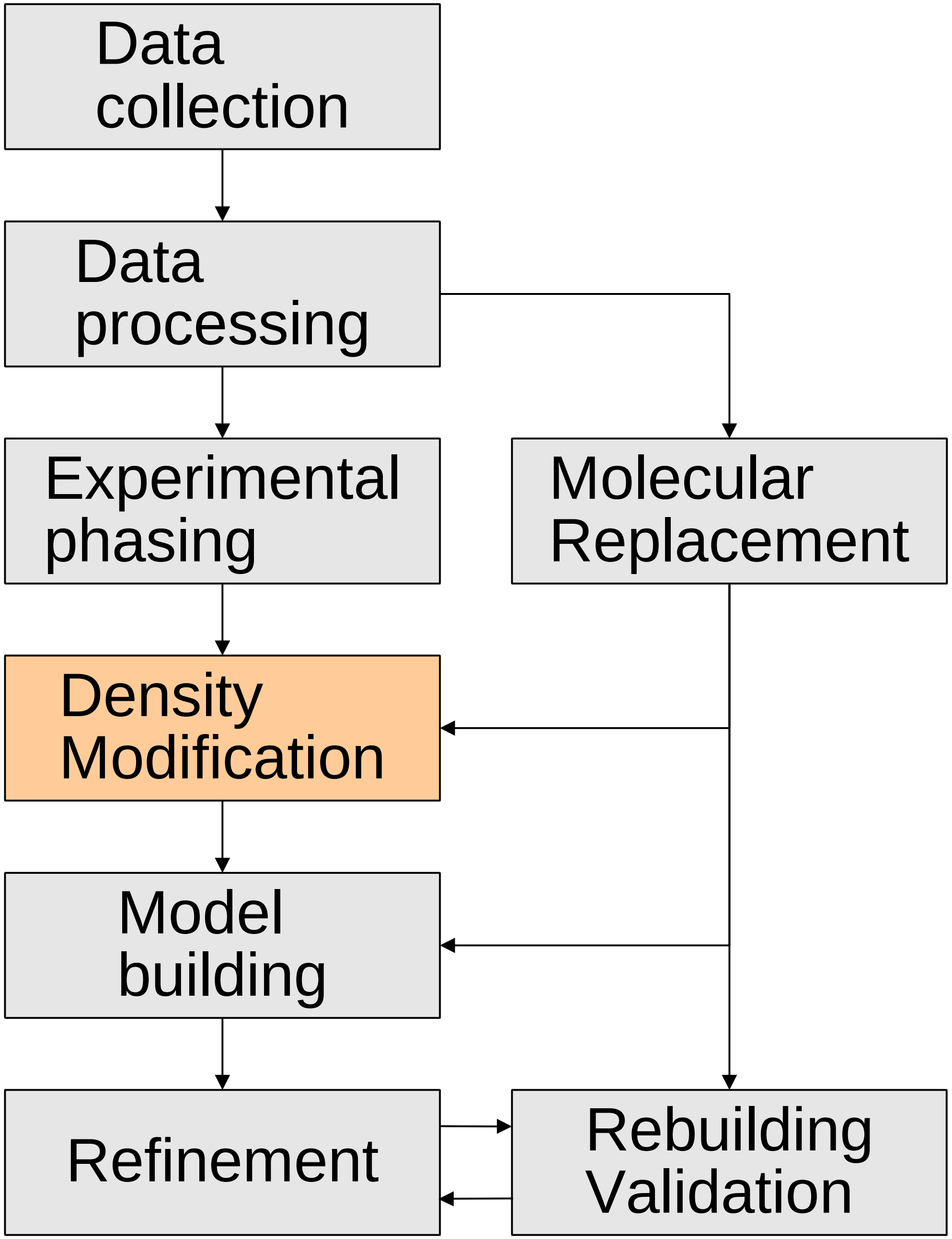
Density Modification

Paul Bond, Kevin Cowtan

kevin.cowtan@york.ac.uk

X-ray structure solution pipeline...





Data collection

Data processing

Experimental phasing

Molecular Replacement

Density Modification

Model building

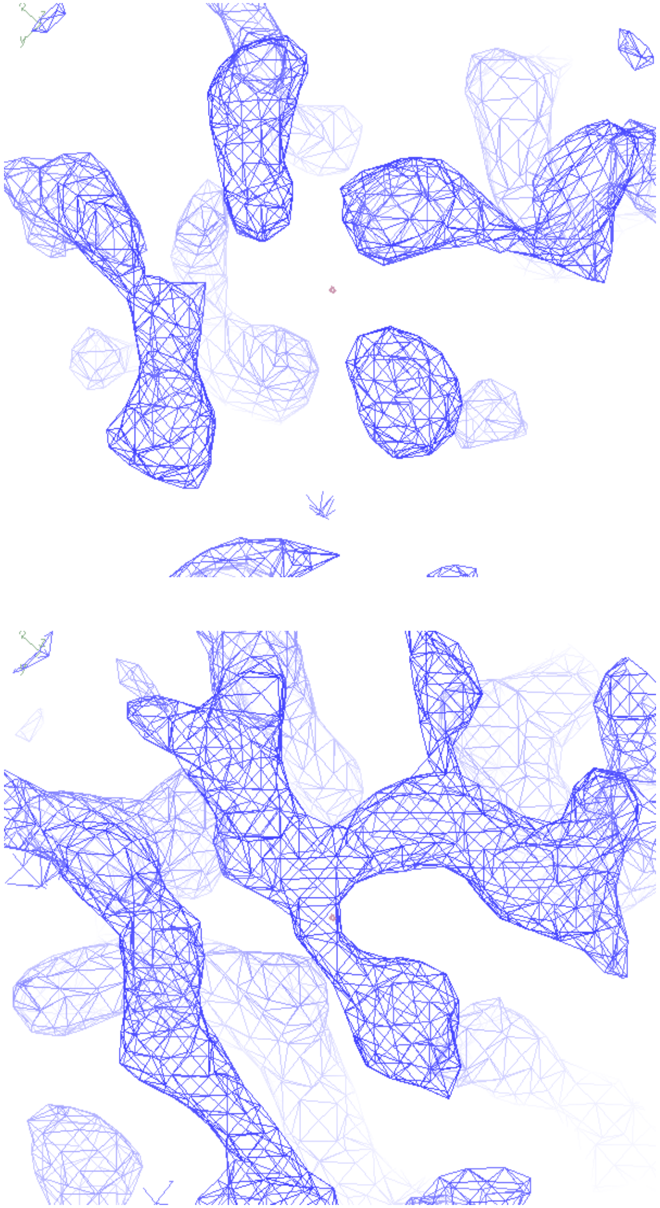
Refinement

Rebuilding Validation

Density modification

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

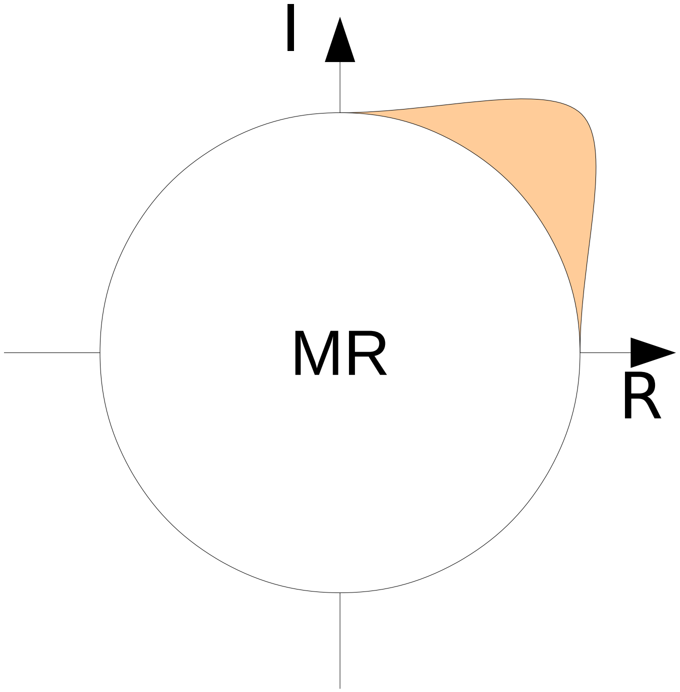
- Statistical density modification: e.g. 'resolve'

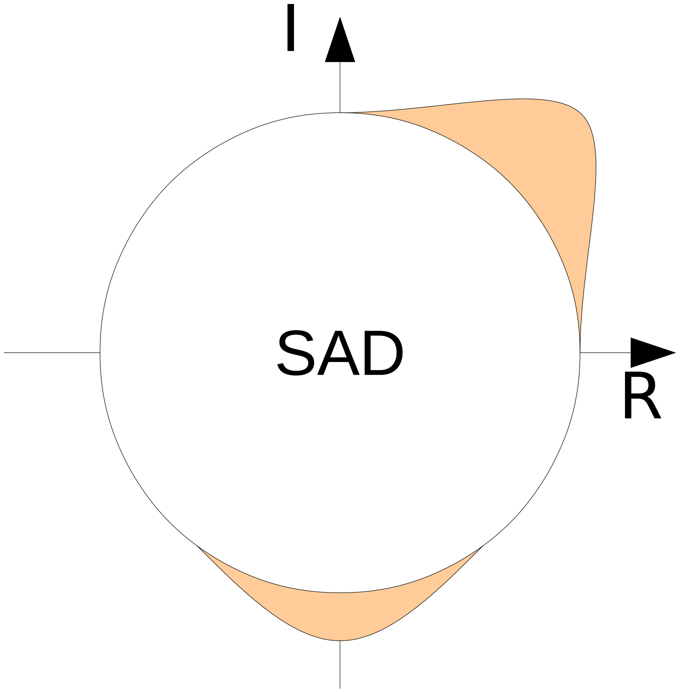


Density modification

Starting point:

- Structure factor amplitudes
- Phase estimates:
    - MR: Unimodel distribution
    - SAD: Bimodal distribution

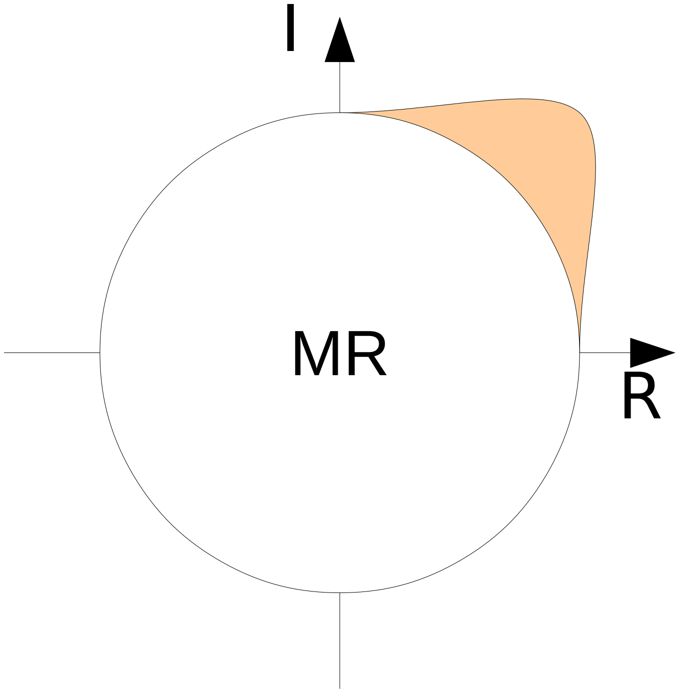


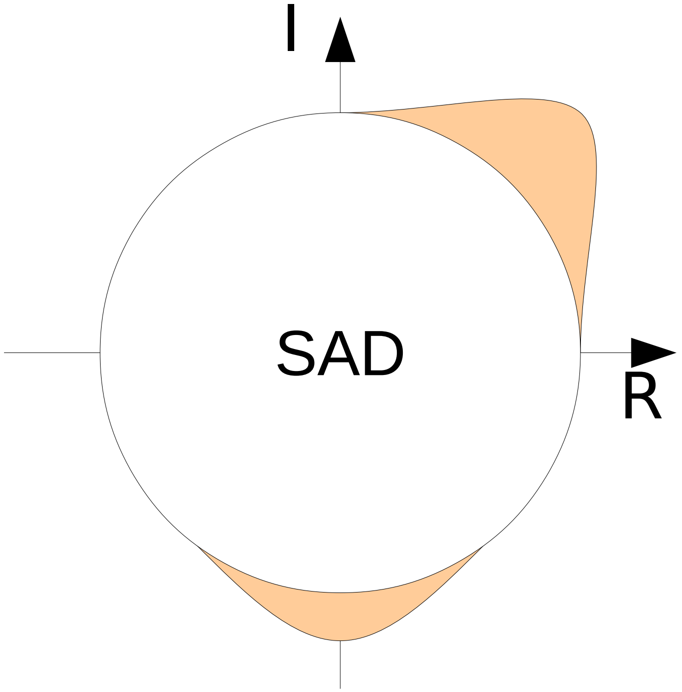


Density modification

How do we represent phase probability distributions?

- Phase/figure of merit - Φ, FOM
  - (unimodel, MR only)
- Hendrickson-Lattman coeffs – ABCD
  - (bimodal or unimodal, general)



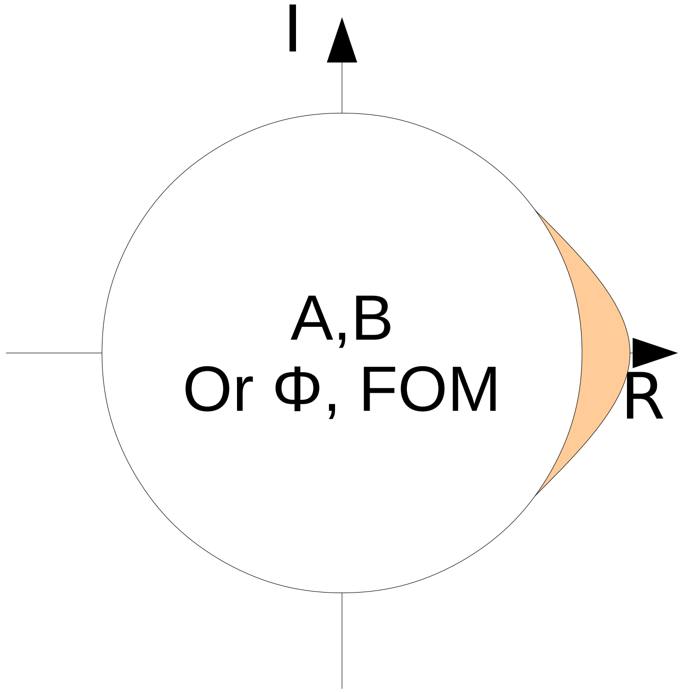


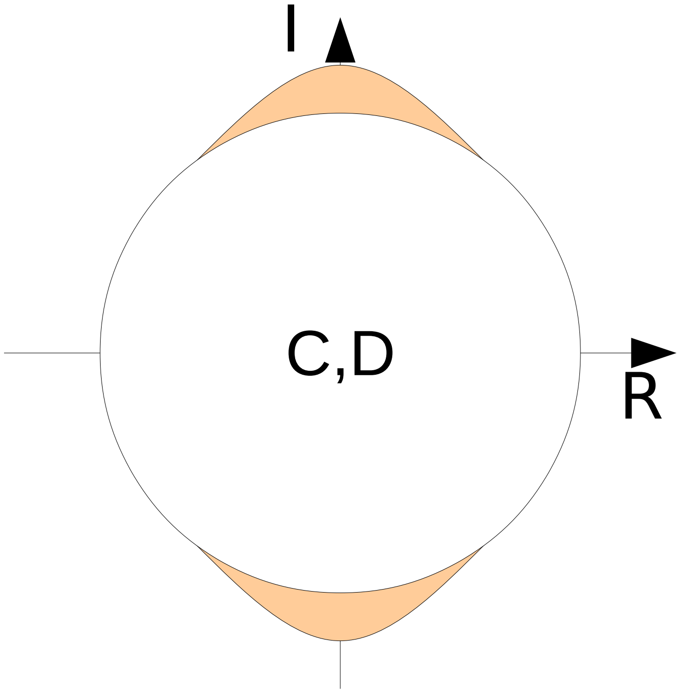
Density modification

A,B represent a unimodal distribution (equivalent to Φ, FOM)

C,D represent the superimposed bimodality.

- Relative size and sign of A,B or C,D control the direction.
- Absolute size $(A^2+B^2)^{1/2}$ controls the sharpness.
- For MR, we get A,B (or Φ, FOM) i.e. C=D=0.
- Together A,B,C,D can describe a bimodal distribution with any combination of peak height and direction.

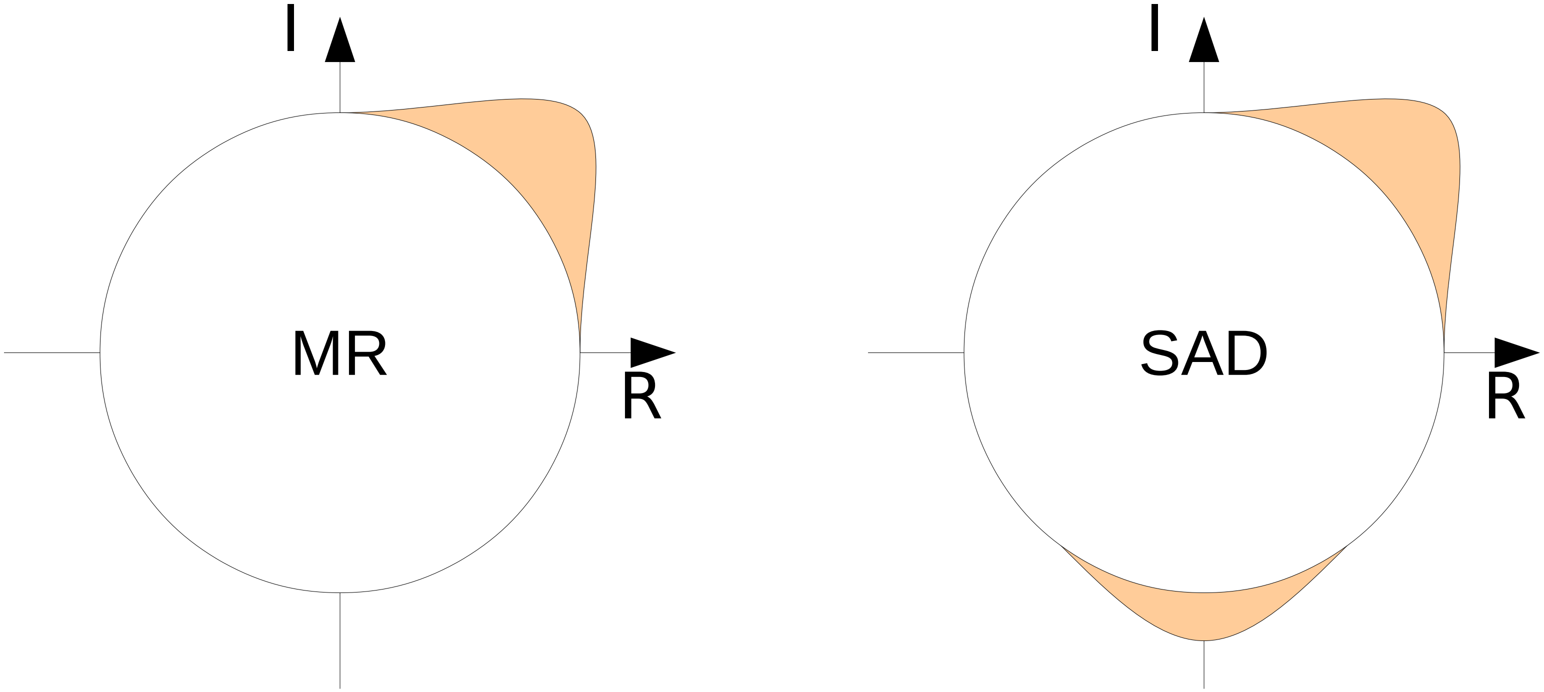




Density modification

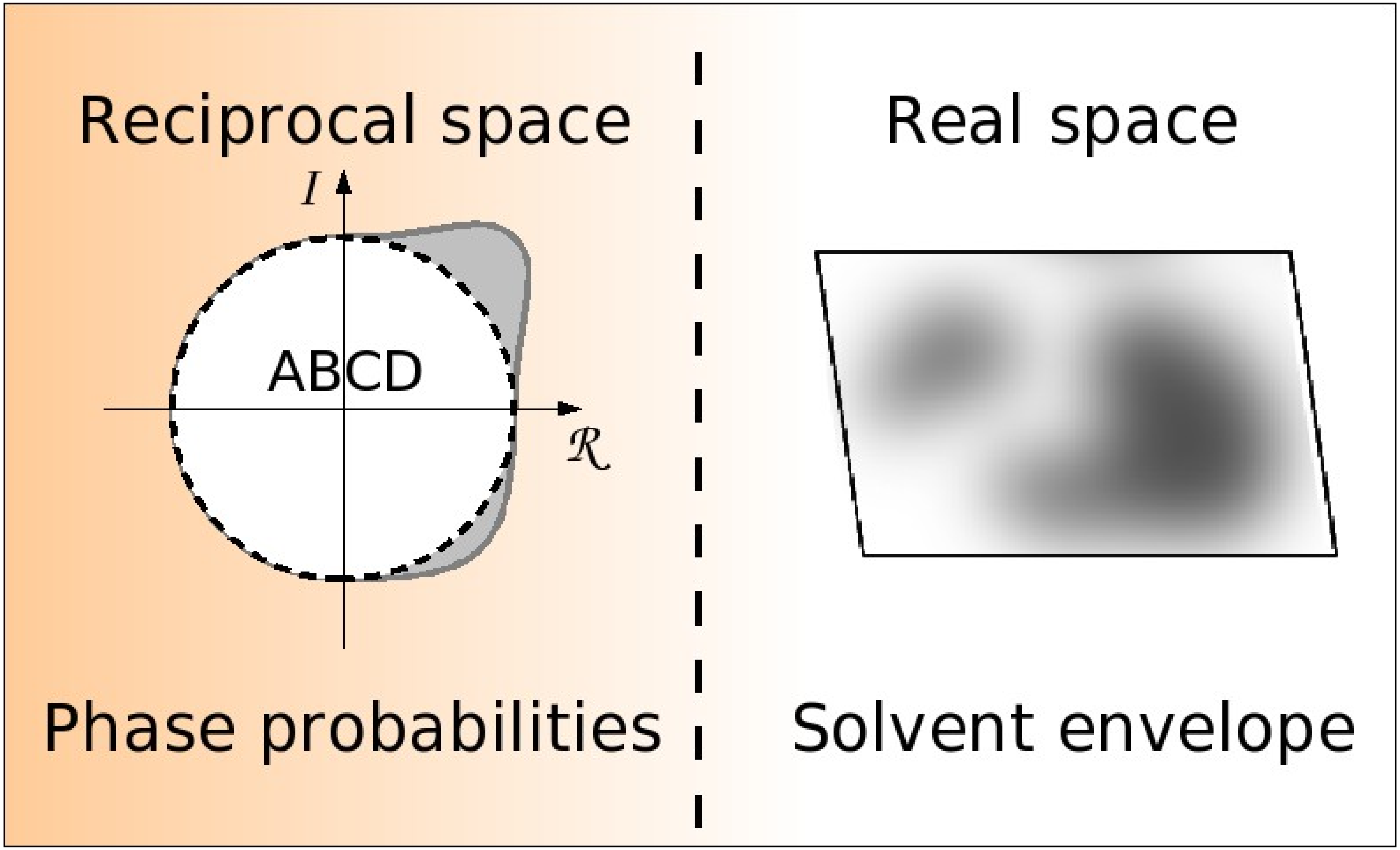
How to calculate a map?

- Need to reduce probability distributions to single phases
- Want to produce the least noisy map
- Use the center of mass



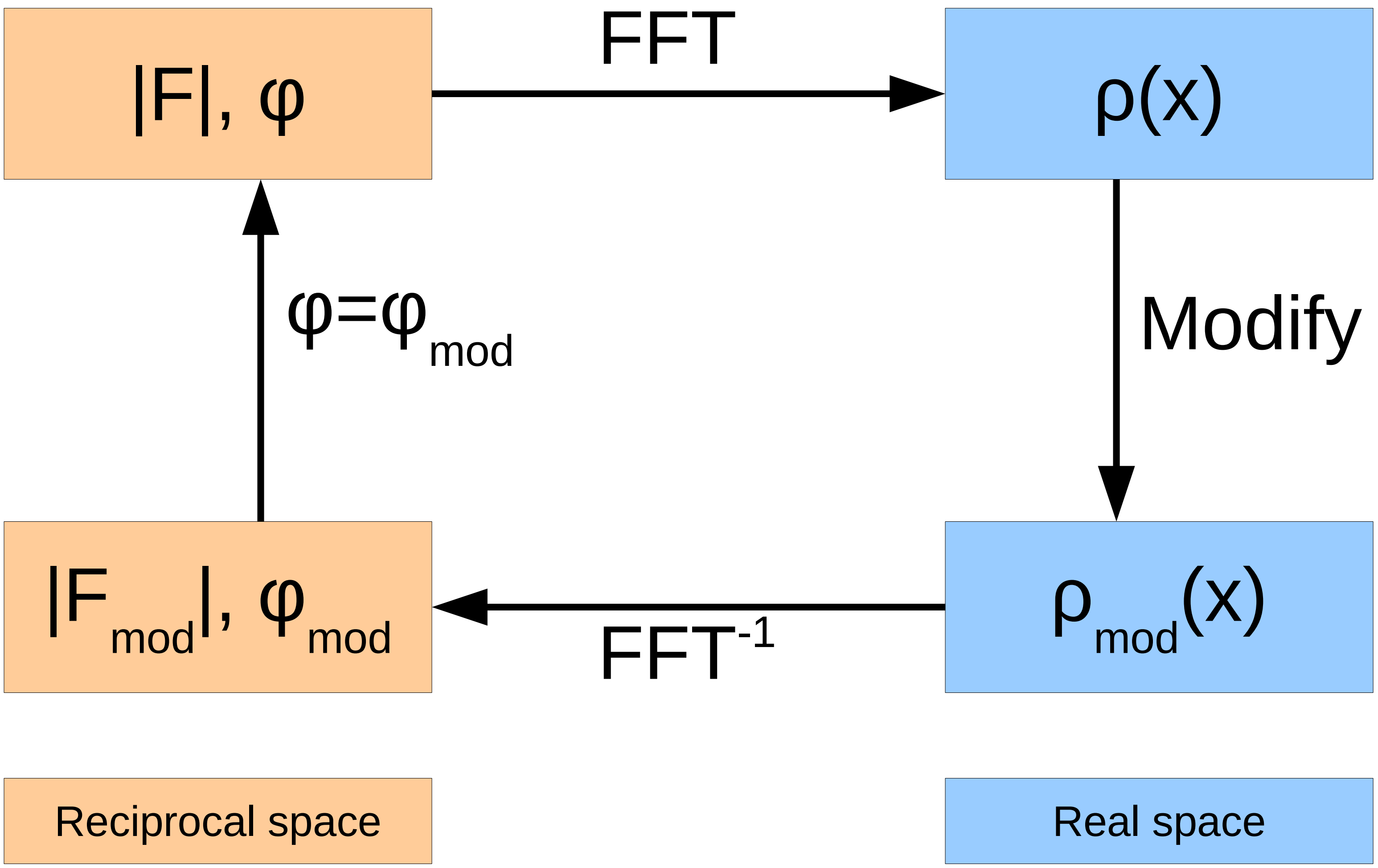
Density modification

- Density modification is a problem in combining information:



Density modification

1. Rudimentary calculation:



$|F|, \varphi$

FFT

$\rho(x)$

Modify ρ

$\rho_{mod}(x)$

FFT^{-1}

$|F_{mod}|, \varphi_{mod}$

$\varphi = \varphi_{mod}$

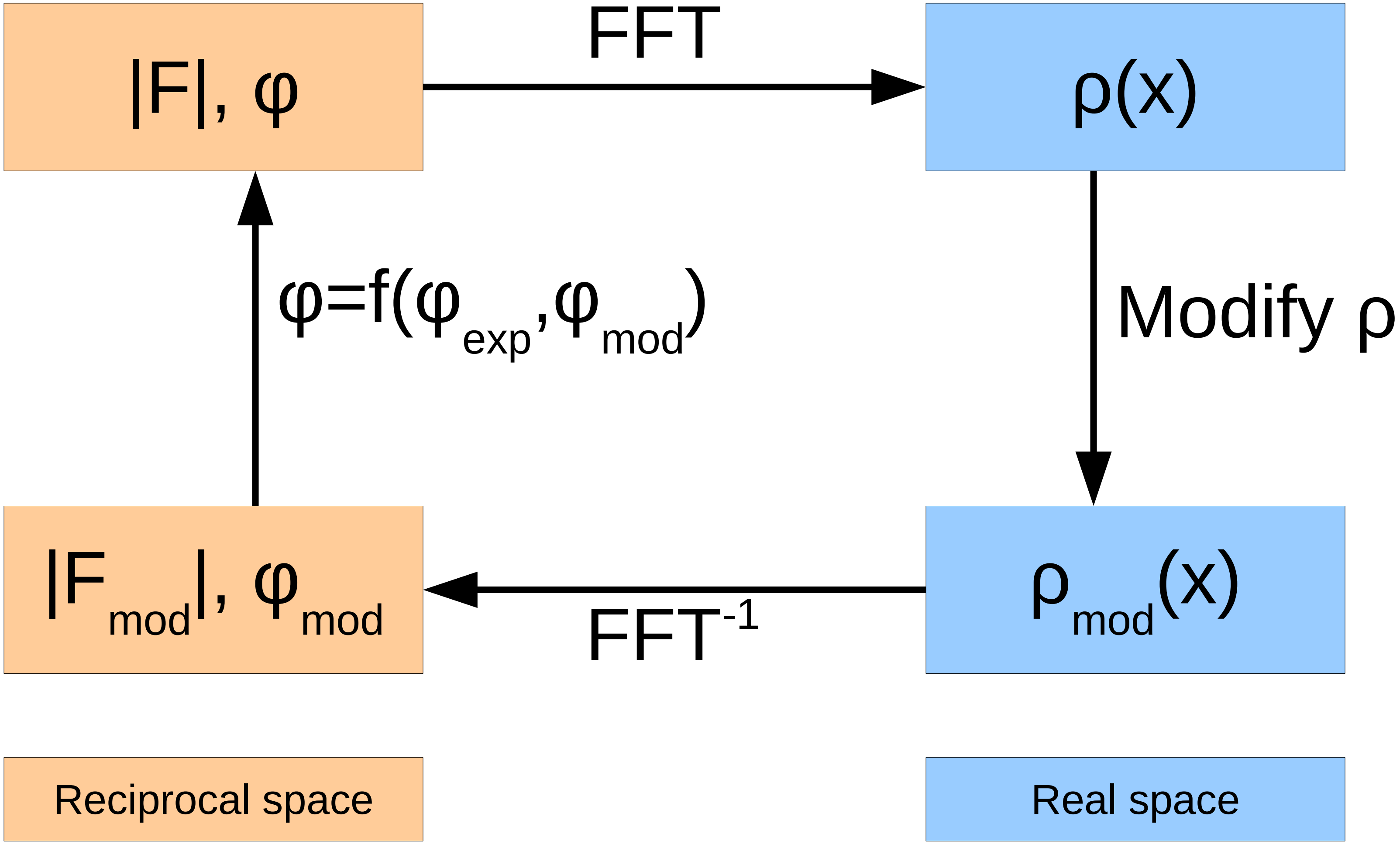
Reciprocal space

Real space

Paul Bond, Kevin Cowtan, kevin.cowtan@york.ac.uk

Density modification

2. Phase weighting:



$|F|, \varphi$

FFT

$\rho(x)$

Modify ρ

$\rho_{mod}(x)$

FFT^{-1}

$|F_{mod}|, \varphi_{mod}$

$\varphi=f(\varphi_{exp},\varphi_{mod})$

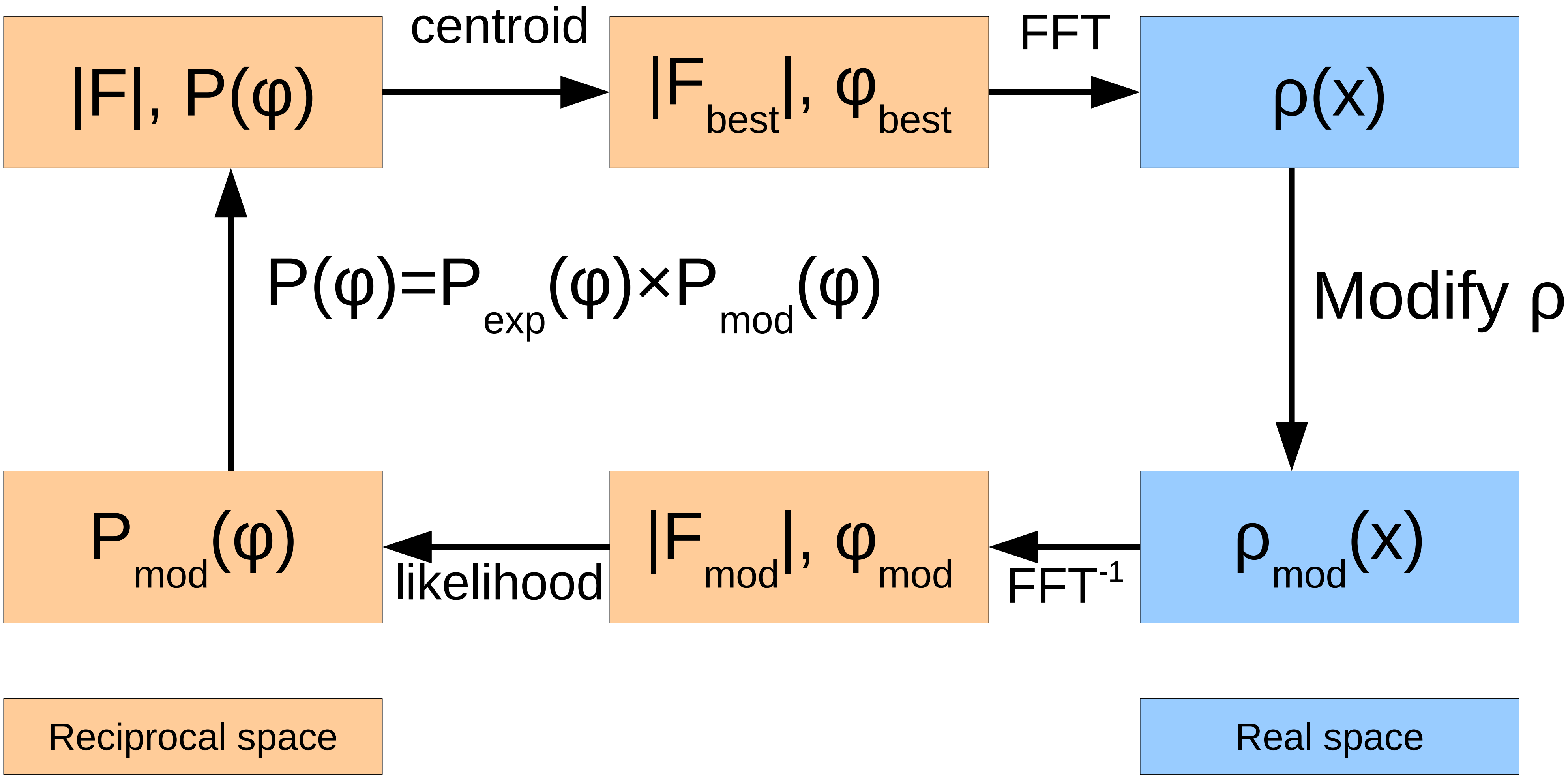
Reciprocal space

Real space

Paul Bond, Kevin Cowtan, kevin.cowtan@york.ac.uk

Density modification

3. Phase probability distributions:



$|F|, P(\varphi)$

centroid

$|F_{best}|, \varphi_{best}$

FFT

$\rho(x)$

Modify ρ

$\rho_{mod}(x)$

FFT^{-1}

$|F_{mod}|, \varphi_{mod}$

likelihood

$P_{mod}(\varphi)$

$P(\varphi)=P_{exp}(\varphi)\times P_{mod}(\varphi)$

Reciprocal space

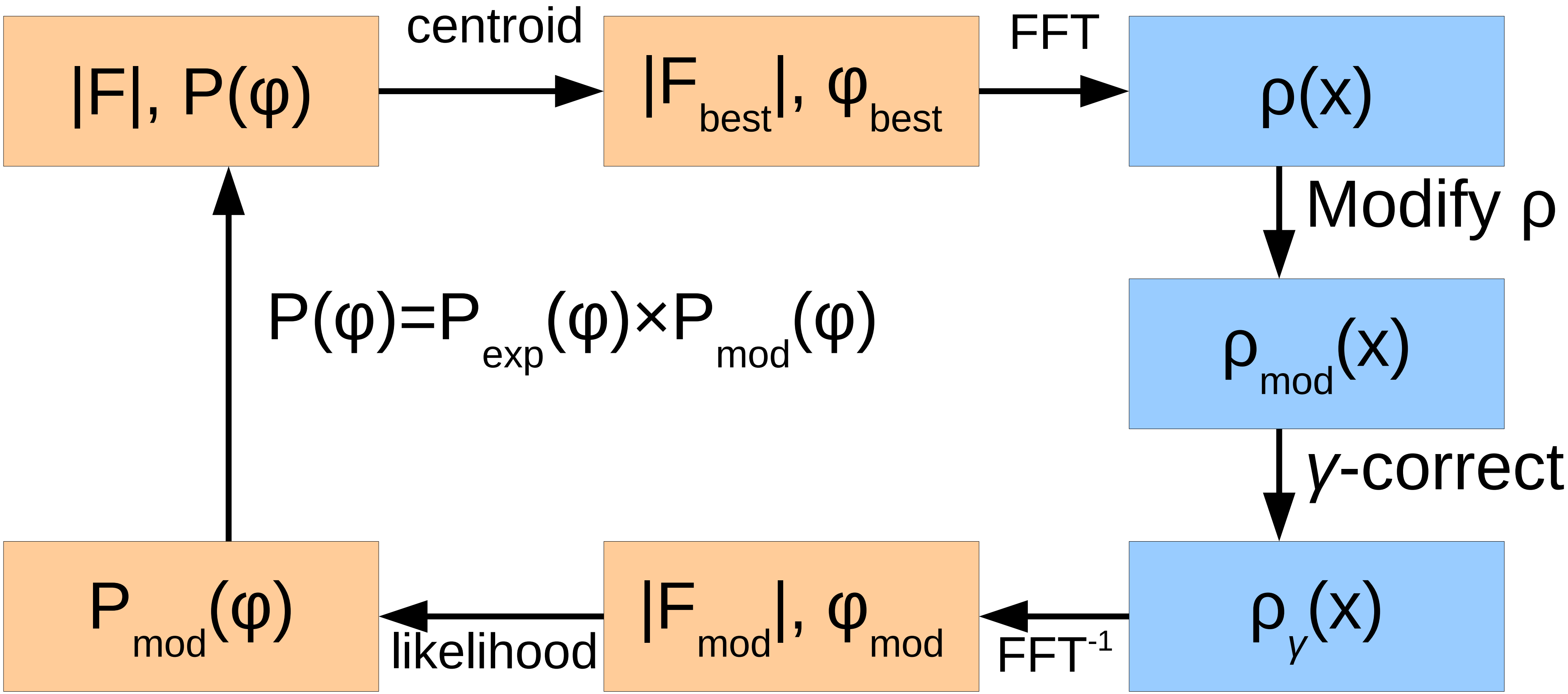
Real space

Paul Bond, Kevin Cowtan, kevin.cowtan@york.ac.uk

Density modification

DM, SOLOMON, (CNS)

4. Bias reduction (gamma-correction):



$|F|, P(\varphi)$

centroid

$|F_{best}|, \varphi_{best}$

FFT

$\rho(x)$

Modify ρ

$\rho_{mod}(x)$

γ-correct

$\rho_{\gamma}(x)$

FFT^{-1}

$|F_{mod}|, \varphi_{mod}$

likelihood

$P_{mod}(\varphi)$

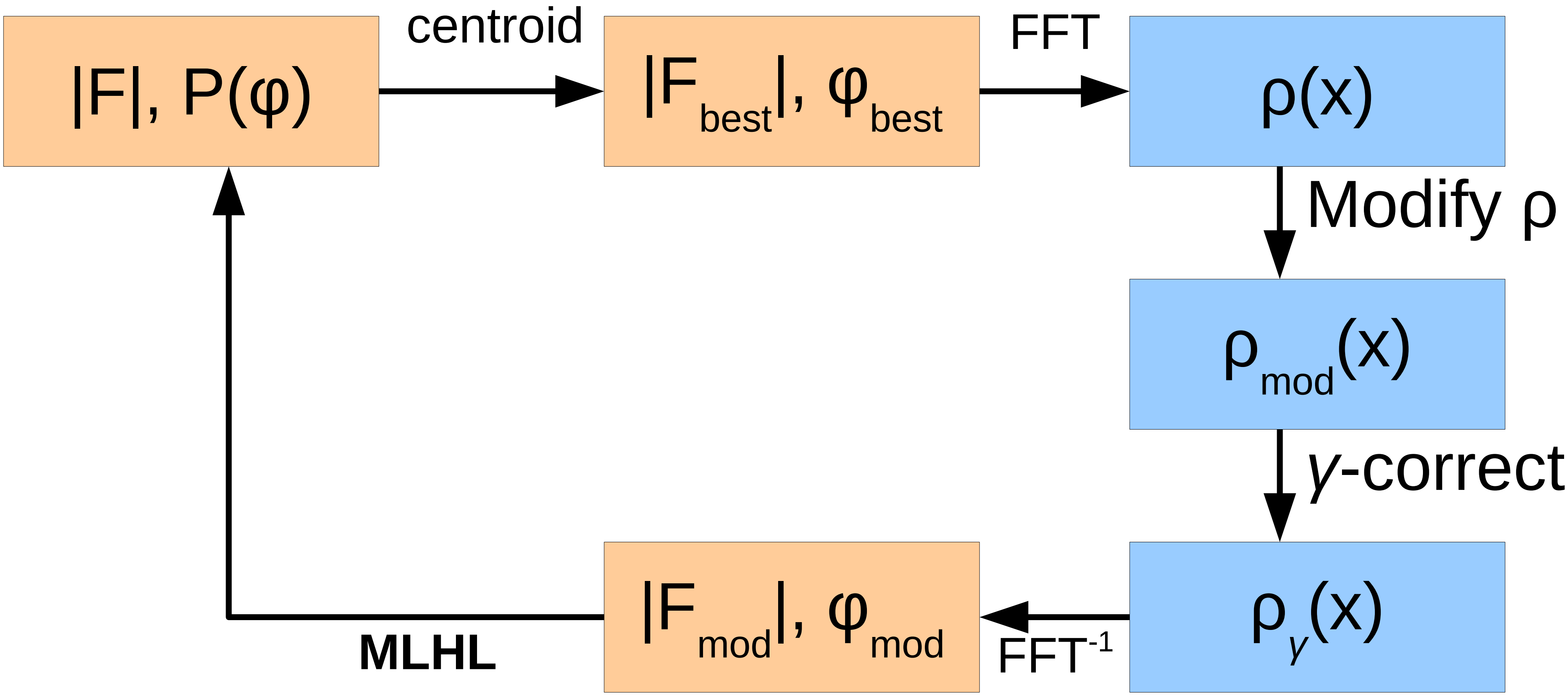
$P(\varphi)=P_{exp}(\varphi)\times P_{mod}(\varphi)$

J.P.Abrahams

Density modification

PARROT

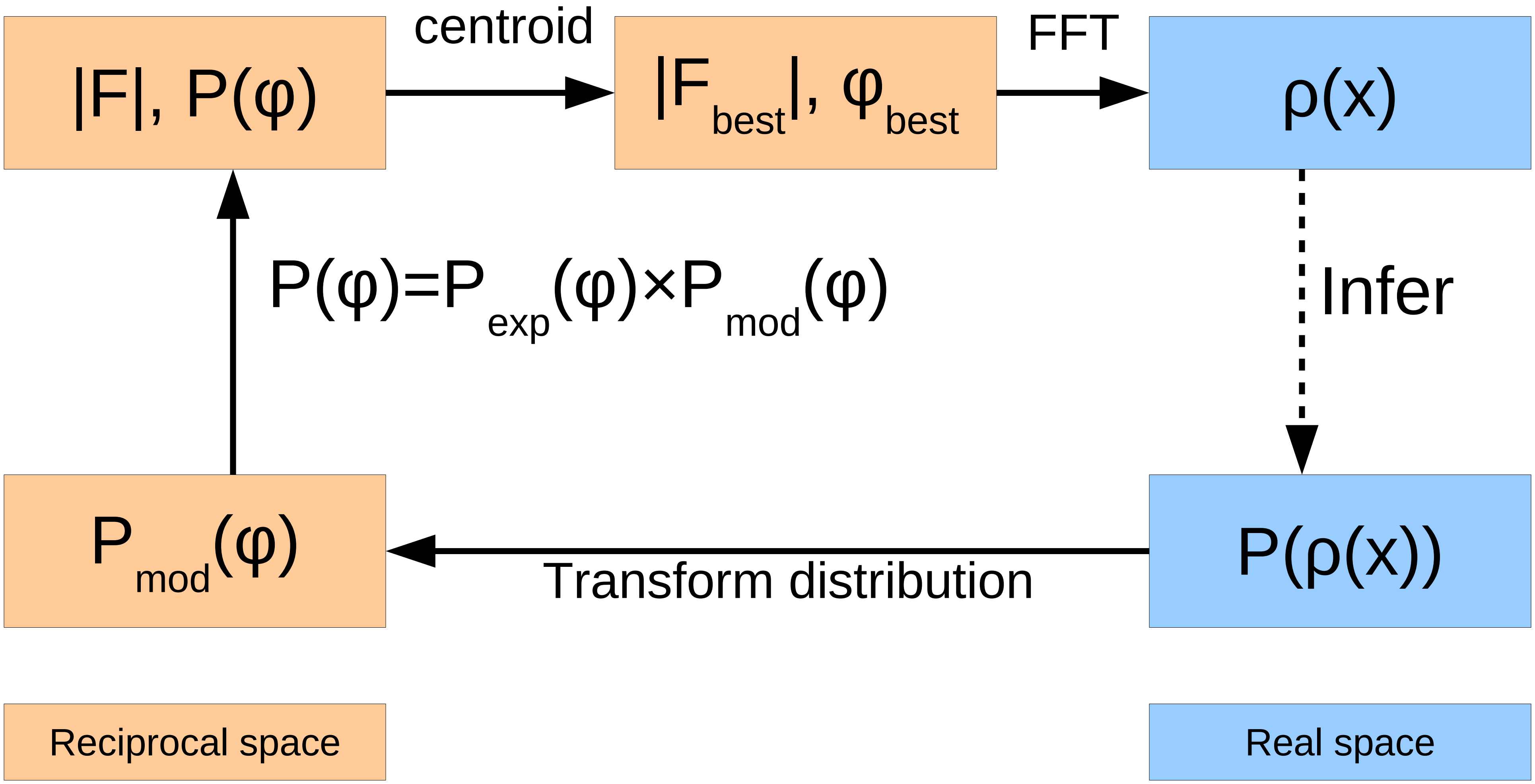
5. Maximum Likelihood H-L:



Density modification

RESOLVE

6. Statistical density modification:



$|F|, P(\varphi)$

centroid

$|F_{best}|, \varphi_{best}$

FFT

$\rho(x)$

Infer

$P(\rho(x))$

Transform distribution

$P_{mod}(\varphi)$

$P(\varphi)=P_{exp}(\varphi)\times P_{mod}(\varphi)$

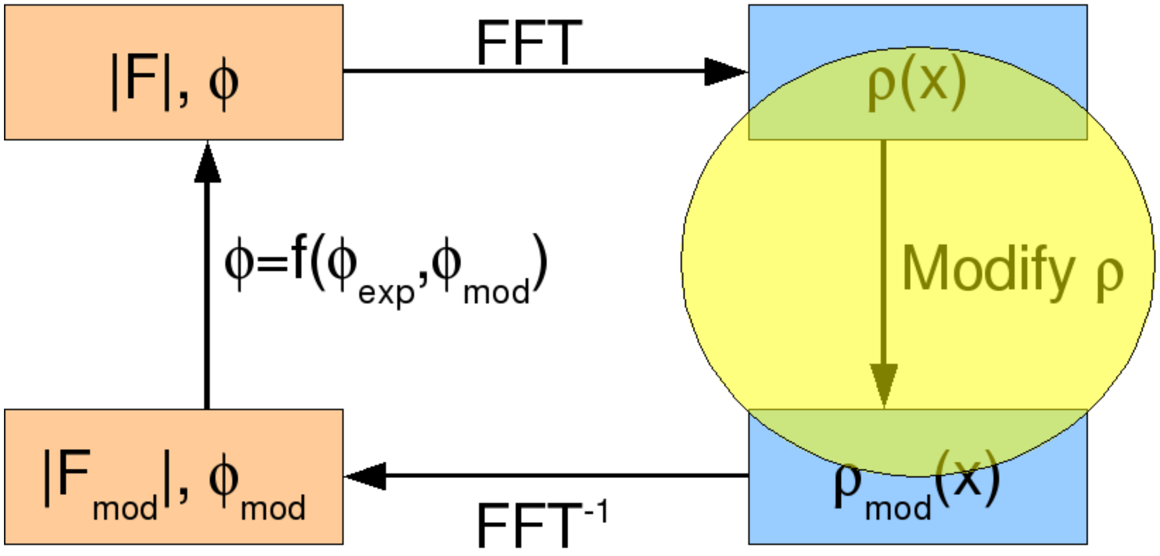
Reciprocal space

Real space

Density modification

Classical density modification techniques:

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



$|F|, \phi$

FFT

$\rho(x)$

Modify ρ

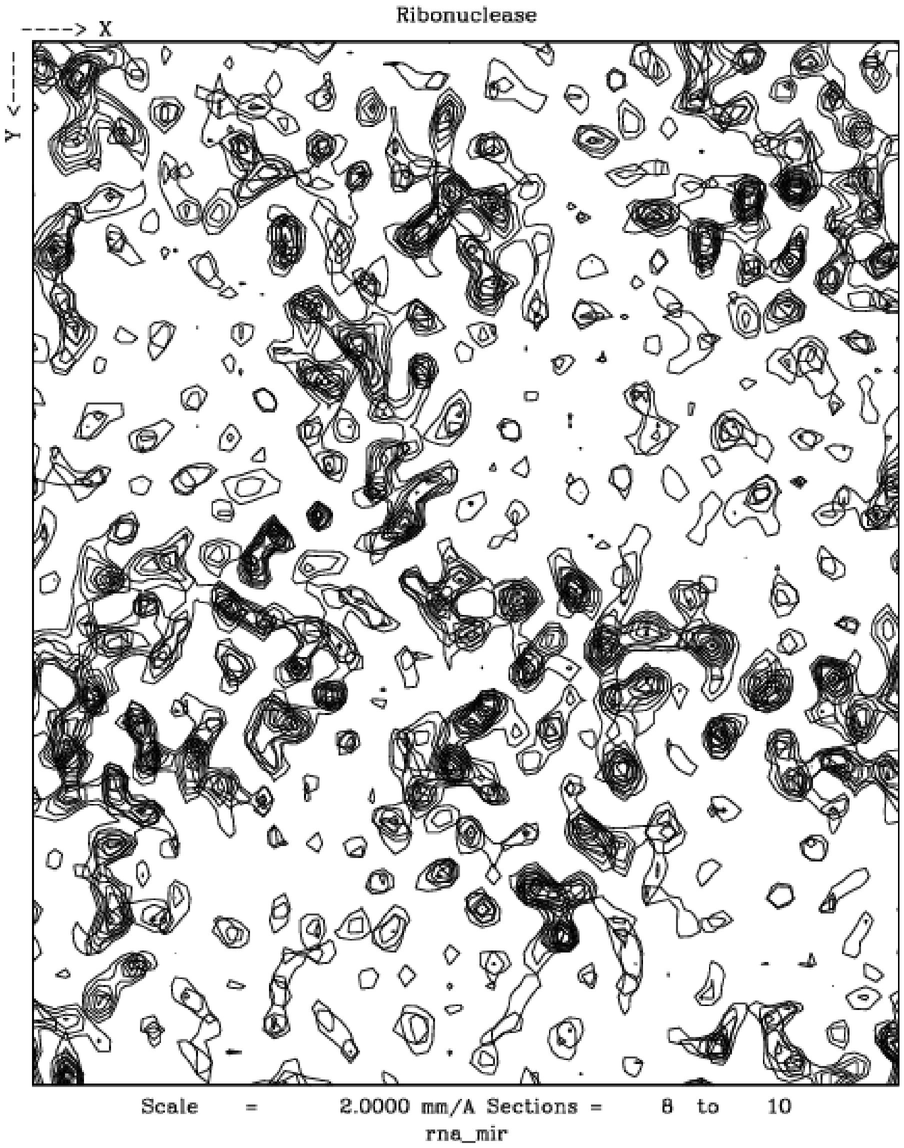
$\rho_{mod}(x)$

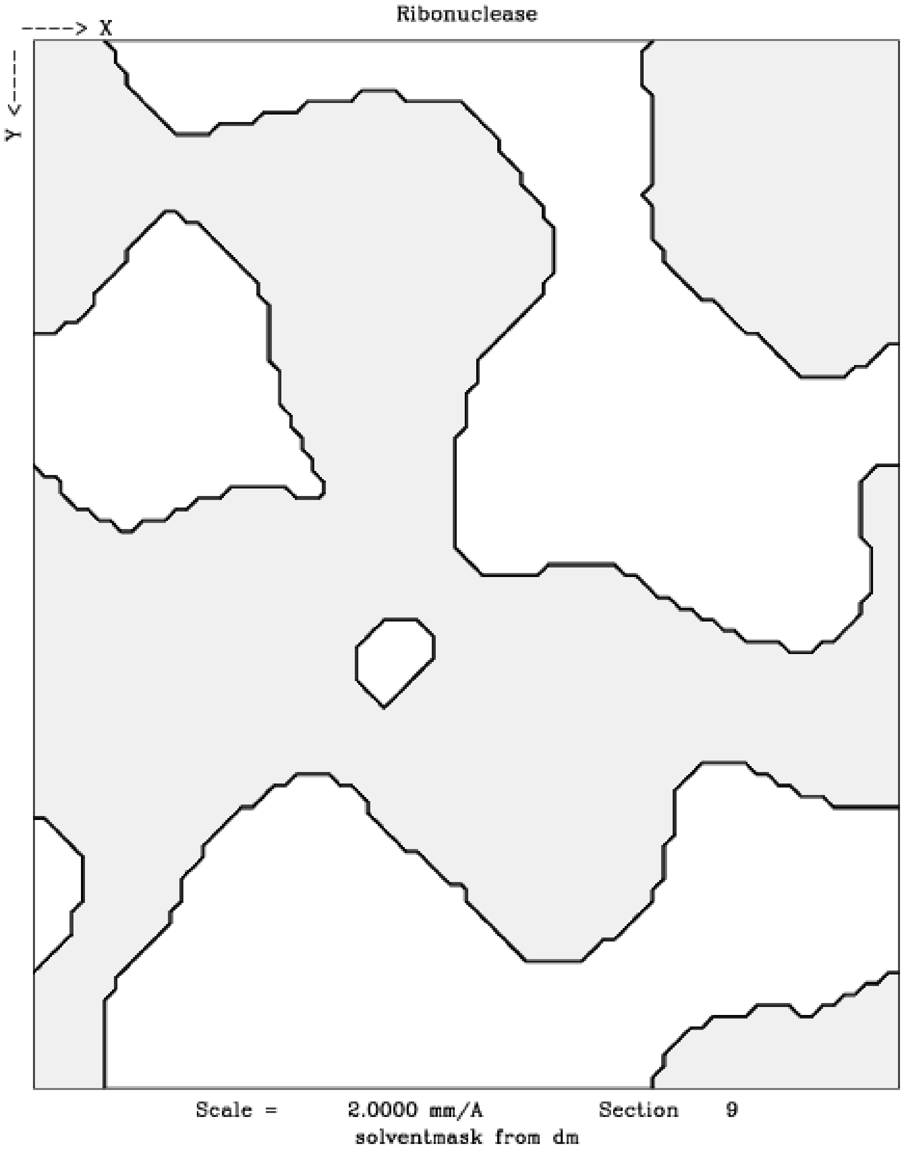
FFT^{-1}

$|F_{mod}|, \phi_{mod}$

$\phi=f(\phi_{exp},\phi_{mod})$

Solvent flattening





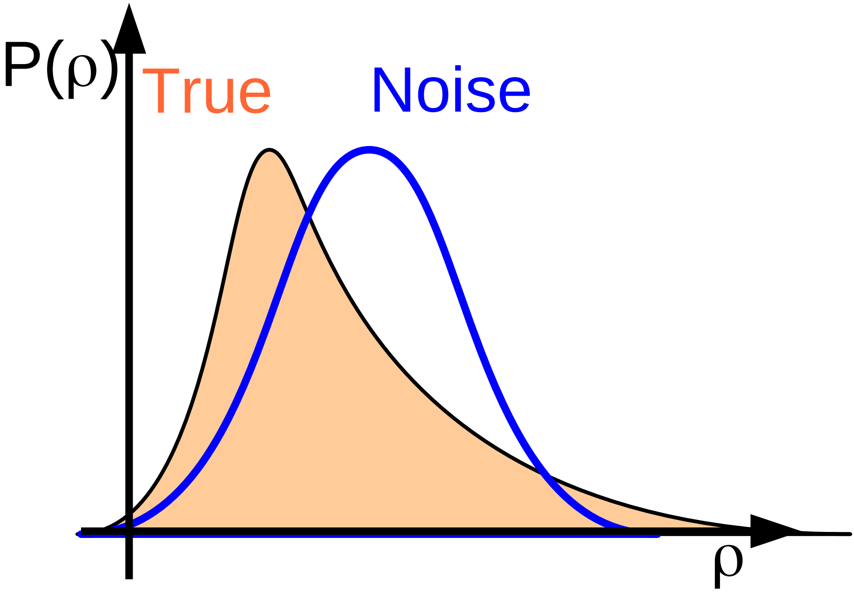
Histogram matching

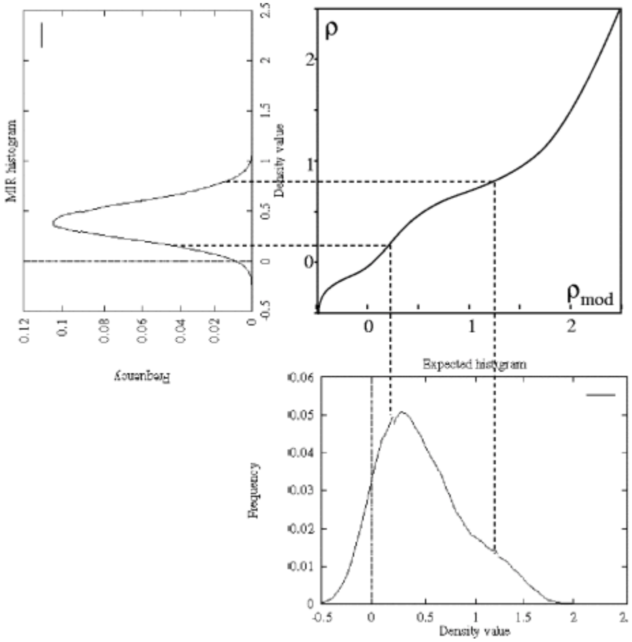
A technique from image processing for modifying the protein region.

- 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.

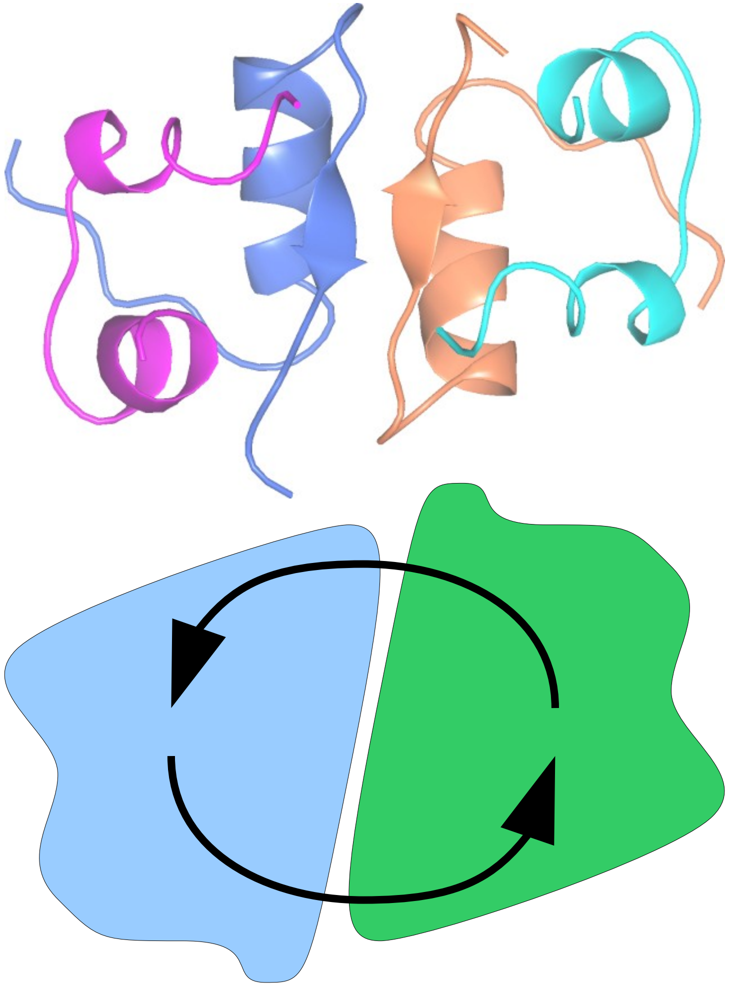
Useful at better than 4A.





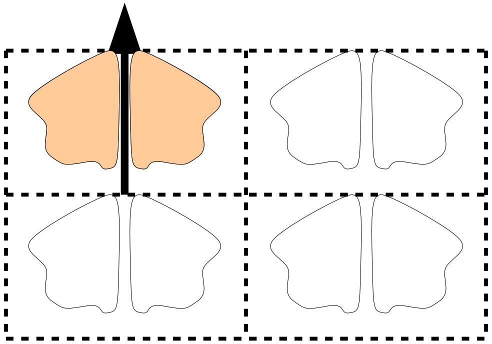
Non-crystallographic symmetry

- 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!

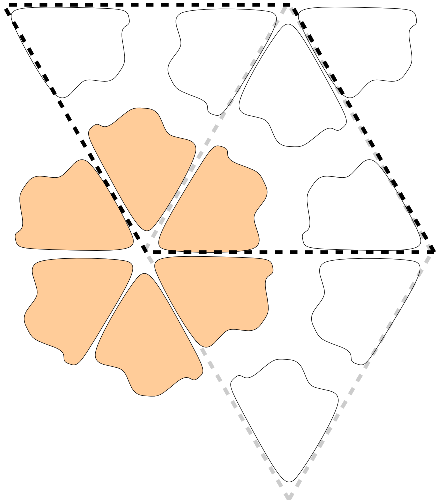


Non-crystallographic symmetry

Crystallographic

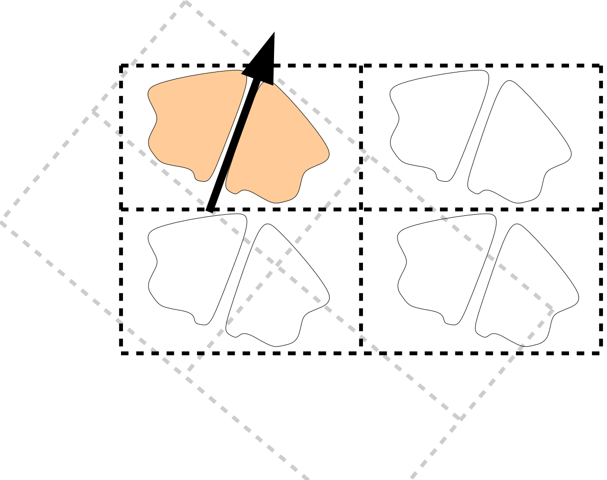


Aligned
2-fold

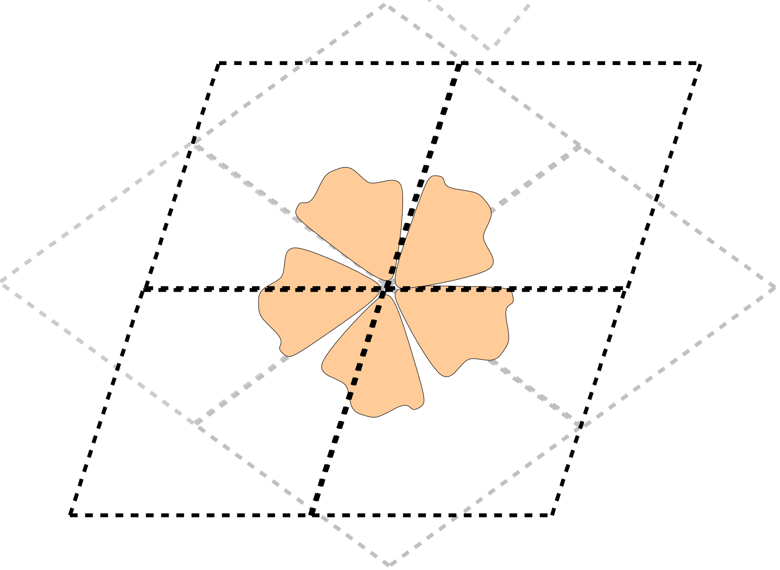


Aligned
6-fold

Non-crystallograpic



Unaligned
2-fold

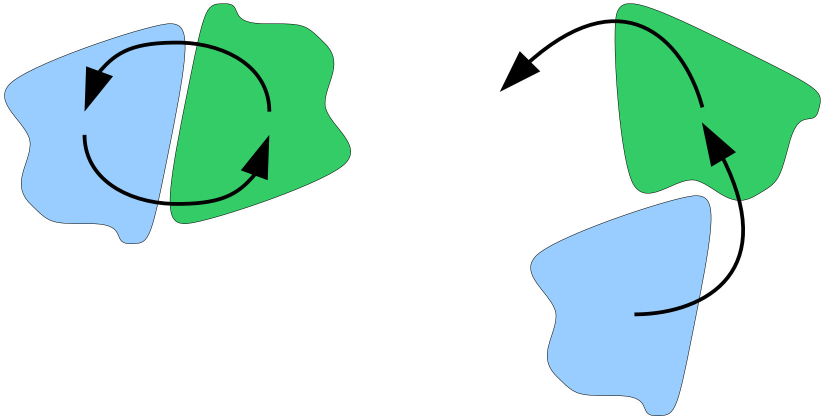


Aligned
5-fold

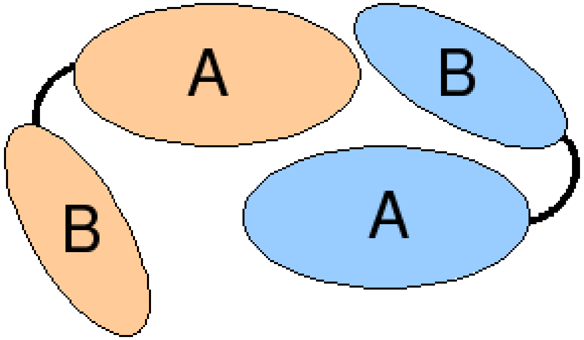
Non-crystallographic symmetry

Useful terms:

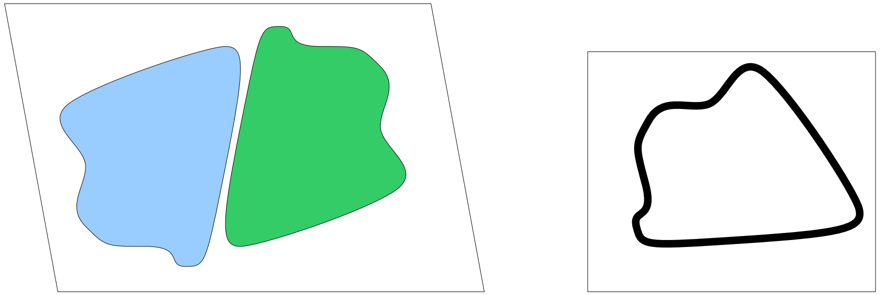
- Proper and improper NCS: (closed and open)



- Multi-domain averaging:

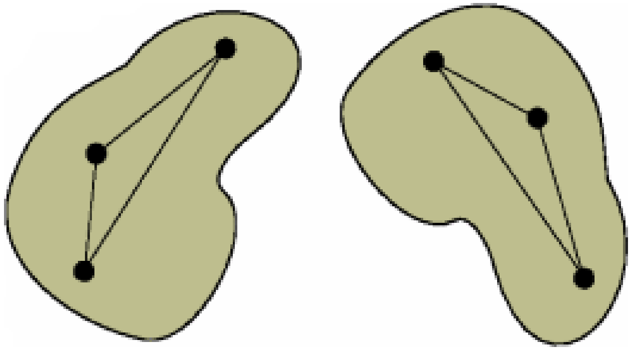


- Multi-crystal averaging:

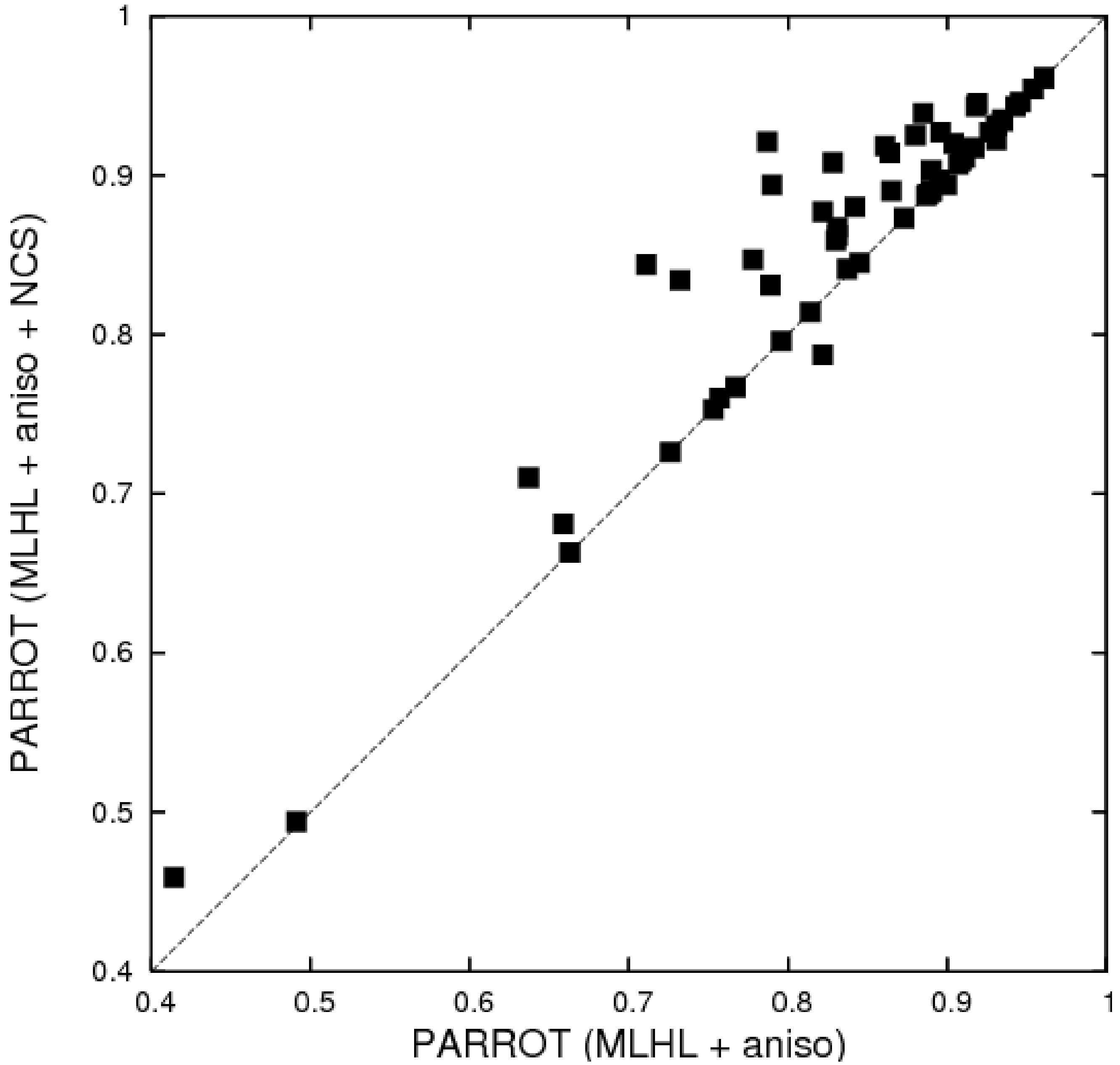


Non-crystallographic symmetry

- 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.



Parrot: simple vs NCS averaged



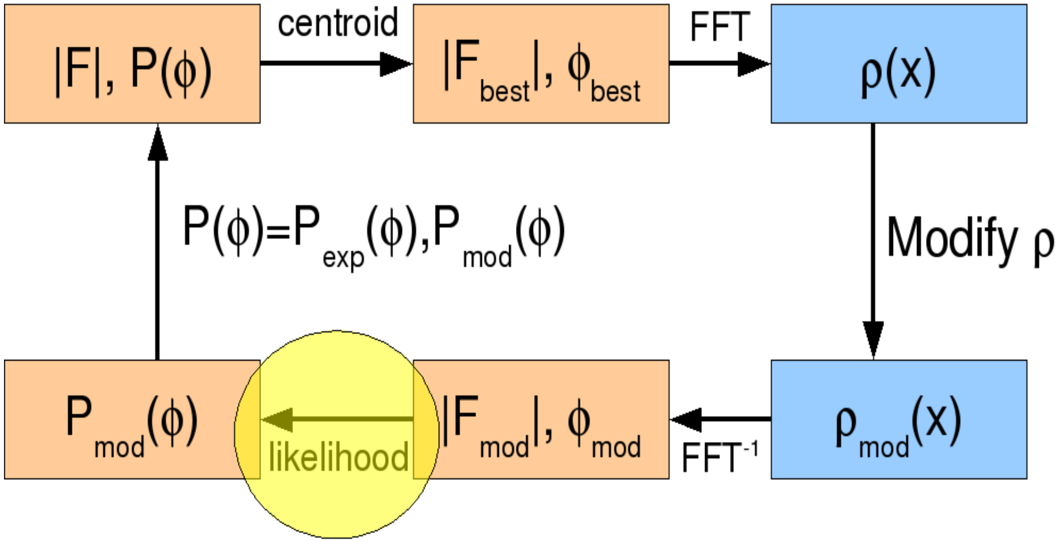
Map correlations

Comparing with and without NCS averaging.

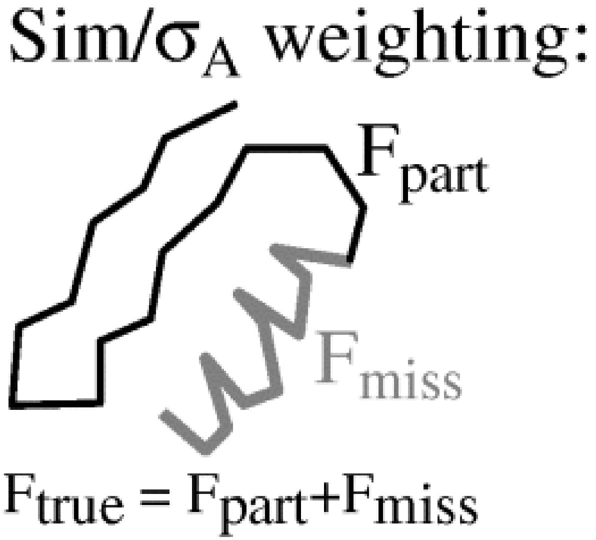
Estimating phase probabilities

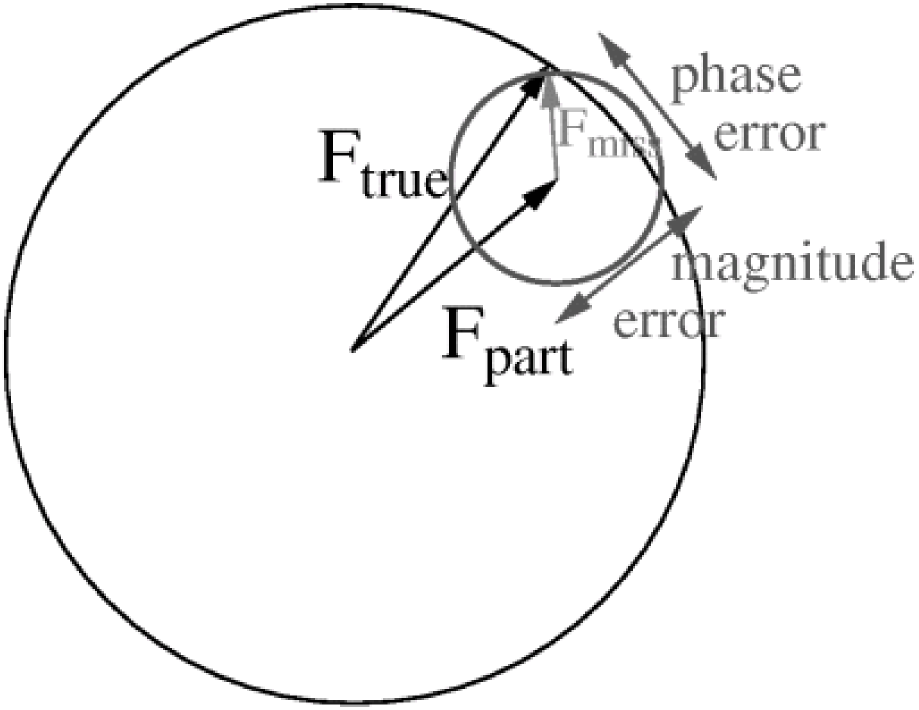
Problem: How do we go from a single phase estimate to a full phase probability distribution?

- 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.



Estimating phase probabilities

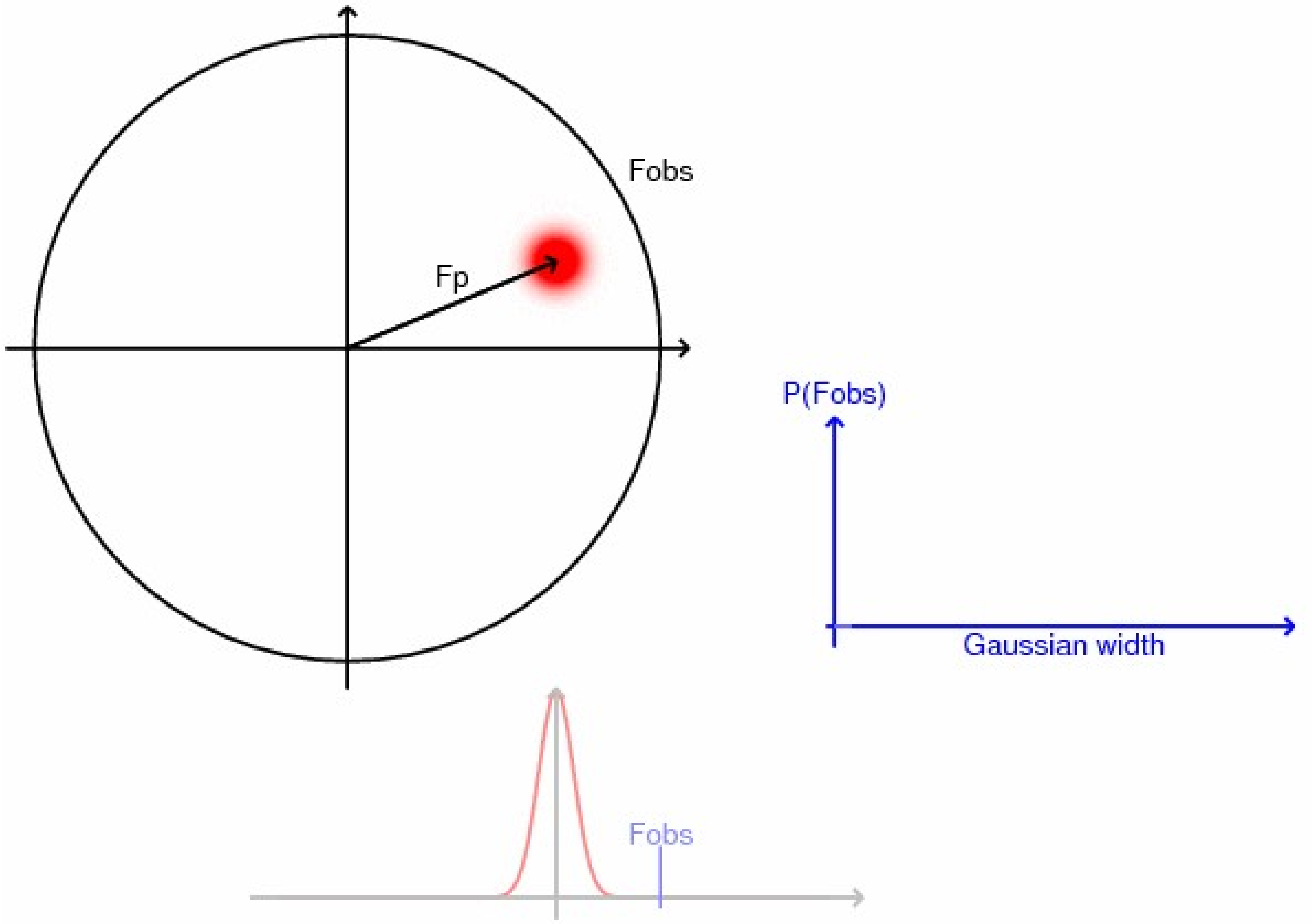




We know $|F_{true}|$, $|F_{part}|$, ϕ_{part}

Assuming ϕ_{part}, ϕ_{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.

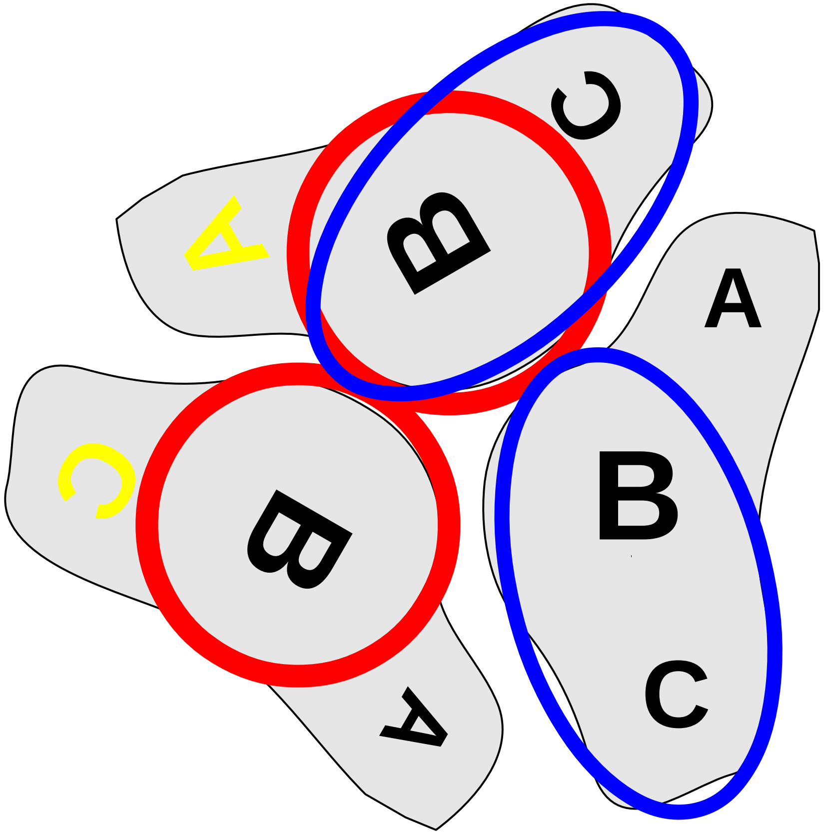
Estimating phase probabilities



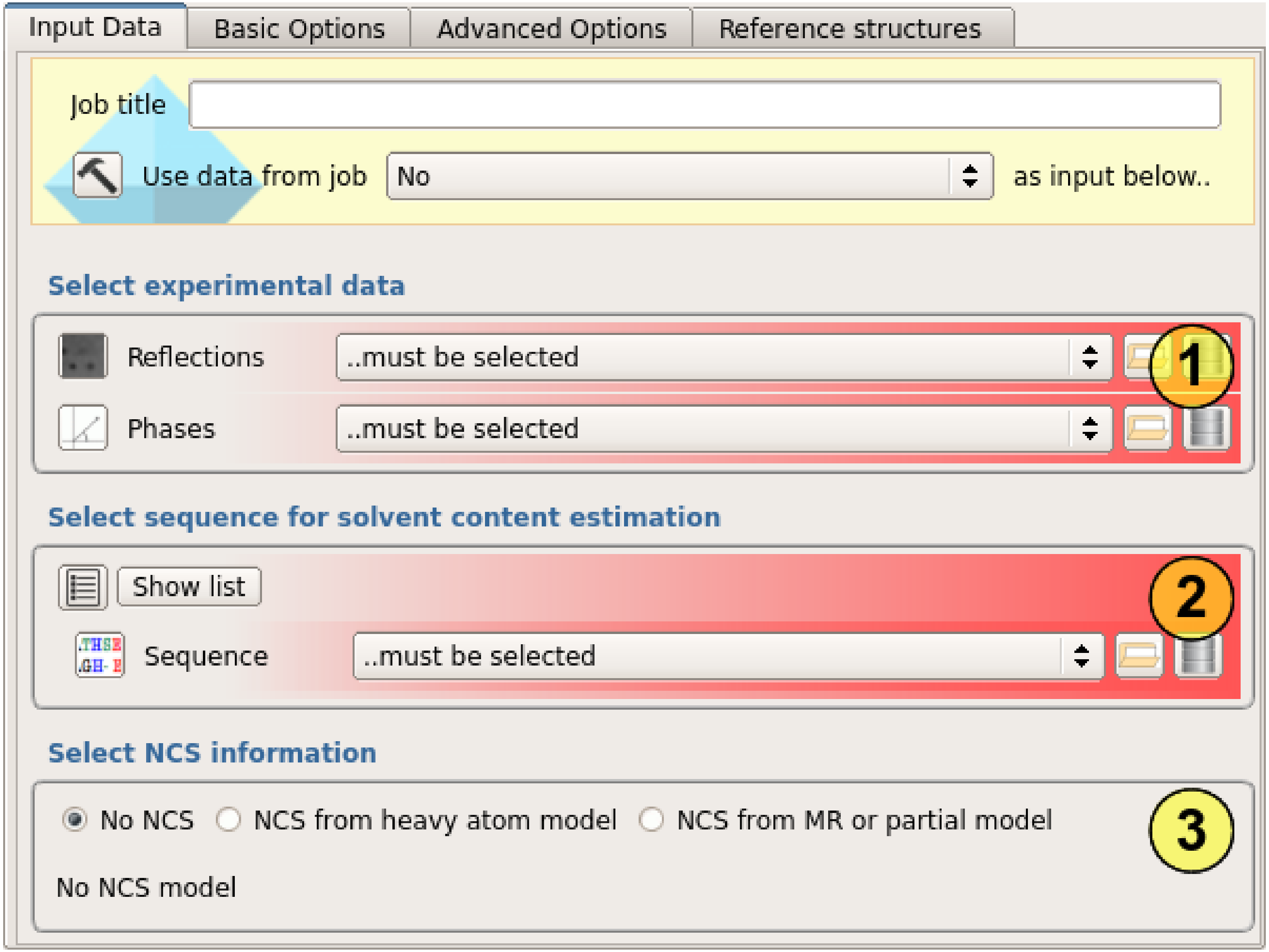
Recent Developments:

Pairwise-weighted NCS averaging:

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



Parrot



Job title

Use data from job No as input below..

Select experimental data

Reflections ..must be selected

Phases ..must be selected

Select sequence for solvent content estimation

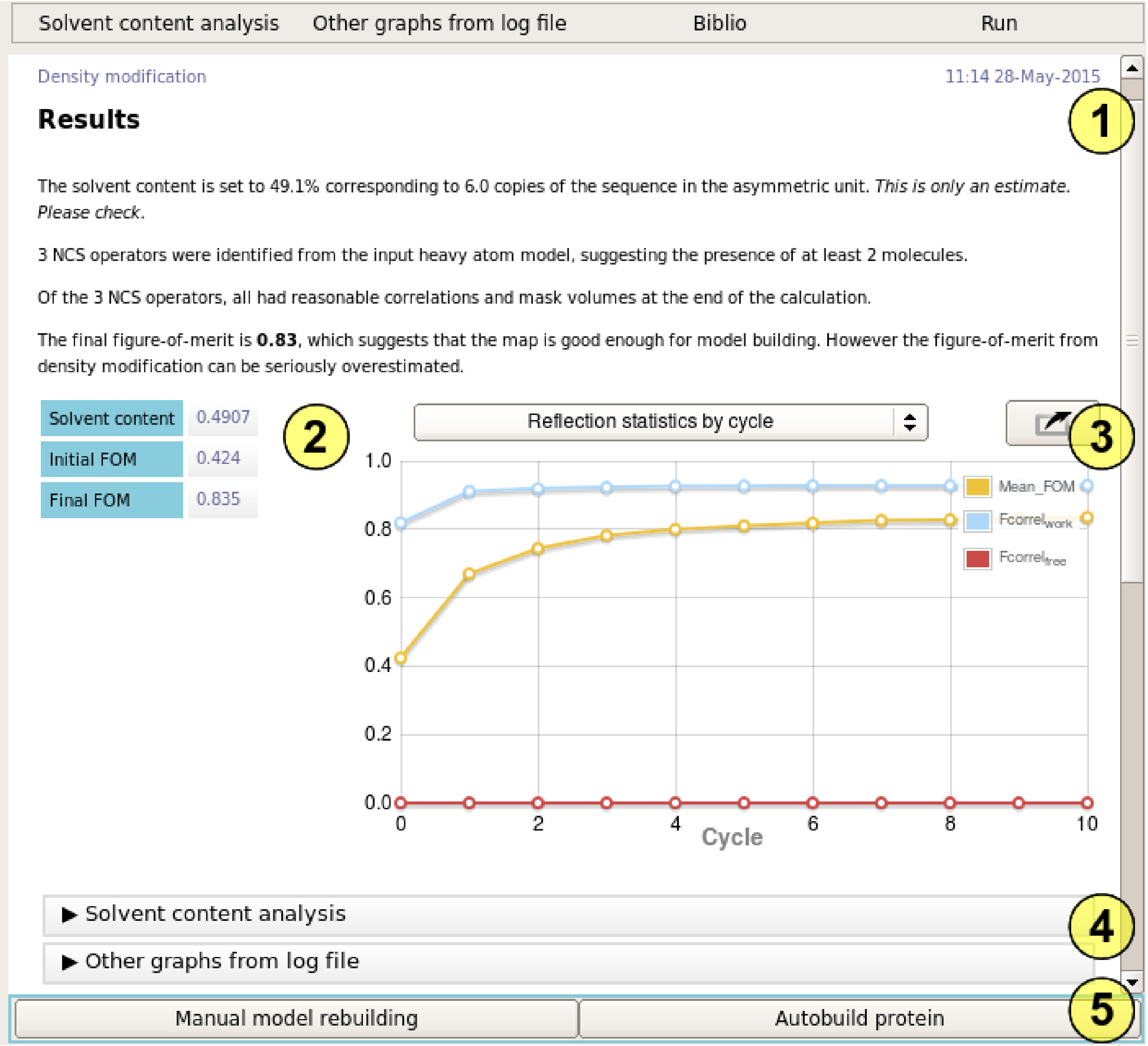
Sequence ..must be selected

Select NCS information

● No NCS ○ NCS from heavy atom model ○ NCS from MR or partial model

No NCS model

Parrot



Density modification

11:14 28-May-2015

Results

The solvent content is set to 49.1% corresponding to 6.0 copies of the sequence in the asymmetric unit. *This is only an estimate. Please check.*

3 NCS operators were identified from the input heavy atom model, suggesting the presence of at least 2 molecules.

Of the 3 NCS operators, all had reasonable correlations and mask volumes at the end of the calculation.

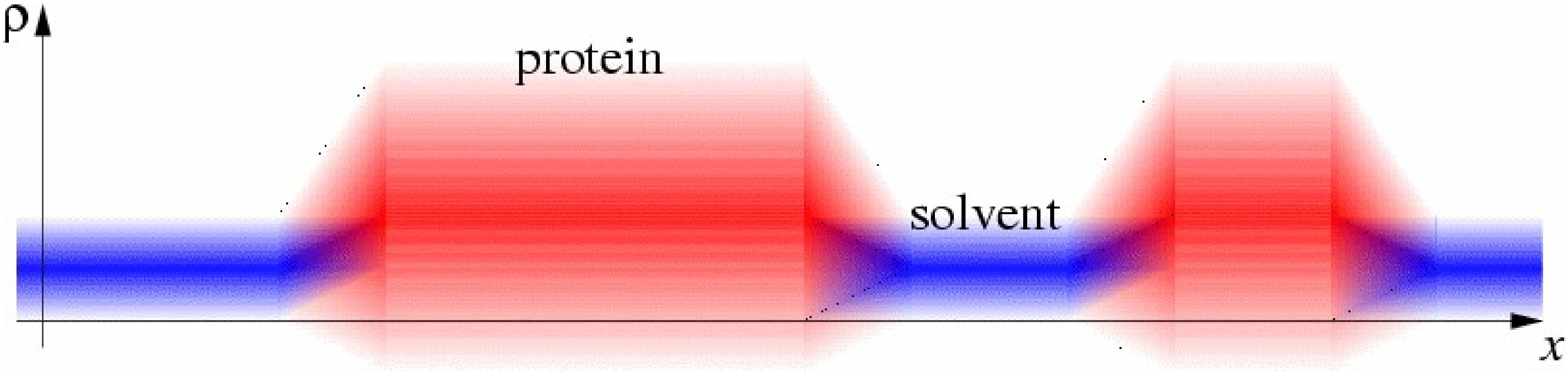
The final figure-of-merit is **0.83**, which suggests that the map is good enough for model building. However the figure-of-merit from density modification can be seriously overestimated.

Solvent content	0.4907
Initial FOM	0.424
Final FOM	0.835

Paul Bond, Kevin Cowtan, kevin.cowtan@york.ac.uk

Statistical density modification

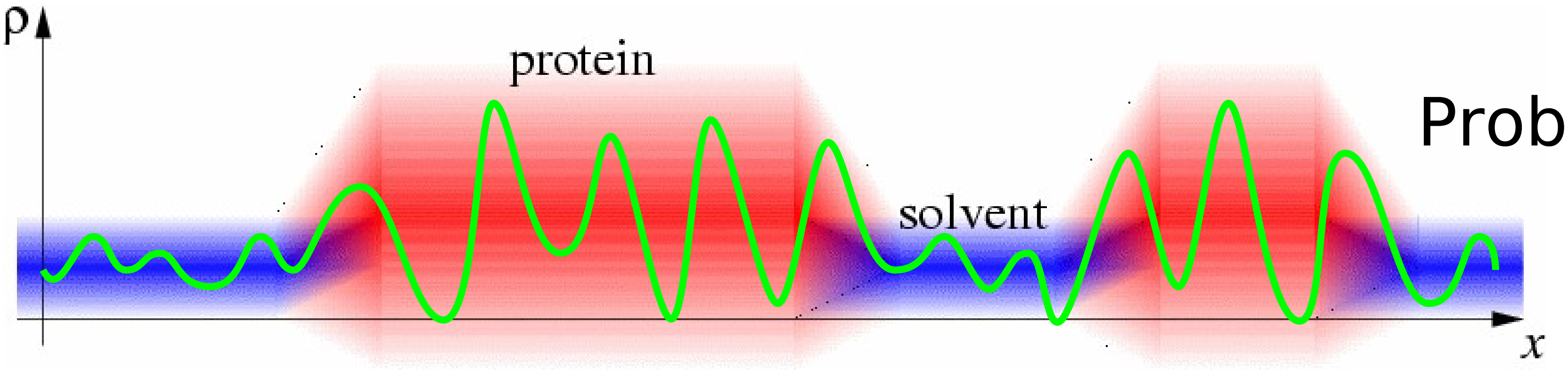
- Form a statistical description of expected map features.



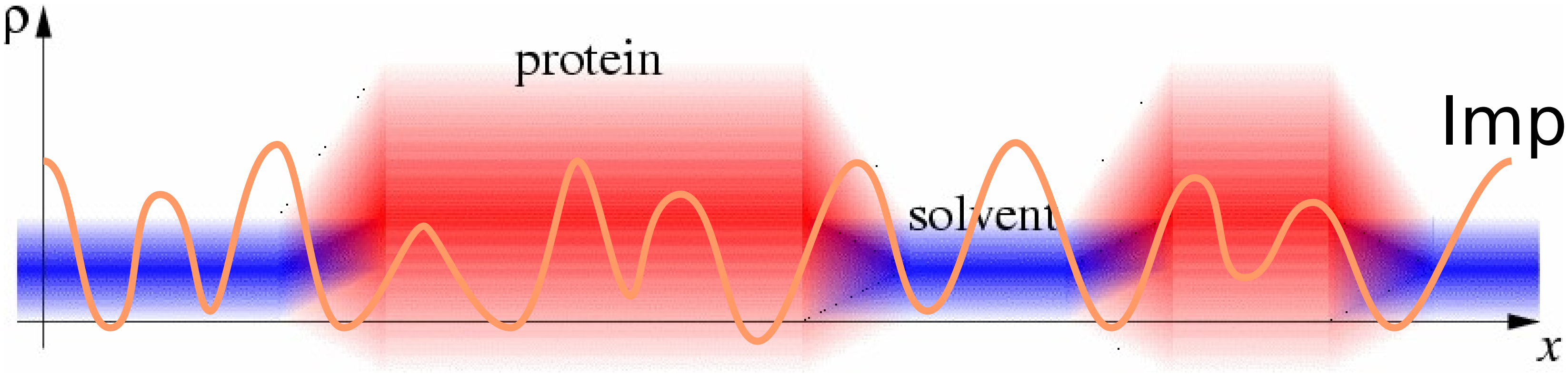
- e.g.
  - Protein has higher mean, and is more peaky (higher variance)
  - Solvent has lower mean, and is flatter (lower variance)

Statistical density modification

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



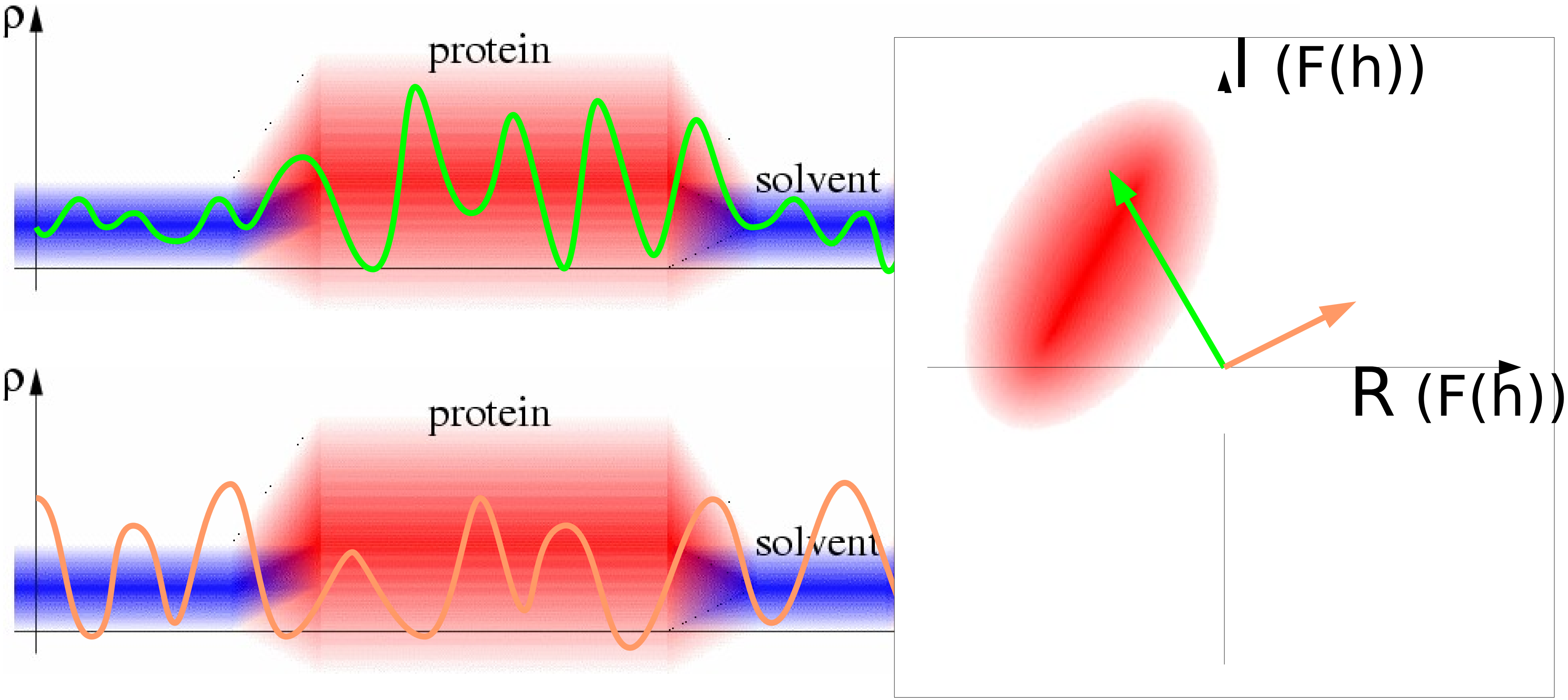
Probable



Improbable

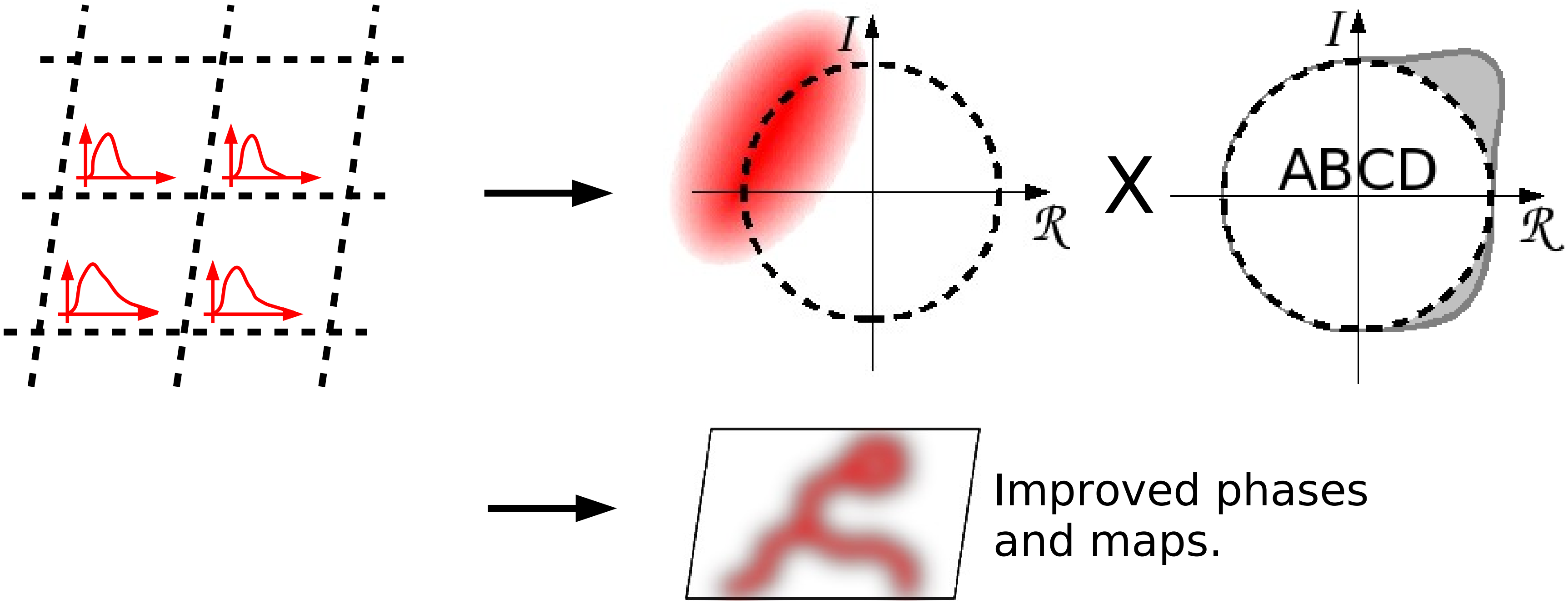
Statistical density modification

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



Statistical density modification

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



Improved phases
and maps.

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

Statistical density modification

Advantages:

- Reduced bias.
- Better phases.

Disadvantages:

- Slow.
- Latest classical methods comparable.

Acknowledgments

Help:

- JCSG data archive: www.jcsg.org
- Garib Murshudov, Raj Pannu, Pavol Skubak
- Eleanor Dodson, Paul Emsley, Randy Read, Clemens Vonrhein

Funding:

- The Royal Society, BBSRC