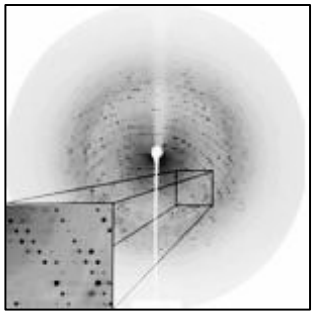
# Density modification

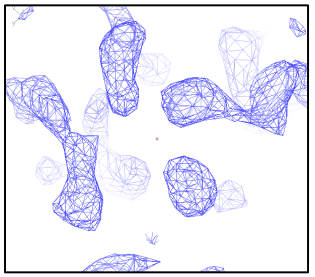
DLS/CCP4 workshop 2014

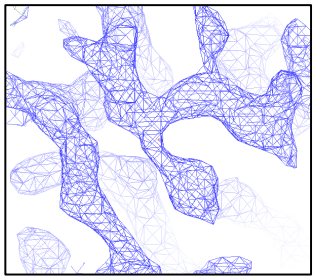
Kevin Cowtan, cowtan@ysbl.york.ac.uk

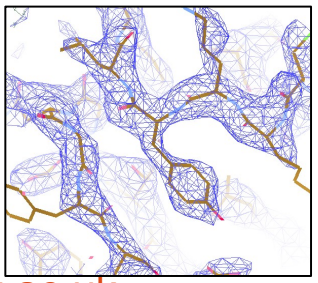
2012

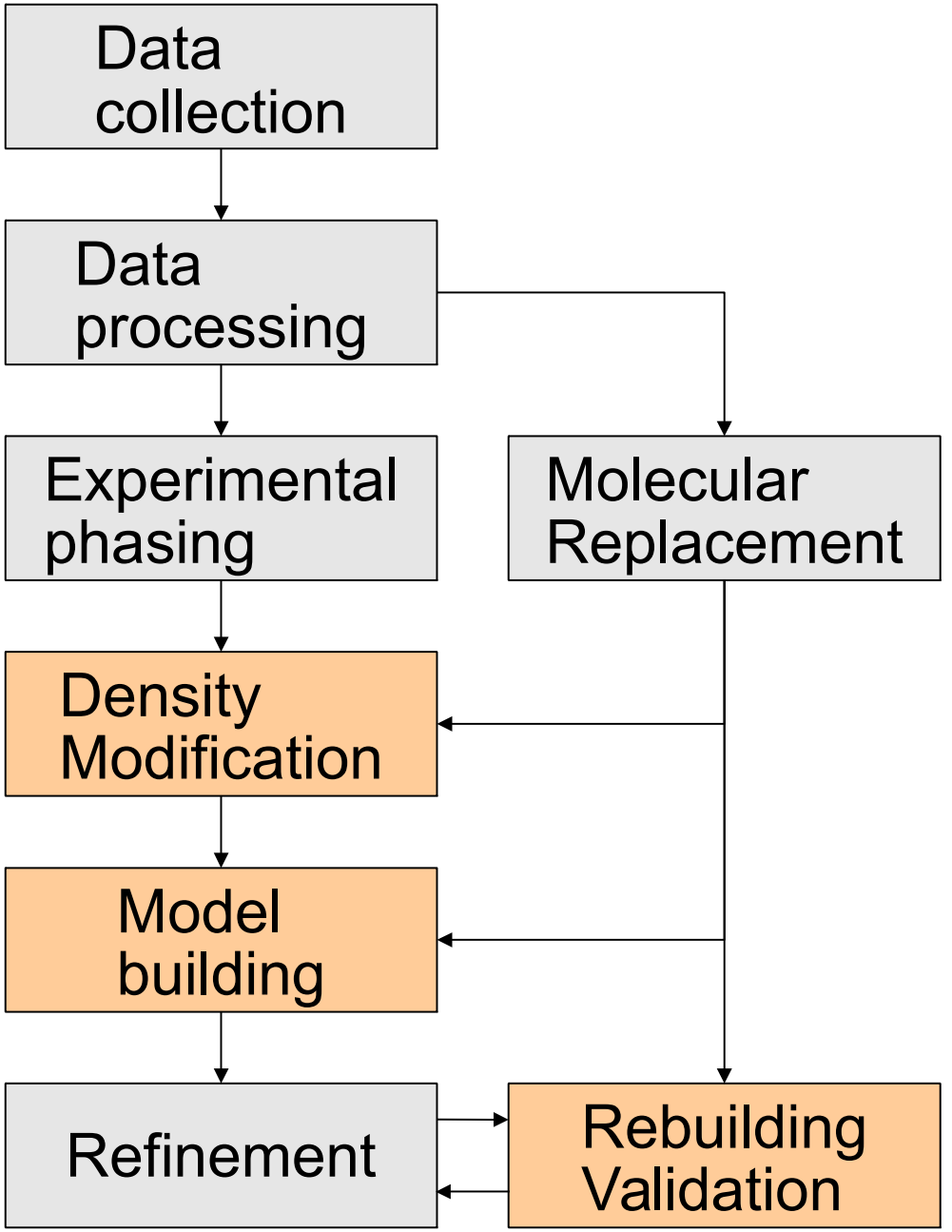
# X-ray structure solution pipeline...







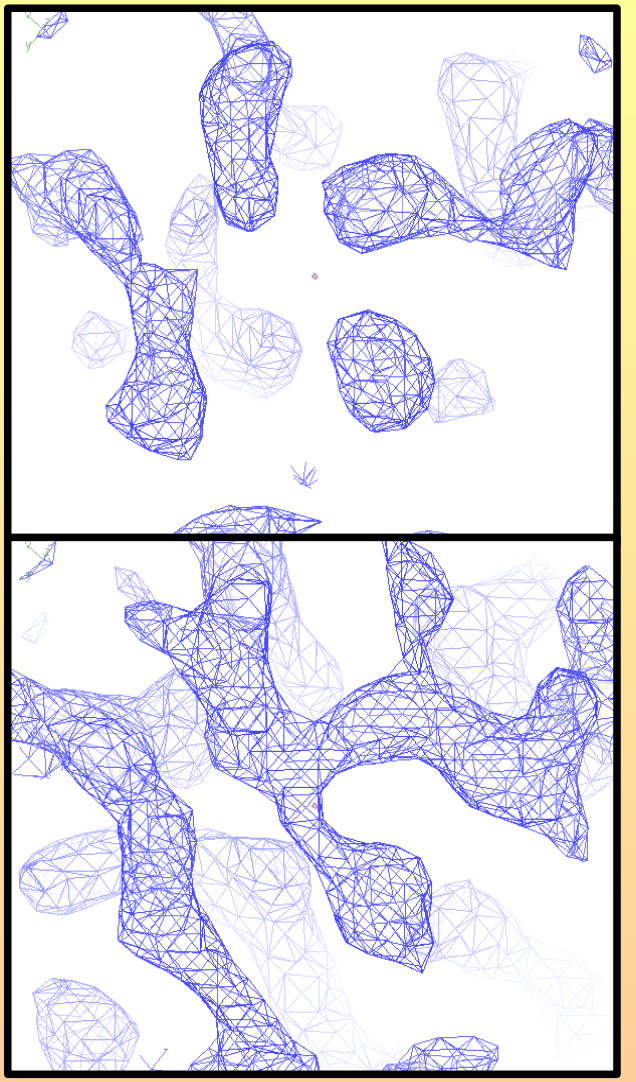





Kevin Cowtan, cowtan@ysbl.york.ac.uk

# Density Modification

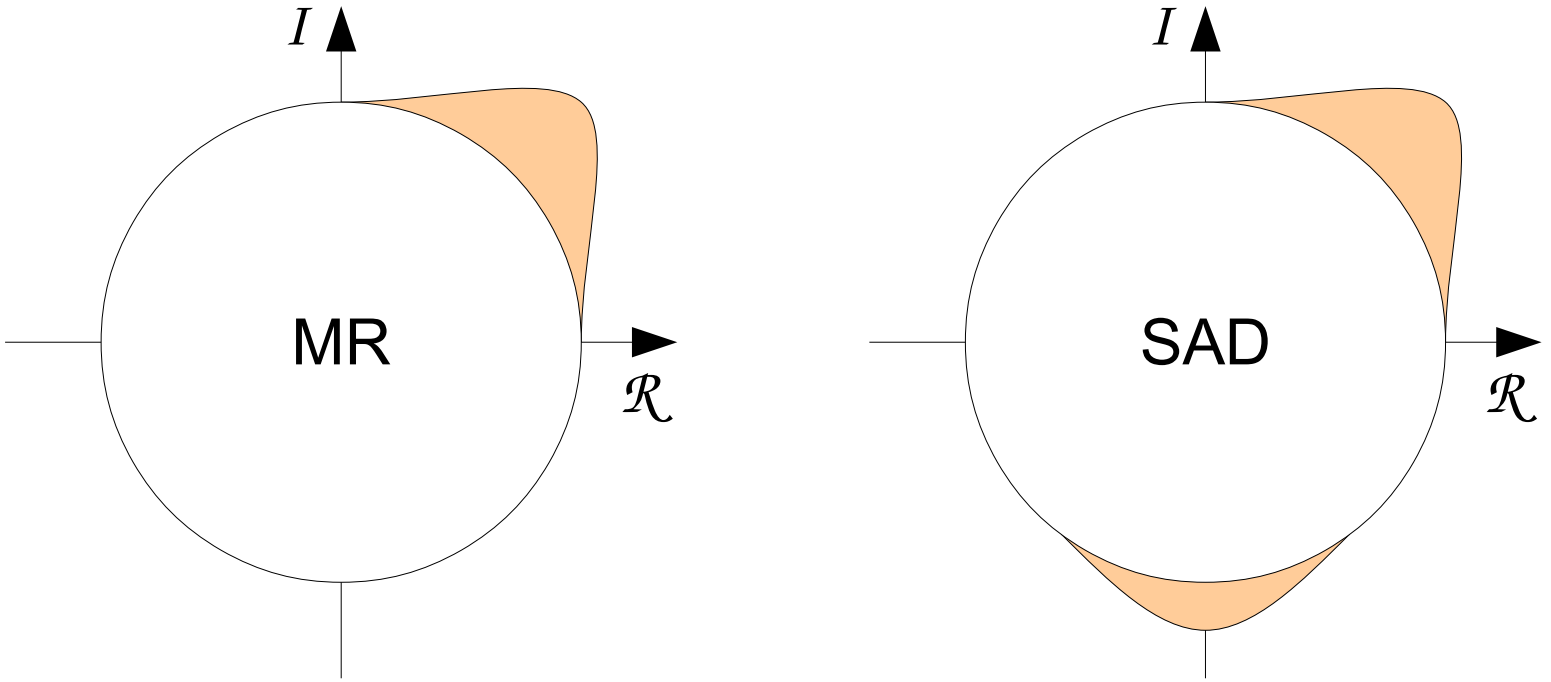
- Traditional density modification: e.g. 'dm', 'solomon', 'parrot', CNS
- Statistical density modification: e.g. 'resolve', 'pirate'



# Density modification

Starting point:

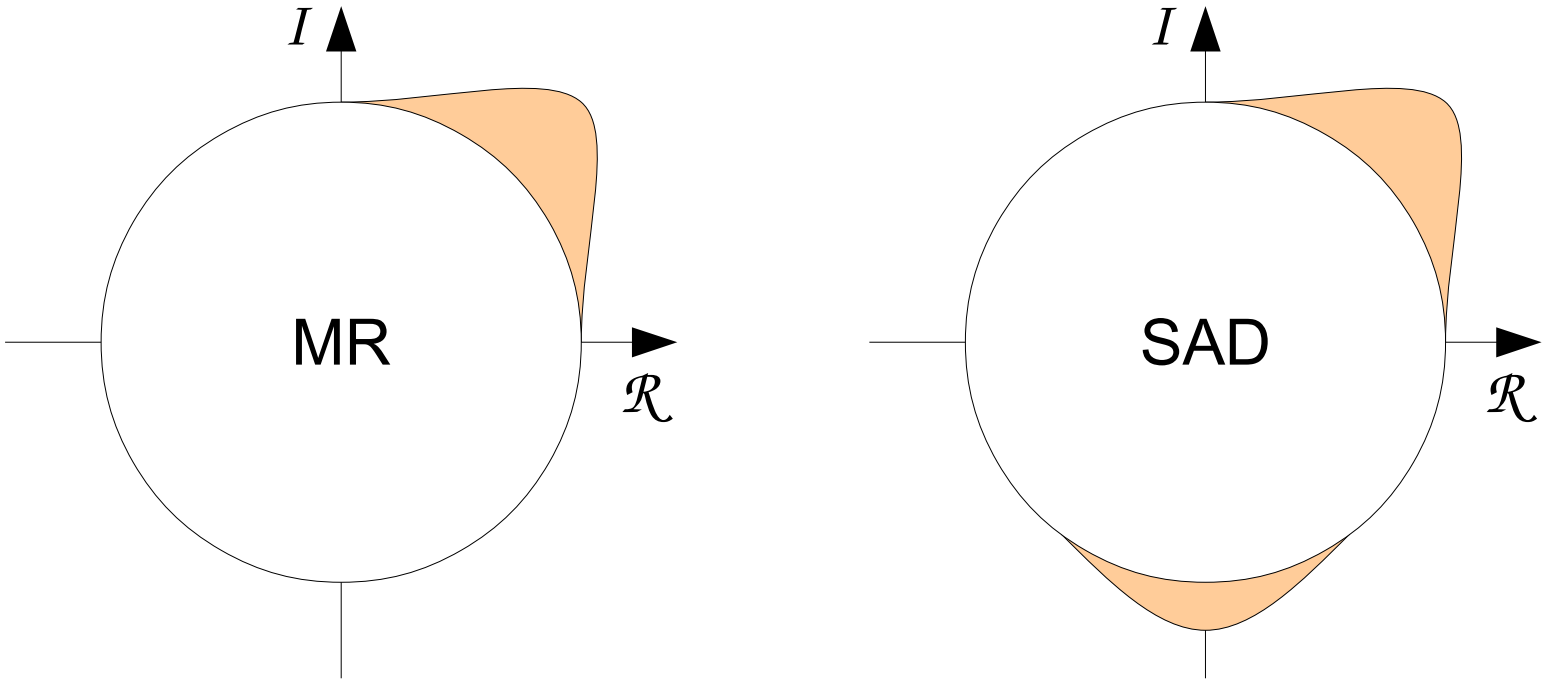
- Structure factor amplitudes
- Phase estimates:
  - MR: Unimodel distribution
  - SAD: Biomodal distribution



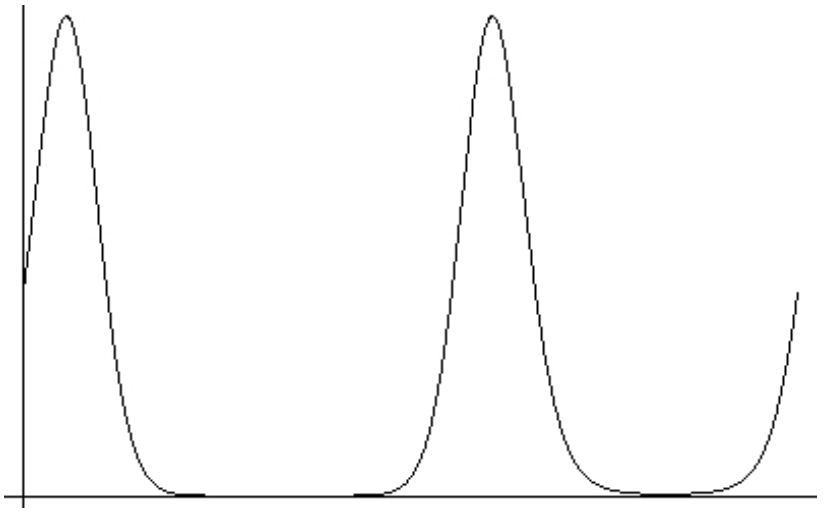
# Density modification

How do we represent phase probability distributions?

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



# A convenient way to represent the phase distribution



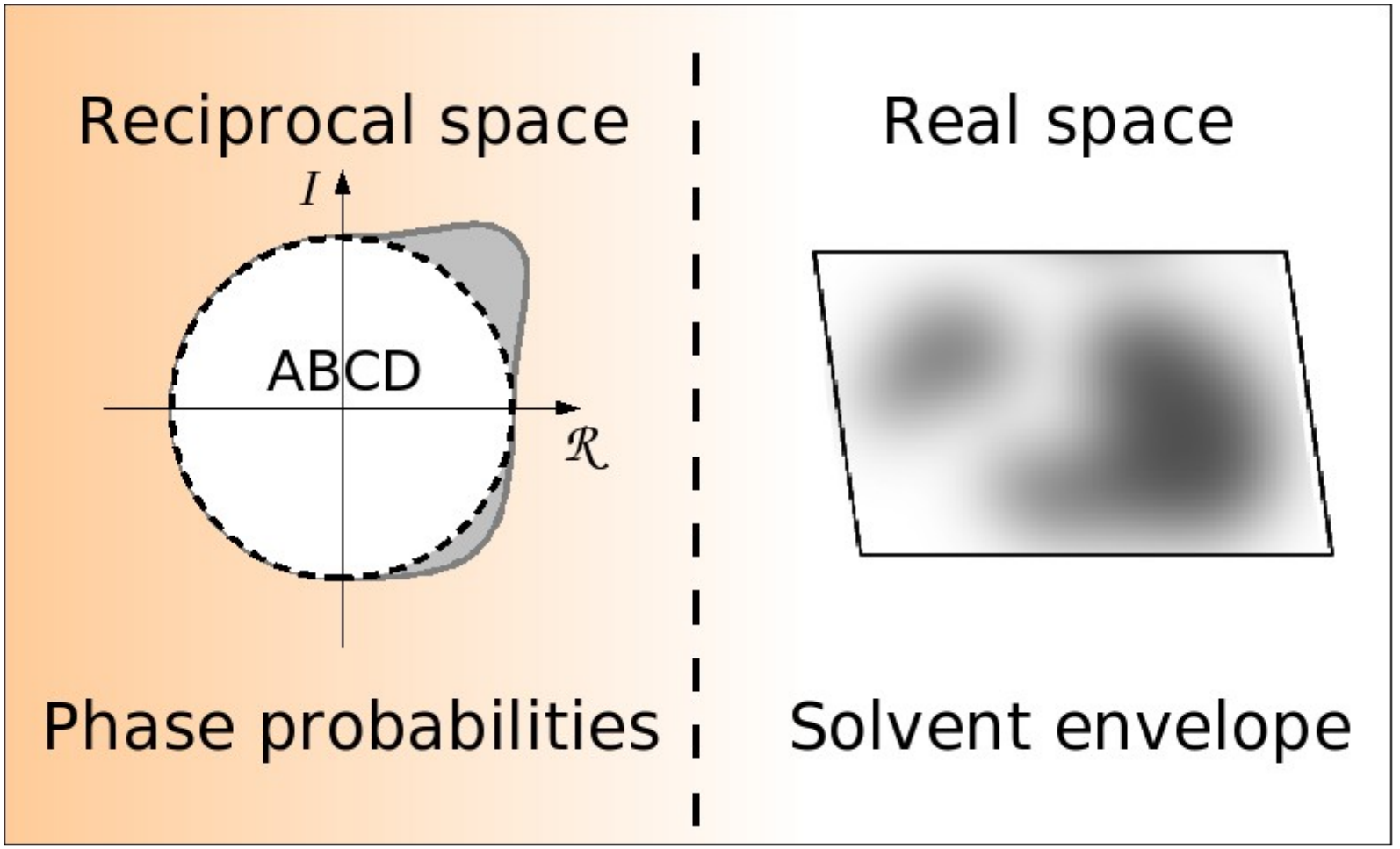
- $P(\theta) = \exp(A\cos(\theta) + B\sin(\theta) + C\cos(2\theta) + D\sin(2\theta))$
- The phase distribution can be represented by “Hendrickson-Lattmann” coefficients, A, B, C, D.

# Phase probabilities

- How do we choose the phase to use for our electron density map?
- The “best phase” corresponds to mean/average/expected value (Blow and Crick).
- What are ways to determine how good and accurate your phases are?
  - FOM: figure of merit: mean cosine of the phase error (a value of 1 means phases are perfect, a value of 0 means phases is the same as random).
  - FOM after density modification may be biased/inaccurate!

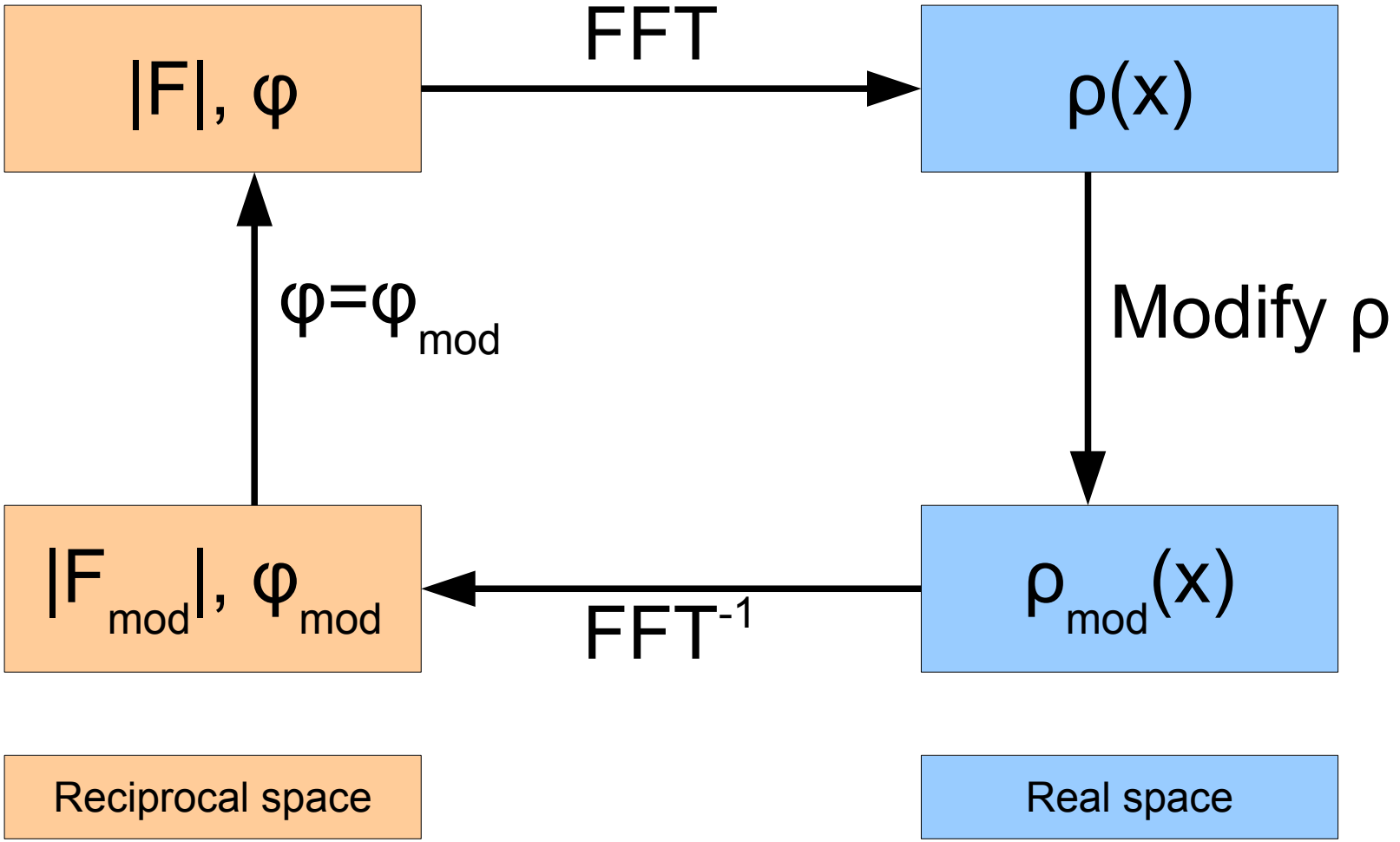
# Density modification

- Density modification is a problem in combining information:



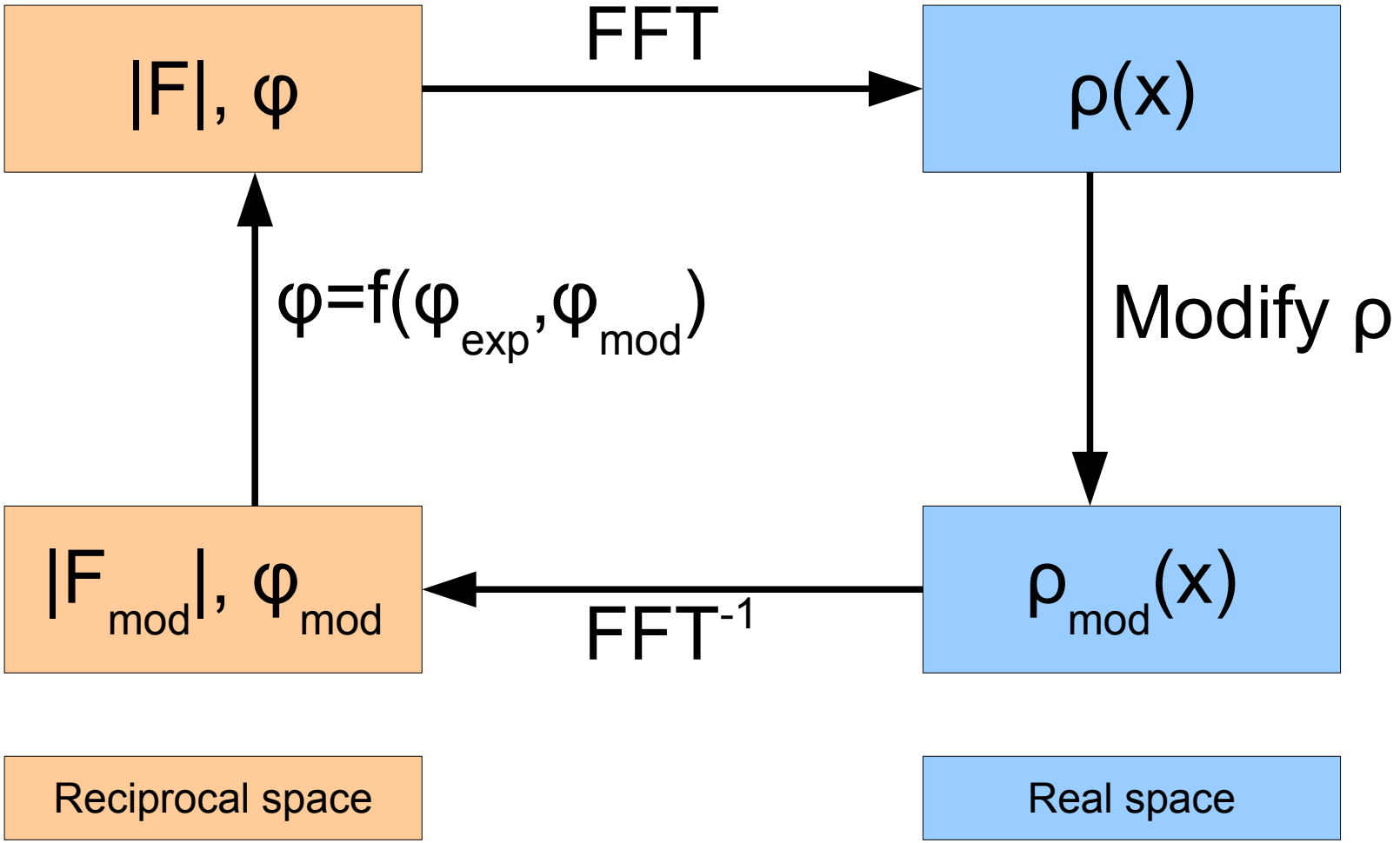
# Density modification

1. Rudimentary calculation:



# Density modification

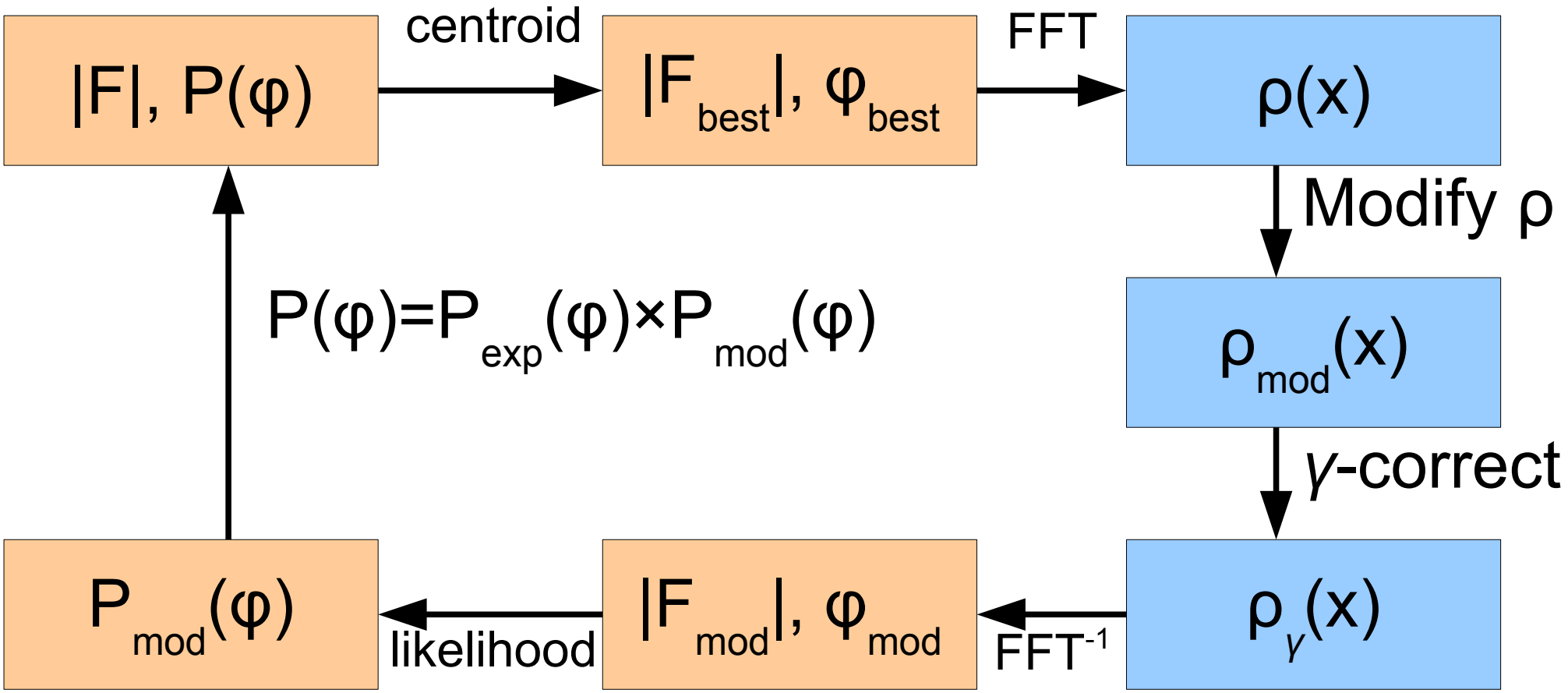
## 2. Phase weighting:



# Density modification

DM, SOLOMON, (CNS)

4. Bias reduction (gamma-correction):




J.P.Abrahams

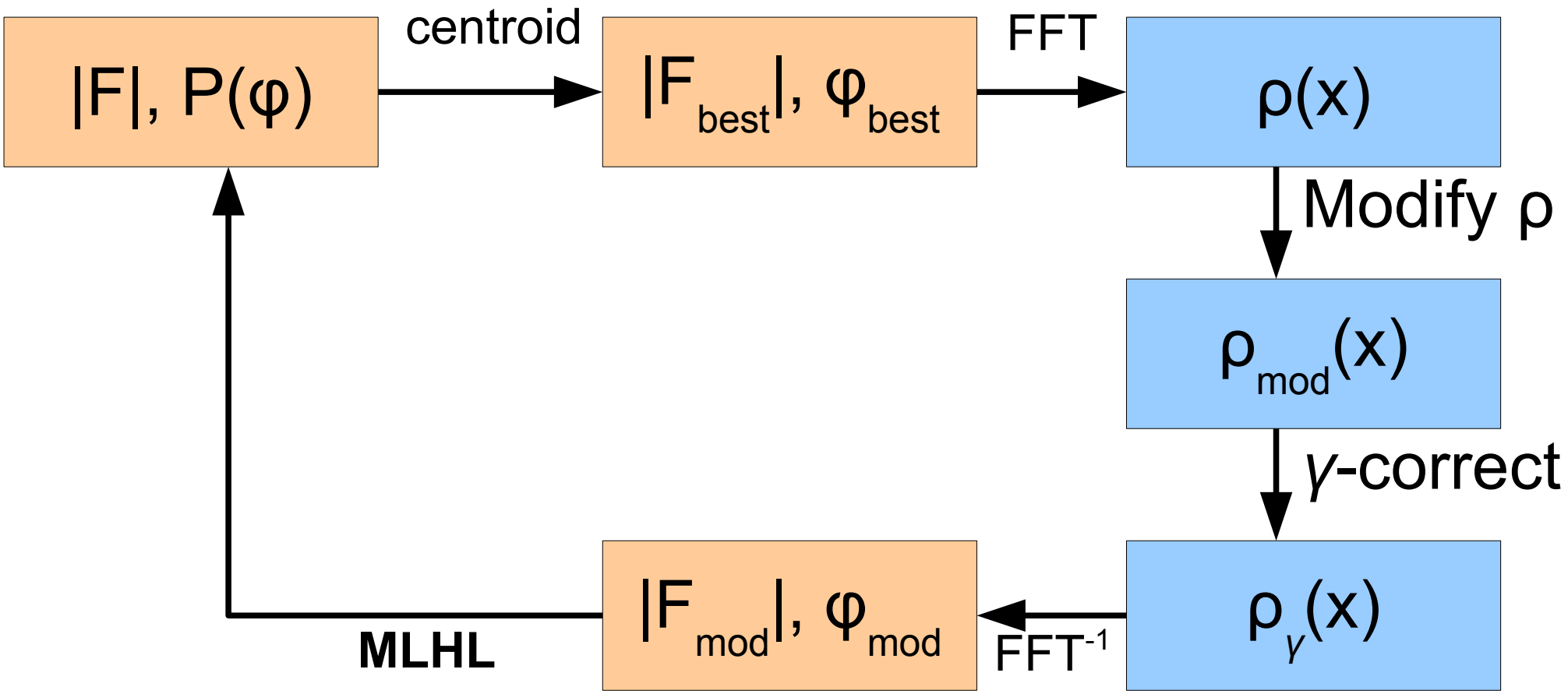
Kevin Cowtan, cowtan@ysbl.york.ac.uk

2012

# Density modification

PARROT

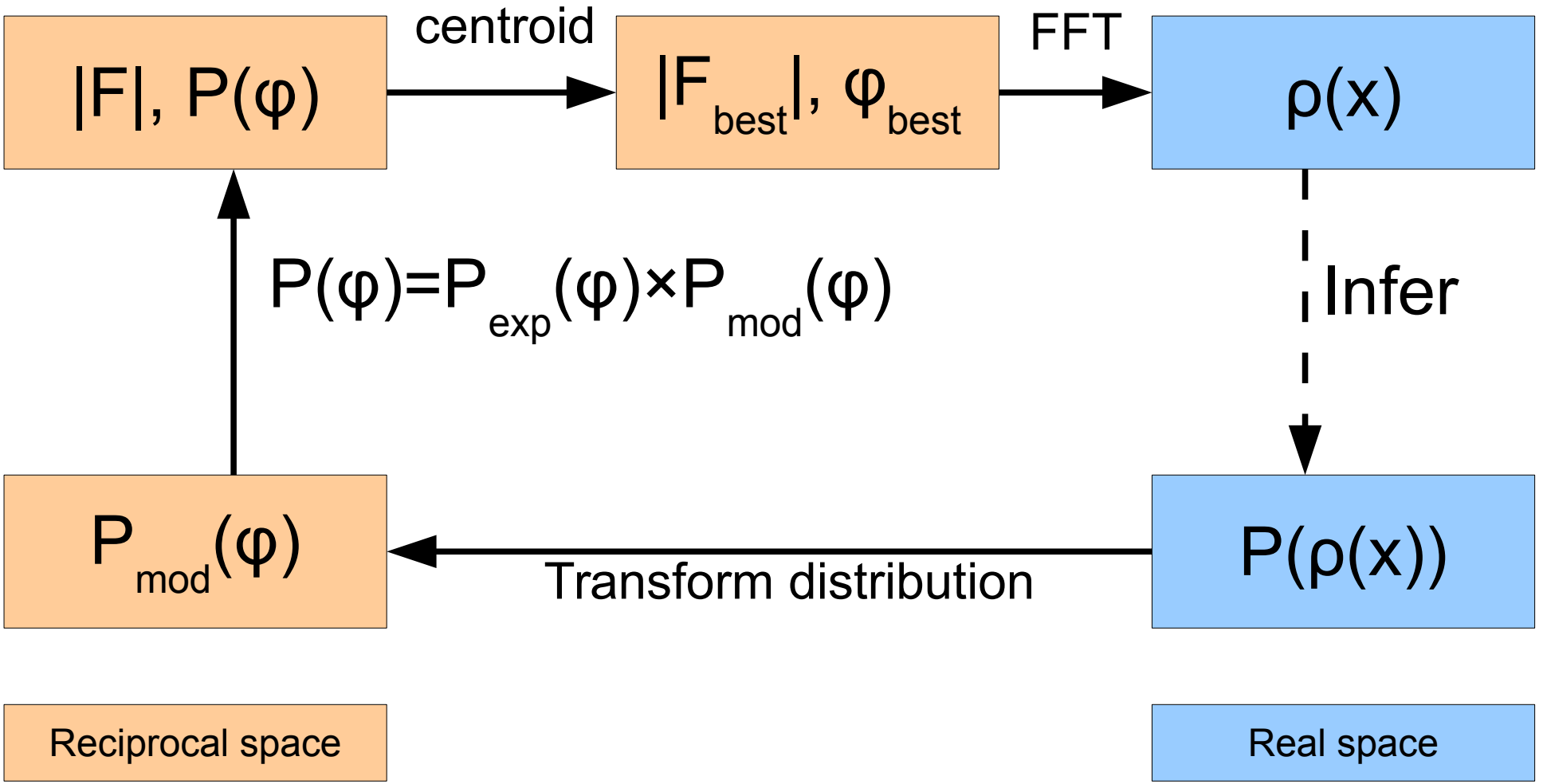
## 5. Maximum Likelihood H-L:



# Density modification

RESOLVE, PIRATE

## 6. Statistical density modification:

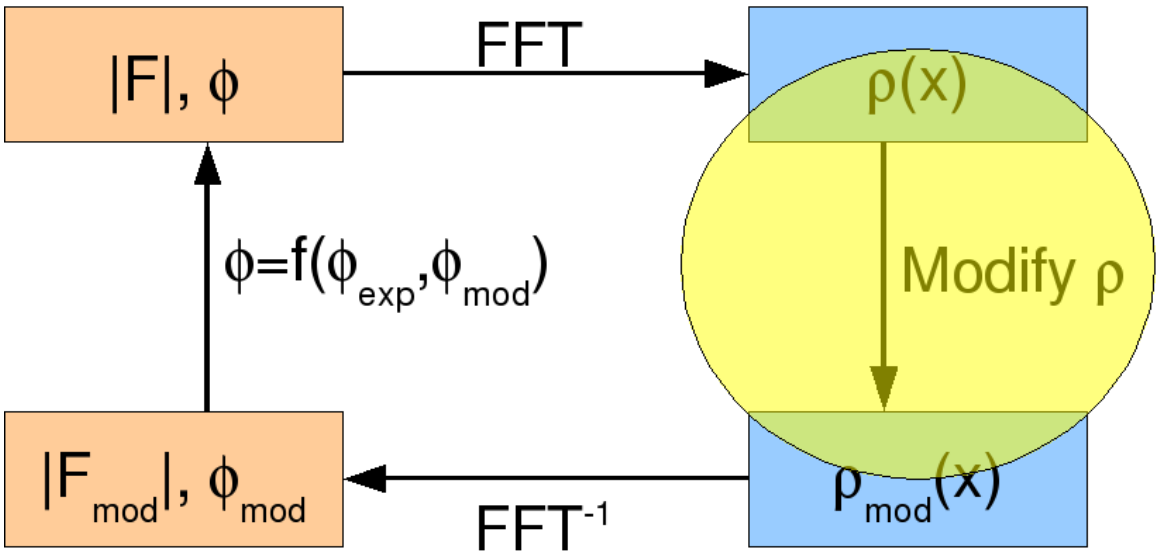



Kevin Cowtan, cowtan@ysbl.york.ac.uk

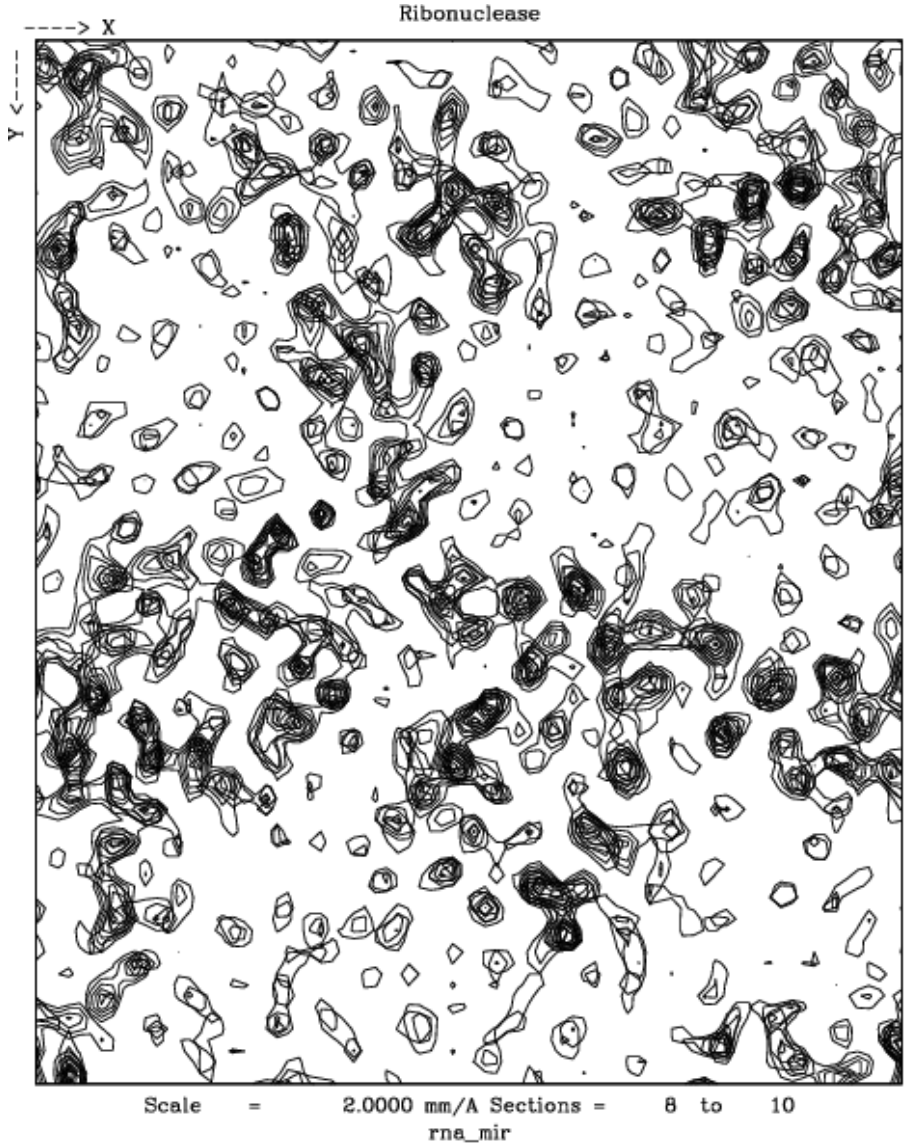
# Density modification

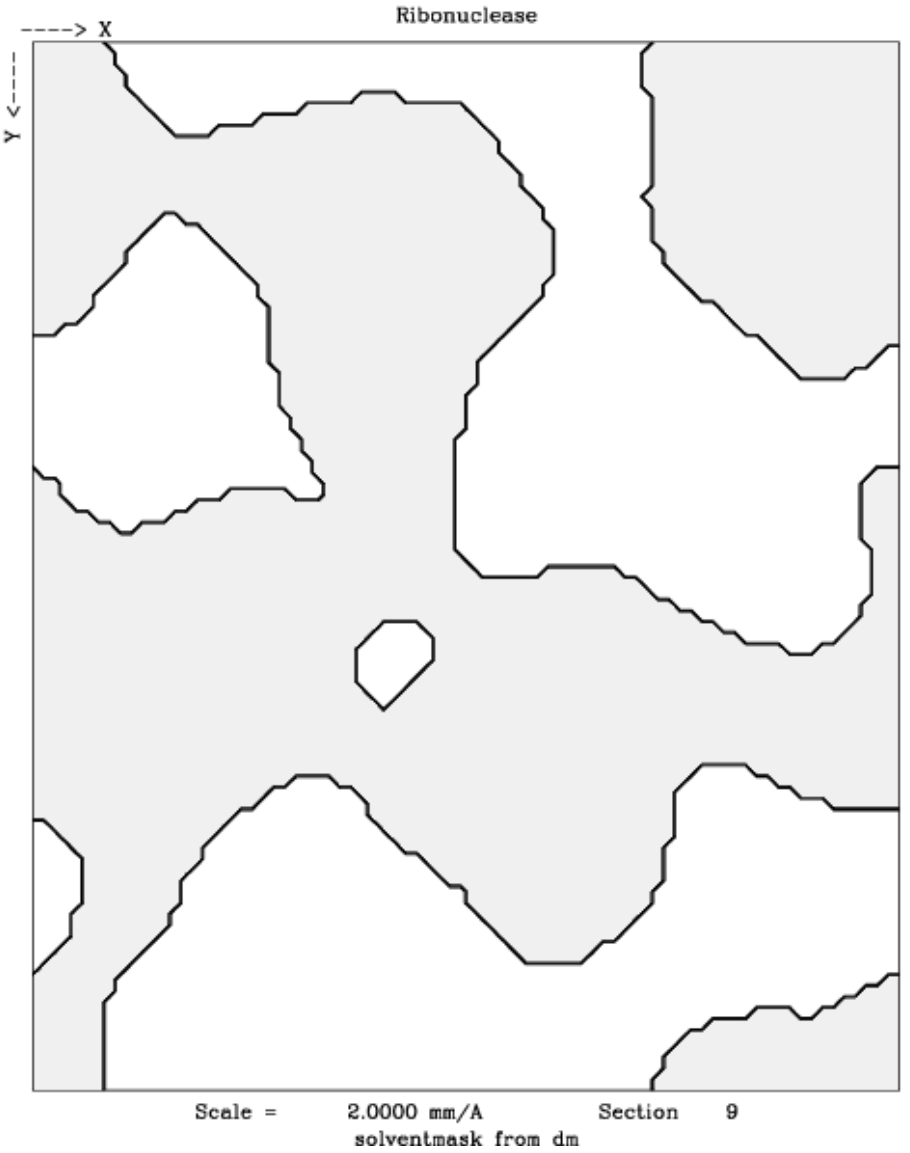
Traditional density modification techniques:

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



# Solvent flattening



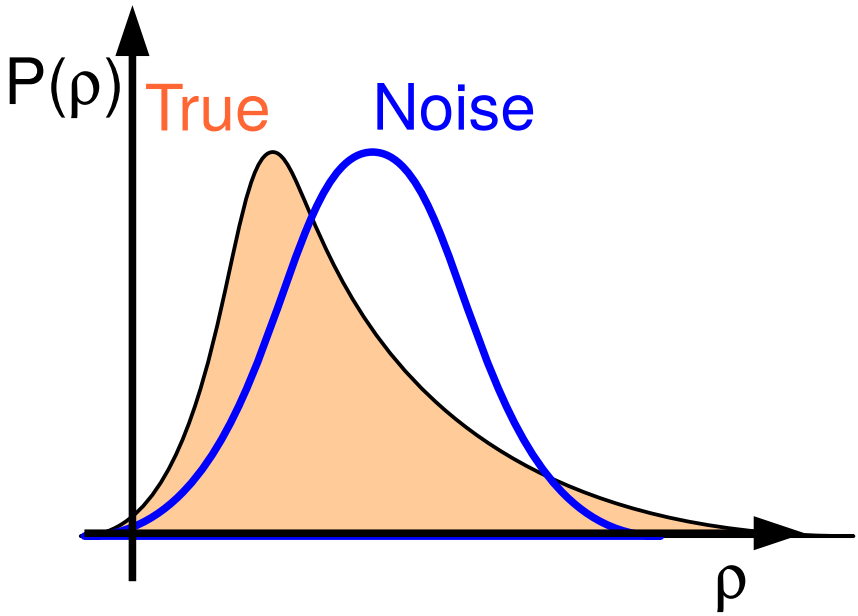
# Histogram matching

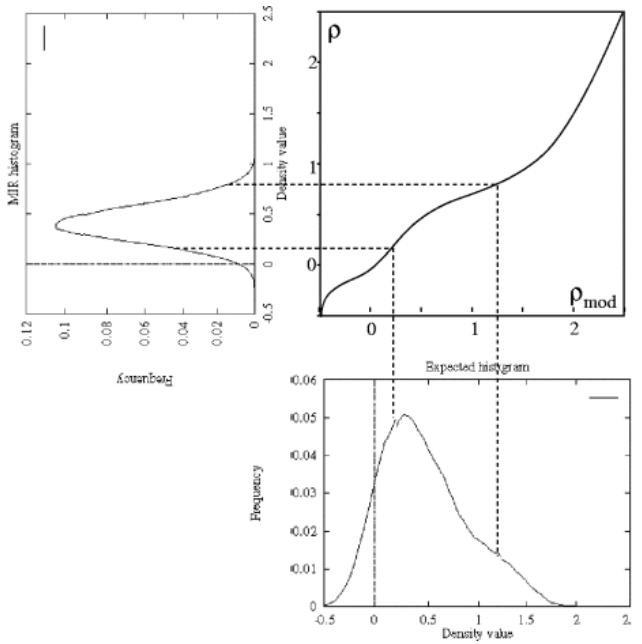
A technique from image processing for modifying the protein region.

- 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.

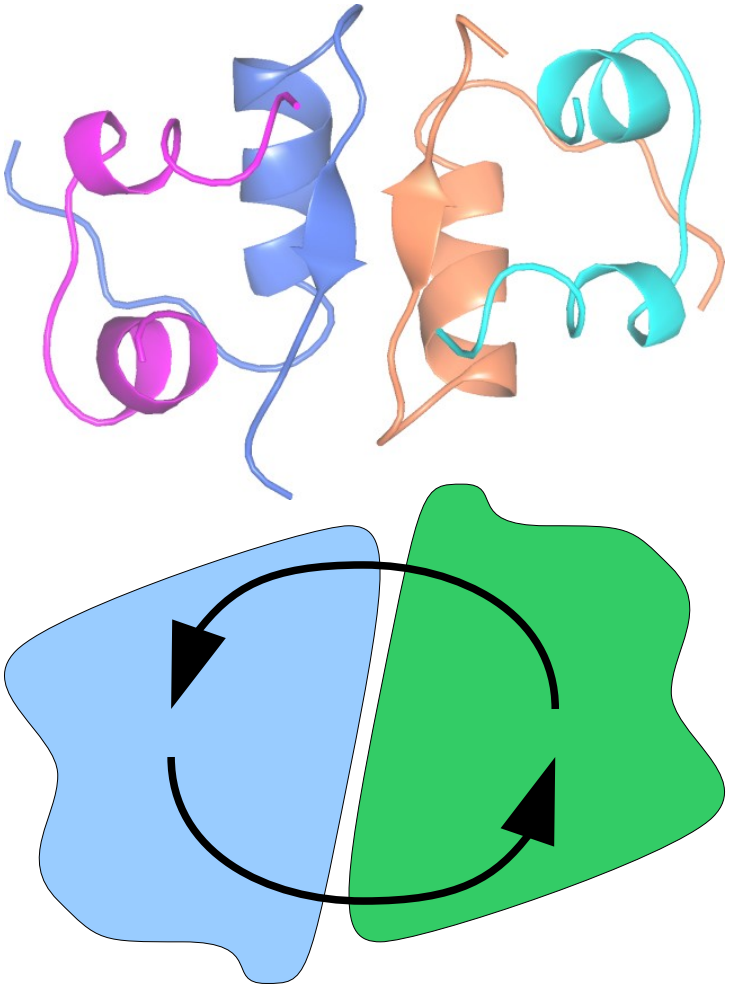



Kevin Cowtan, cowtan@ysbl.york.ac.uk

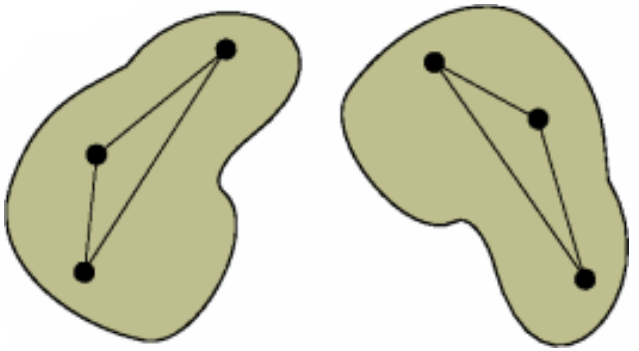
# Non-crystallographic symmetry

- 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!



# 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.



# Combining phase probabilities

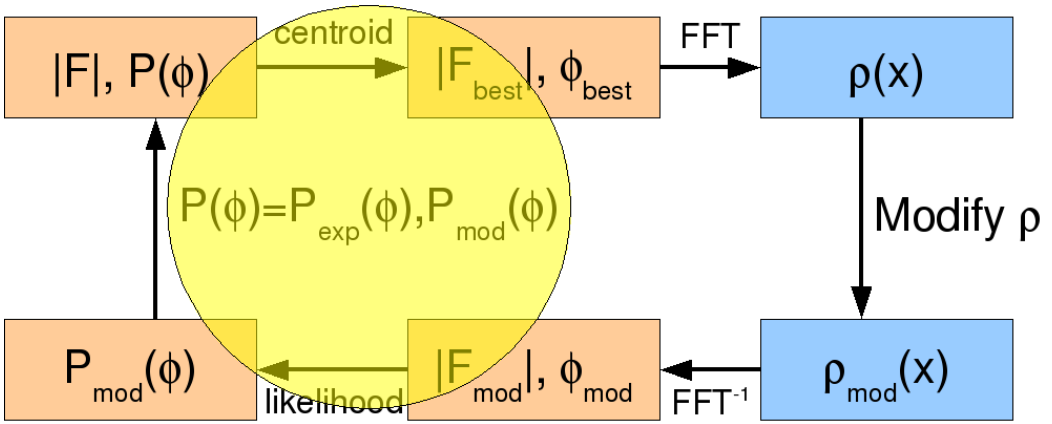
Once we have an estimate for the error in $\phi_{mod}$, we can construct a probability distribution $P_{mod}(\phi)$.

The the next cycle can be started with

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

**Problem**: $P_{exp}(\phi)$ and $P_{mod}(\phi)$ are not independent.

The result is bias, increasing with cycle.



# 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

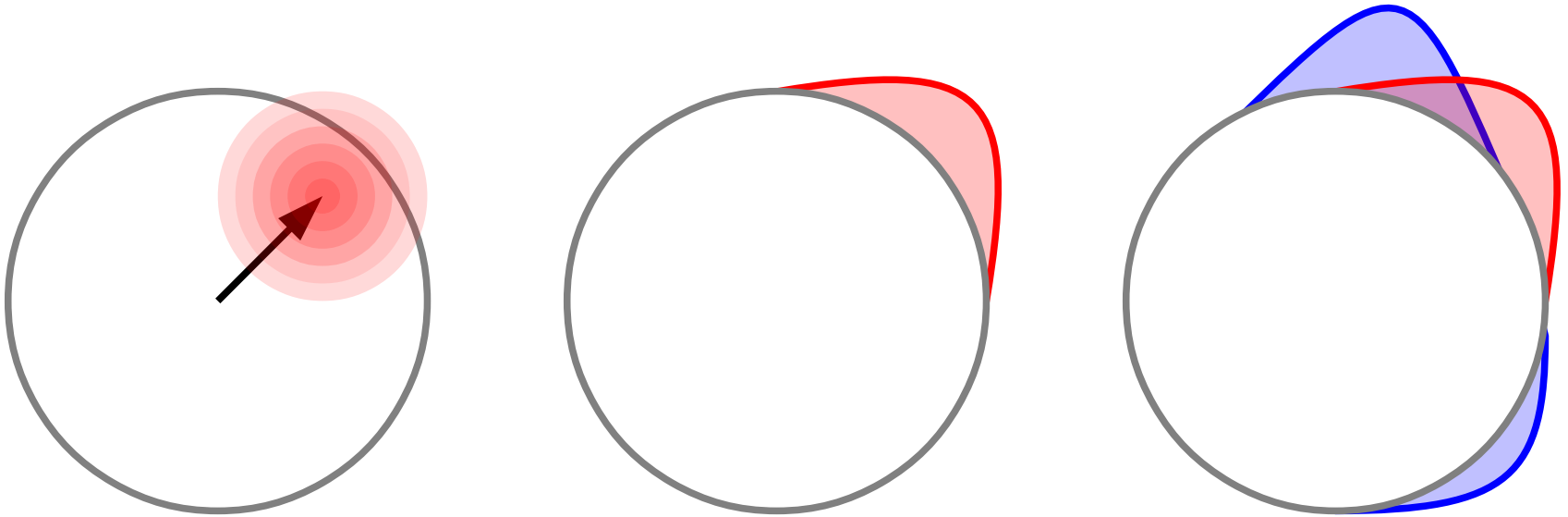
# Density modification in Parrot

## 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

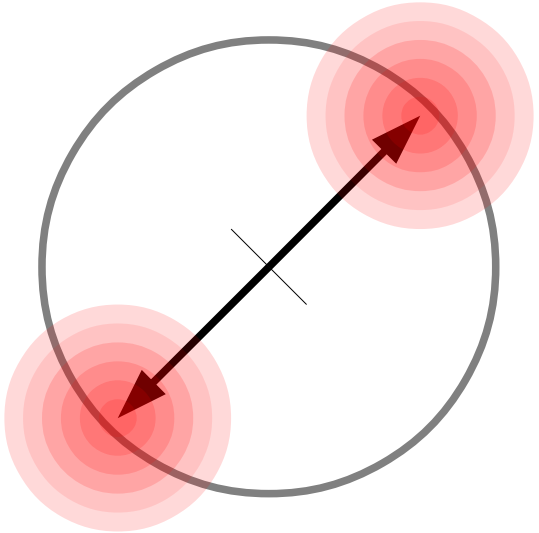
Turn this into a phase probability distribution

Combine with the experimental phase probability

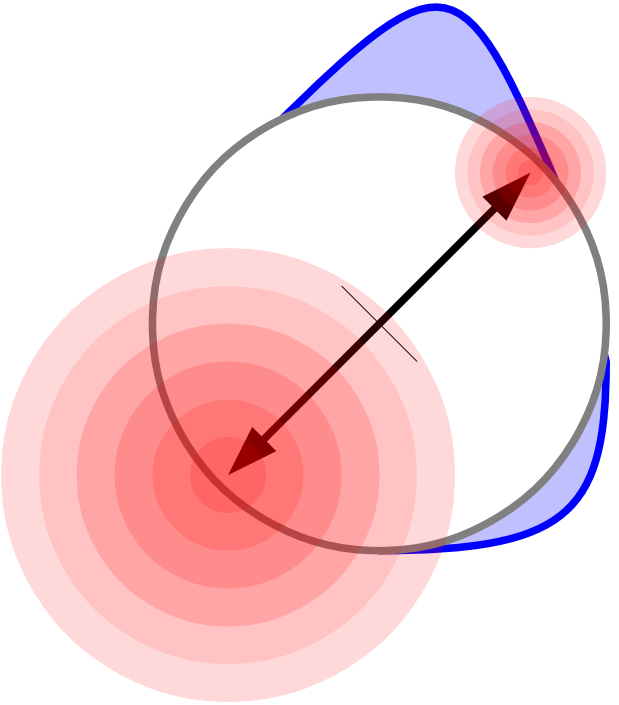
The estimate for the accuracy of the modified F/phase come from the agreement between the modified F and the observed F. **Source of bias.**

# Estimating phase probabilities

Problem:



Error estimation does not take into account experimental phase information

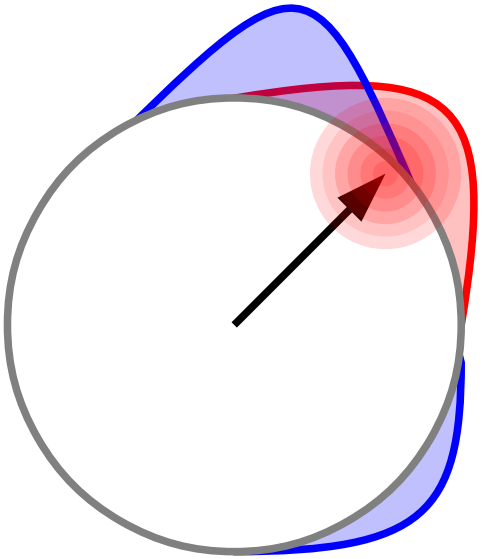


The experimental data tells us that the probable error is different in the two cases

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

# Estimating phase probabilities

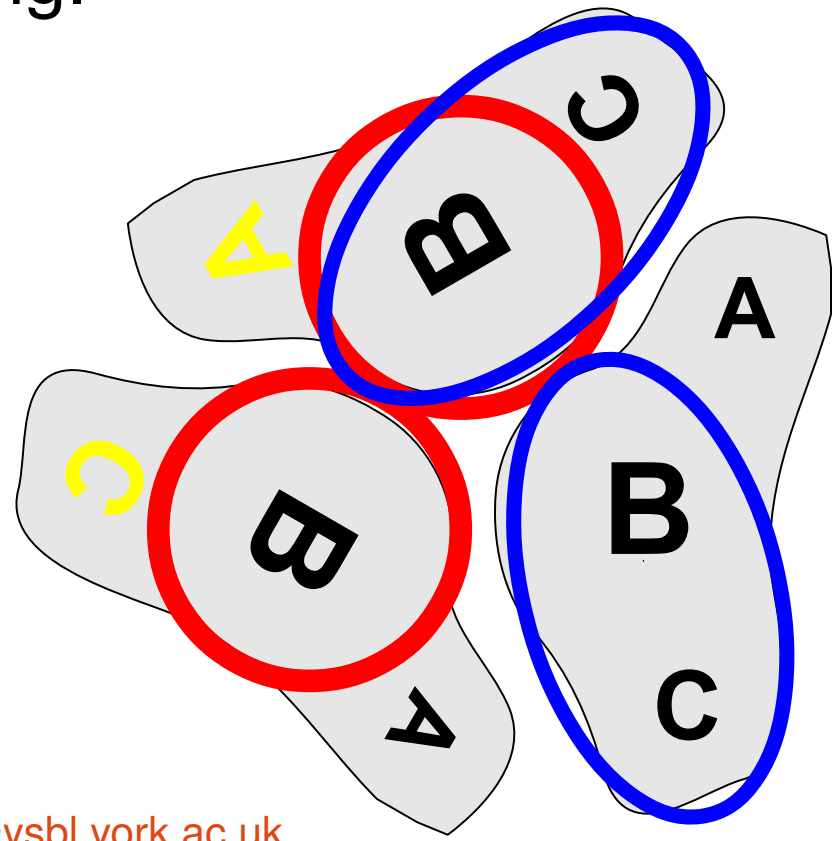
Solution:
MLHL-type likelihood target function.



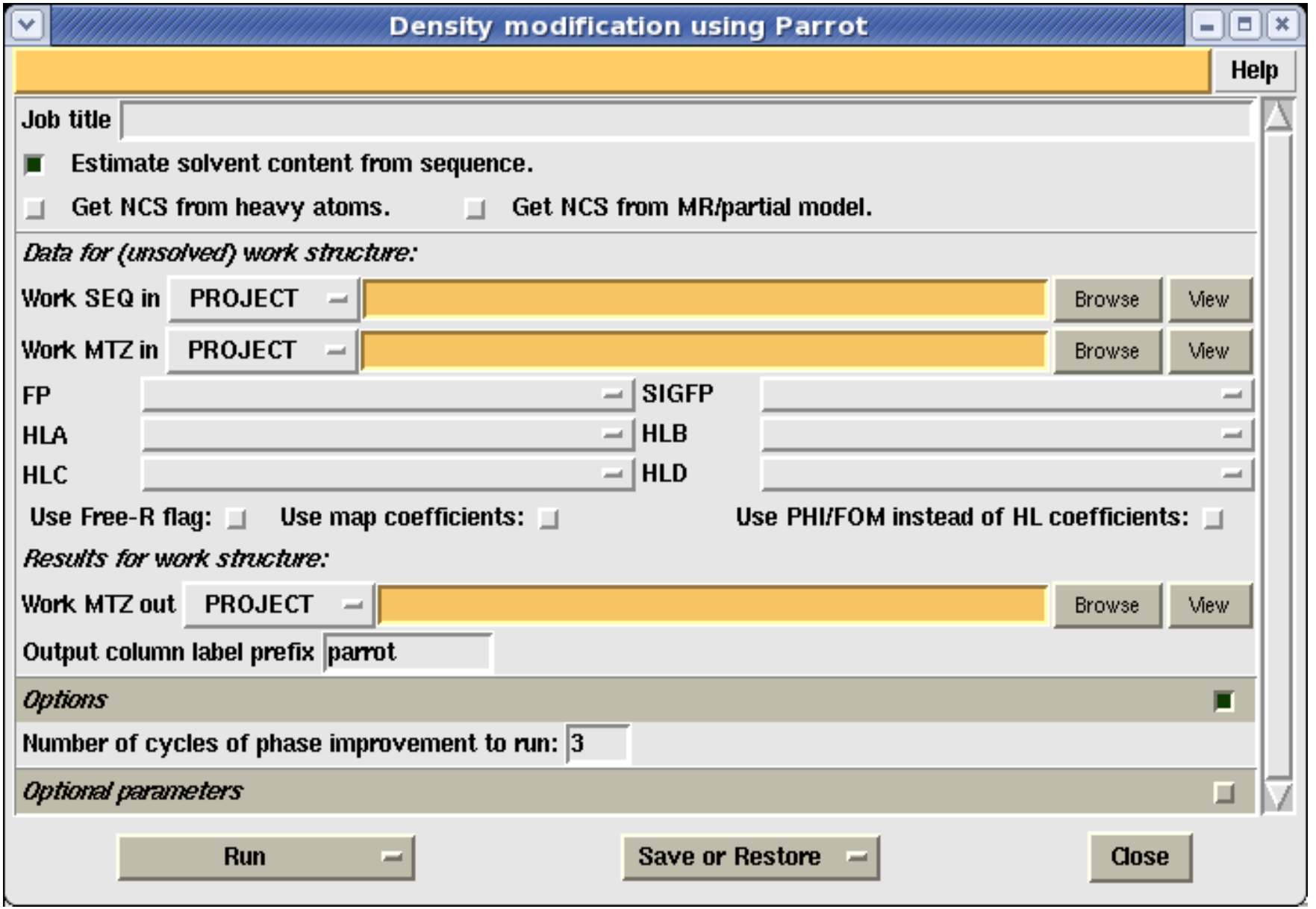
Perform the error estimation and phase combination in a single step, using a likelihood function which incorporates the experimental phase information as a prior.
This is the same MLHL-type like likelihood refinement target used in modern refinement software such as *refmac* or *phenix.refine*.

# Pairwise-weighted NCS averaging

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

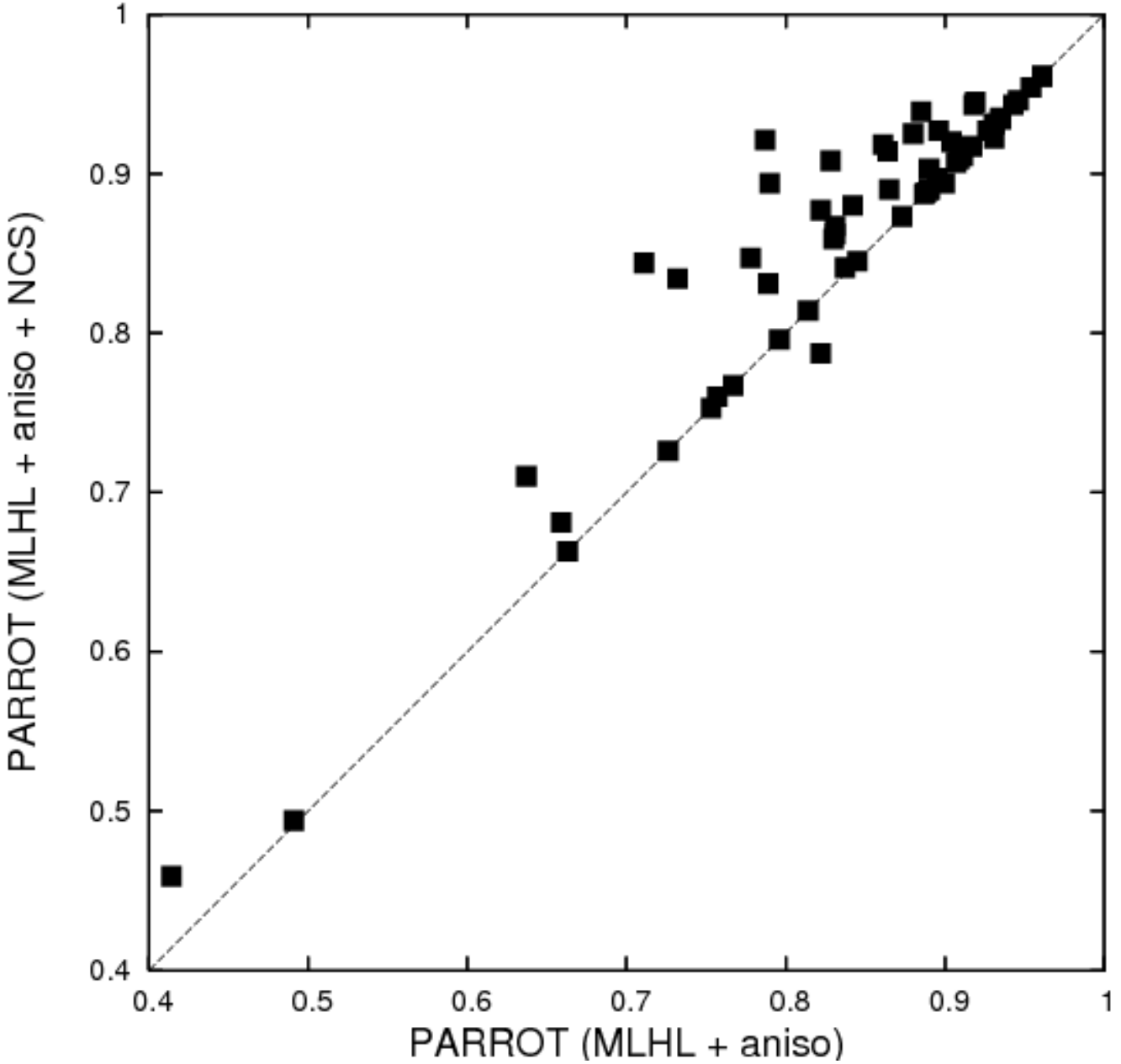


# Parrot




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

# Parrot: simple vs NCS averaged



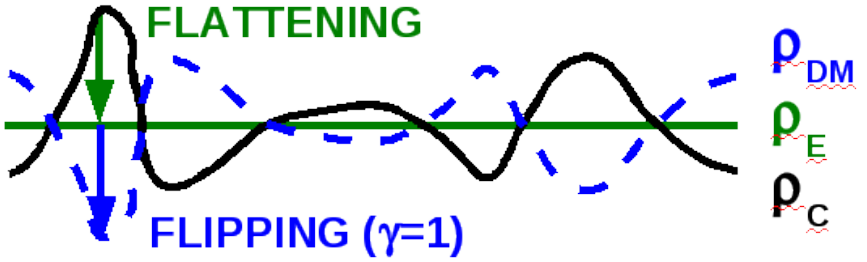

Map correlations

Comparing with and without NCS averaging.

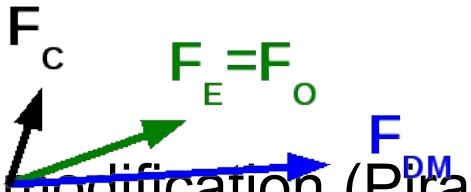
Kevin Cowtan, cowtan@ysbl.york.ac.uk

# Bias reduction methods

- Map flipping (gamma correction): The (likely) solvent regions are not flattened but flipped




- Structure factor flipping:

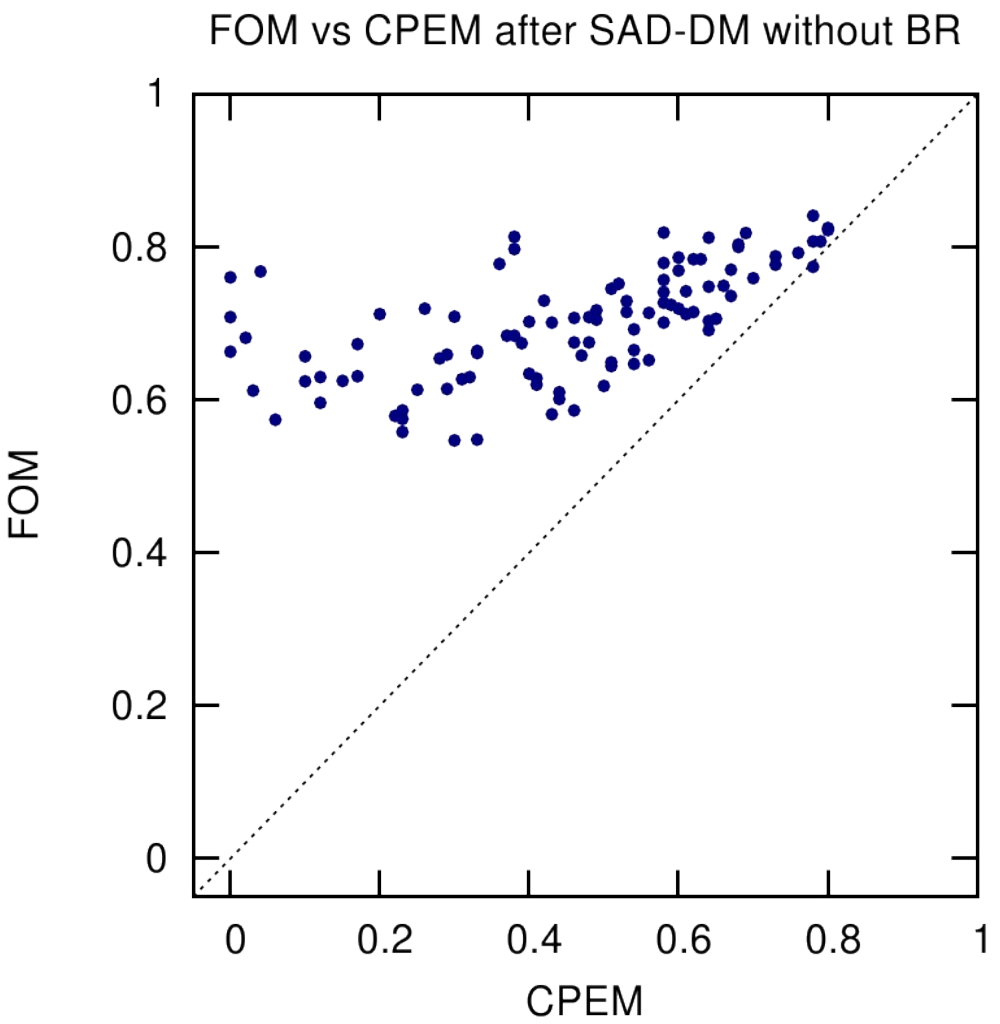


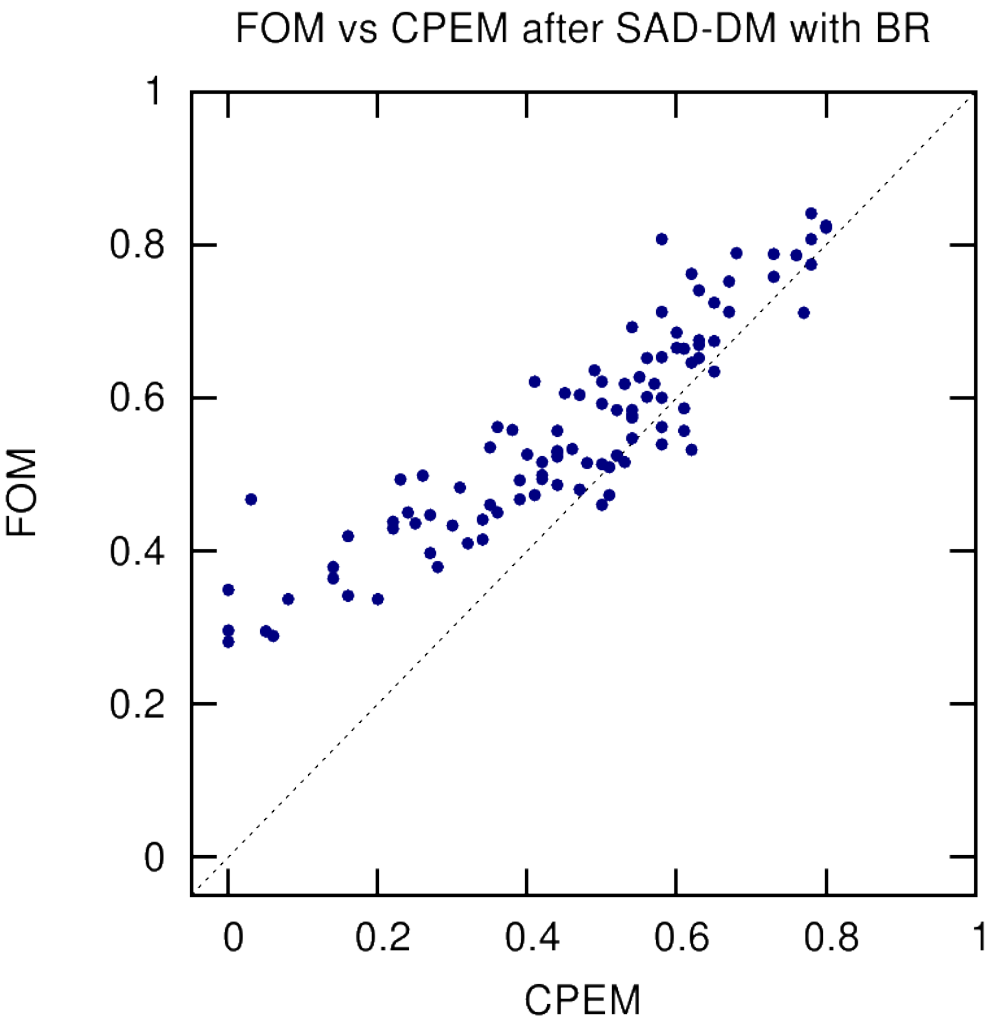

- Statistical density modification (Pirate)

# Bias reduction in density modification

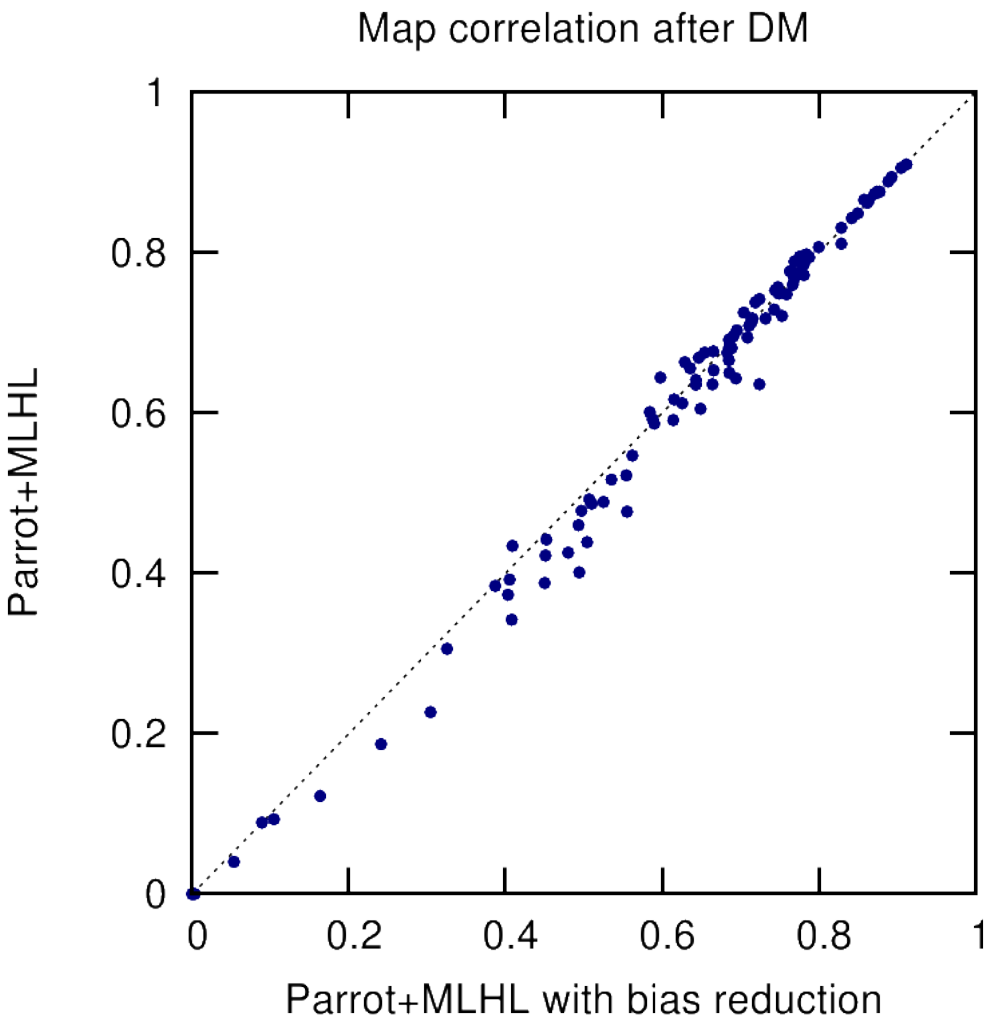
- Density modified map is obtained from experimental map leading to artificially high correlations between the observed and modified amplitudes.
- β correction is applied to the Luzzati error parameter to reduce bias of modified data.

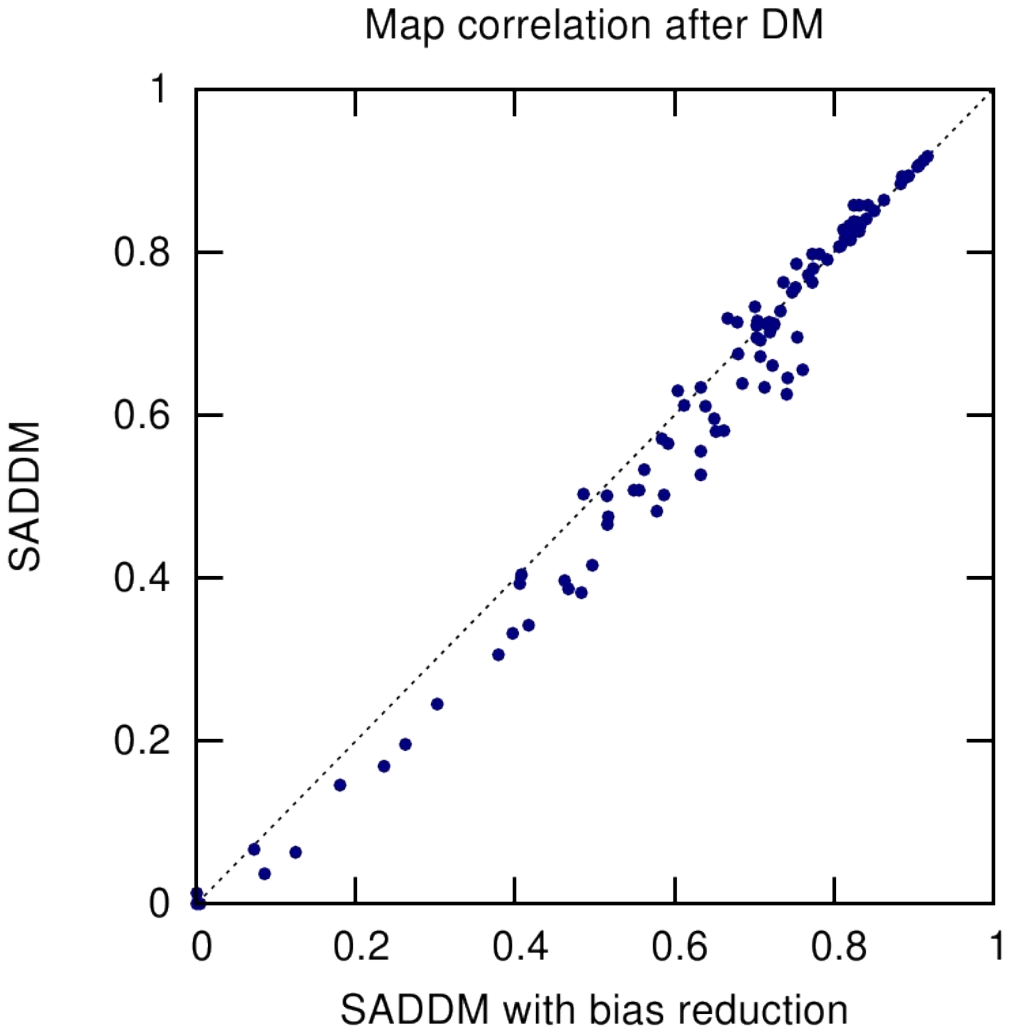
# FOM and phase error after DM with/without bias reduction



# Map correlation after DM with/without bias reduction



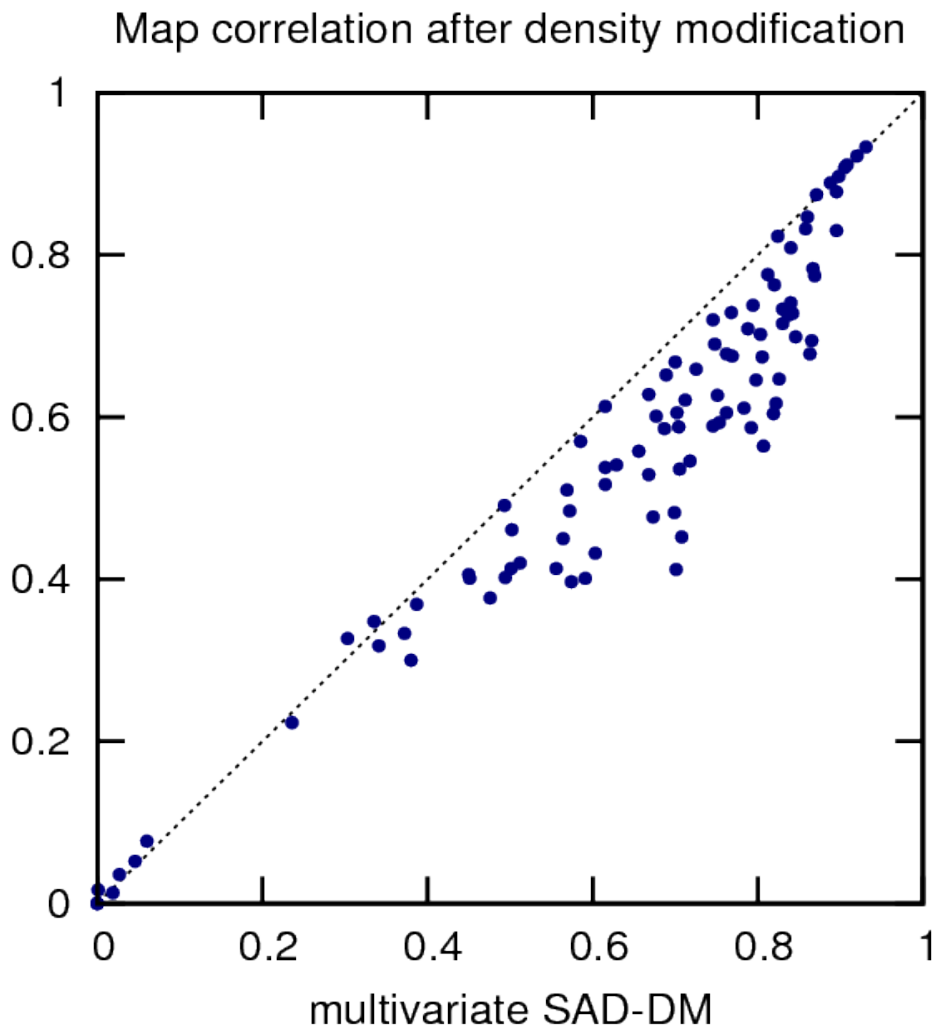
# Multivariate phase combination for density modification

- Phase combination procedures assume independence between the original map and the density modified map
- Multivariate SAD-DM probability distribution:

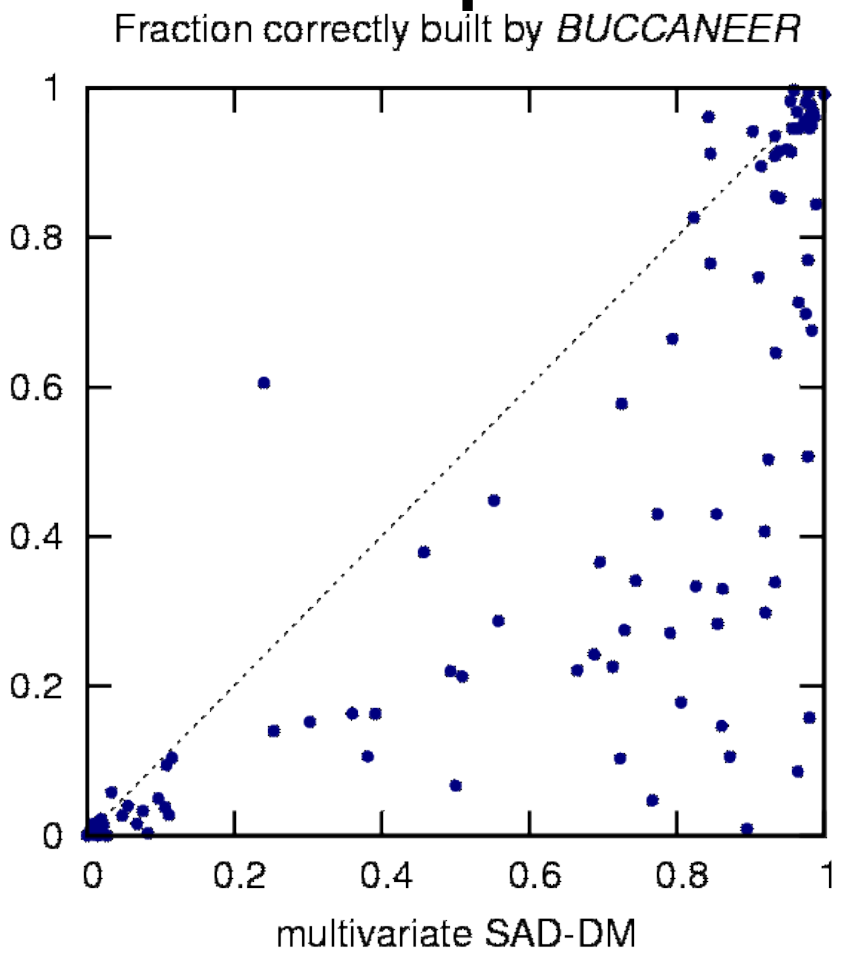
$$P_{DM}\left(F_o^+, F_o^- \middle| F_H^+, \phi_H^+, F_H^-, \phi_H^-, F_{DM}, \phi_{DM}\right)$$

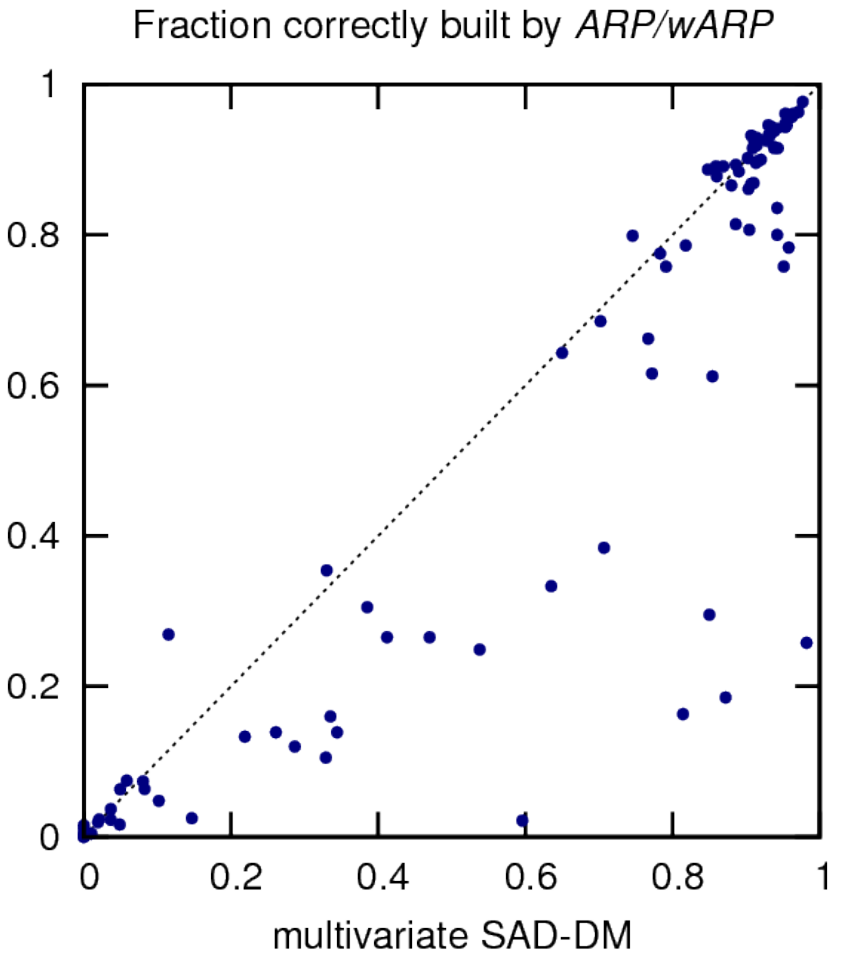
- Advantages: no independence assumption, dynamic construction of phase probabilities rather than static HL
- Implementation: MULTICOMB, REFMAC

# Comparison of Rice vs. multivariate function



# Results of model building after DM with Rice and SAD-DM phase combination



# Acknowledgements

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