Molecular Replacement
(Alexei Vagin's lecture)

Contents

- What is Molecular Replacement
- Functions in Molecular Replacement
- Weighting scheme
- Information from data and model
- Some special techniques of Molecular Replacement

Molecular replacement programs and systems

Programs

- Amore
- MOLREP
- QS
- PHASER
- EPMR
- CNS
- URO

Systems

- MrBump
- Phenix
- BALBES
- Many others

Molecular replacement programs and systems

Programs

- Amore
- MOLREP
- QS
- PHASER
- EPMR
- CNS
- URO

Systems

- MrBump
- Phenix
- BALBES
- Many others

Introduction: Where MR could help?

Same protein could be crystallised in different space groups

Mutants

Complexes

Homologous proteins

Some structure could be derived using NMR

Homology modeling

MR works best when similarity (3D similarity) between search and target molecules is high and the search model is relatively big.

Introduction

Molecular replacement (MR) is a phasing technique. It may help to derive initial phases. If the MR is successful then you need to do many cycles of refinement and model building.

Its attractive side is that it produces initial atomic model also. However avoiding bias towards model may be difficult especially at low resolution. If there are more than one copies of the molecule in the asymmetric unit then non-crystallographic (NCS) averaging may improve phases and maps.

If the resolution high enough (e.g. 2.5 or better) then automatic model building (arp/warp, solve/resolve, buccaneer) may help in model rebuilding.

Overall results reported in PDB

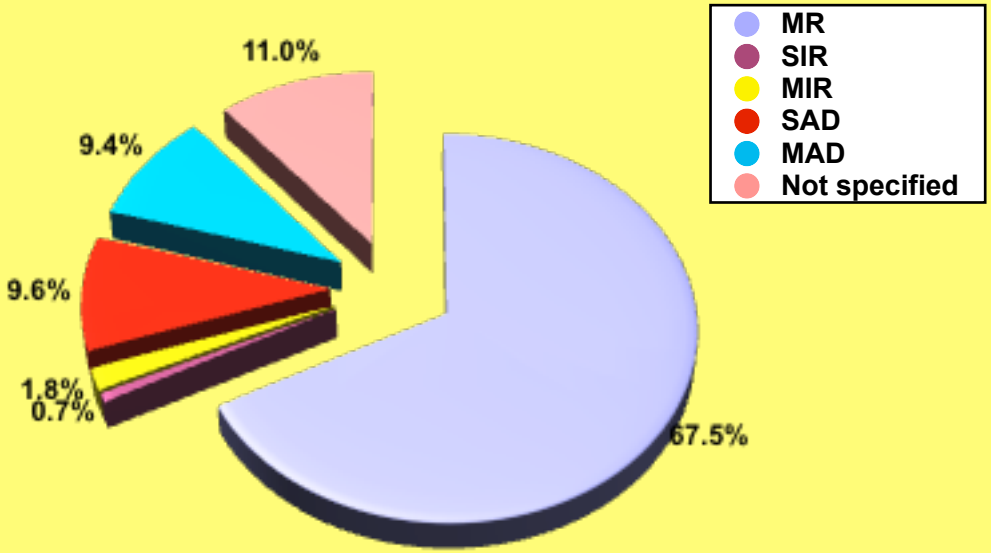


Diagram showing the percentage of structures in the PDB solved by different techniques

67.5% of structures are solved by Molecular Replacement (MR)

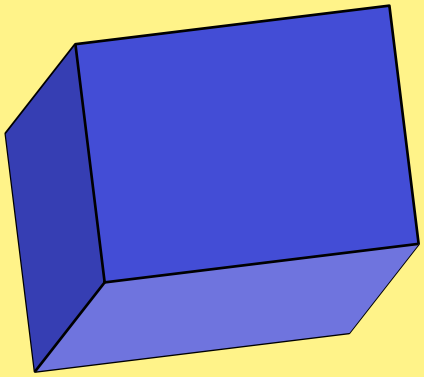
21% of structures are solved by experimental phasing

Molecular Replacement

If we can find the rotation and translation that puts the model in the correct position in the crystal cell, THEN we can calculate phases.

unknown structure

```
MGDKPIWEQIGSSFIQHYYQLFDNDRTQLGAIY
IDASCLTWEGQQFQGKAAIVEKLSSLPFQKIQH
SITAQDHQPTPDSCIISMVVGQLKADEDPIMGF
HQMFLLKNINDAWVCTNDMFRLALHNFG
```



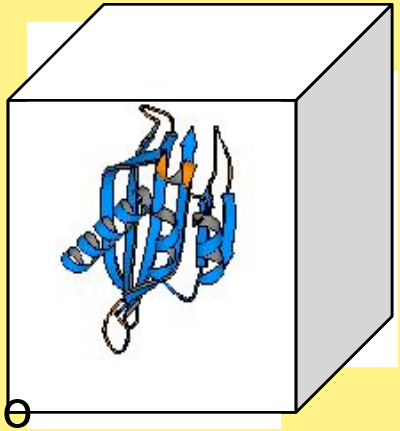
origin o

H K L	F	ϕ
0 0 1	2.5	30
0 0 2	72.1	85
0 0 3	26.9	310

etc...

known structure

```
PSPLLVGREFVRQYYTLLNKAPEYLHRFYGRNSSY
VHGGVDASGKPQEAVYGQNDIHHKVLSLNFSECHT
KIRHVDAHATLSDGVVVQVMGLLSNSGQPERKFMQ
TFVLAPEGSVPNKFYVHNDMFRYEDE
```



origin o

H K L	F	ϕ
0 0 1	10.4	120
0 0 2	3.1	10
0 0 3	52.2	280

etc...

Molecular Replacement

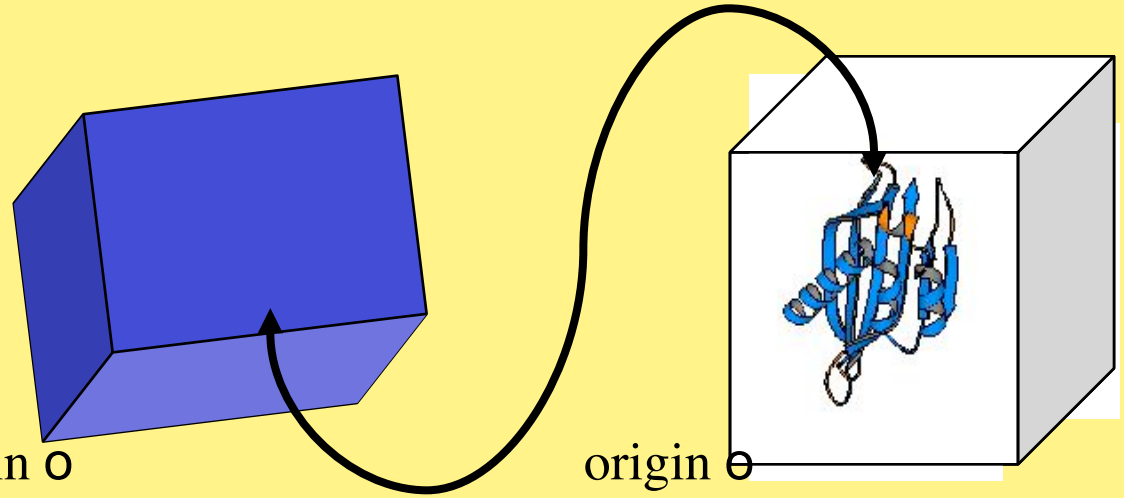
If we can find the rotation and translation that puts the model in the correct position in the crystal cell, THEN we can calculate phases.

unknown structure

```
MGDKPIWEQIGSSFIQHYYQLFDNDRTQLGAIY
IDASCLTWEGQQFQGKAAIVEKLSSLPFQKIQH
SITAQDHQPTPDSCIISMVVGQLKADEDPIMGF
HQMFLLKNINDAWVCTNDMFRLALHNFG
```

known structure

```
PSPLLVGREFVRQYYTLLNKAPEYLHRFYGRNSSY
VHGGVDASGKPQEAVYGQNDIHHKVLSLNFSECHT
KIRHVDAHATLSDGVVVQVMGLLSNSGQPERKFMQ
TFVLAPEGSVPNKFYVHNDMFRYEDE
```



origin o

H K L	F	ϕ
0 0 1	2.5	30
0 0 2	72.1	85
0 0 3	26.9	310

etc...

origin o

H K L	F	ϕ
0 0 1	10.4	120
0 0 2	3.1	10
0 0 3	52.2	280

etc...

Molecular Replacement

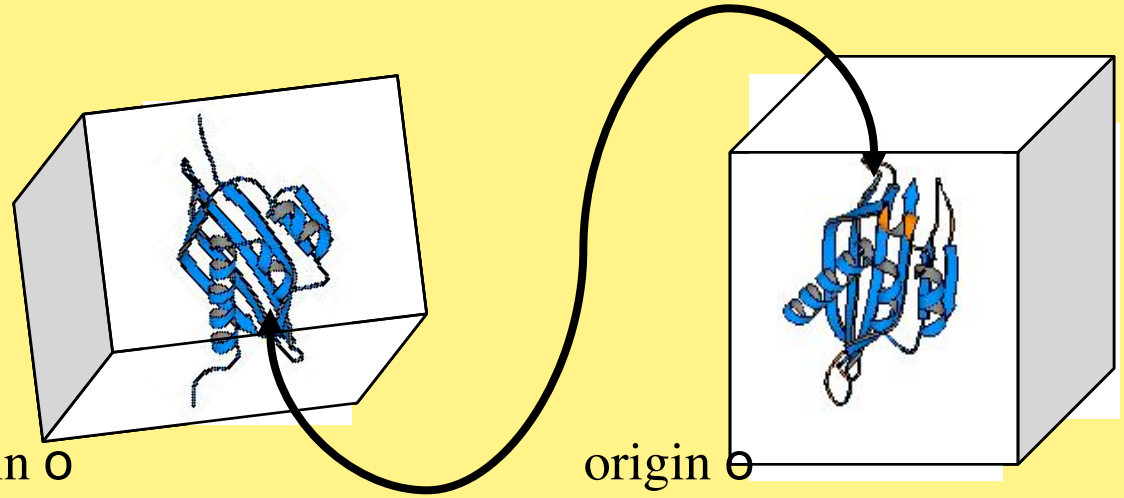
If we can find the rotation and translation that puts the model in the correct position in the crystal cell, THEN we can calculate phases.

unknown structure

```
MGDKPIWEQIGSSFIQHYYQLFDNDRTQLGAIY
IDASCLTWEGQQFQGKAAIVEKLSSLPFQKIQH
SITAQDHQPTPDSCIISMVVGQLKADEDPIMGF
HQMFLLKNINDAWVCTNDMFRLALHNFG
```

known structure

```
PSPLLVGREFVRQYYTLLNKAPEYLHRFYGRNSSY
VHGGVDASGKPQEAVYGQNDIHHKVLSLNFSECHT
KIRHVDAHATLSDGVVVQVMGLLSNSGQPERKFMQ
TFVLAPEGSVPNKFYVHNDMFRYEDE
```



origin o

H	K	L	F	ϕ
0	0	1	2.5	30
0	0	2	72.1	85
0	0	3	26.9	310

etc...

origin o

H	K	L	F	ϕ
0	0	1	10.4	120
0	0	2	3.1	10
0	0	3	52.2	280

etc...

Molecular replacement

place a homologous model into the crystal with unknown structure

or

Atomic Model --> EM map

Molecular replacement

place a homologous model into the crystal with unknown structure

or

Atomic Model --> EM map

1) 6 - dimensional search
   check all orientations and positions

Molecular replacement

place a homologous model into the crystal with unknown structure

or

Atomic Model --> EM map

1) 6 - dimensional search

check all orientations and positions

2) 3-d + 3-d search

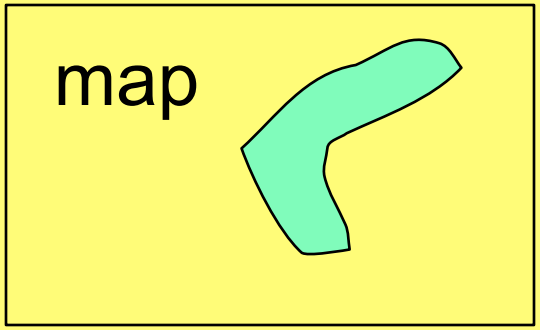
orientations positions

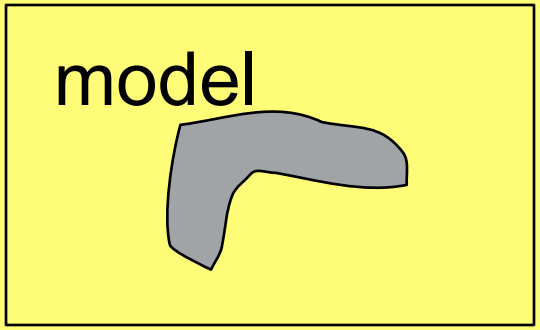
Conventional Molecular Replacement

Functions of molecular replacement

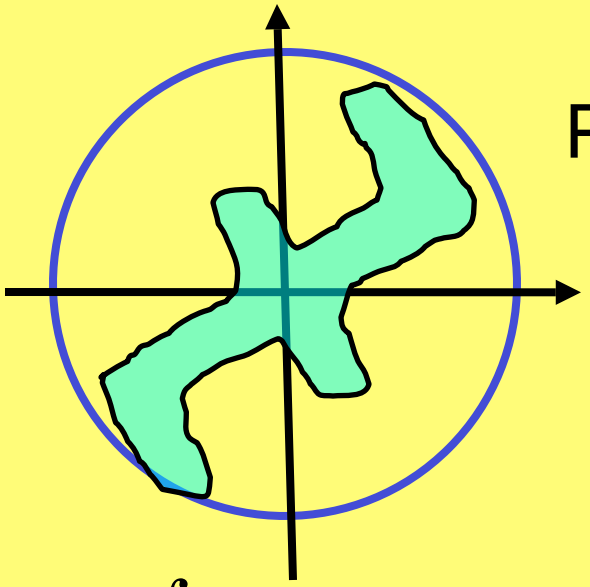
- Cross Rotation function
- Self Rotation function
- Translation function
- Phased Translation function
- Fast Packing function

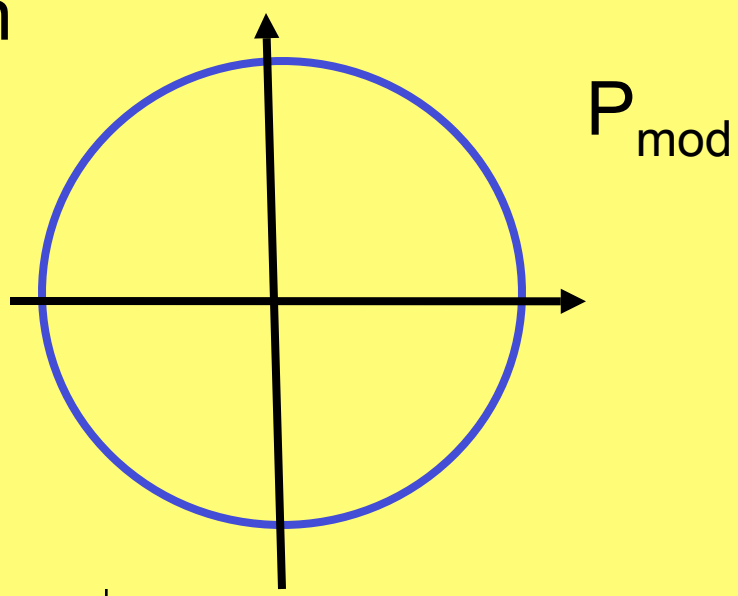
Cross Rotation Function





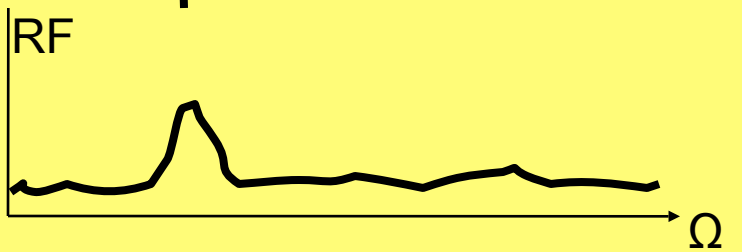
Patterson



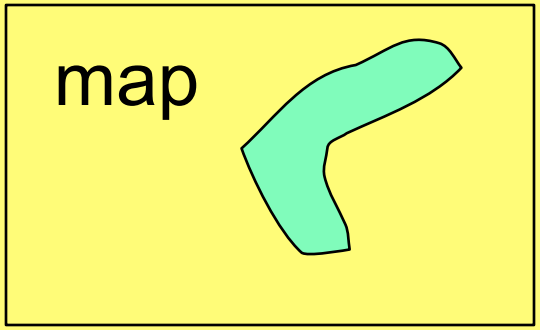


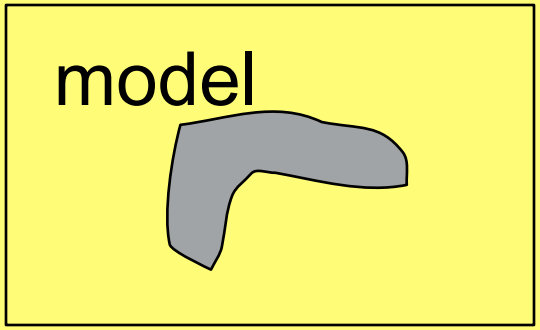
$$RF(\Omega) = \int P_{obs}(r)\ \Re_{\Omega}\{ P_{mod}(r)\}\ dr$$

$\Re_{\Omega}$ - rotation operator

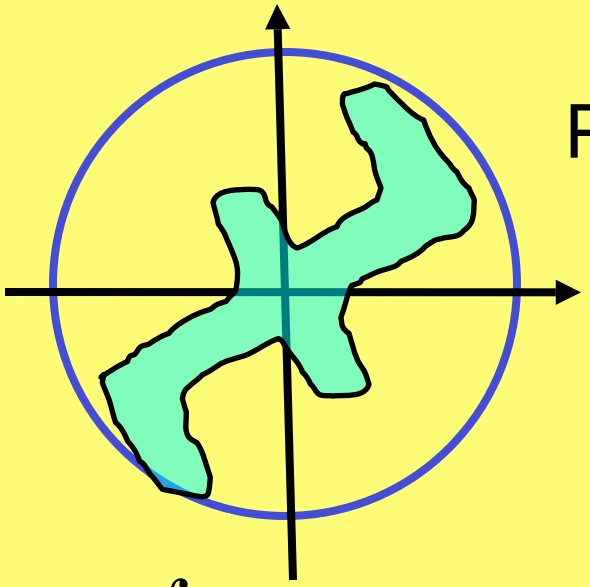


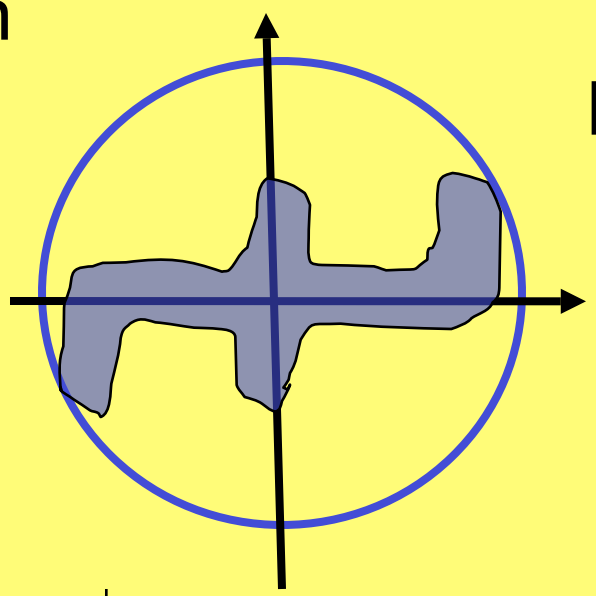
Cross Rotation Function





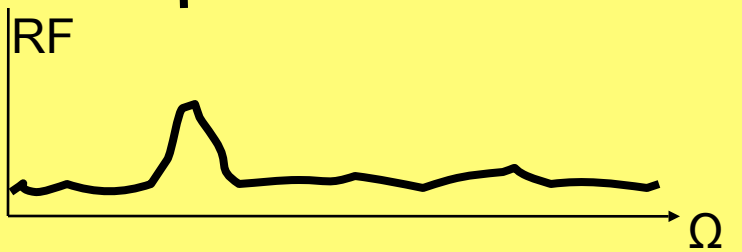
Patterson



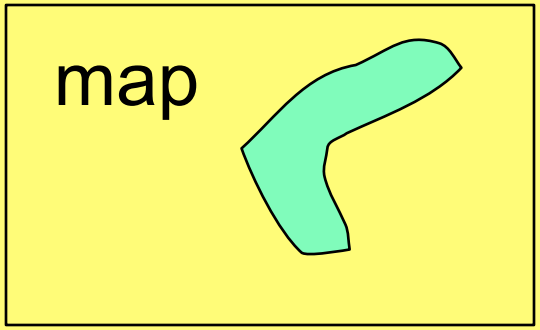


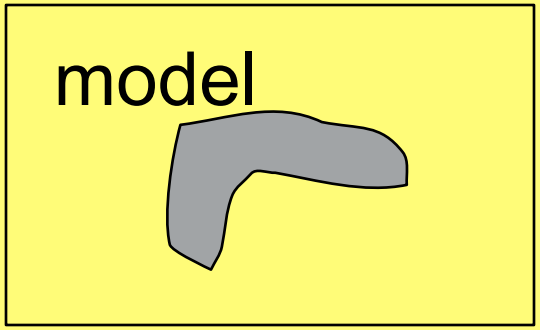
$$RF(\Omega) = \int P_{obs}(r)\ \Re_{\Omega}\{ P_{mod}(r)\}\ dr$$

$\Re_{\Omega}$ - rotation operator

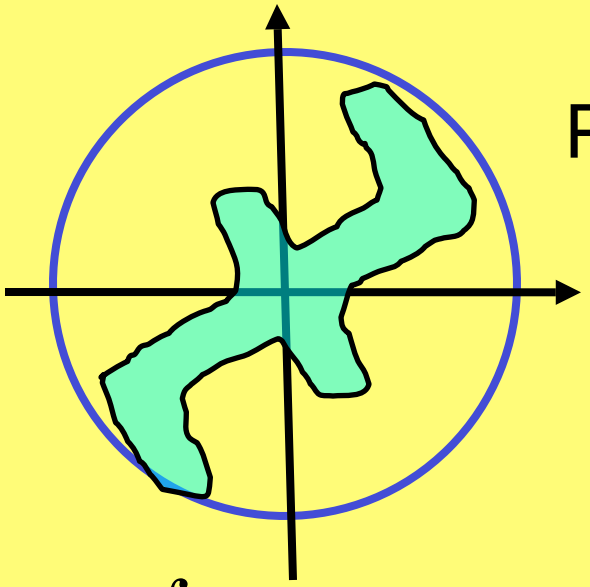


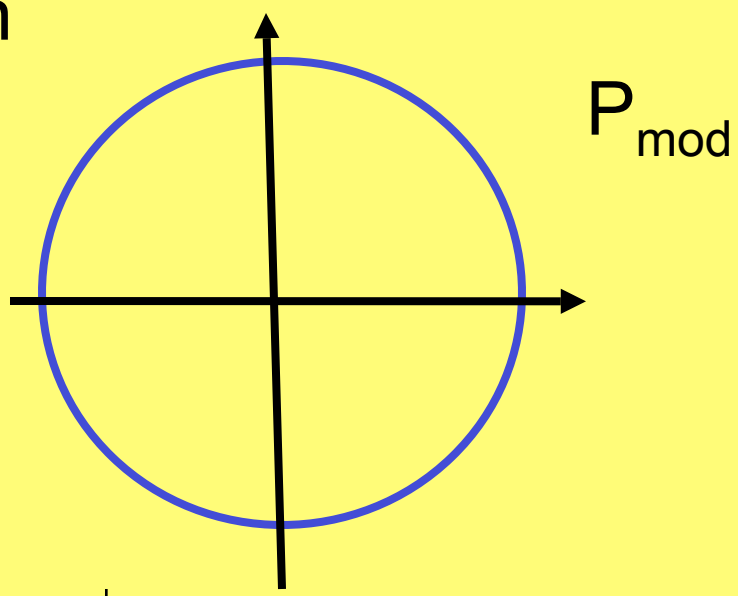
Cross Rotation Function





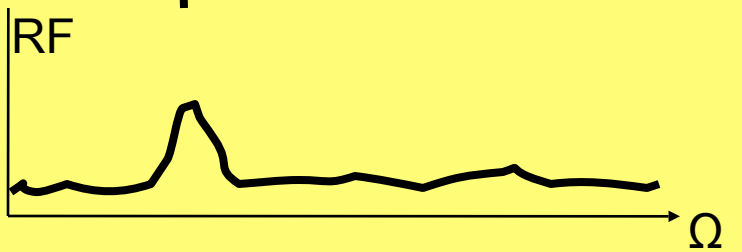
Patterson



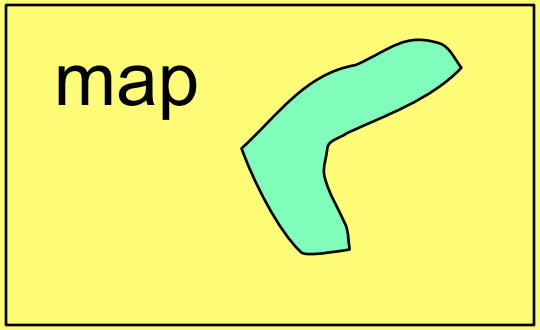


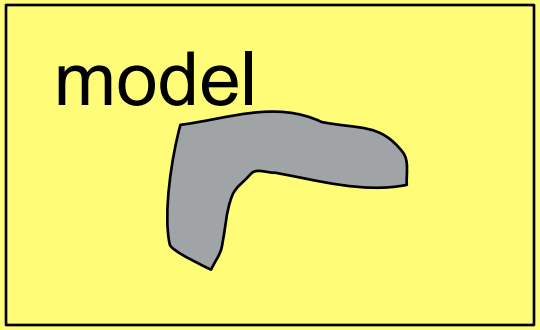
$$RF(\Omega) = \int P_{obs}(r)\ \Re_{\Omega}\{ P_{mod}(r)\}\ dr$$

$\Re_{\Omega}$ - rotation operator

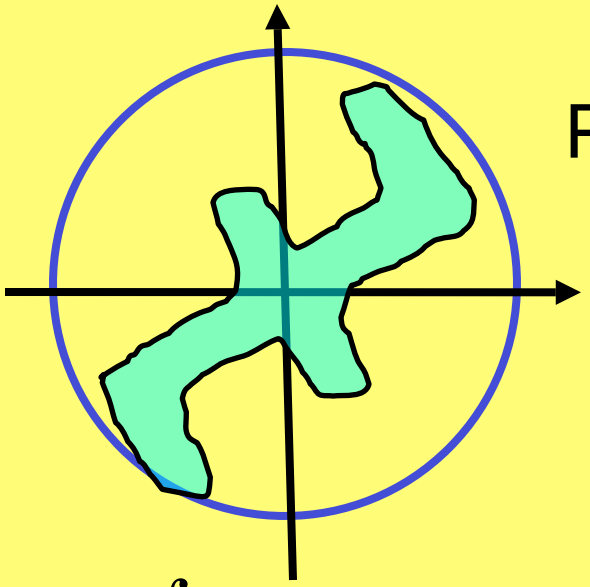


Cross Rotation Function

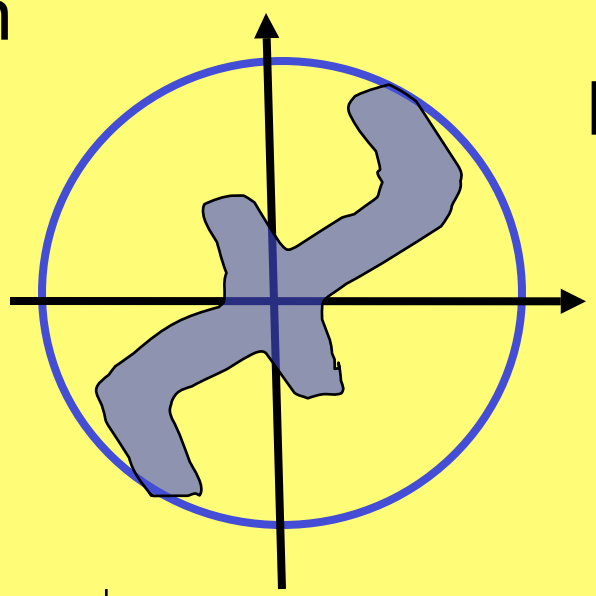




Patterson



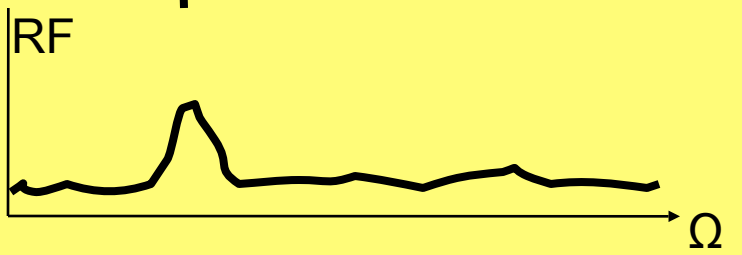
P_{obs}



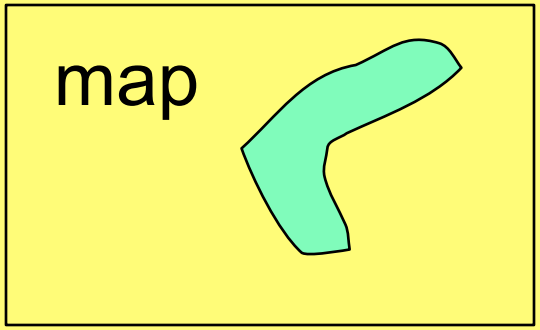
P_{mod}

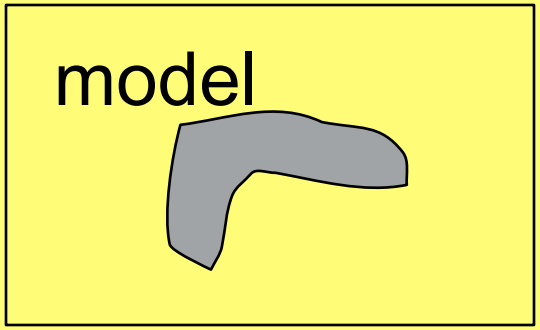
$$RF(\Omega) = \int P_{obs}(r)\ \Re_{\Omega}\{ P_{mod}(r)\}\ dr$$

$\Re_{\Omega}$ - rotation operator

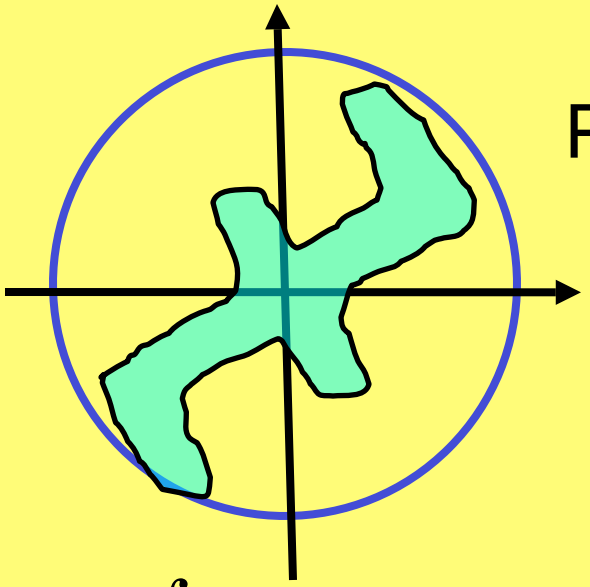


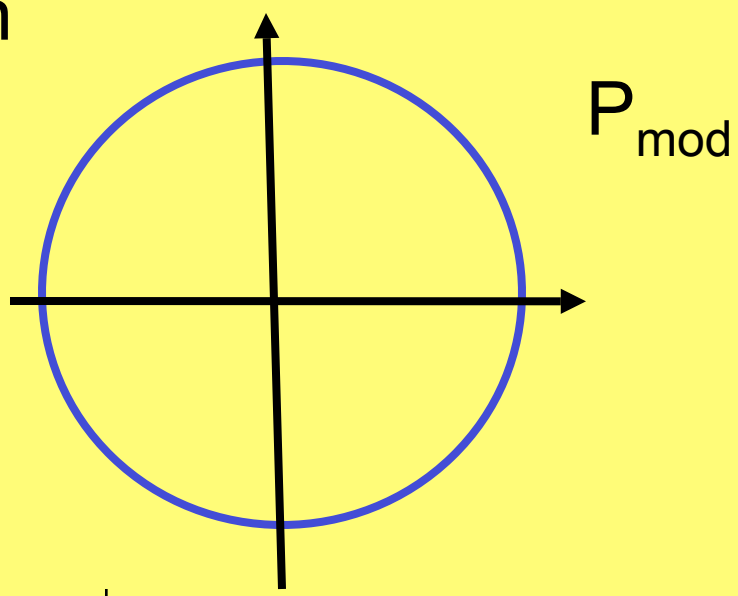
Cross Rotation Function





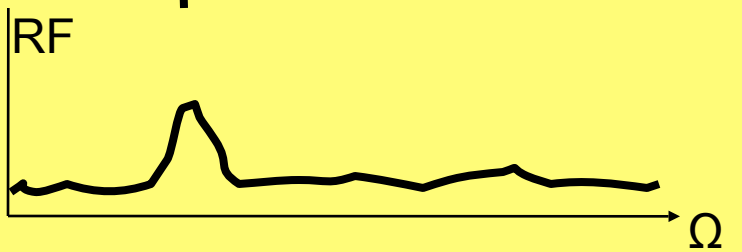
Patterson



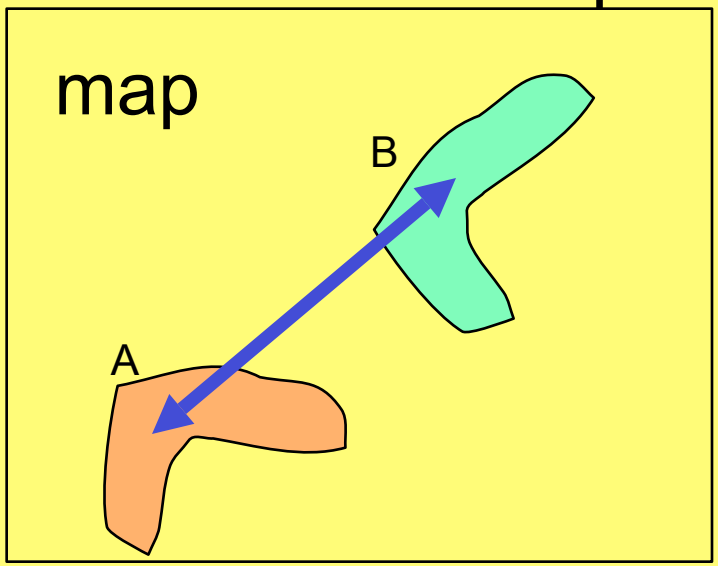


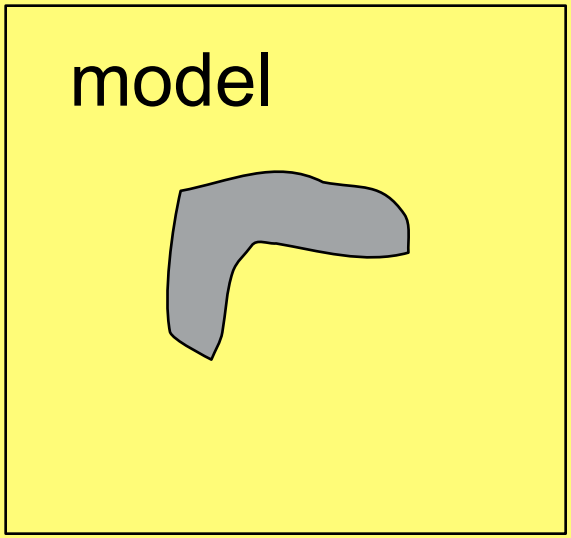
$$RF(\Omega) = \int P_{obs}(r)\ \Re_{\Omega}\{ P_{mod}(r)\}\ dr$$

$\Re_{\Omega}$ - rotation operator

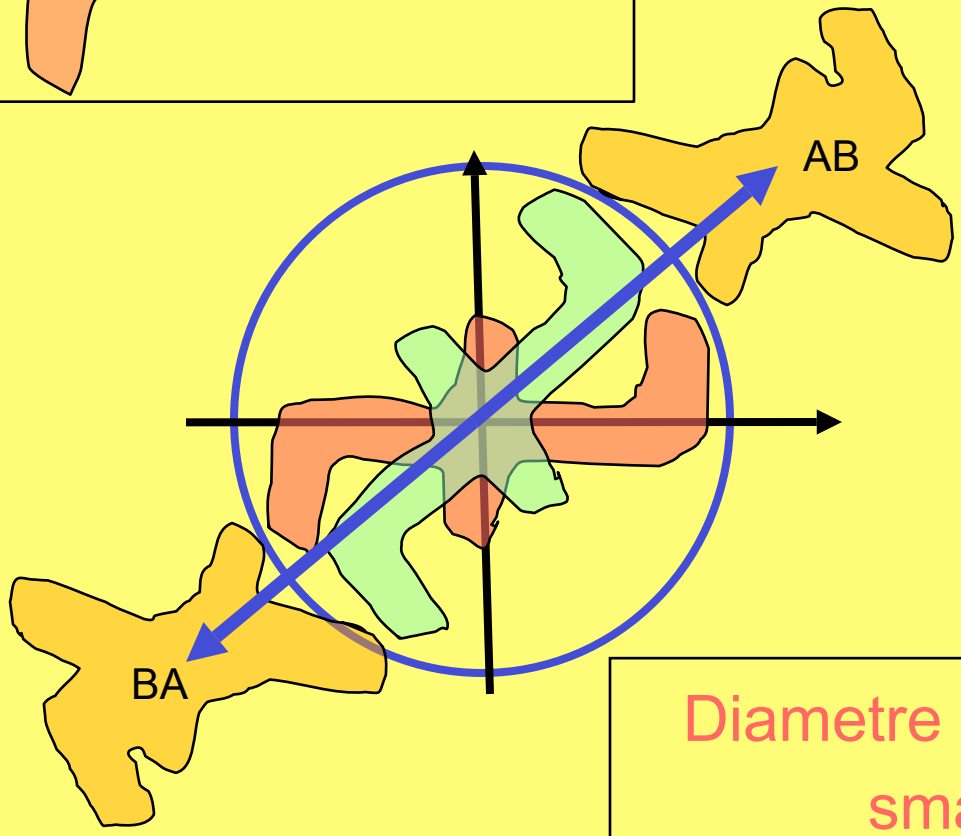


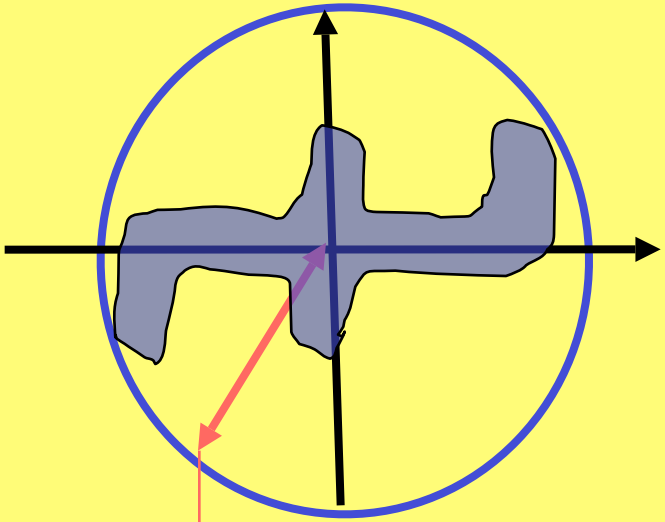
Optimal radius of integration





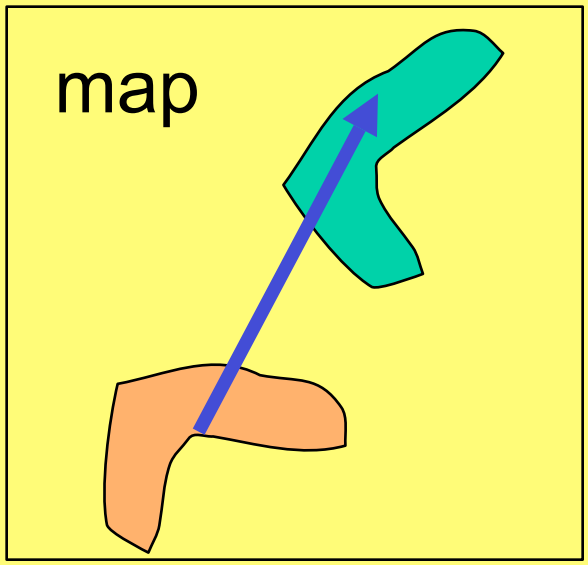
Vinter



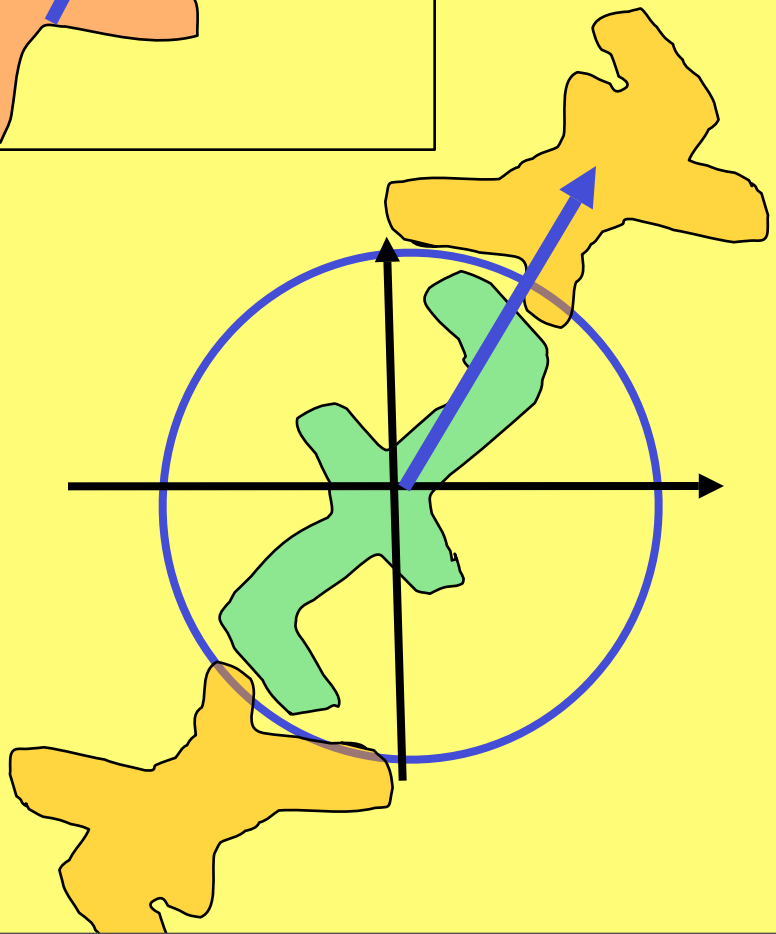


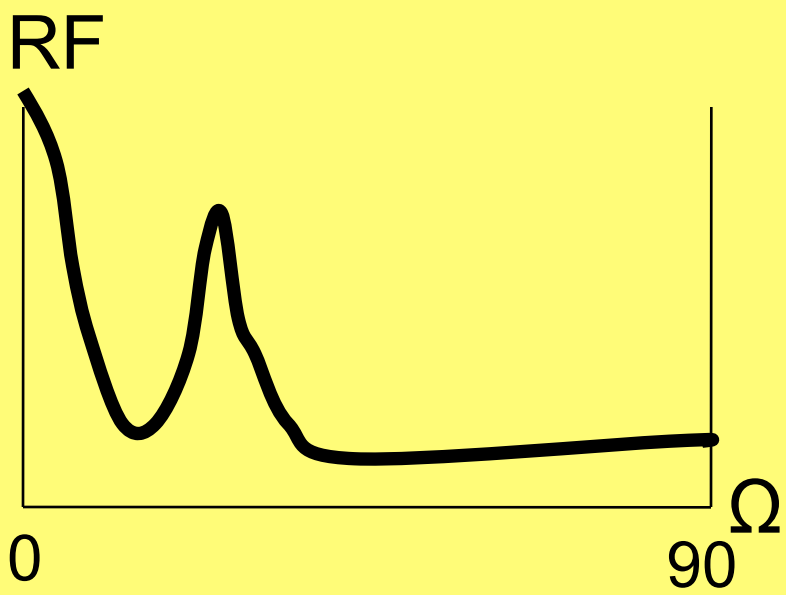
Diametre of the model and less than smallest cell dimension

Self Rotation Function

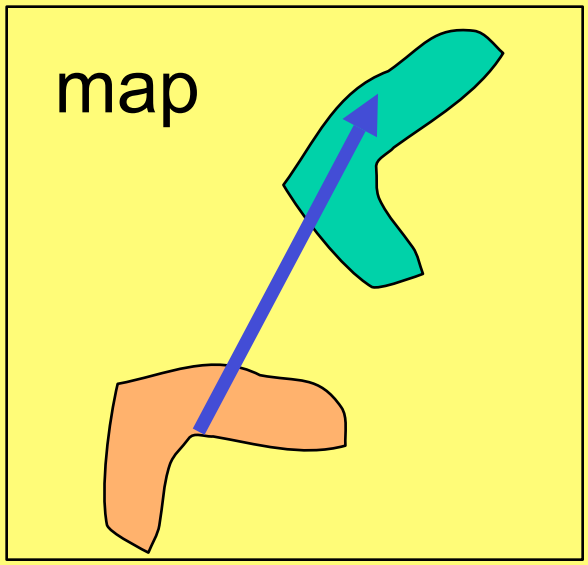


$$RF(\Omega) = \int P_{obs}(r)\ \ \Re_{\Omega}\{ P_{obs}(r)\}\ \ dr$$

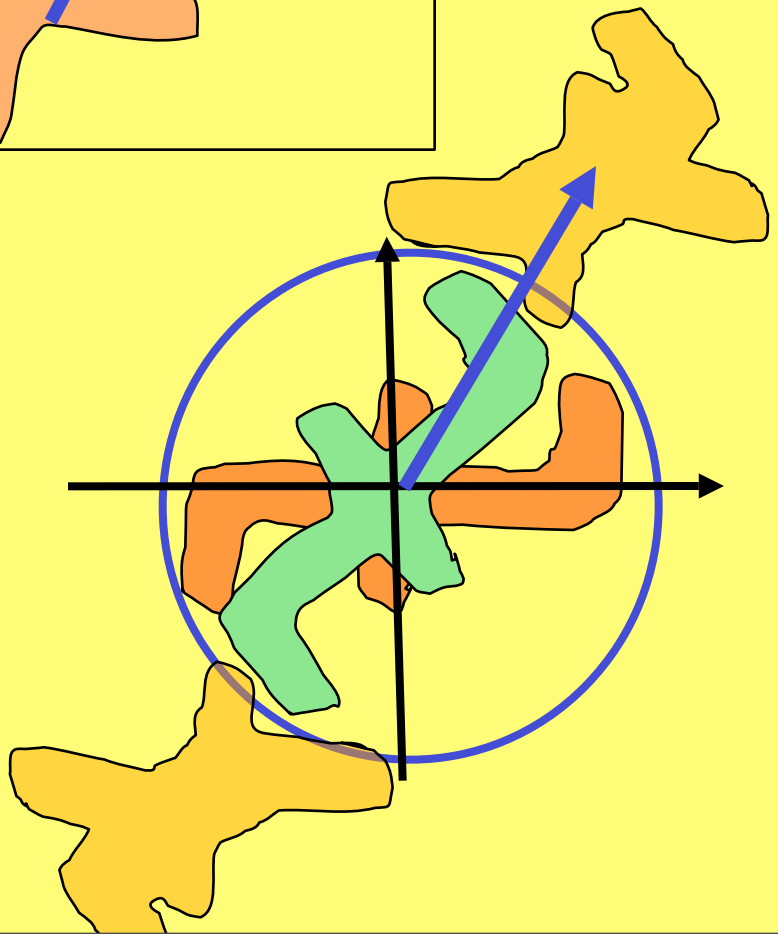


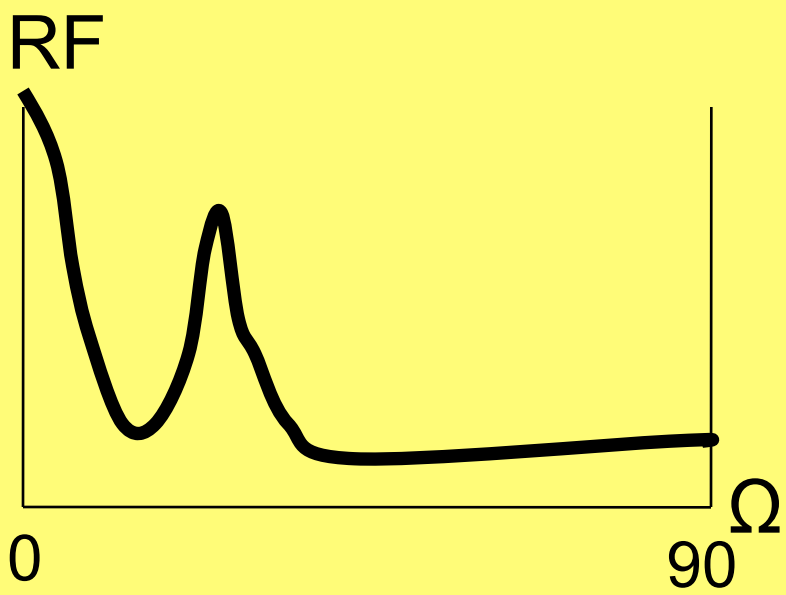


Self Rotation Function

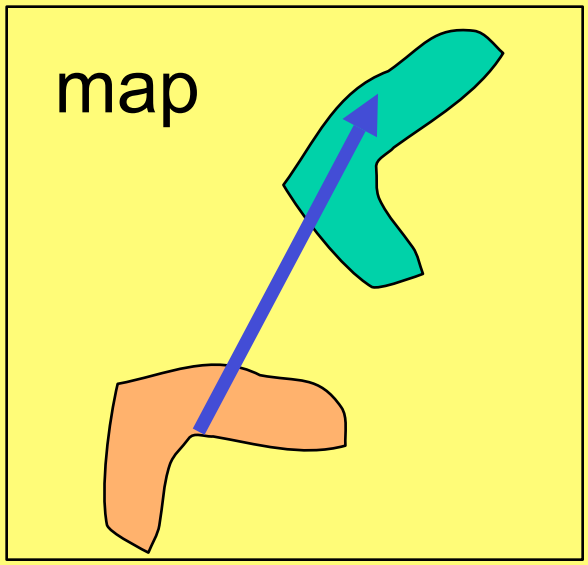


$$RF(\Omega) = \int P_{obs}(r)\ \ \Re_{\Omega}\{ P_{obs}(r)\}\ \ dr$$

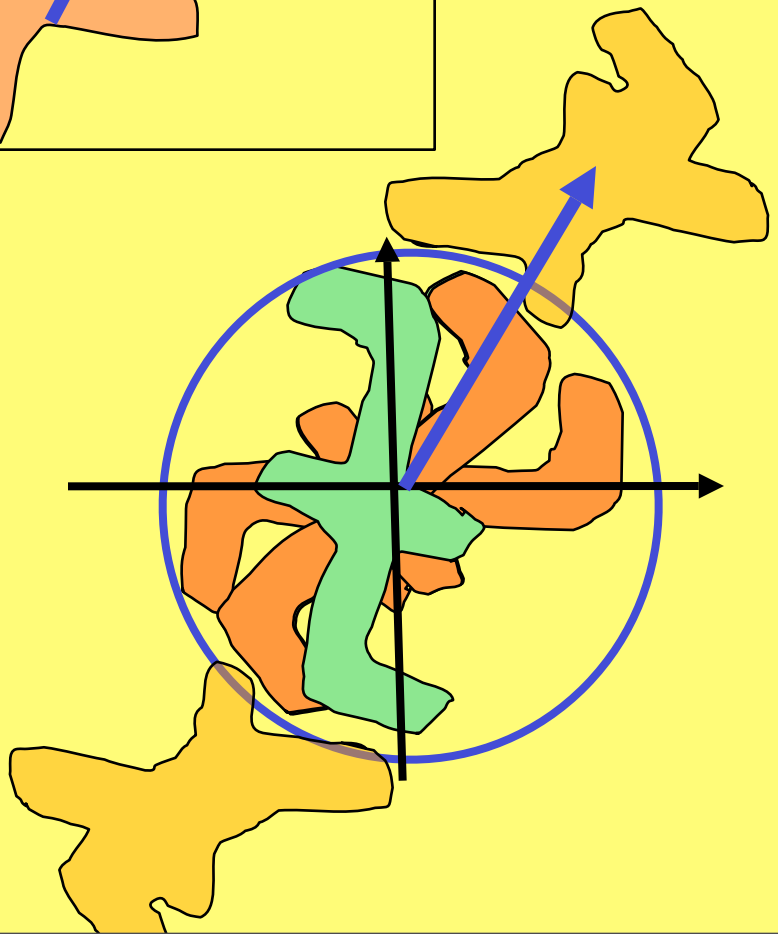


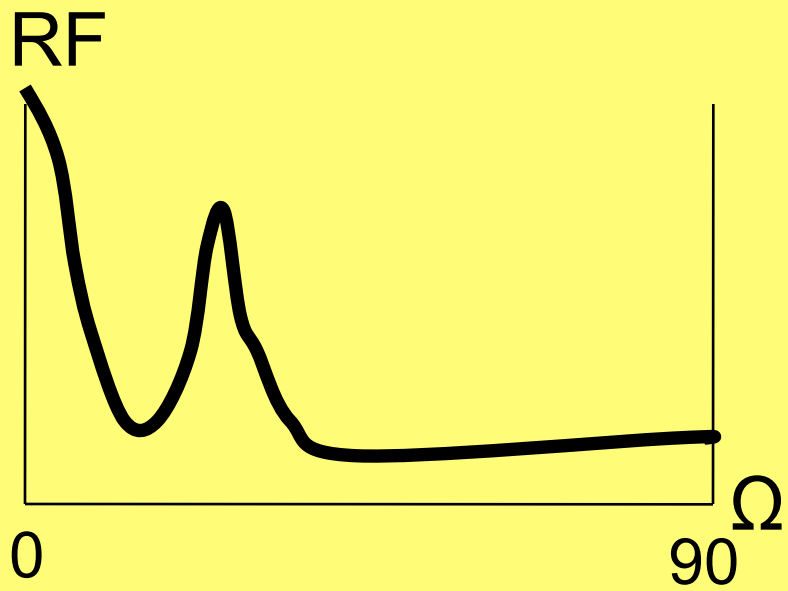


Self Rotation Function

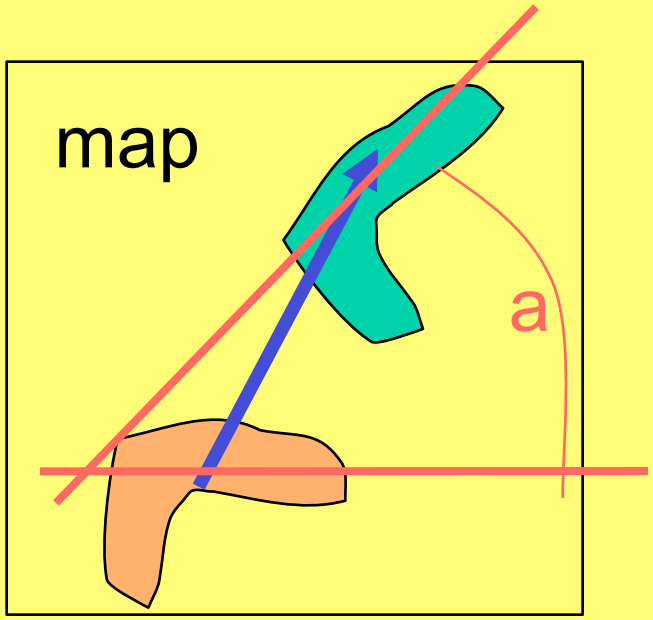


$$RF(\Omega) = \int P_{obs}(r)\ \Re_{\Omega}\{ P_{obs}(r)\}\ dr$$

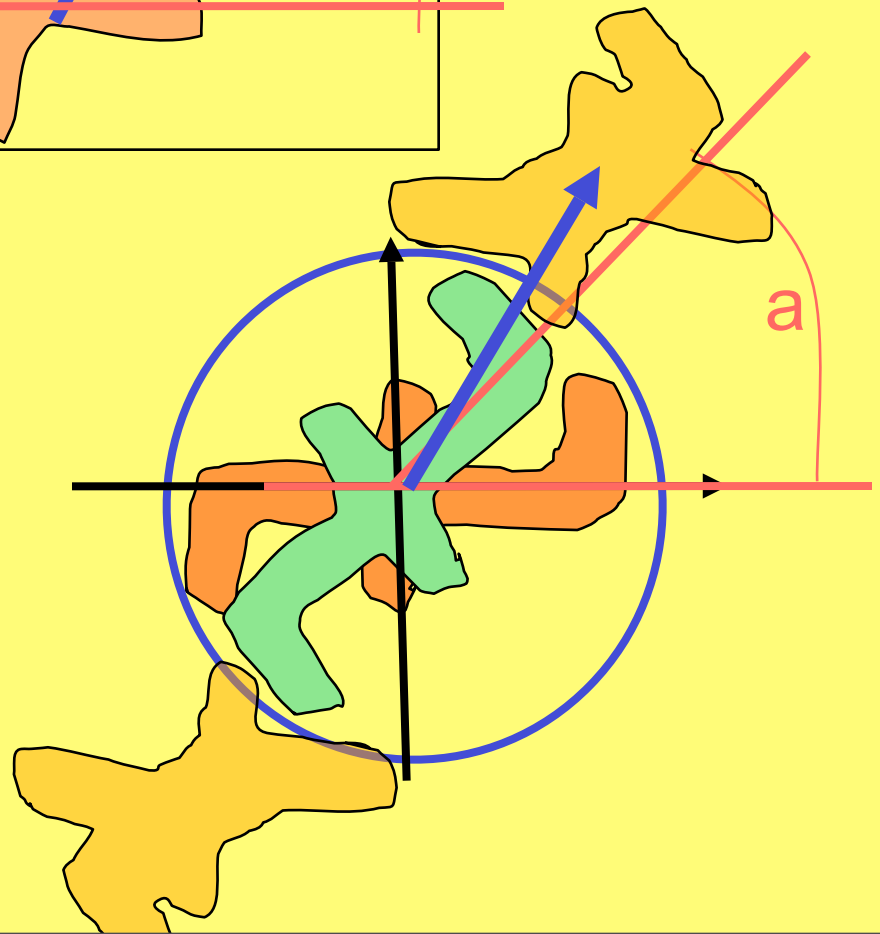


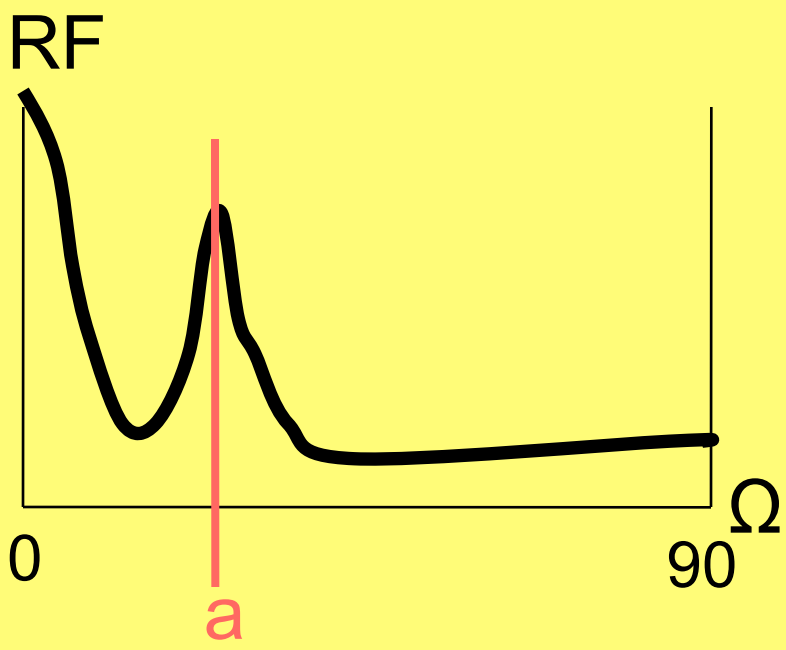


Self Rotation Function

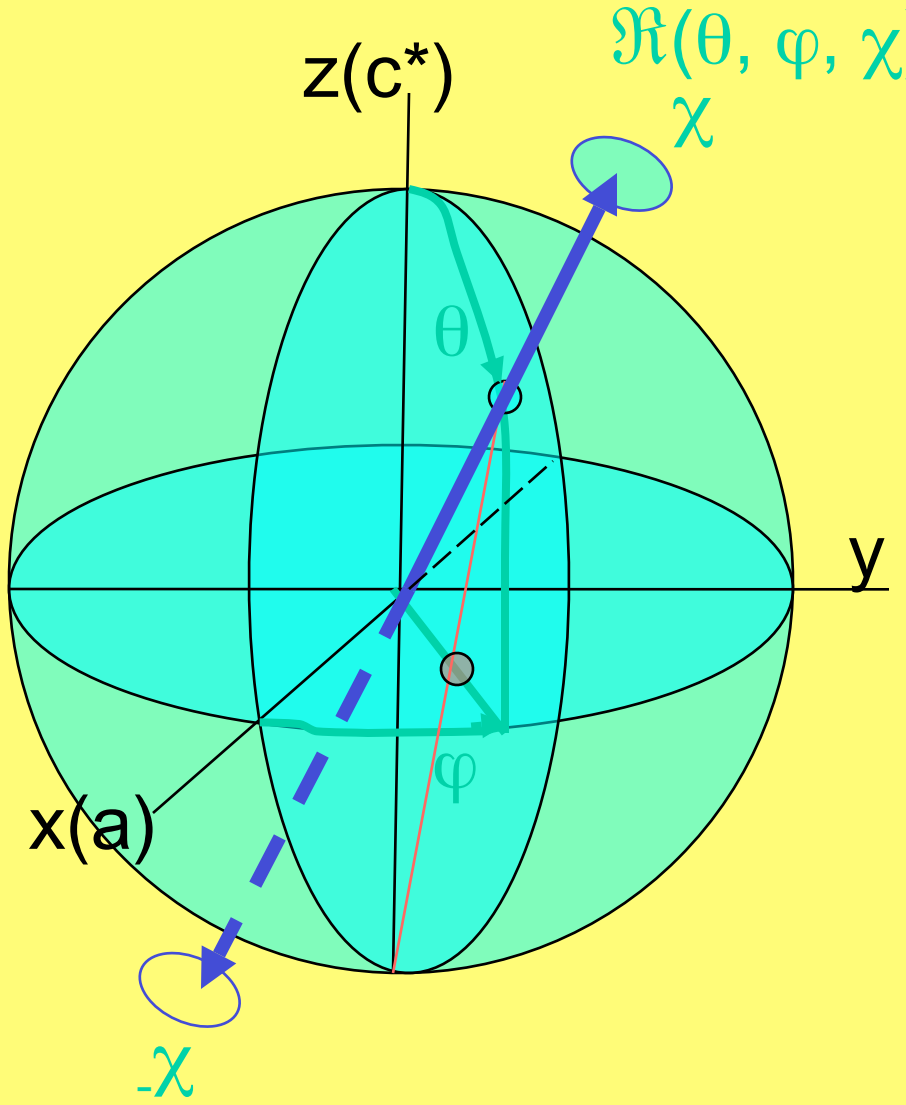


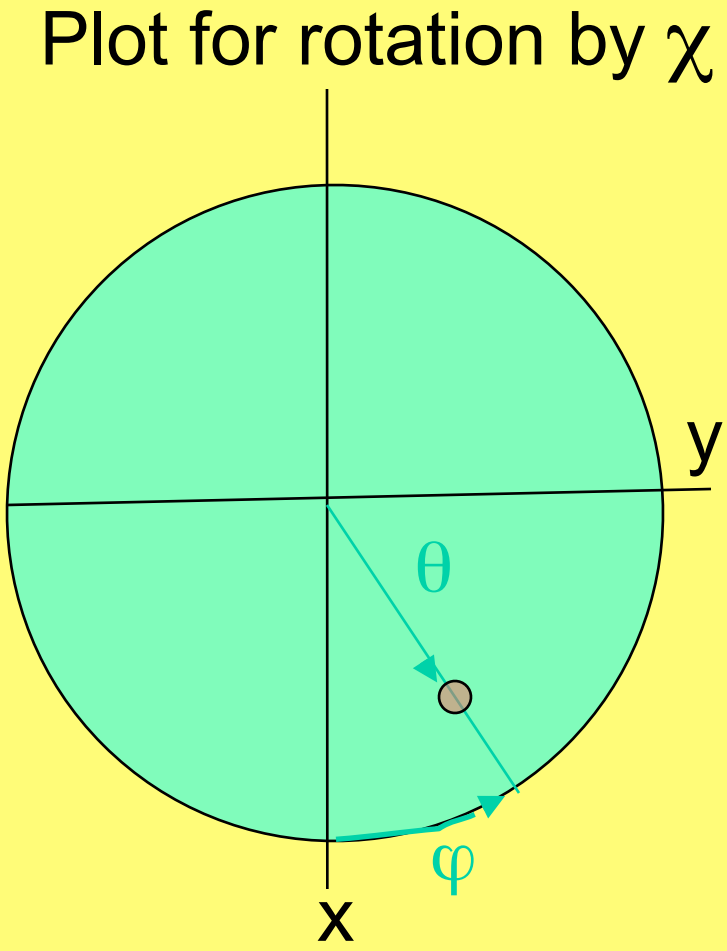
$$RF(\Omega) = \int P_{obs}(r)\ \Re_{\Omega}\{ P_{obs}(r)\}\ dr$$





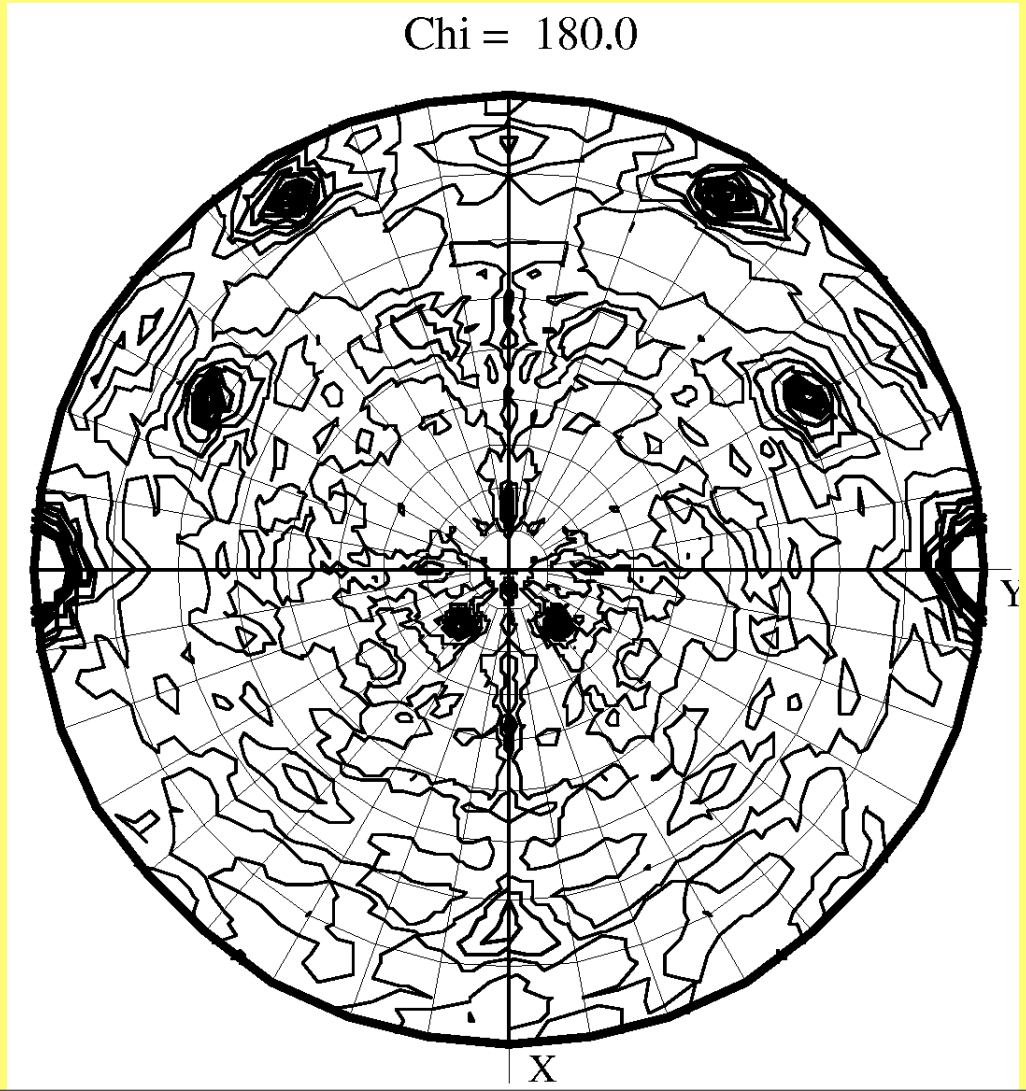
Stereographic projection





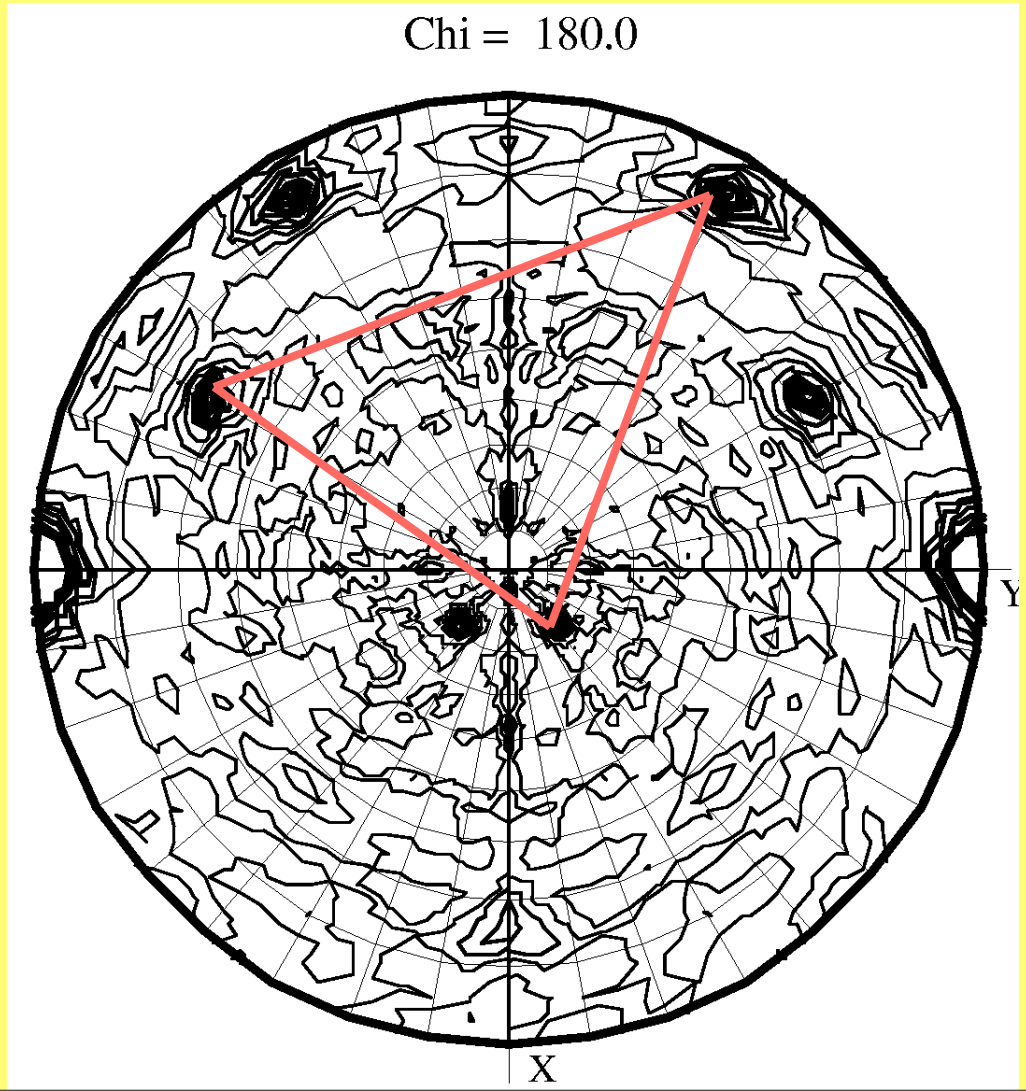
Self Rotation Function

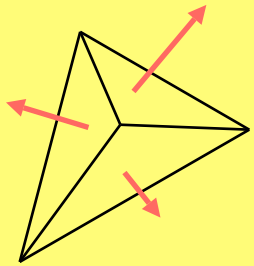
Space group $P2_1$ one tetramer



Self Rotation Function

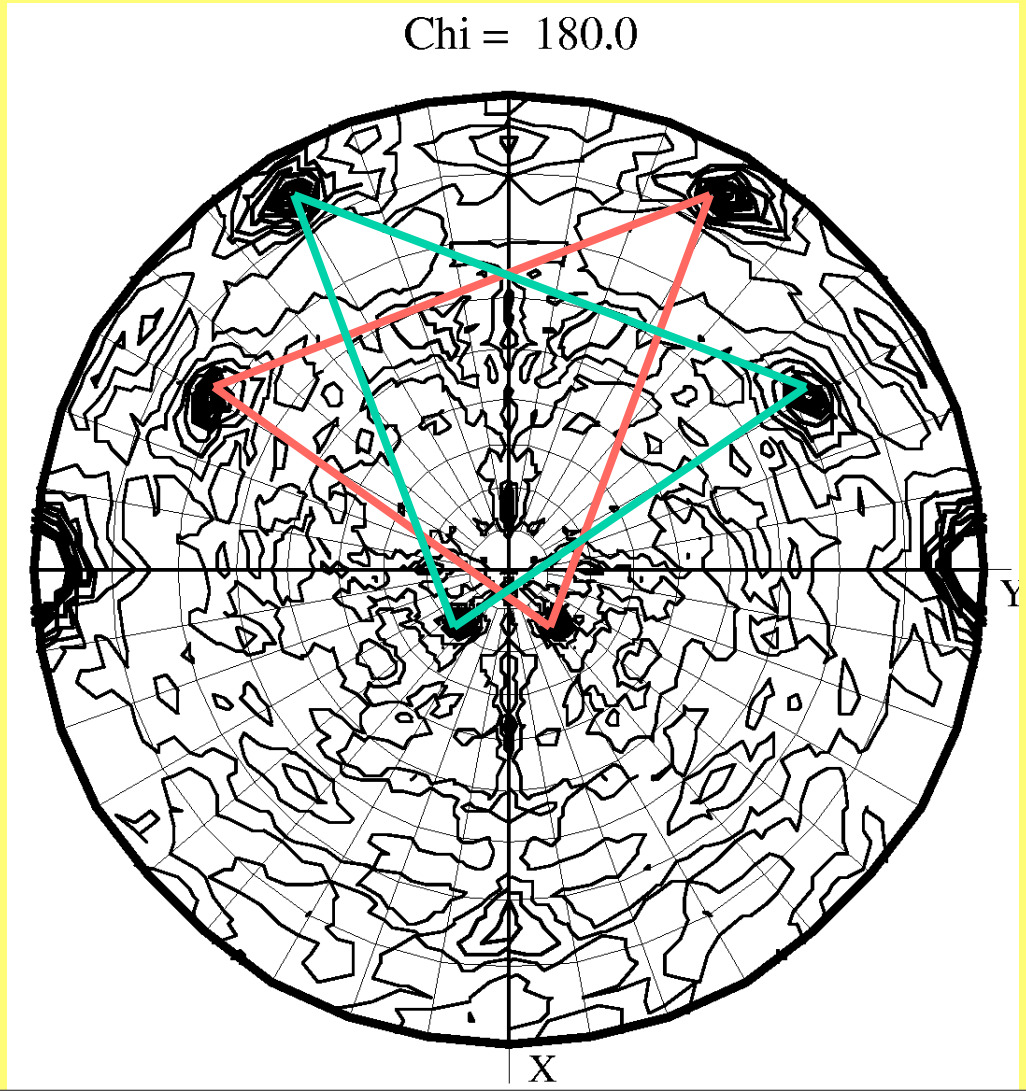
Space group $P2_1$ one tetramer

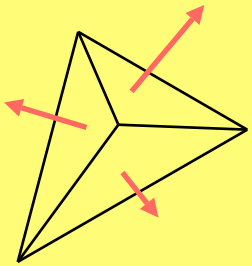


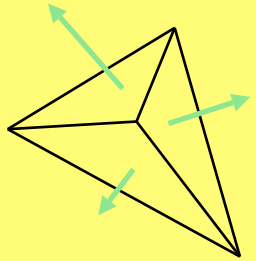


Self Rotation Function

Space group $P2_1$ one tetramer







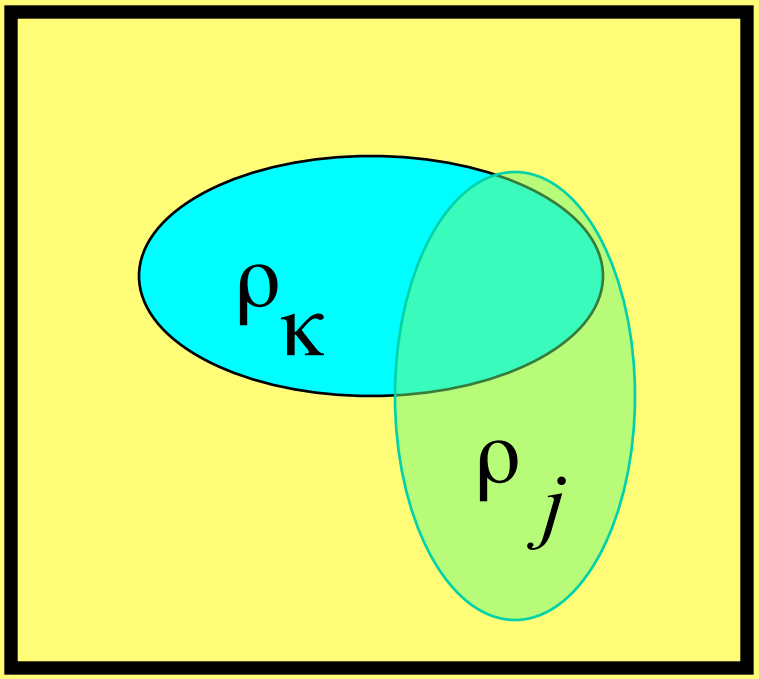
Translation Function

To find relative position of molecules again Patterson function is used. “Correctly oriented” molecules are shifted to position r, corresponding Patterson is calculated and it is compared with observed Patterson. Maximum correspondence between two Pattersons indicate potentially correct position.

$$TF(s) = \int P_{obs}(r)\ P_{calc}(s,r)\ dr$$

s - vector of translation

Fast Packing Function



Estimation of overlap:

$$\int \rho_k(\mathbf{r},\mathbf{s})\rho_j(\mathbf{r},\mathbf{s})d\mathbf{r}$$

Packing function:

$$P(\mathbf{s}) = \mathbf{1} - \sum_k \sum_j \int \rho_k(\mathbf{r},\mathbf{s})\rho_j(\mathbf{r},\mathbf{s})d\mathbf{r}$$

K # j

Questions

How to use X-ray data

- Maximum resolution limit ?
- Minimal resolution limit ?
- Weighting scheme ?

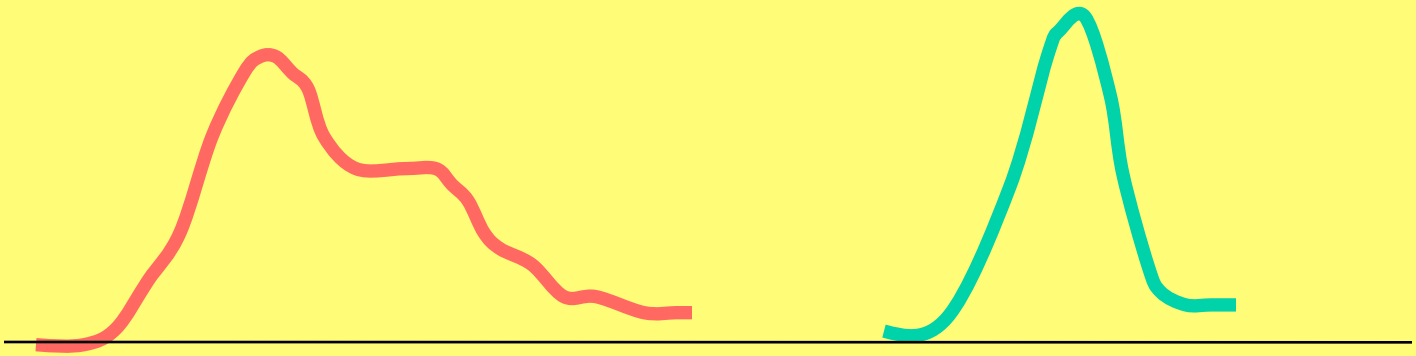
Short introduction to Fourier Transformation

Operators:

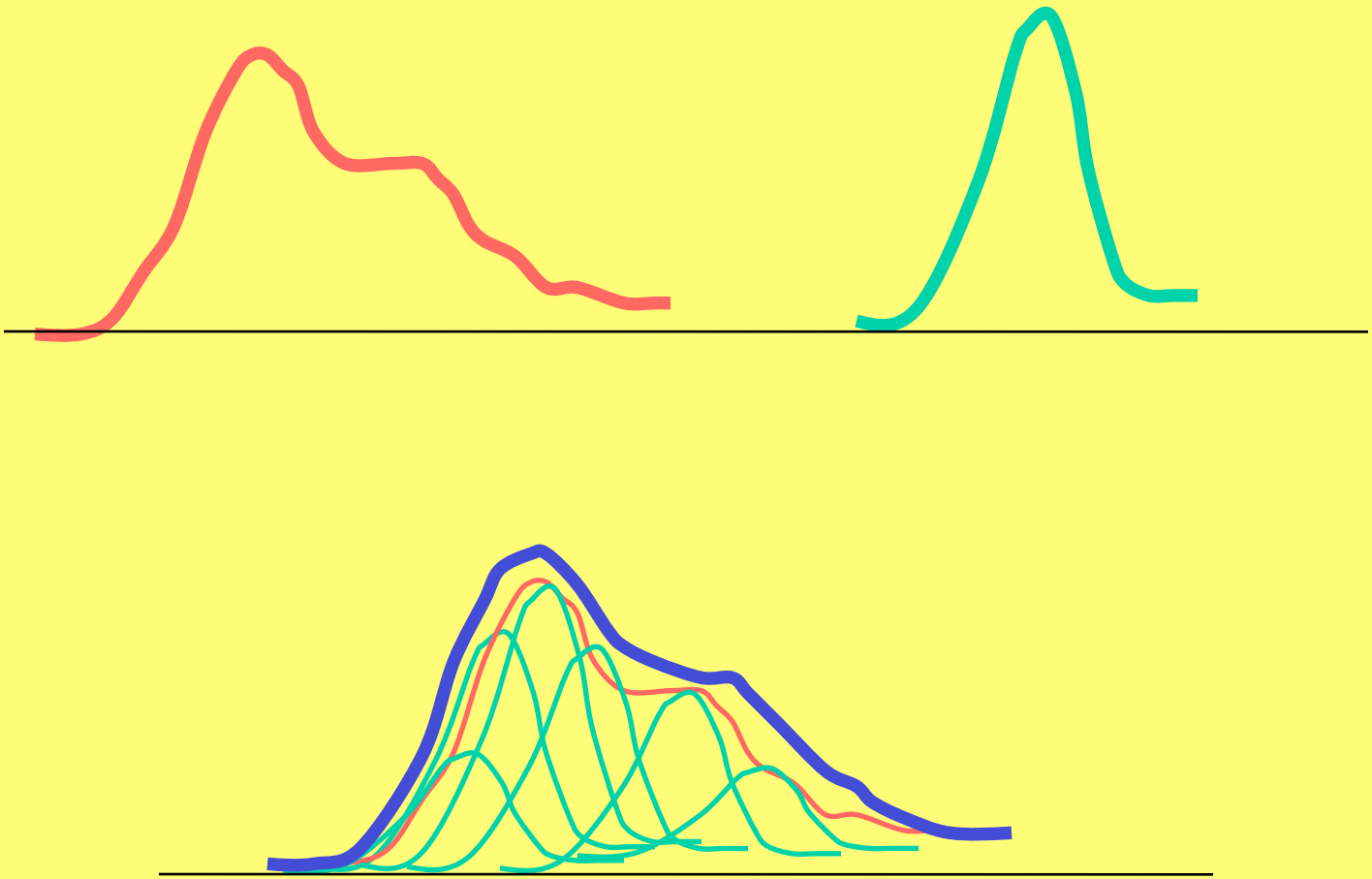
addition $\leftarrow \mathcal{F} \rightarrow$ addition

convolution $\leftarrow \mathcal{F} \rightarrow$ product

Convolution

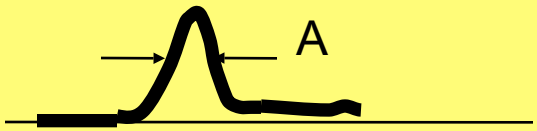


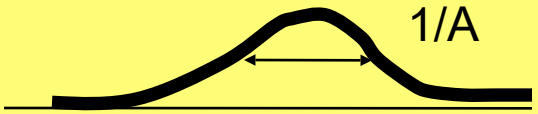
Convolution

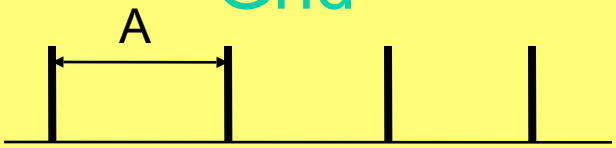


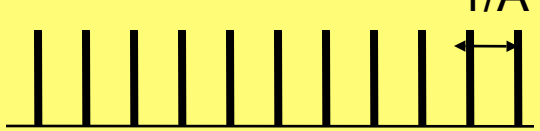
Functions

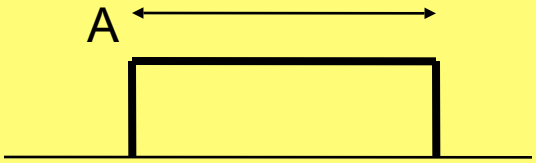
Real space		Reciprocal space
Real function	$\leftarrow \mathcal{F} \rightarrow$	Complex function
Gaussian	$\leftarrow \mathcal{F} \rightarrow$	Gaussian
Grid	$\leftarrow \mathcal{F} \rightarrow$	Grid
Step function	$\leftarrow \mathcal{F} \rightarrow$	Interference function

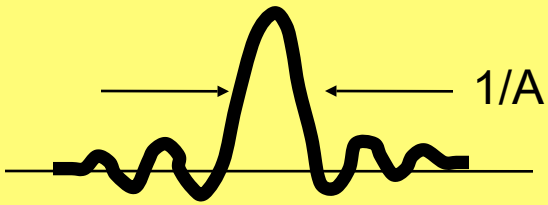






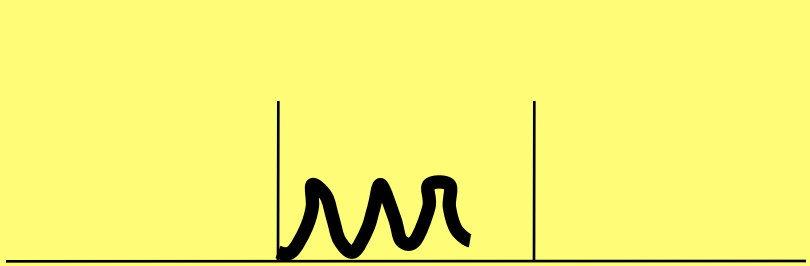


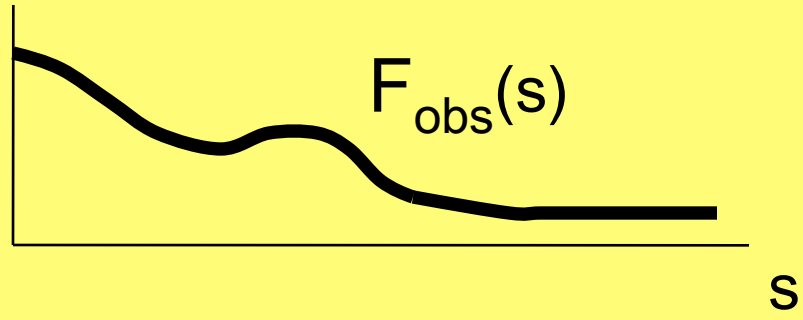


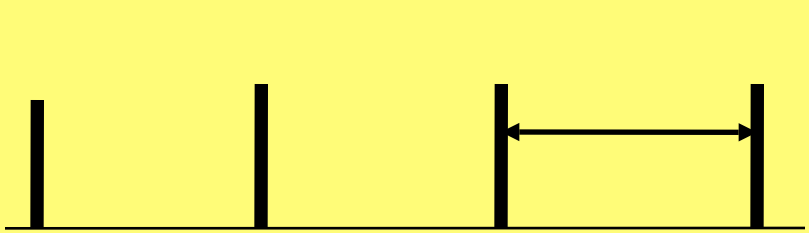


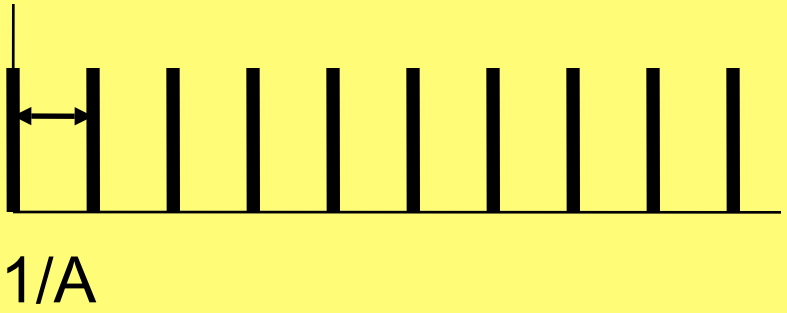
Crystal and Structure Factors

Real space ← $\mathcal{F}$ → Reciprocal space

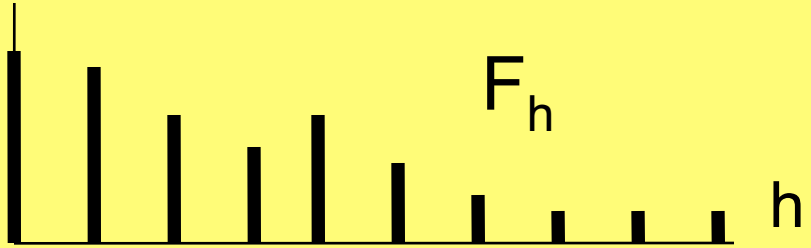


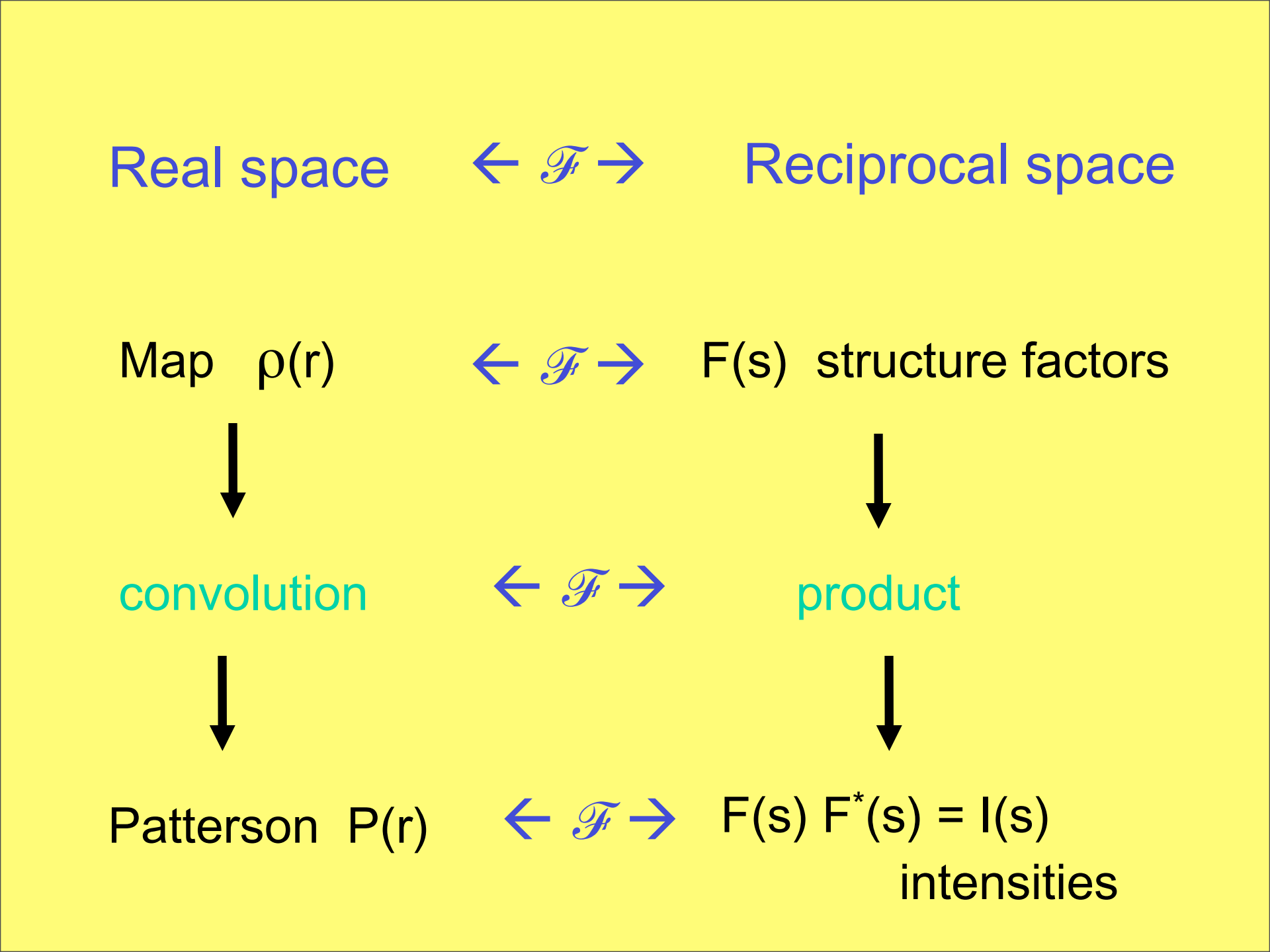






Convolution ← $\mathcal{F}$ → Product



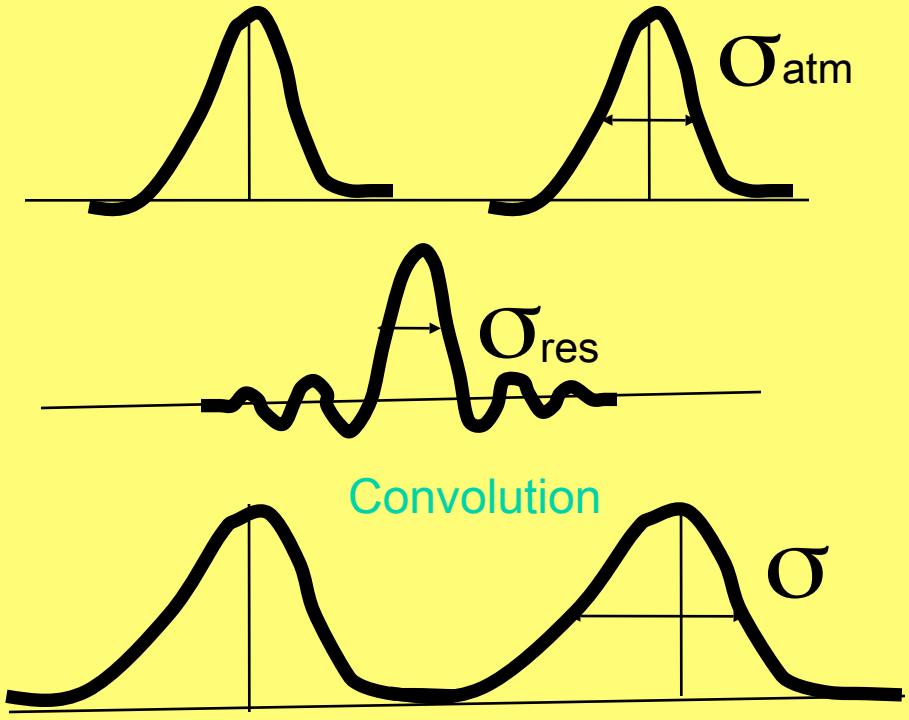

Real space
← 𝓕 →
Reciprocal space
Map ρ(r)
← 𝓕 →
F(s) structure factors
convolution
← 𝓕 →
product
Patterson P(r)
← 𝓕 →
F(s) F*(s) = I(s)
intensities

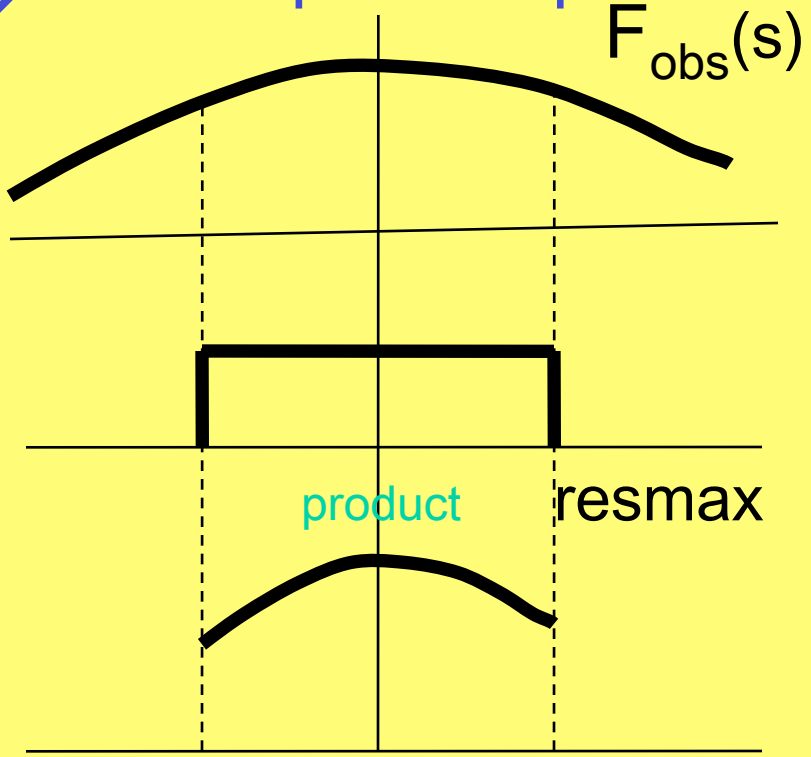
High resolution data

- High resolution limit from Optical resolution
- Weights for high resolution data

Optical resolution

Real space ← $\mathcal{F}$ → Reciprocal space



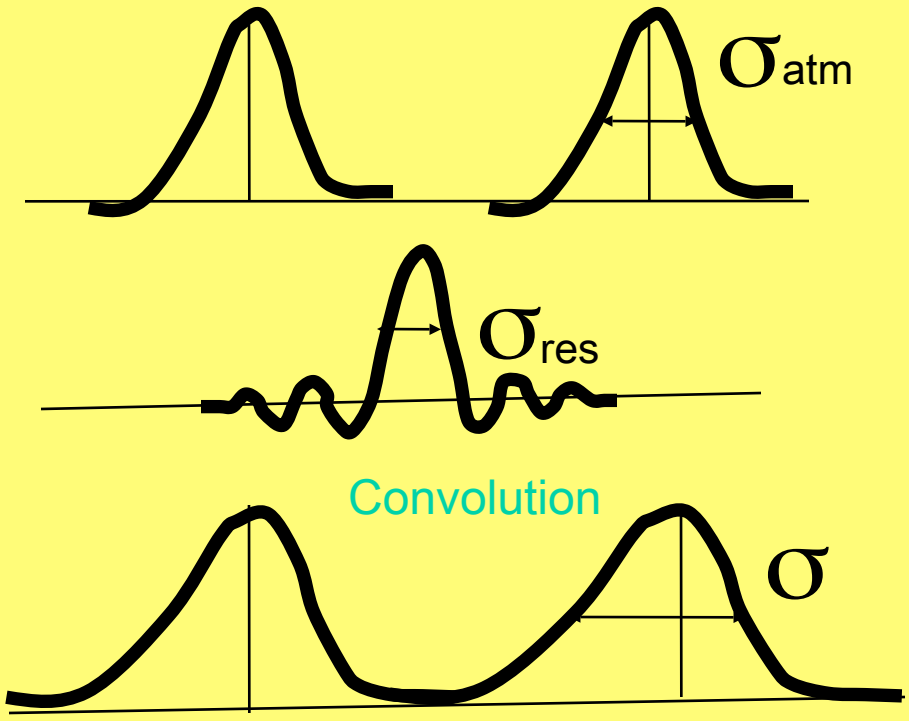


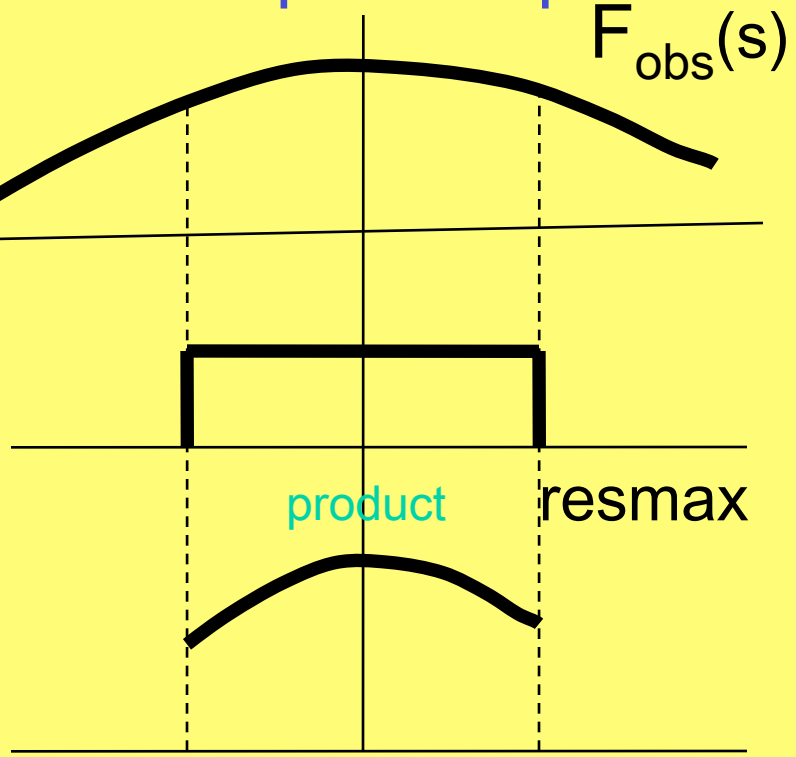
$$\sigma_{res} = 0.356\ resmax$$

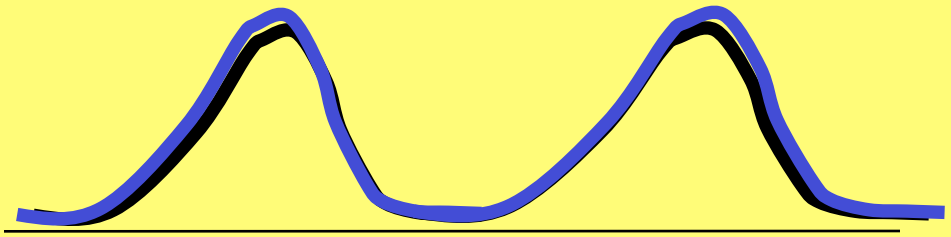
$$\sigma^2 = \sigma_{atm}^2 + \sigma_{res}^2$$

Optical resolution

Real space ← $\mathcal{F}$ → Reciprocal space





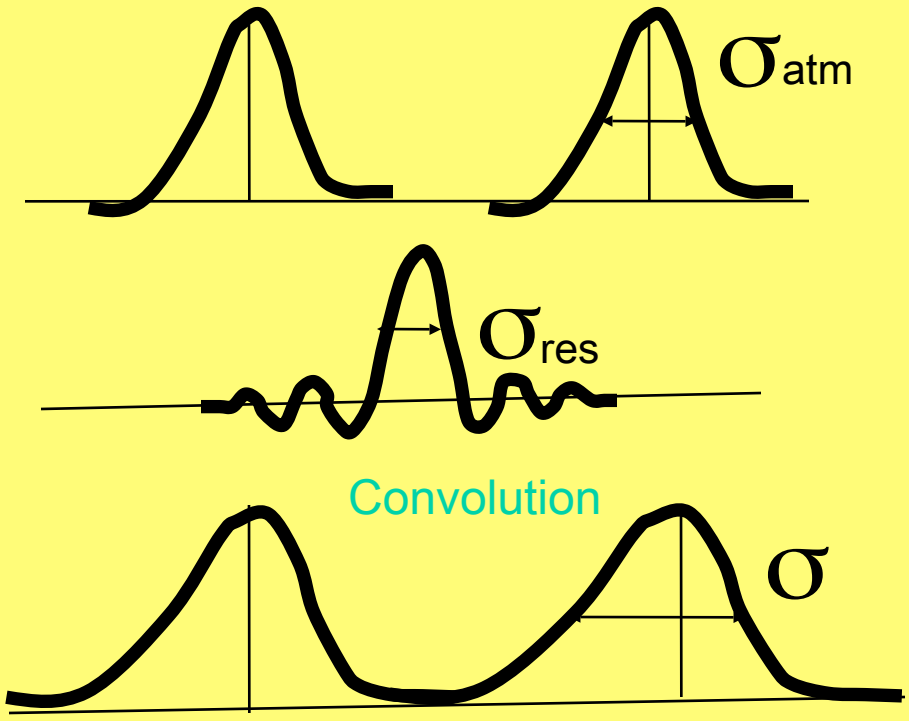


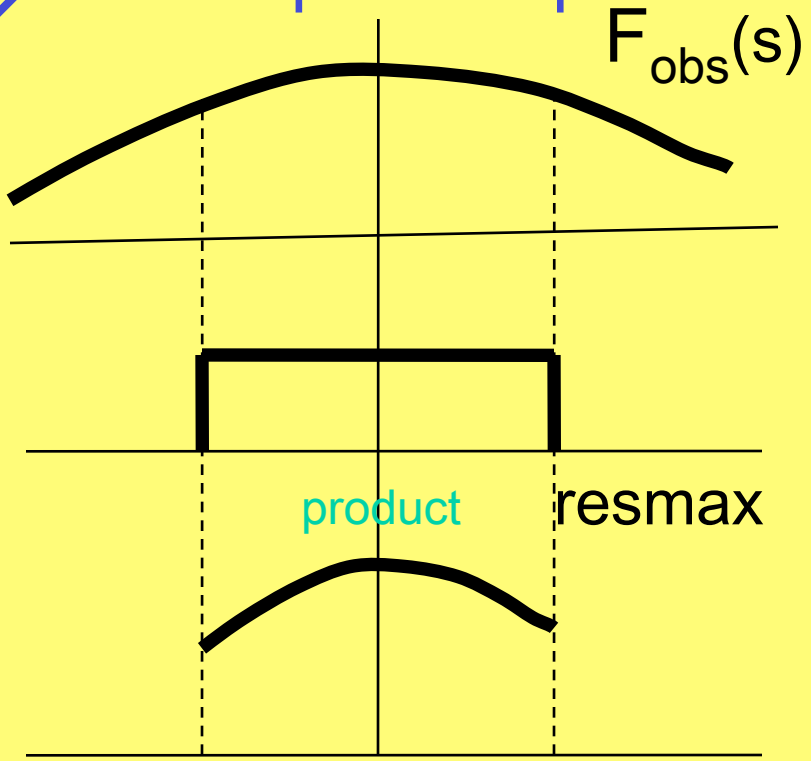
$$\sigma_{res} = 0.356\ \text{resmax}$$

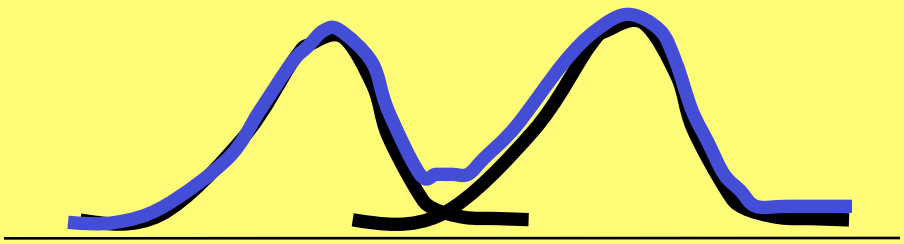
$$\sigma^2 = \sigma_{atm}^2 + \sigma_{res}^2$$

Optical resolution

Real space ← $\mathcal{F}$ → Reciprocal space





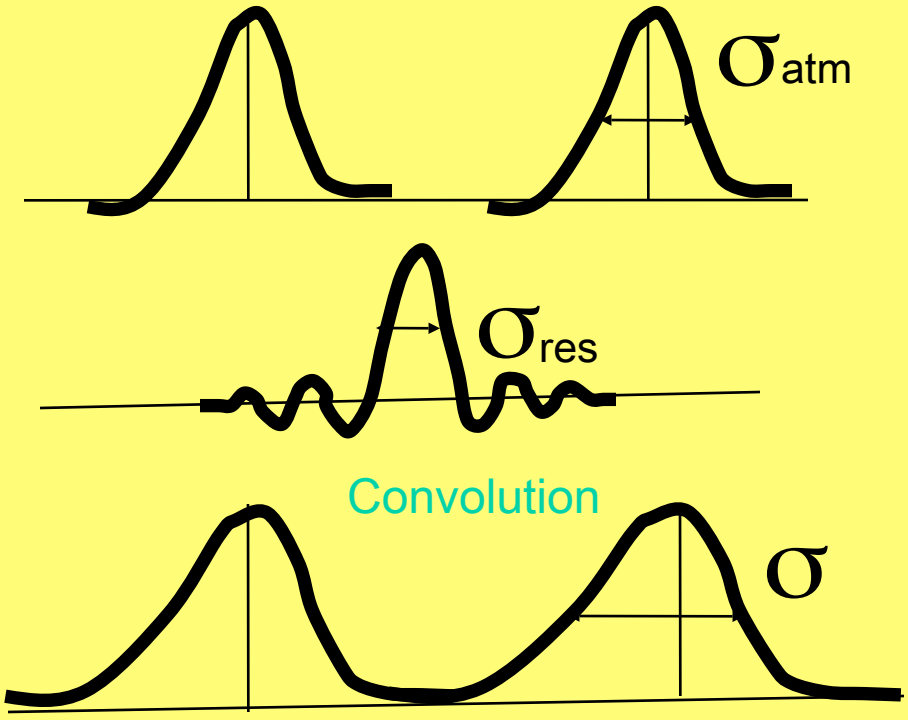


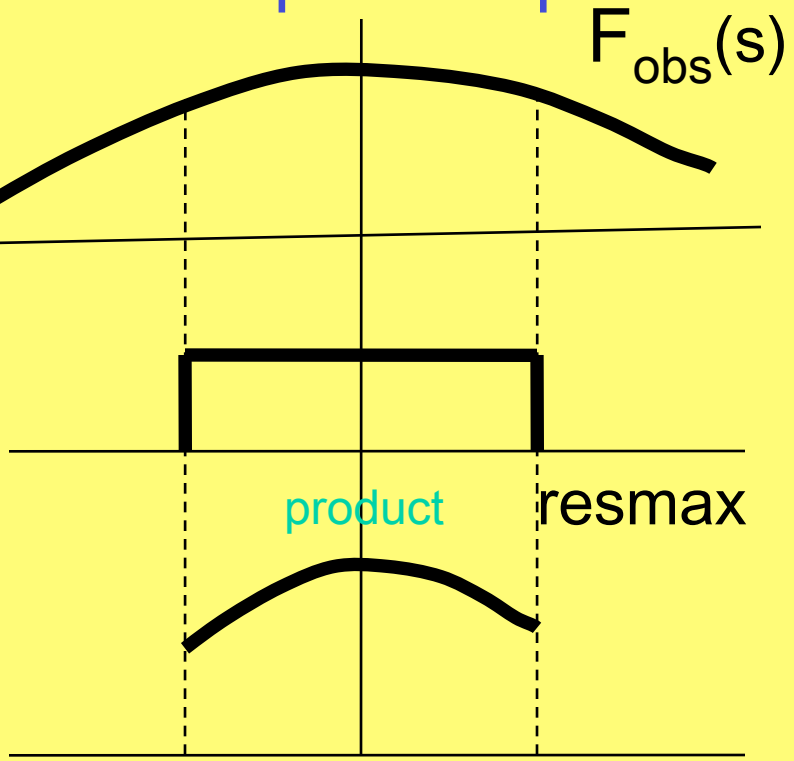
$$\sigma_{res} = 0.356\ resmax$$

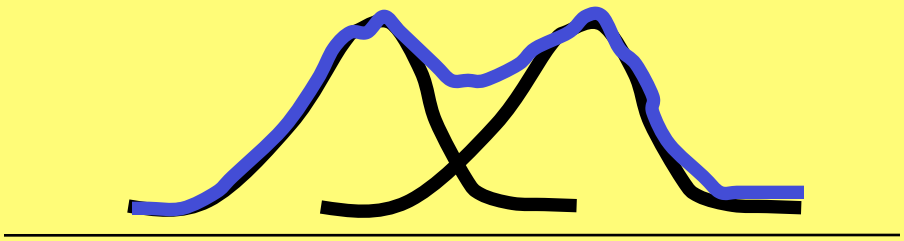
$$\sigma^2 = \sigma_{atm}^2 + \sigma_{res}^2$$

Optical resolution

Real space ← $\mathcal{F}$ → Reciprocal space





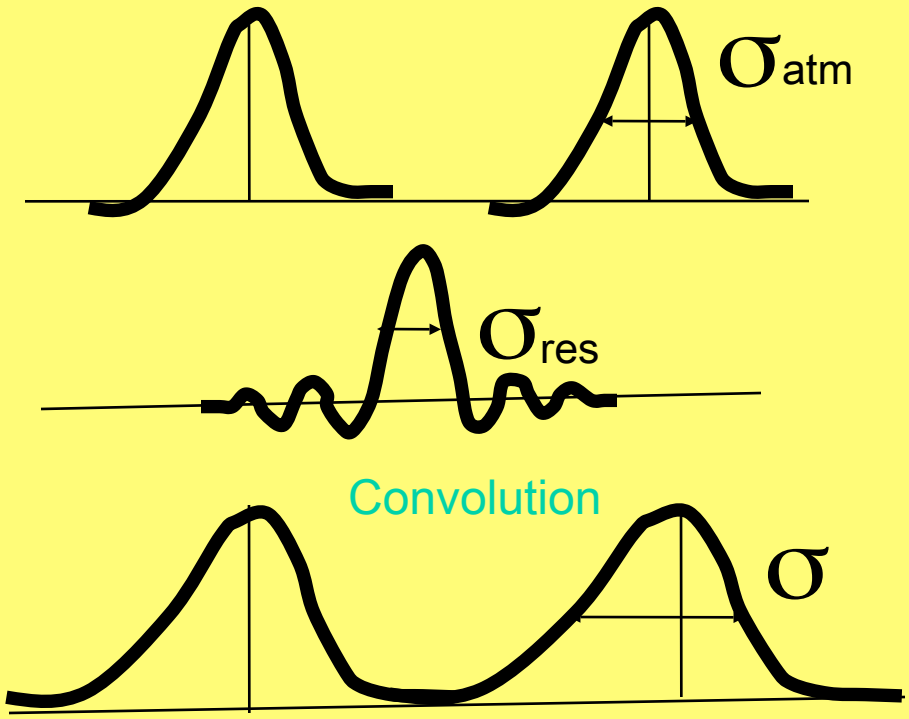


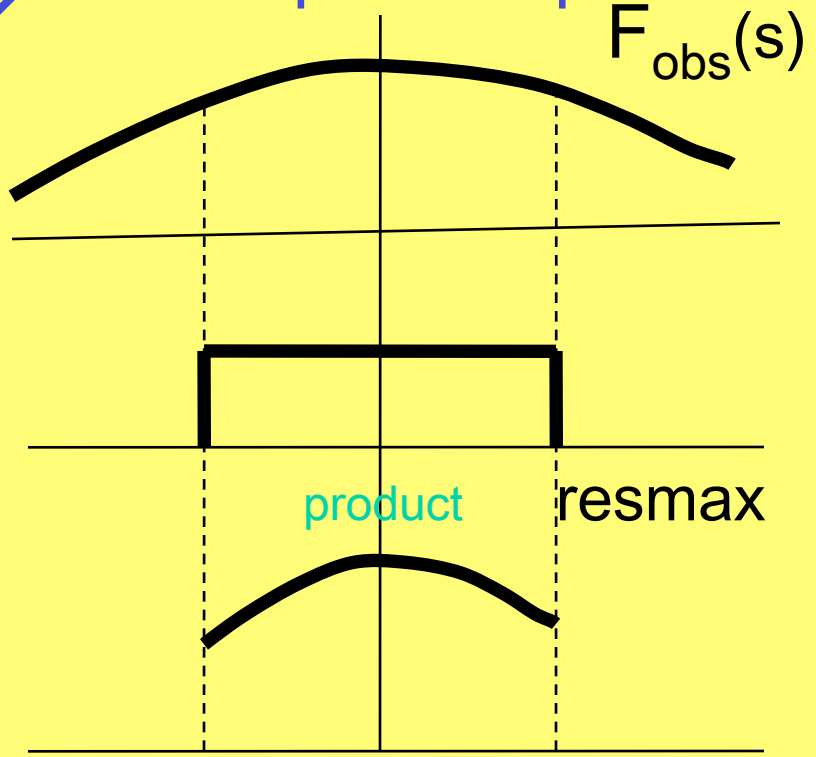
$$\sigma_{res} = 0.356\ resmax$$

$$\sigma^2 = \sigma_{atm}^2 + \sigma_{res}^2$$

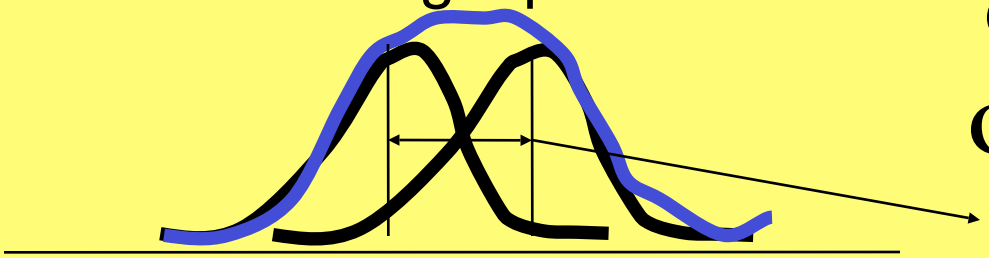
Optical resolution

Real space ← $\mathcal{F}$ → Reciprocal space





Single peak



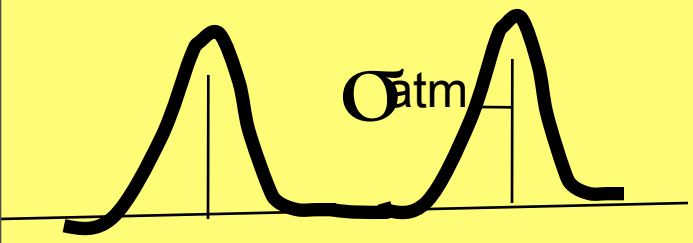
$\sigma_{res} = 0.356\ resmax$

$\sigma^2 = \sigma_{atm}^2 + \sigma_{res}^2$

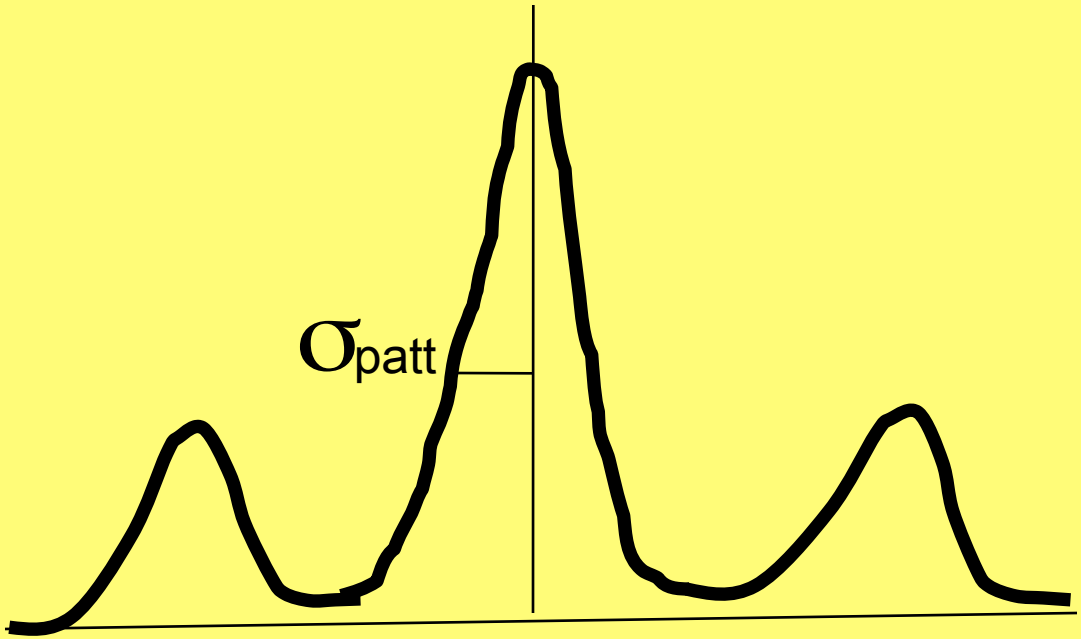
$Opt_{res} = 2\ \sigma$

Optical resolution from origin peak of Patterson

Real space



Patterson

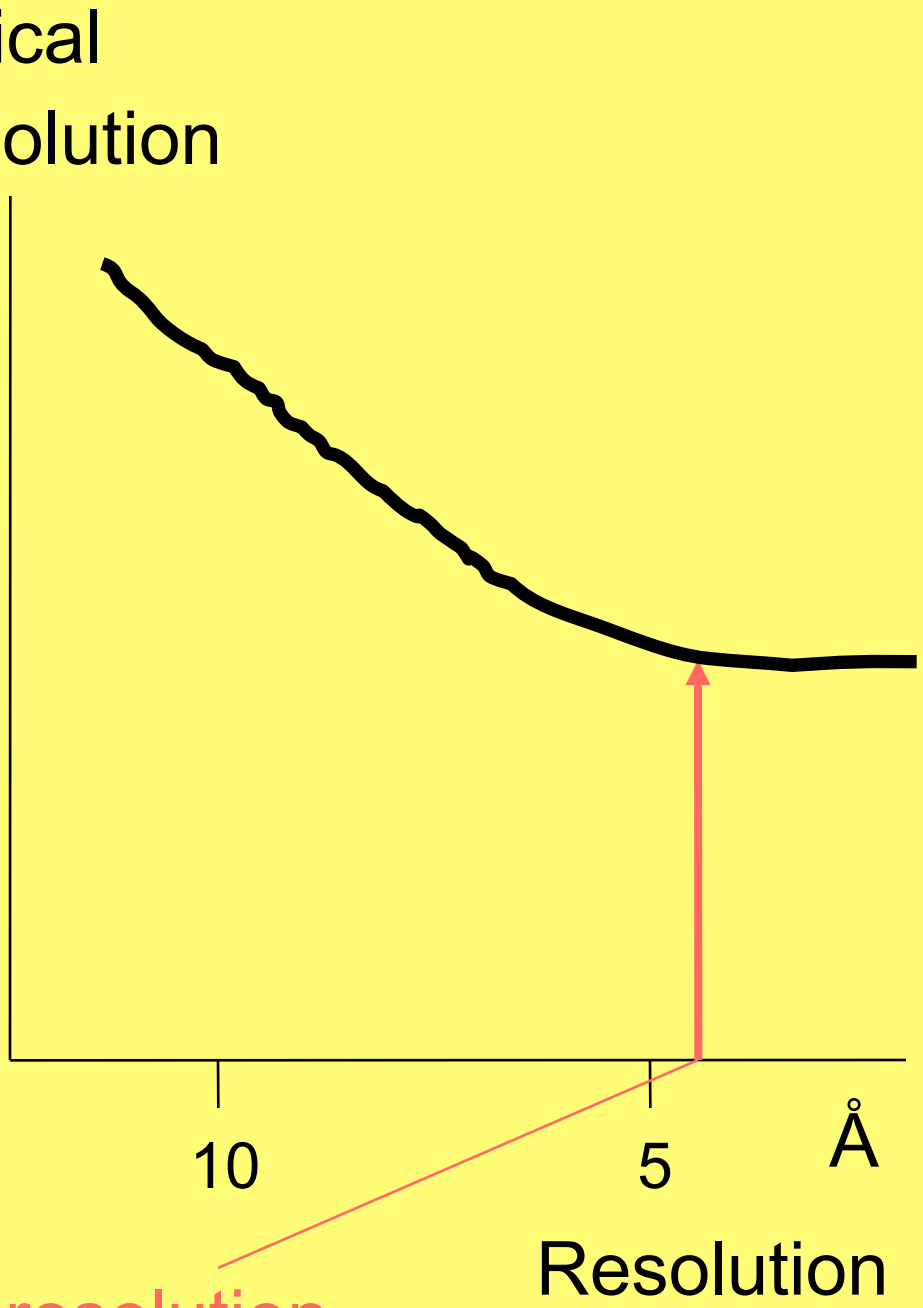


$$Opt_{res} = 2\,\sigma_{atm}$$

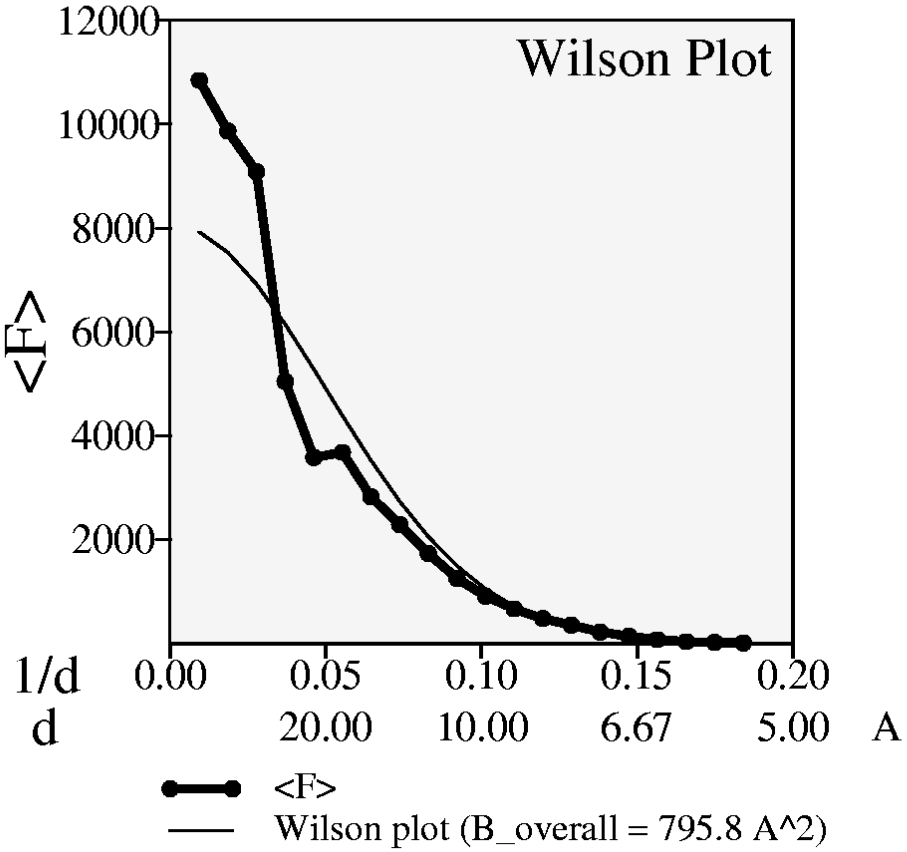
$$\sigma_{atm}^{2} = (\sigma_{patt}^{2} + \sigma_{res}^{2})/2$$

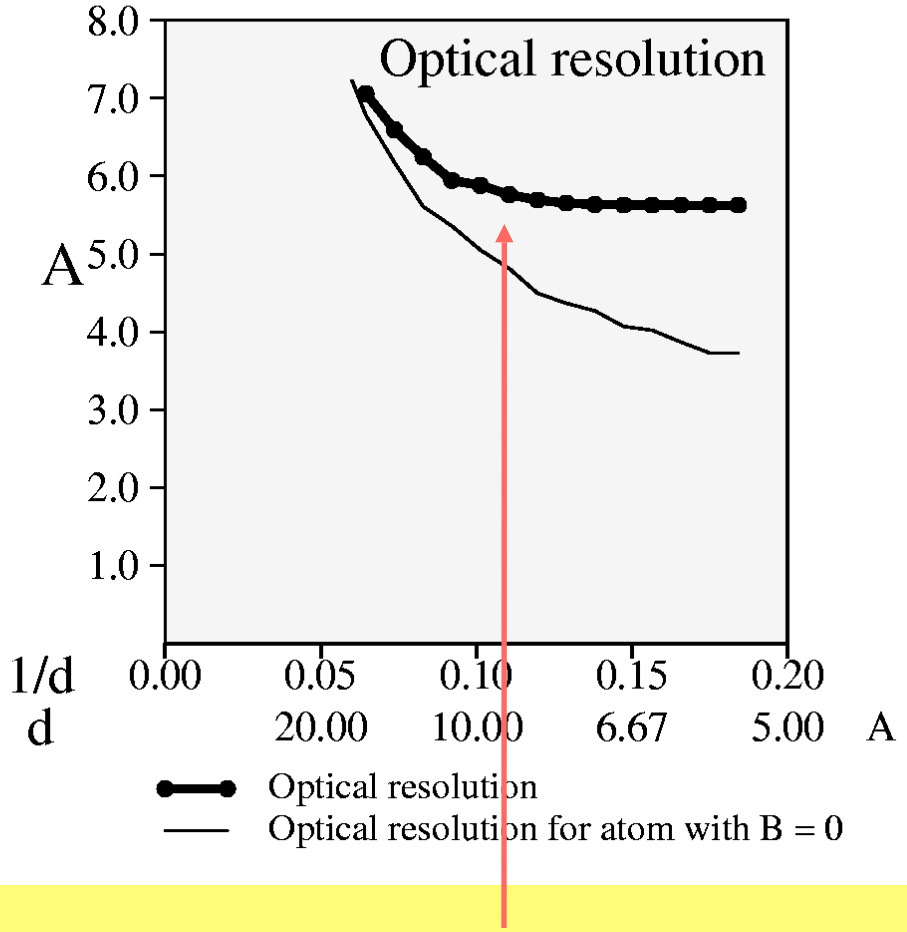
$$Opt_{res} = 2\sigma$$

$$\sigma^2 = (\sigma_{atm}^2 + \sigma_{res}^2)/2$$

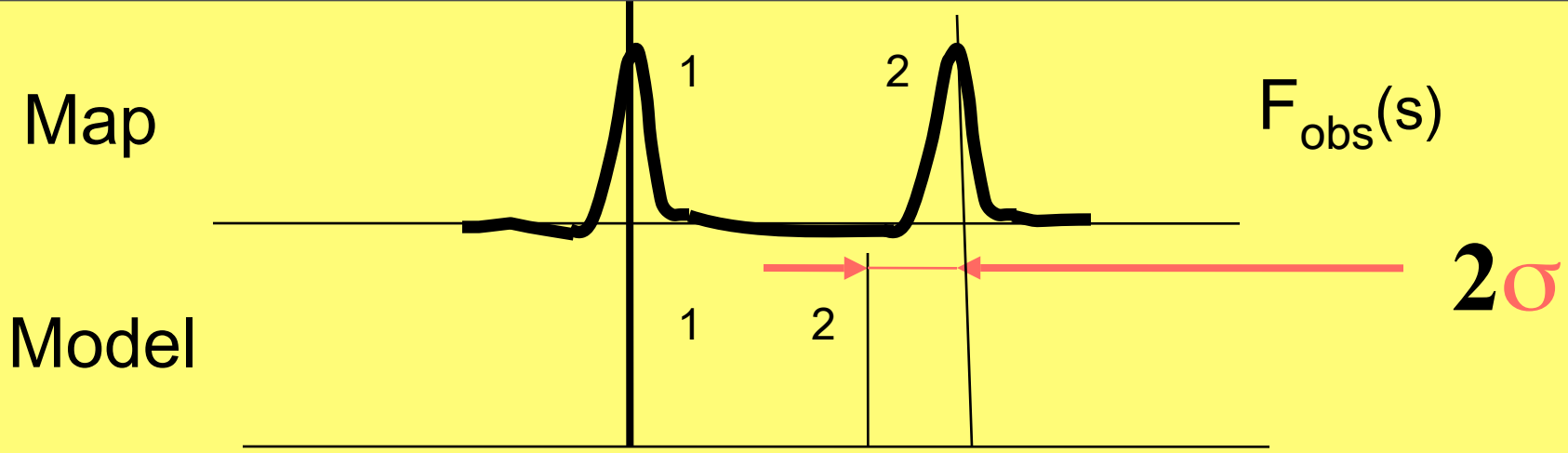


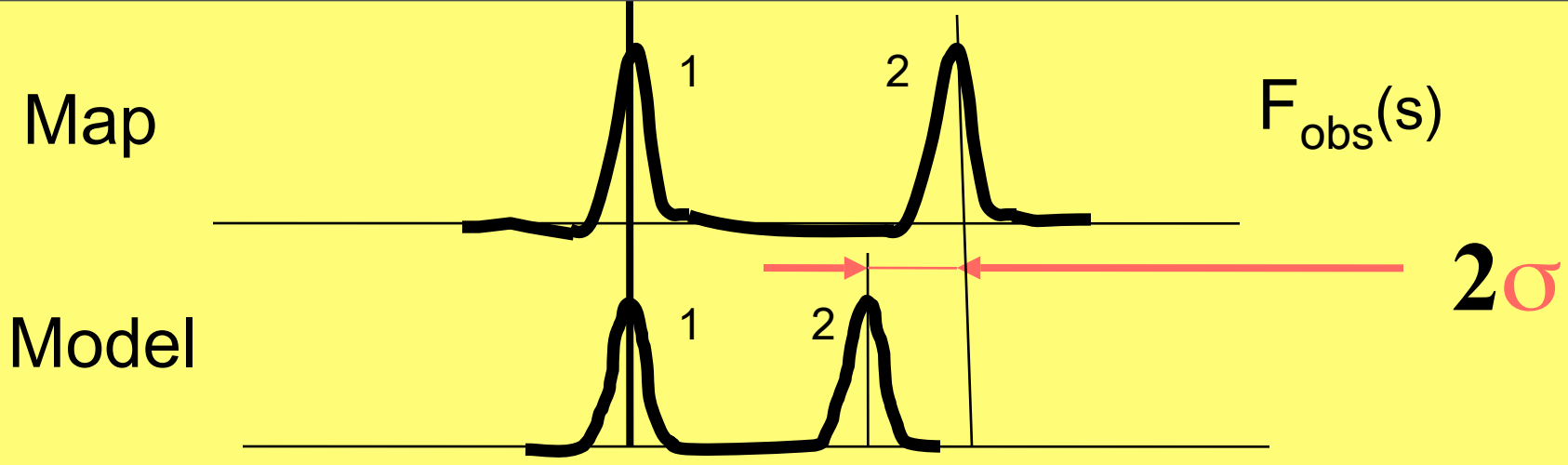
Optical Resolution (by sfcheck)

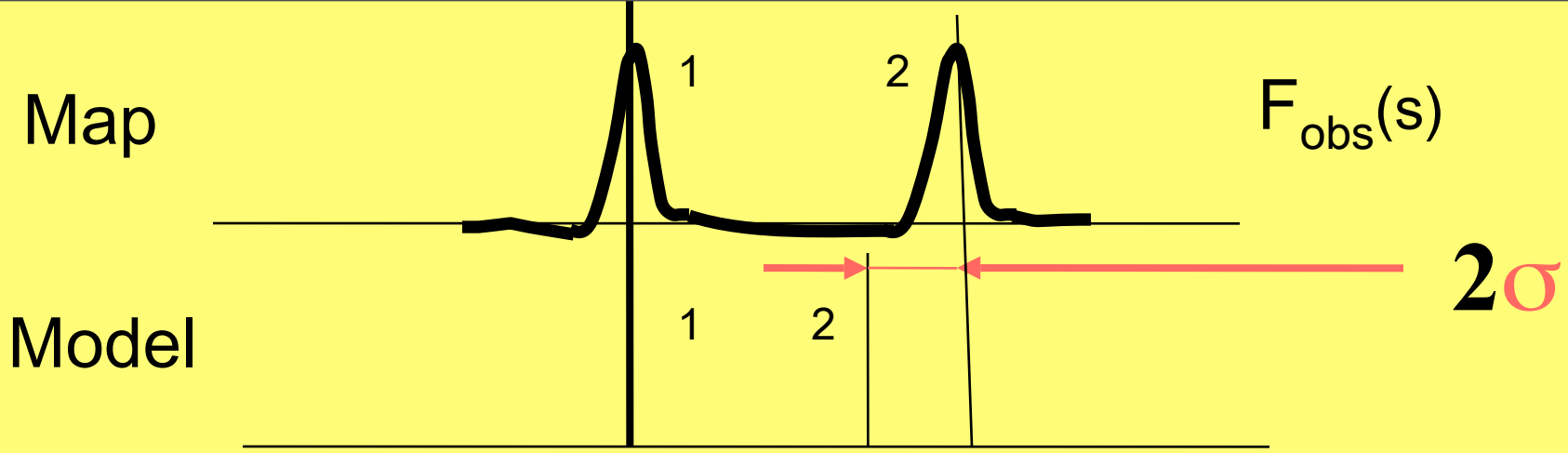


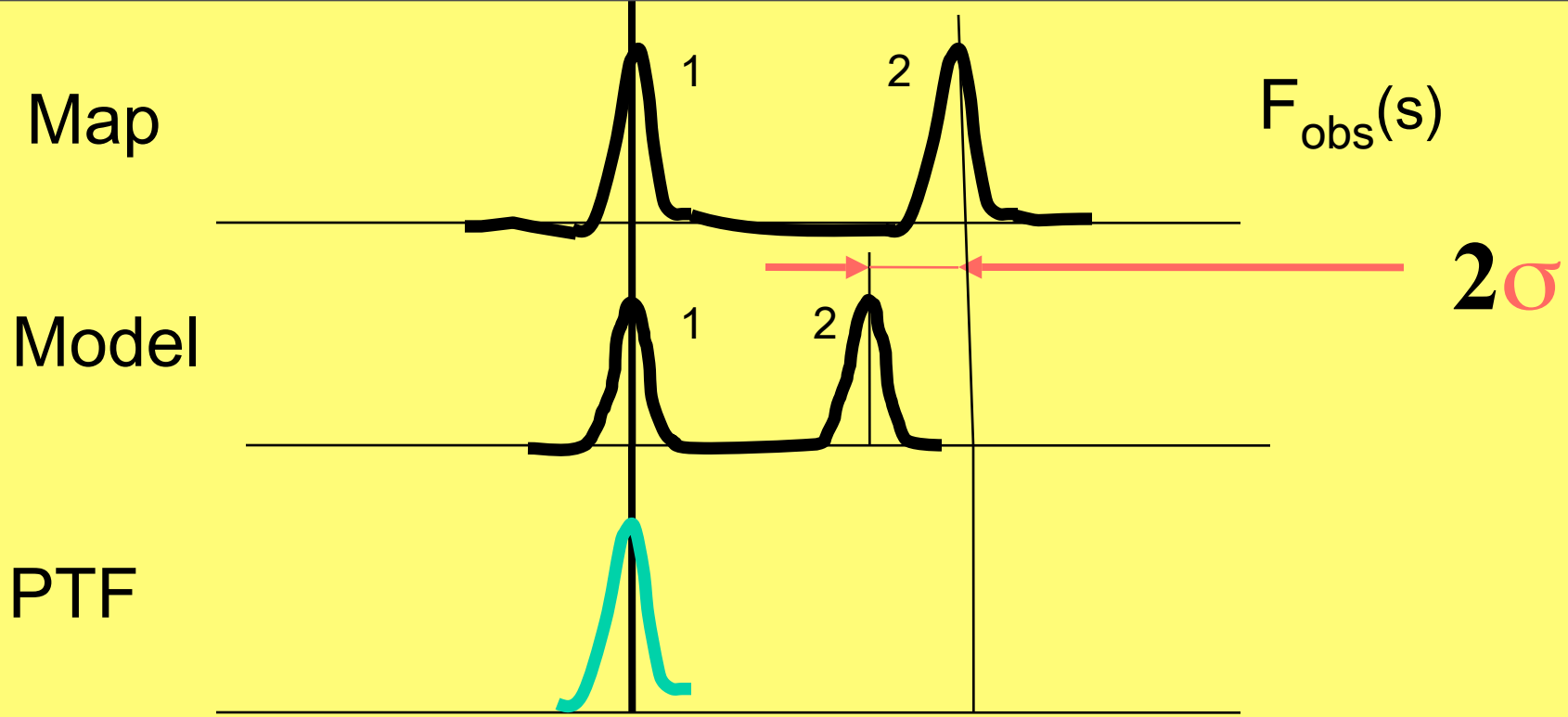


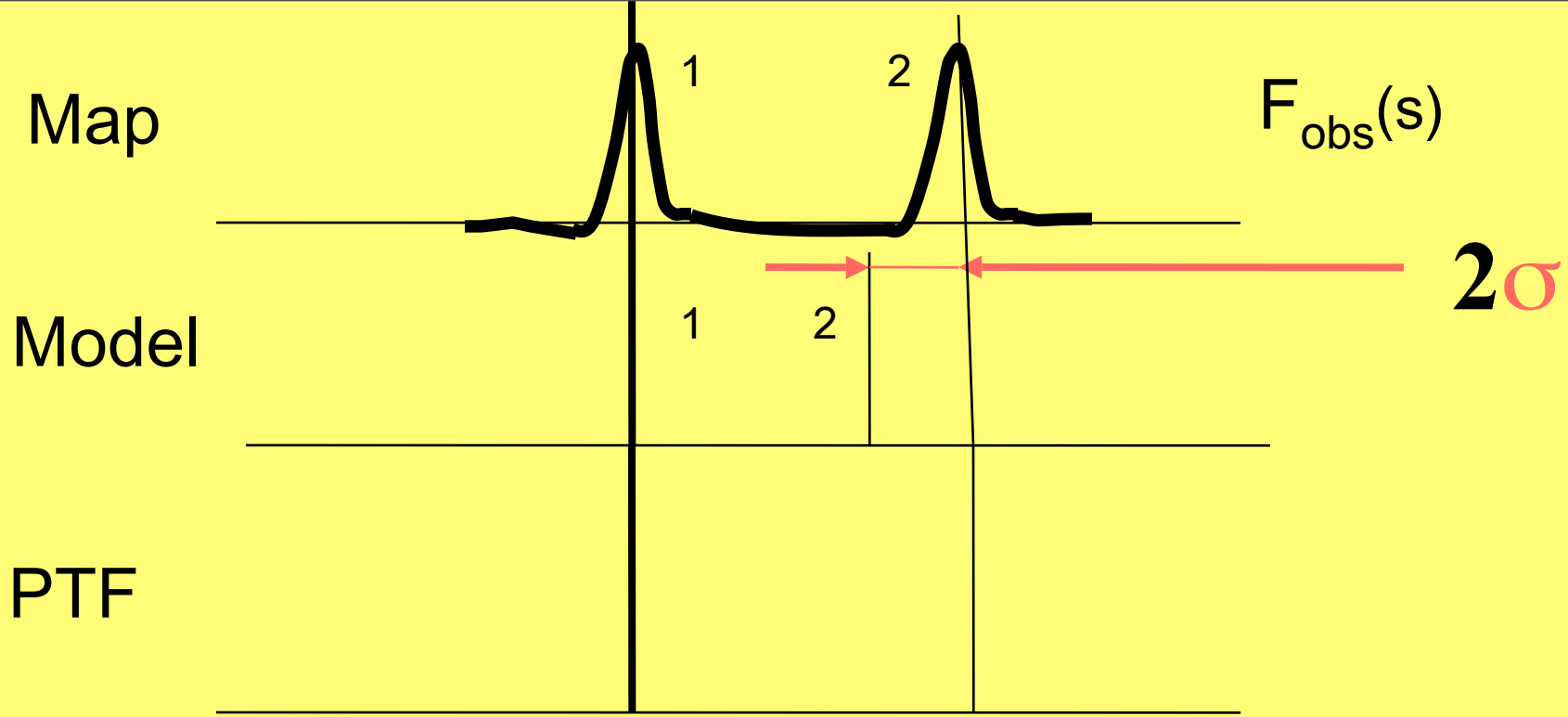
Weights for high resolution data and similarity

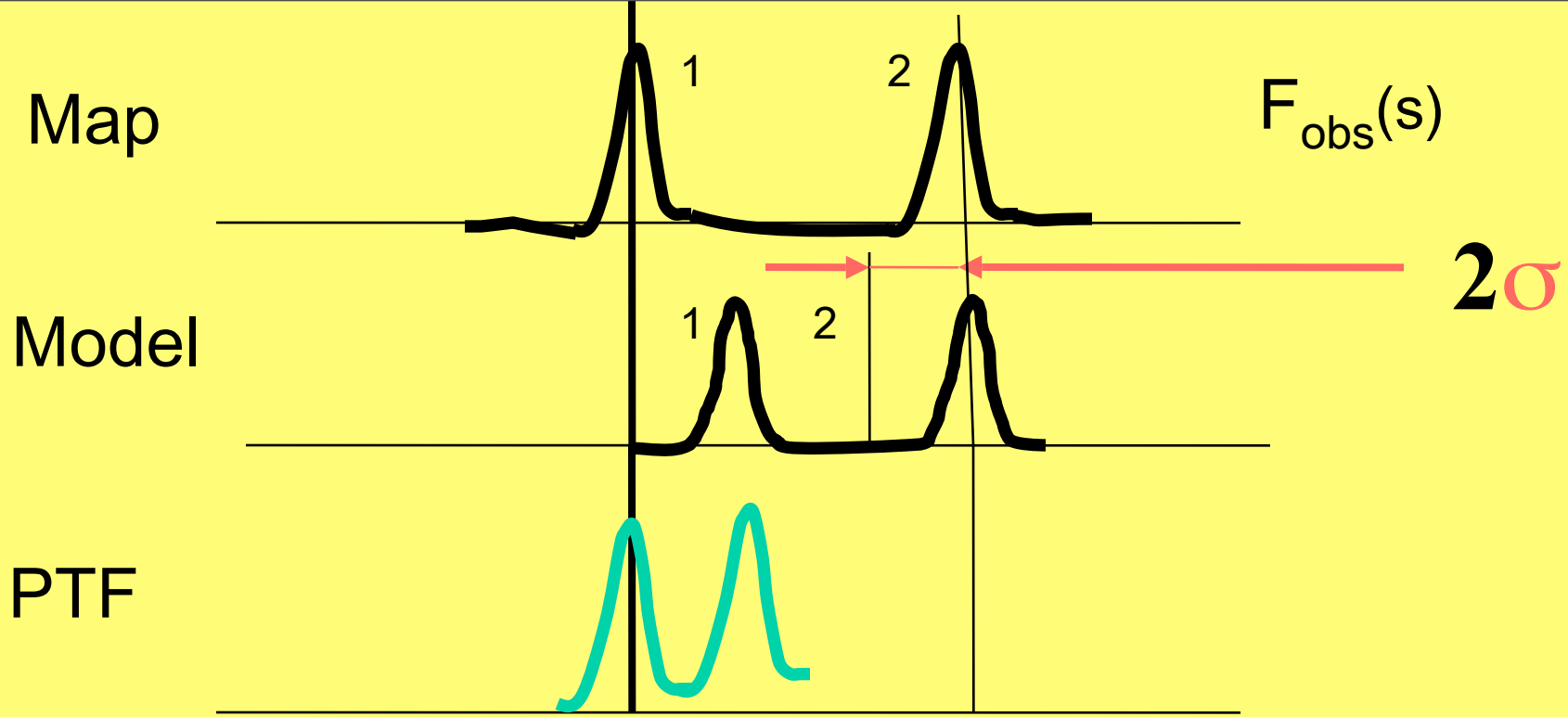

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2

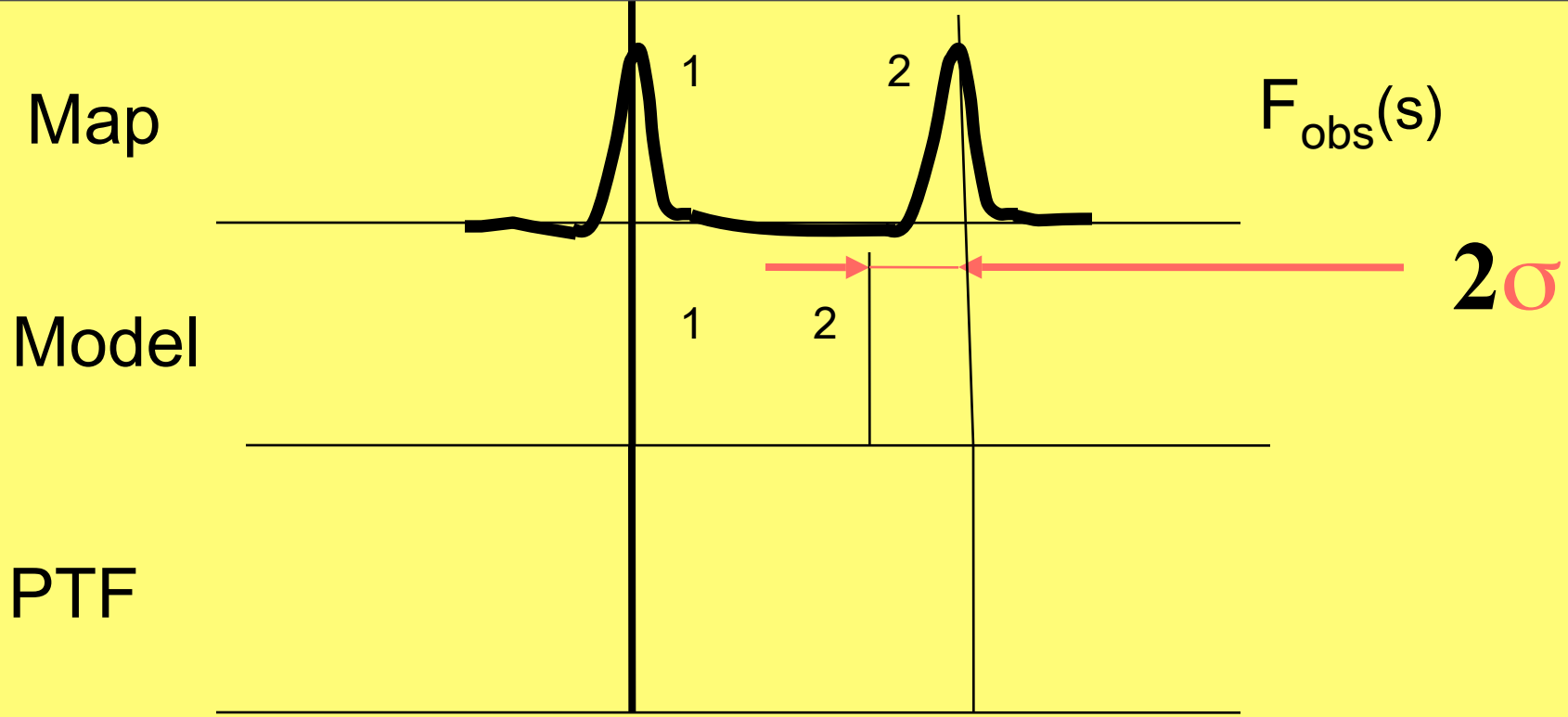

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2

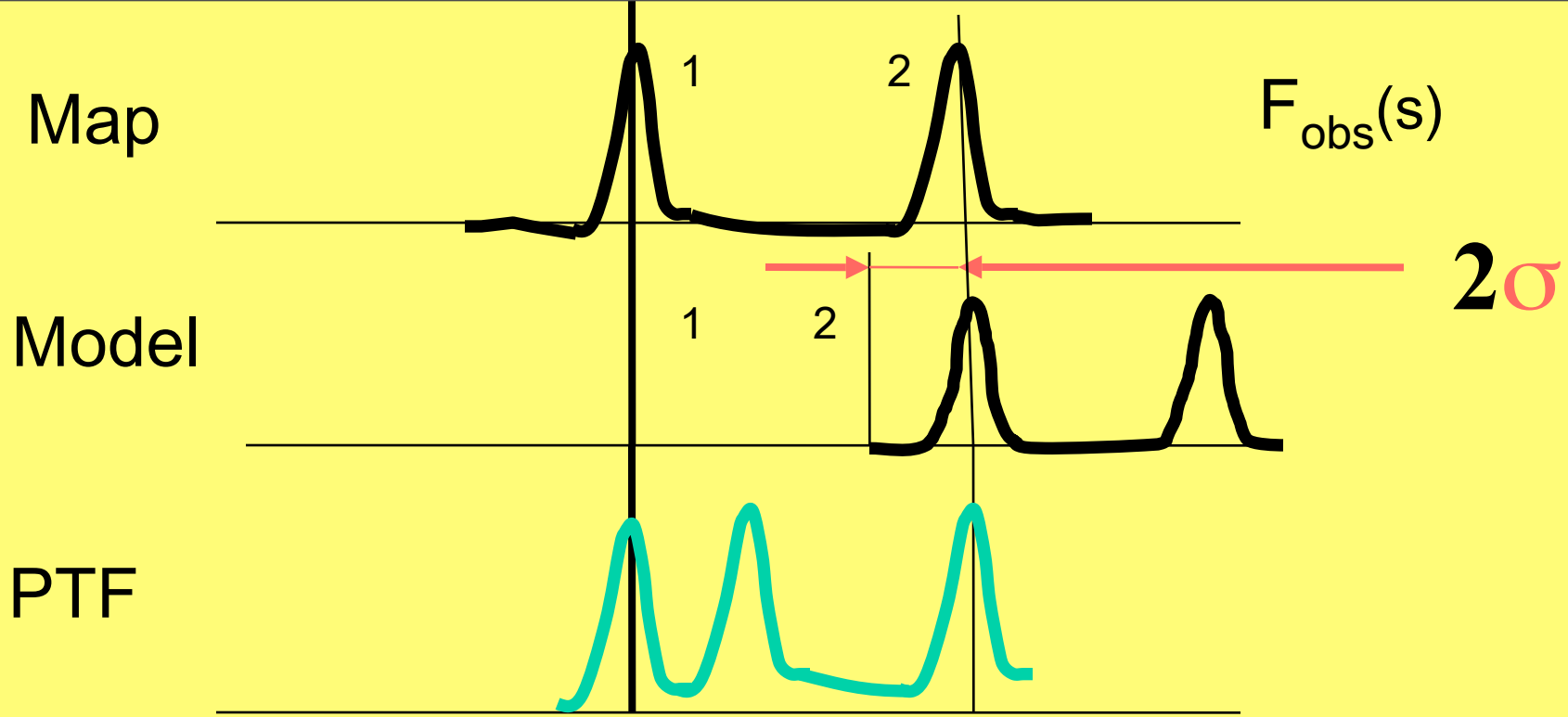

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2

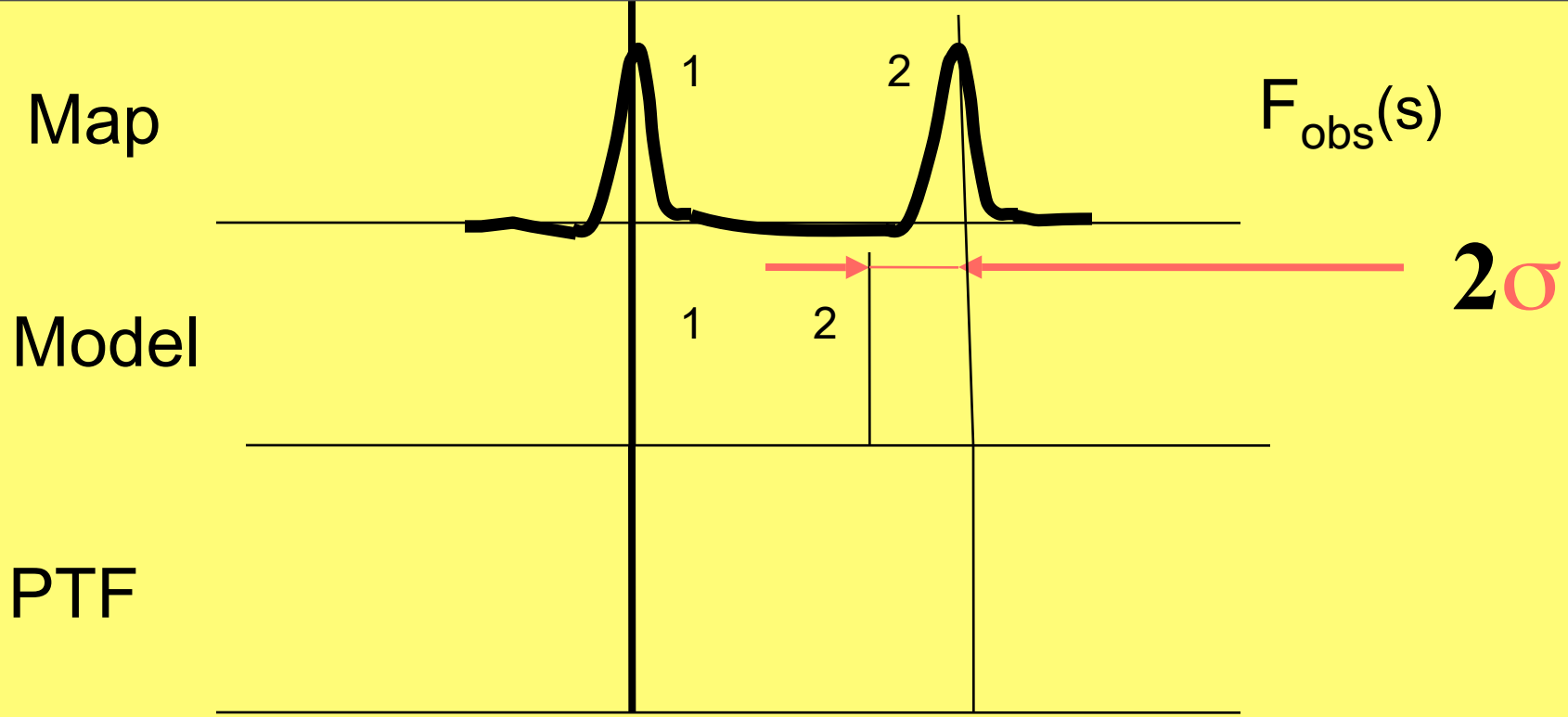

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF

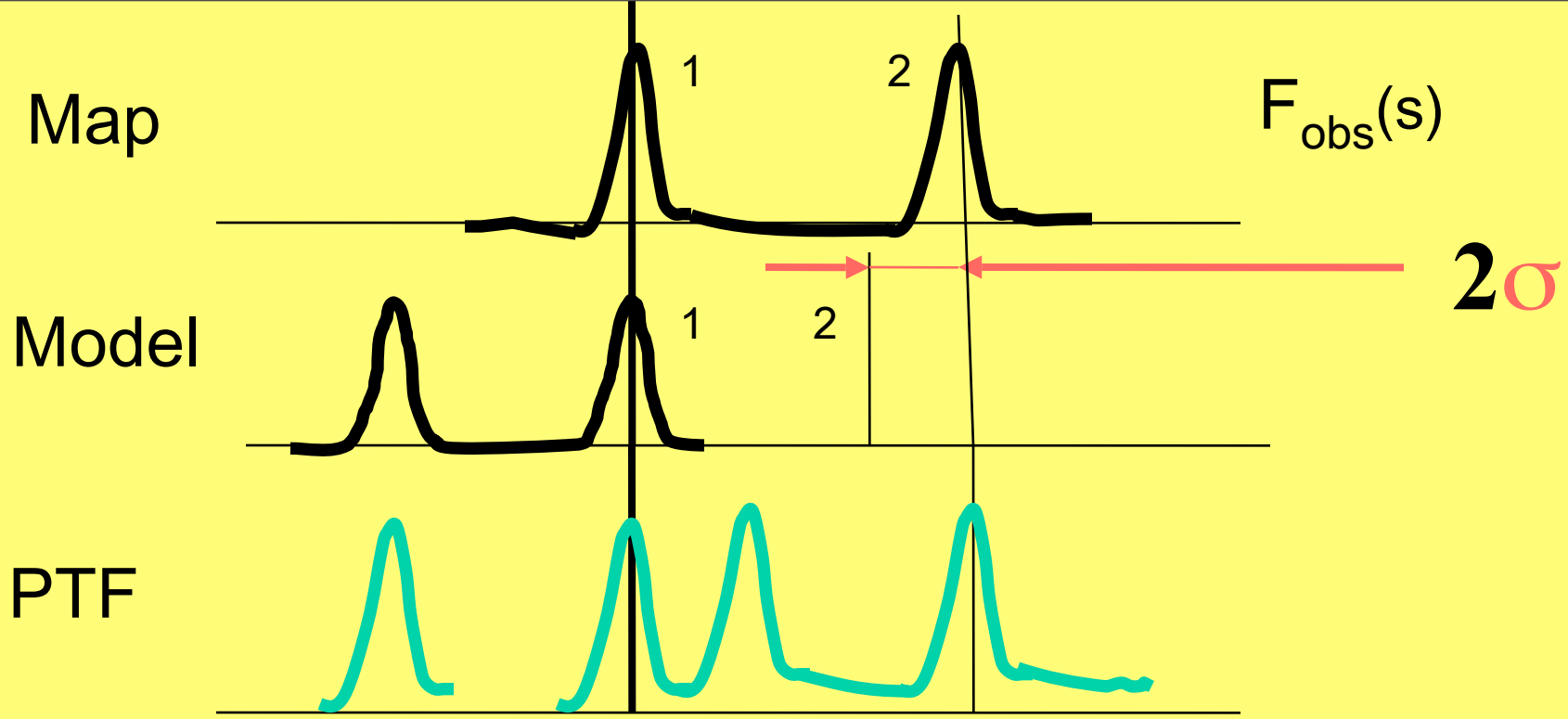

Map
Model
PTF
1
2
$F_{obs}(s)$
1
2
2σ

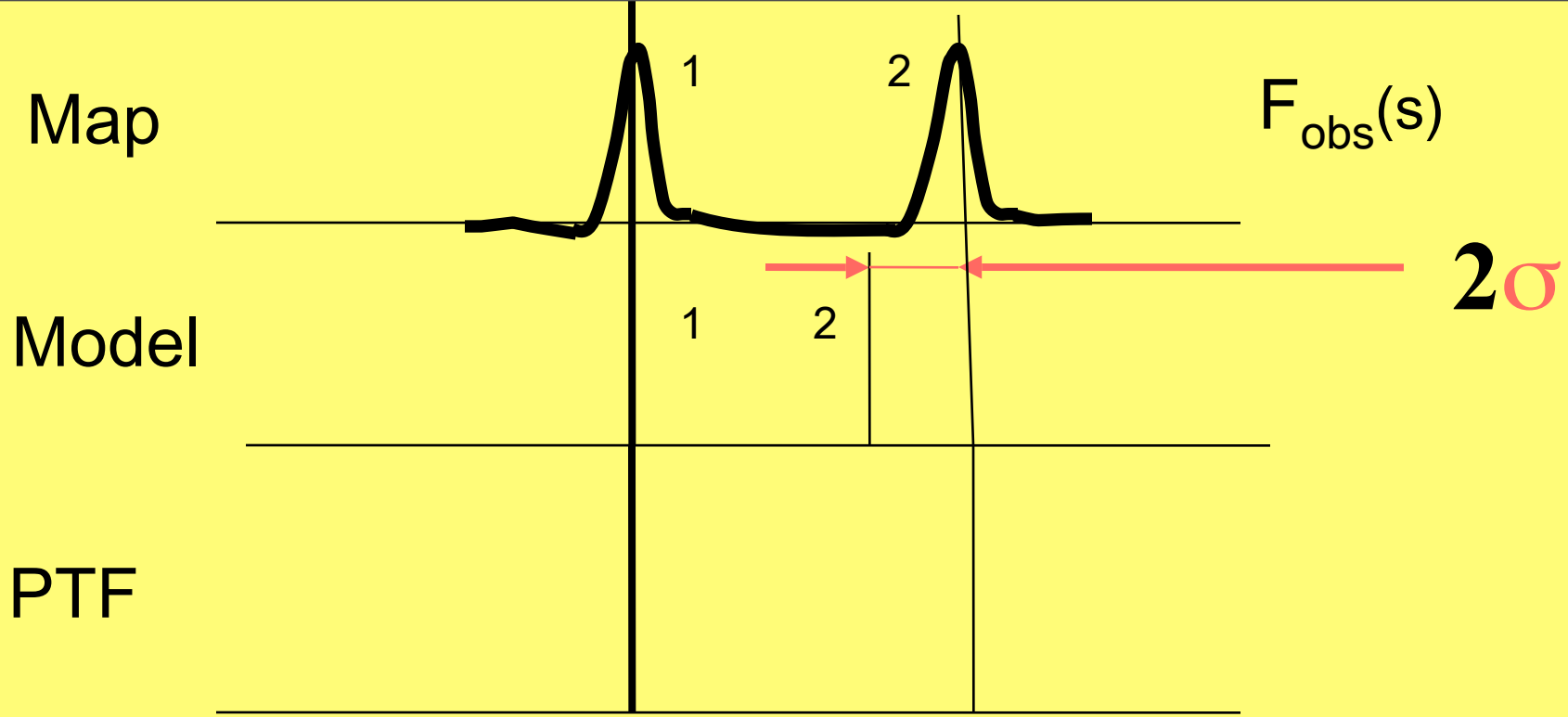

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF

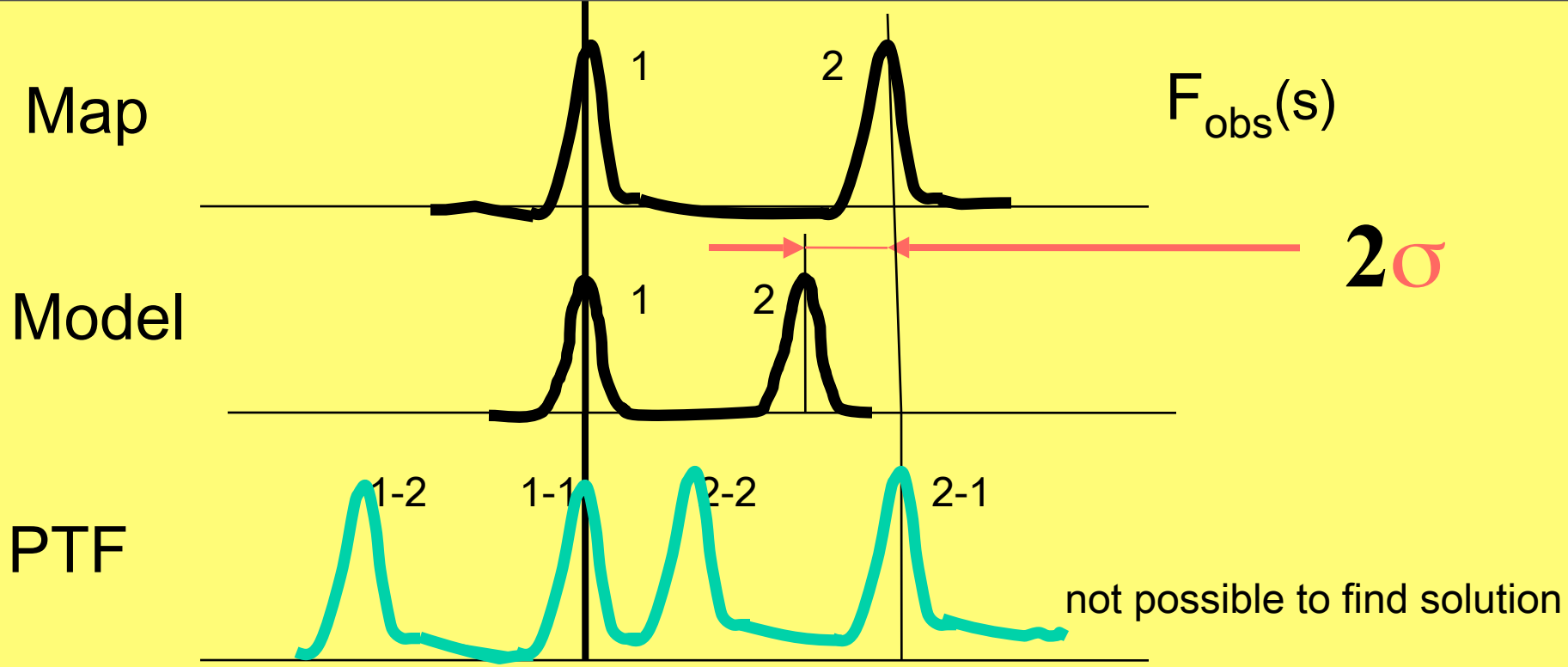

Map
Model
PTF
1
2
$F_{obs}(s)$
1
2
2σ

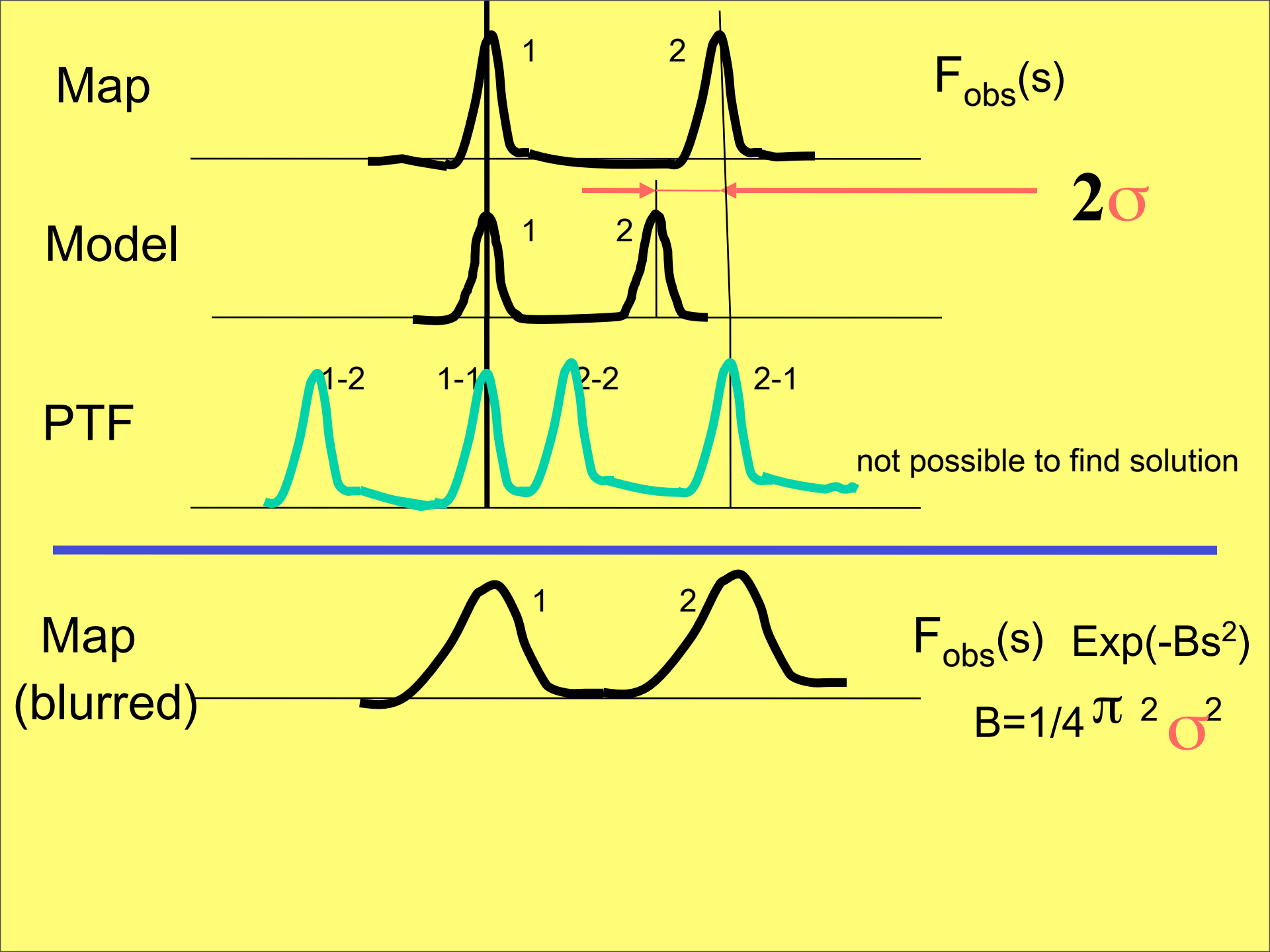

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF

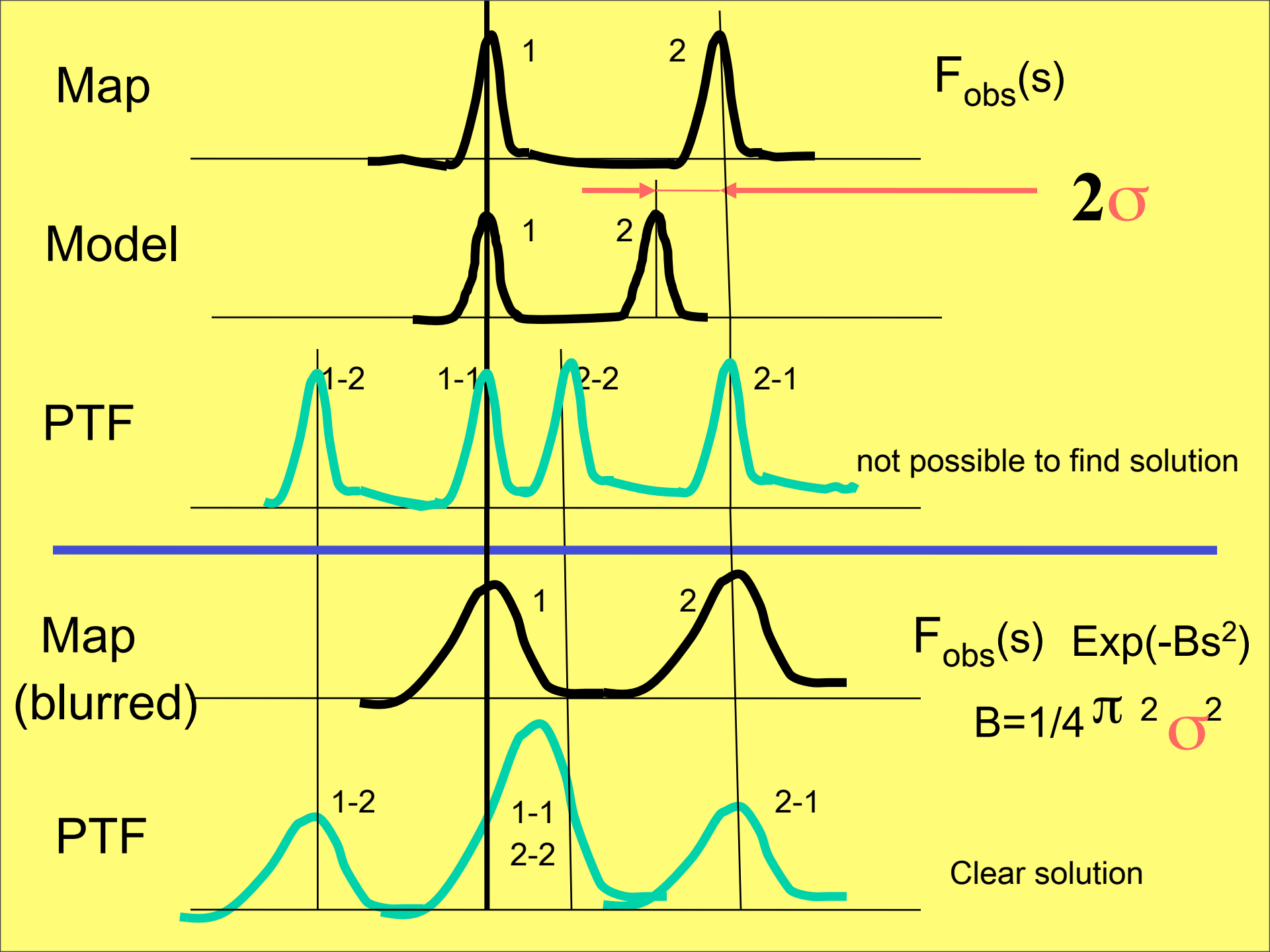

Map
Model
PTF
1
2
$F_{obs}(s)$
1
2
2σ


Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF


Map
Model
PTF
1
2
$F_{obs}(s)$
1
2
2σ


Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF
1-2
1-1
2-2
2-1
not possible to find solution


Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF
1-2
1-1
2-2
2-1
not possible to find solution
Map
(blurred)
1
2
$F_{obs}(s)$ $Exp(-Bs^2)$
$B=1/4\,\pi^2\,\sigma^2$

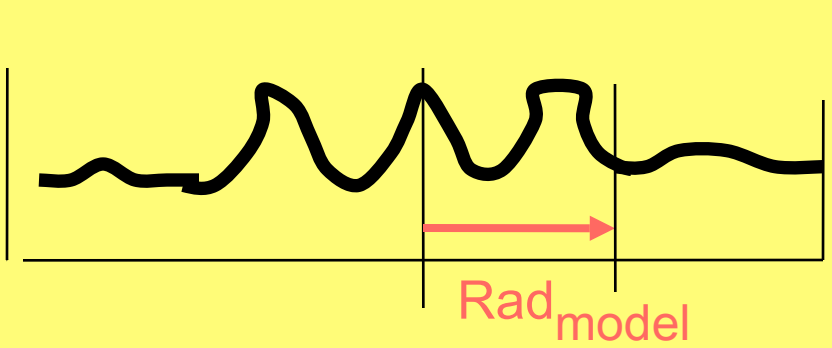

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF
1-2
1-1
2-2
2-1
not possible to find solution
Map
(blurred)
1
2
$F_{obs}(s)$ $Exp(-Bs^2)$
$B=1/4\pi^2\sigma^2$
PTF
1-2
1-1
2-2
2-1
Clear solution

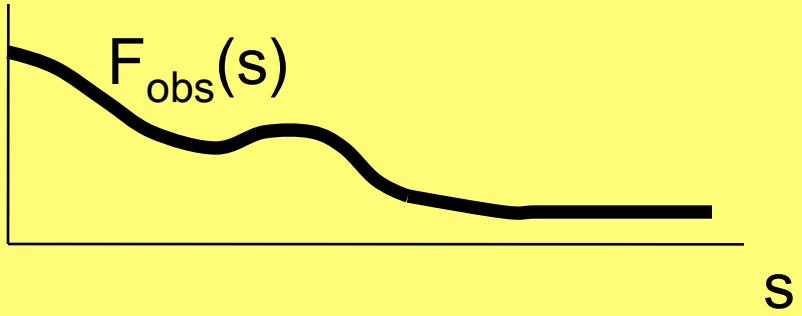
Low resolution data

Weights for low resolution data and size of model

Soft minimal resolution cut-off

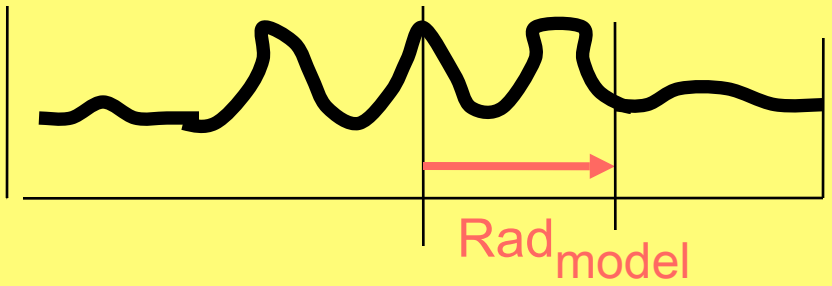
Real space ← $\mathscr{F}$ → Reciprocal space

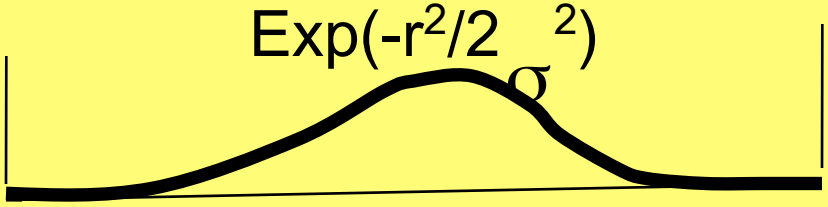


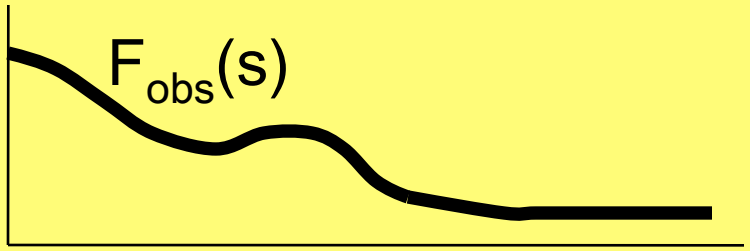


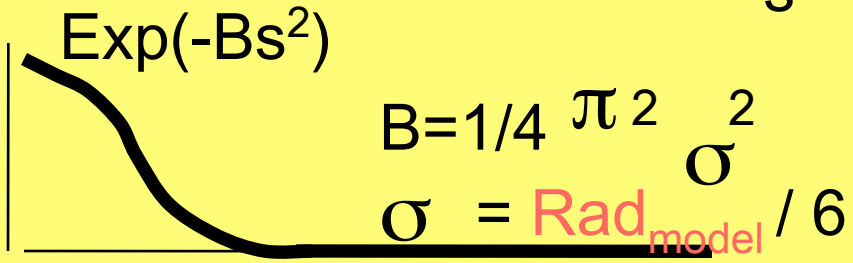
Soft minimal resolution cut-off

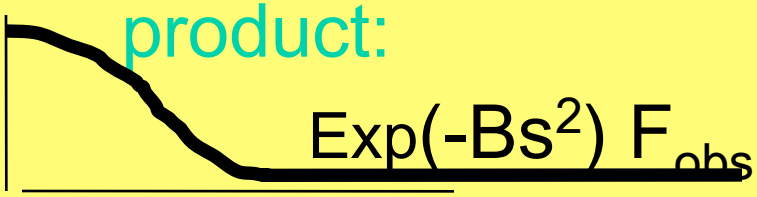
Real space ← $\mathscr{F}$ → Reciprocal space



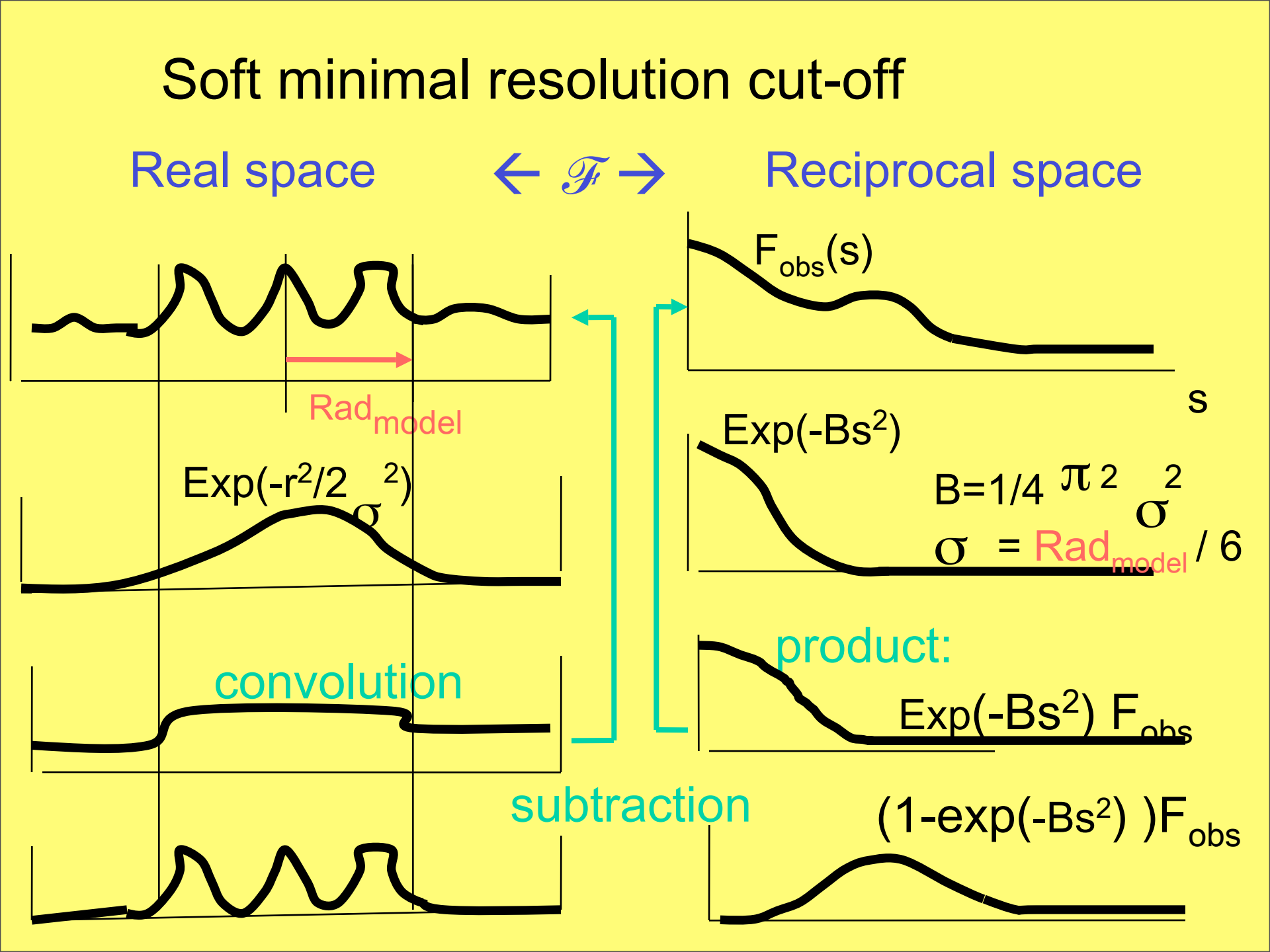






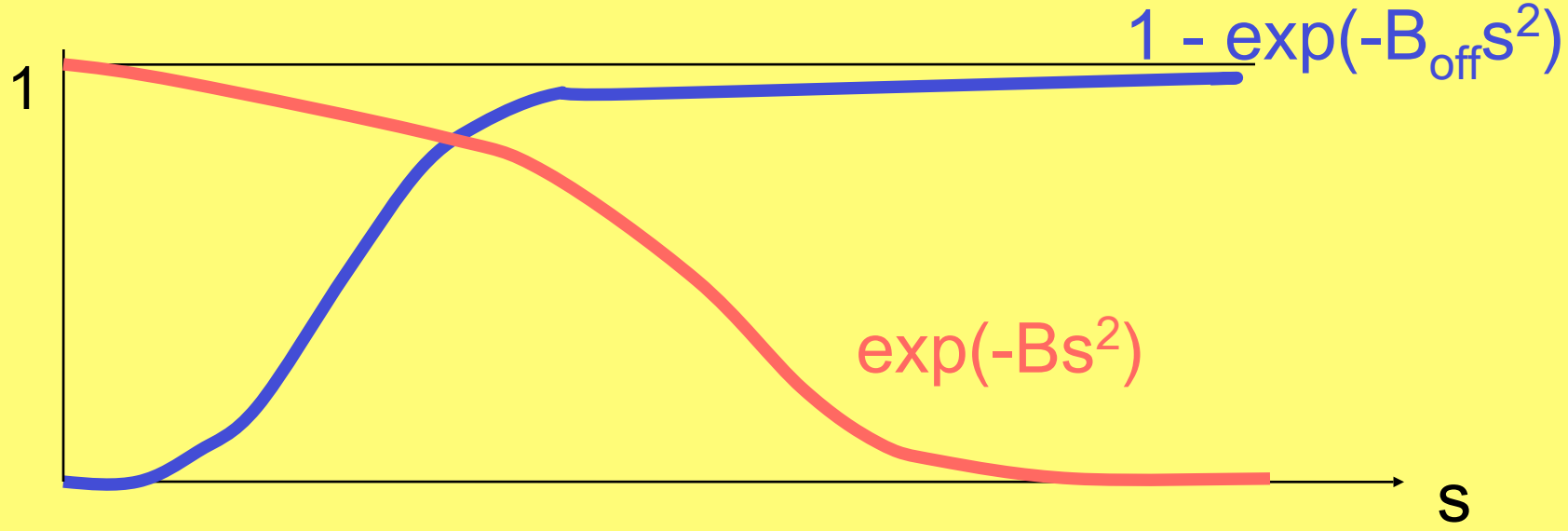


Soft minimal resolution cut-off



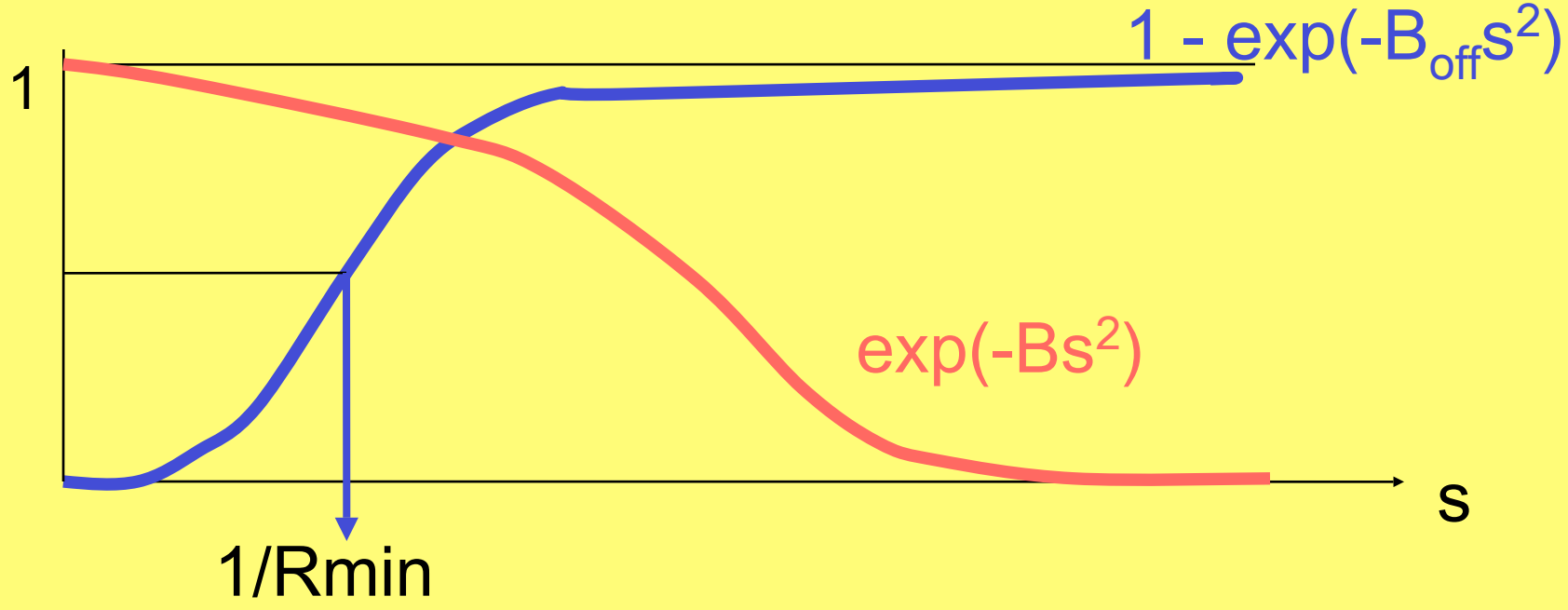
Weighting scheme

$F_{used} = (1 - \exp(-B_{off}s^2))\ F_{obs}\ \exp(-Bs^2)$



Weighting scheme

$F_{used} = (1 - \exp(-B_{off}s^2)) \; F_{obs} \; \exp(-Bs^2)$



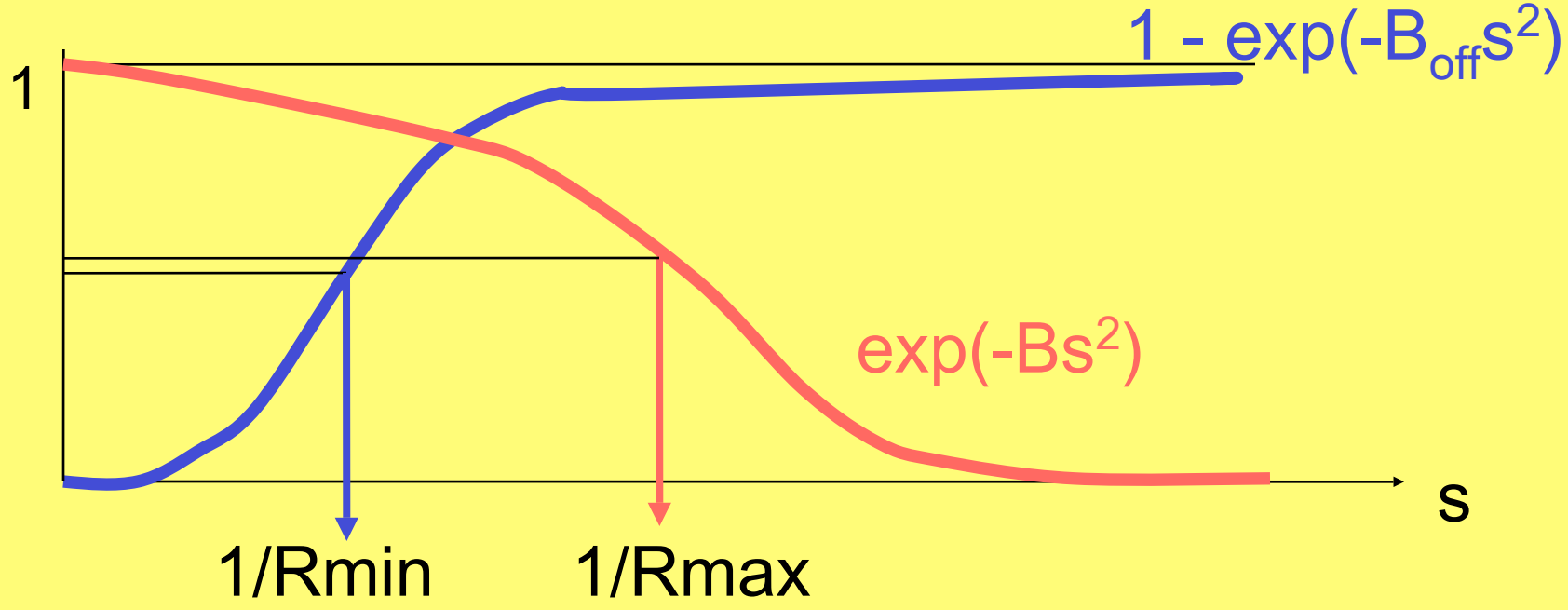
$Rmin = \sqrt{2B_{off}} = 2\pi\sigma_{model}$

$\sigma_{model} = Rad_{model} / 6$

$Rmin \approx Rad_{model}$

Weighting scheme

$F_{used} = (1 - \exp(-B_{off}s^2))\ F_{obs}\ \exp(-Bs^2)$



$Rmin = \sqrt{2B_{off}} = 2\ \pi\ \sigma_{model}$

$\sigma_{model} = Rad_{model} / 6$

$Rmin \approx Rad_{model}$

$B = 1/4\ \pi^2\ \sigma^2_{Model\ similarity}$

$Rmax = \sqrt{2B}$

Weighting scheme

Two filters in Image processing:

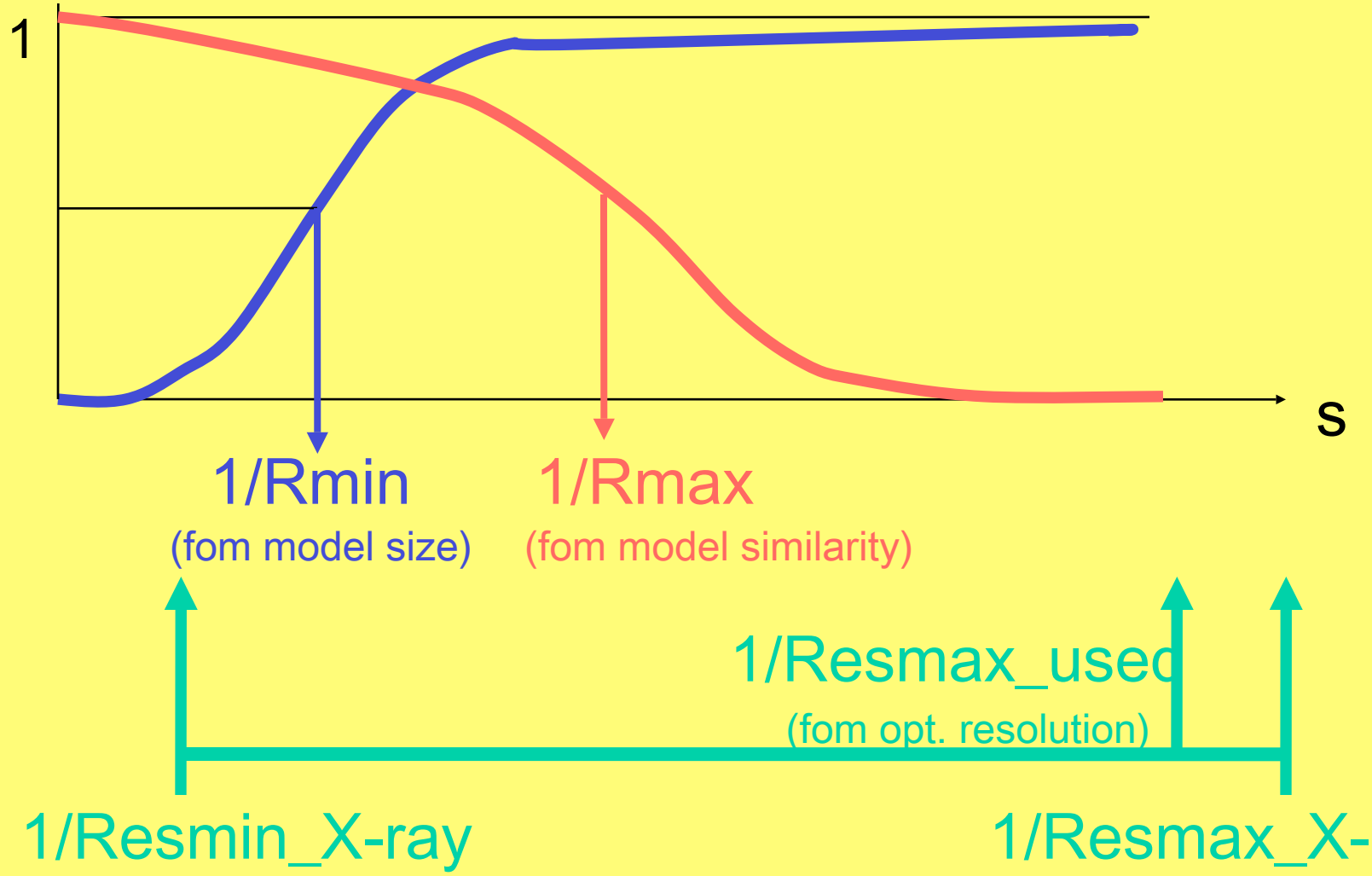
Gaussian highpass filter

Gaussian lowpass filter

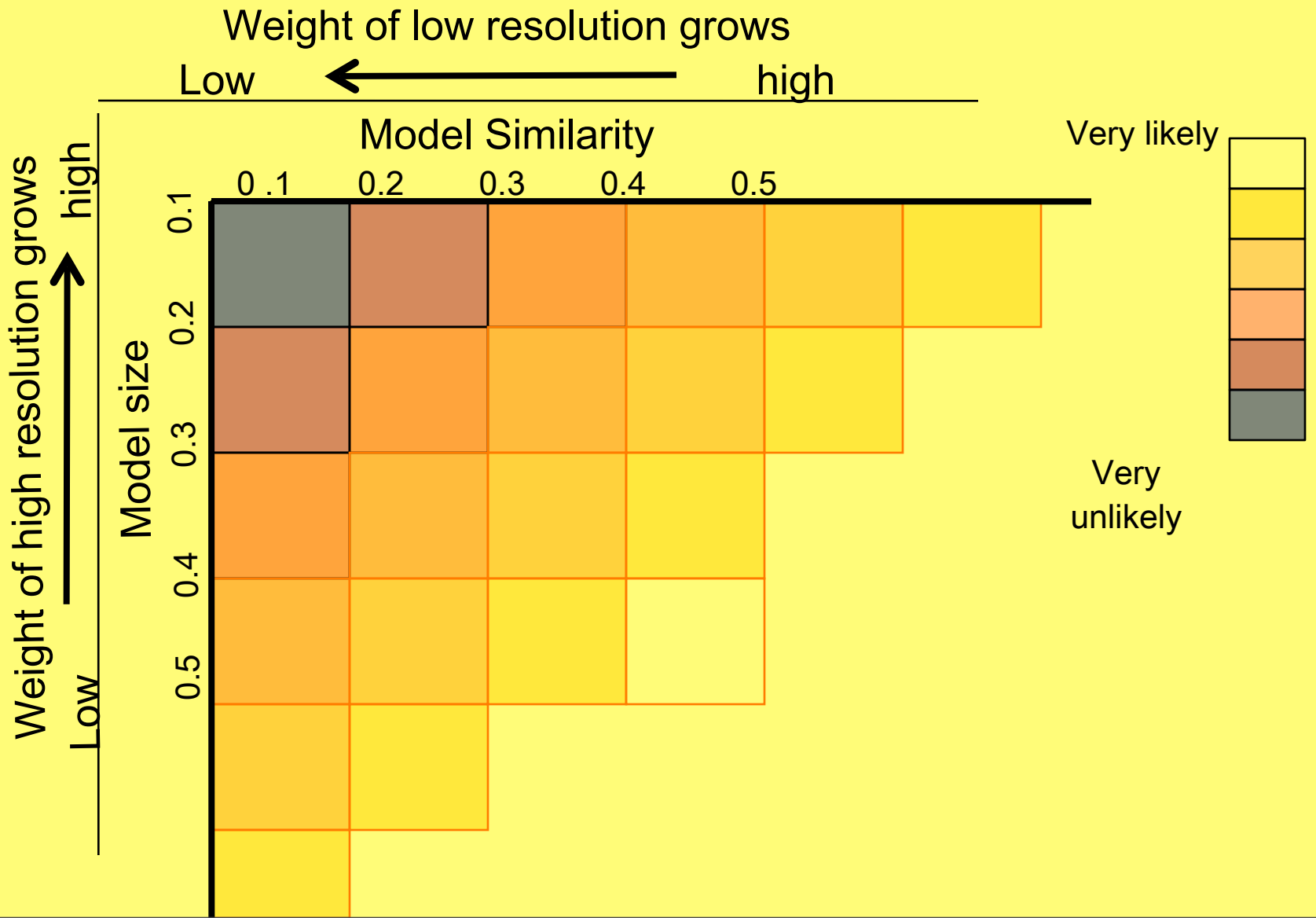
$$F_{used} = (1 - \exp(-B_{off}s^2)) \; F_{obs} \; \exp(-Bs^2)$$

We can consider this weighting scheme as an approximation to the likelihood approach

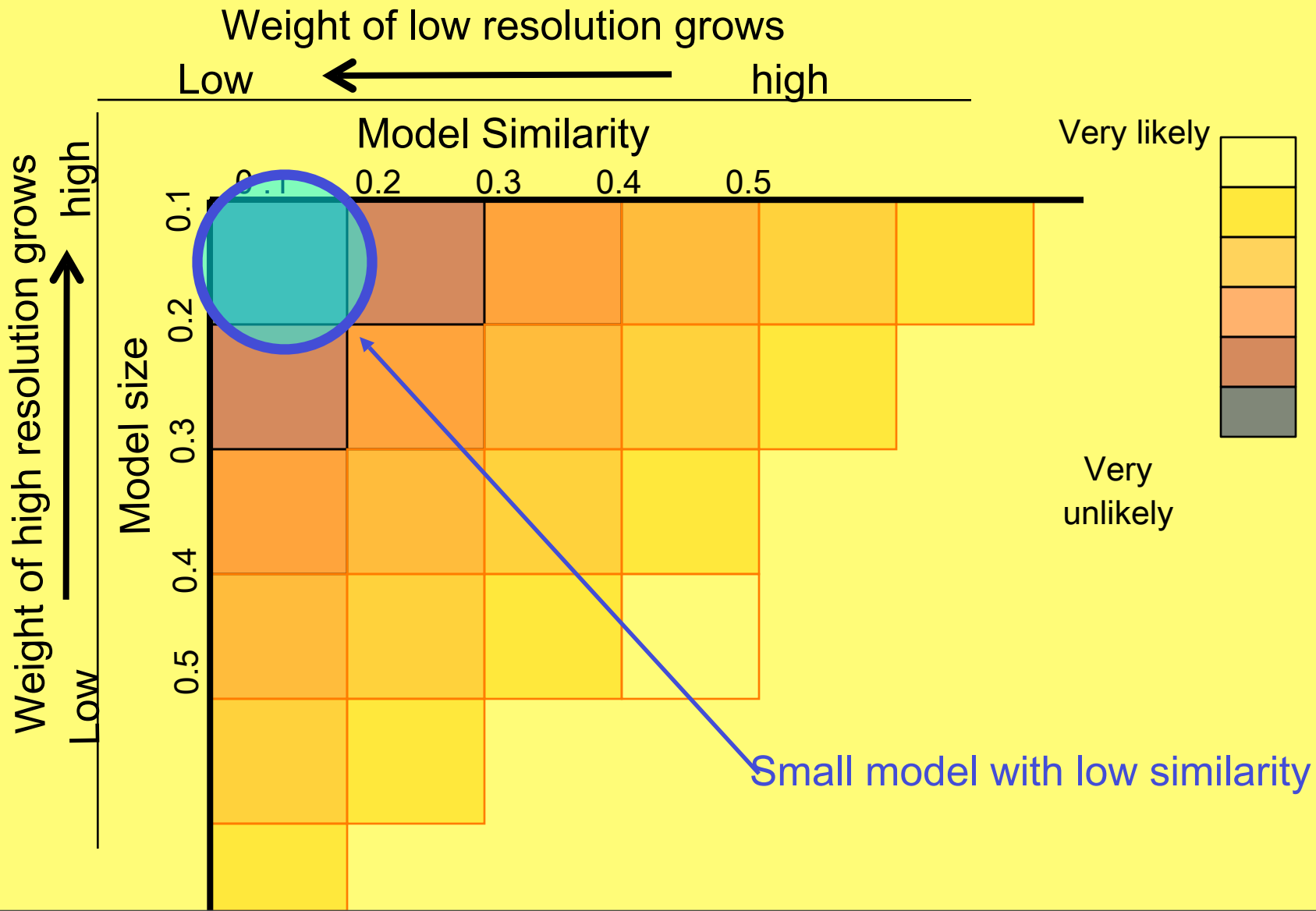
Information in X-ray and Model must overlap



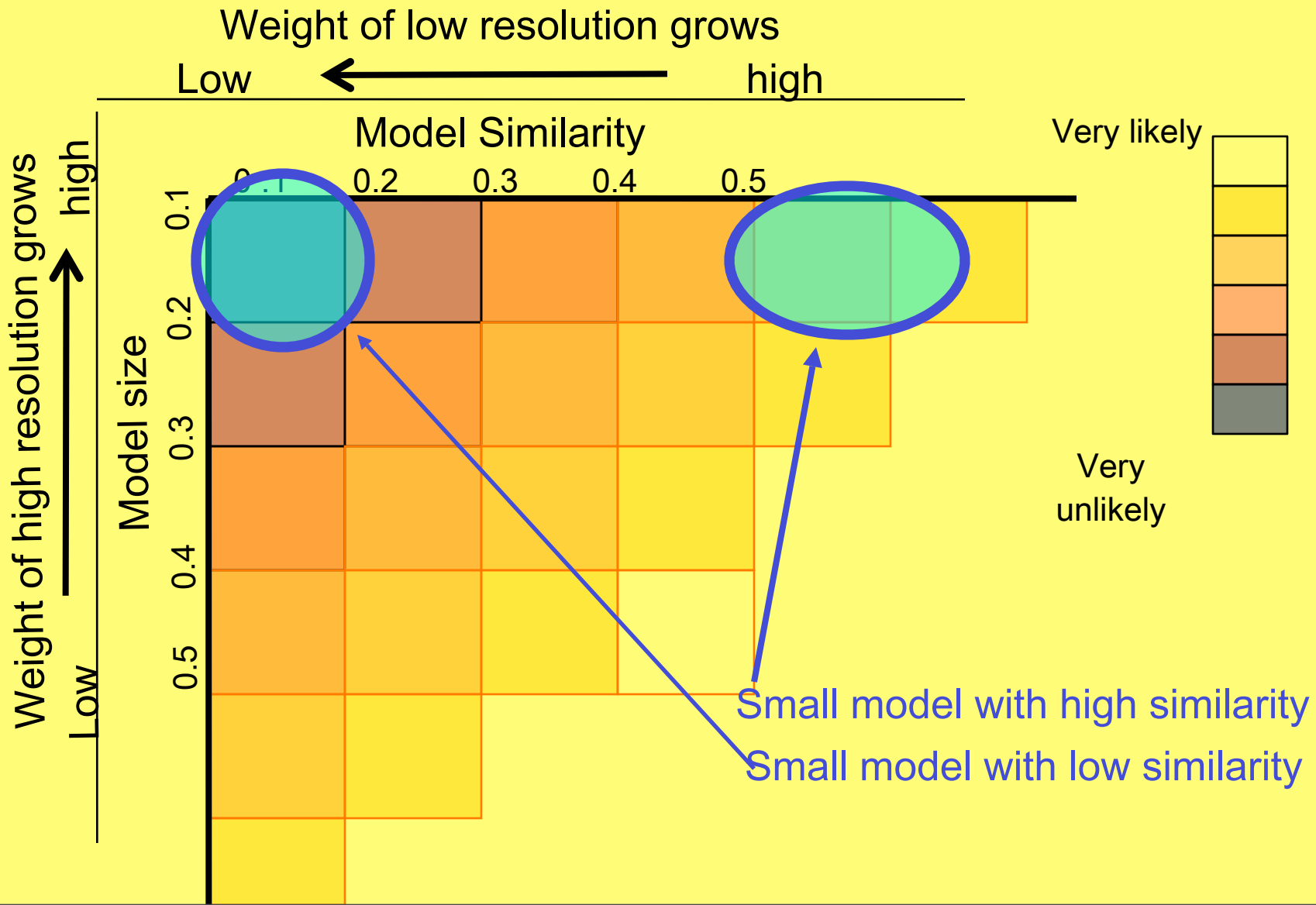
Can we find solution?



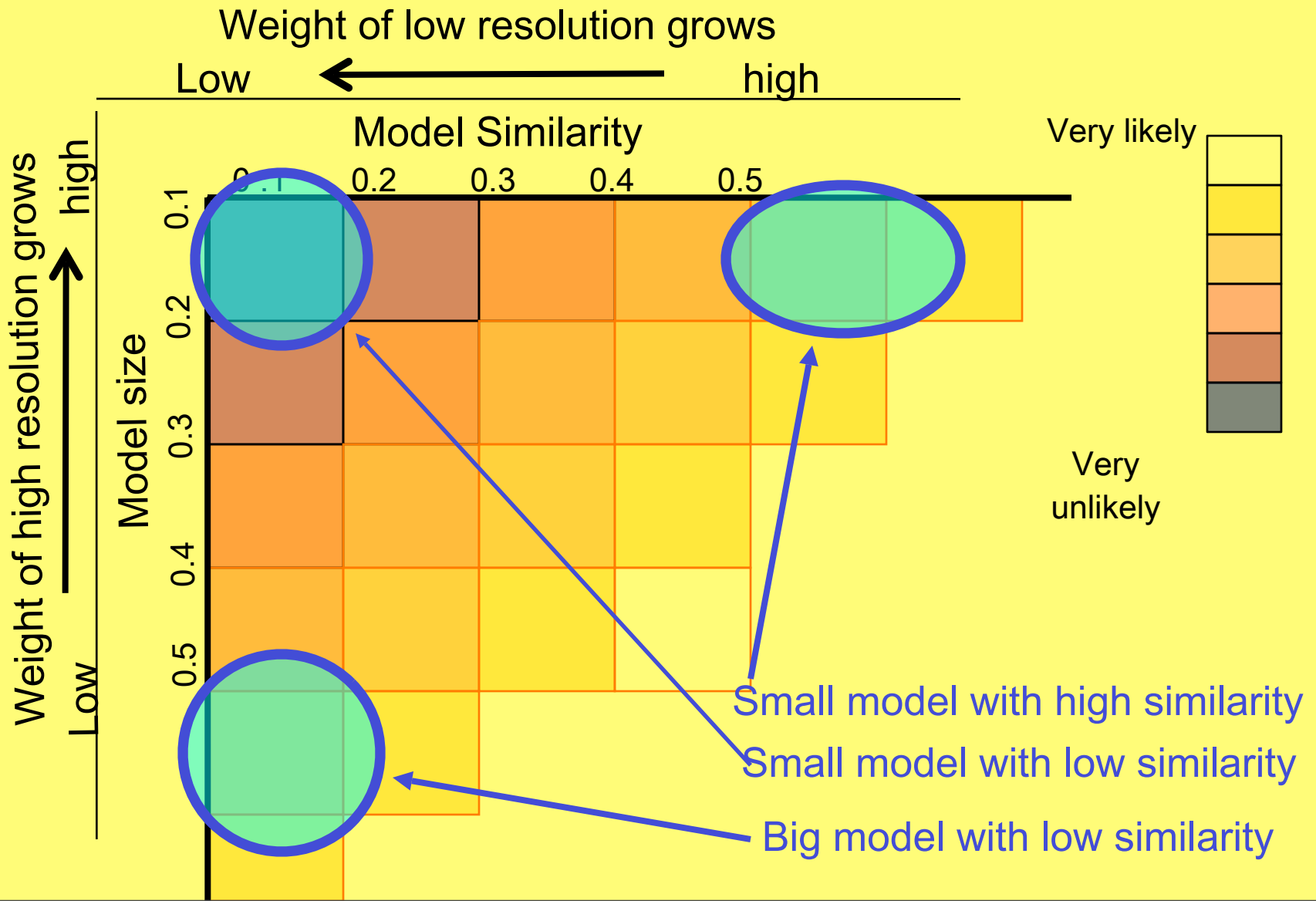
Can we find solution?



Can we find solution?



Can we find solution?



What do you need to do before MR

1) Examine the data
2) Examine the model

Examine the data (e.g by sfcheck)

- Completeness of data
- Signal-to-noise
- Anisotropy (make correction?)
- Pseudo-translation
- Twinning
- Resolution

Sfcheck 1

Title: XXXXXXXX ?
Date: XX–XXX–XX
PDB code: XXXX

Crystal

Cell parameters:

a: 99.66 A b: 99.66 A c: 64.33 A
α: 90.00 β: 90.00 γ: 120.00

Space group: H 3

Structure Factors

Input

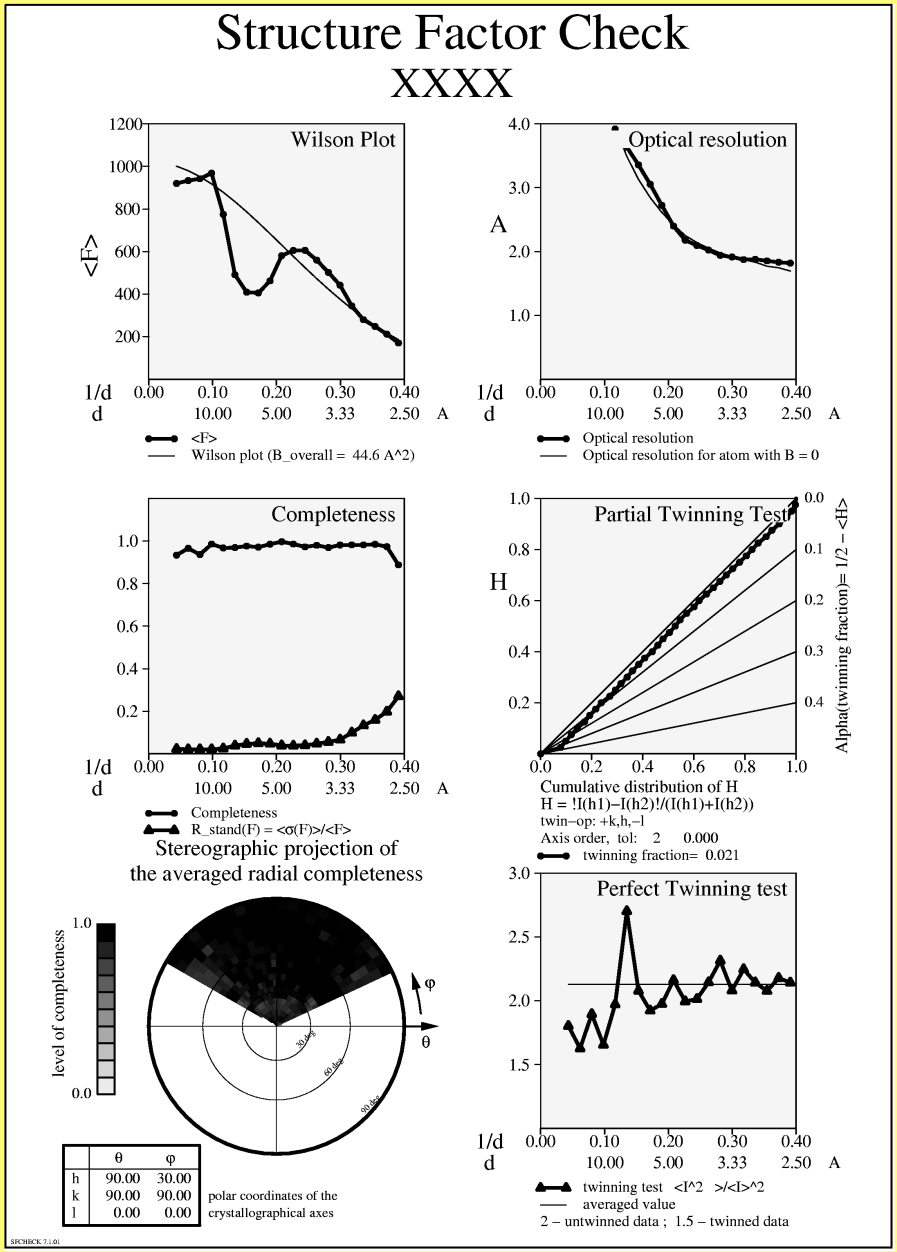
Nominal resolution range:		29.1 – 2.50 A
Reflections in file:		7974
Unique reflections	above 0:	7974
	above 1σ:	7973
	above 3σ:	5026

SFCHECK

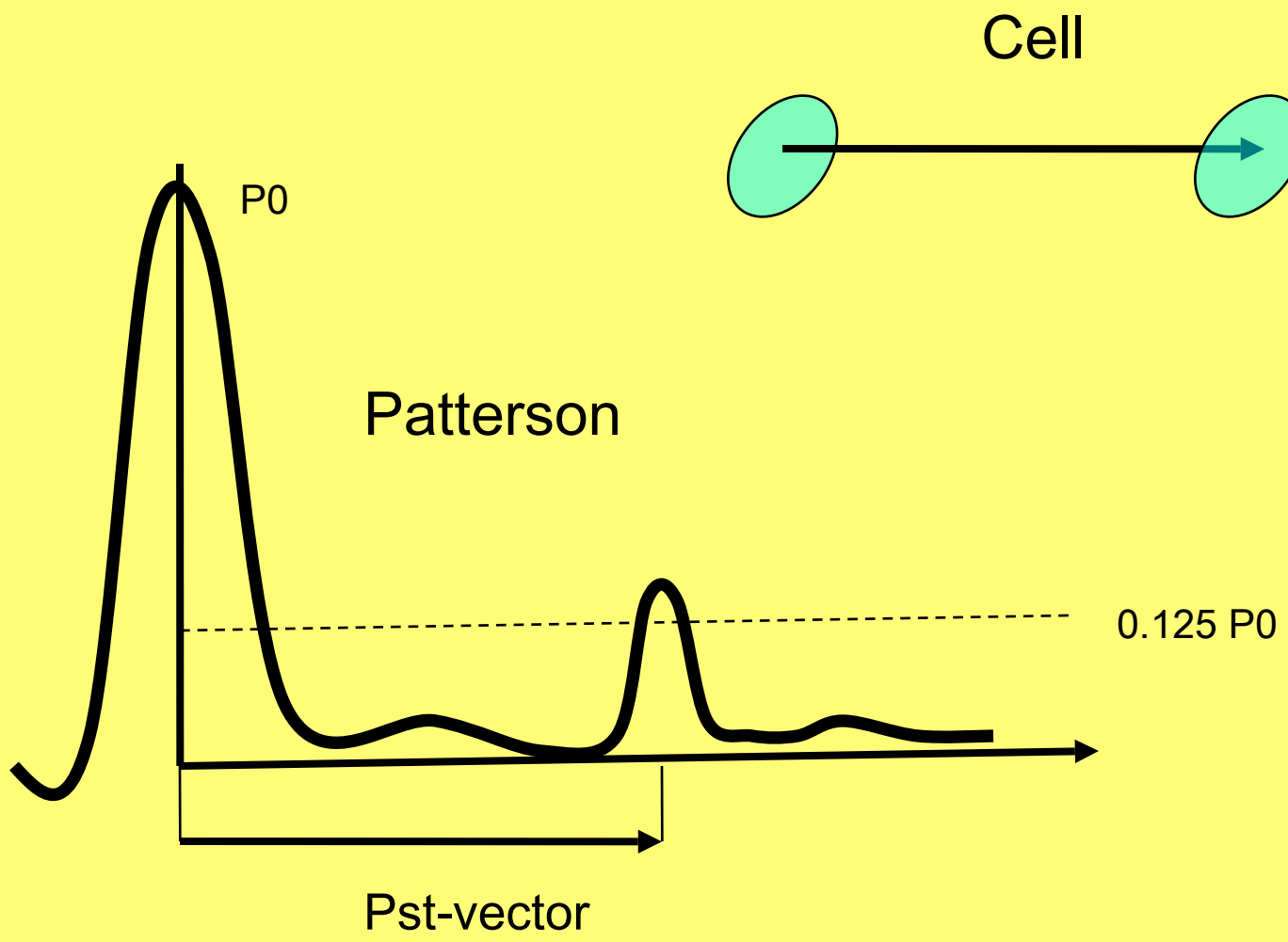
Nominal resolution range: 29.1 – 2.50 A
\05max. from input data, min. from author\05

Used reflections: 7974
Completeness: 96.7 %
R_stand(F) = <σ(F)>/<F> : 0.087
Anisotropic distribution of Structure Factors
ratio of eigen values: 0.6510 0.6510 1.0000
B_overall (by Patterson): 34.A^2
Optical resolution: 1.82 A
Expected opt. resol. for complete data set: 1.82 A
Estimated minimal error: 0.202 A
Pseudo–translation is not detected

Sfcheck 2


Structure Factor Check
XXXX
Wilson Plot
<F>
1/d
d
<F>
Wilson plot (B_overall = 44.6 A^2)
Optical resolution
A
Optical resolution
Optical resolution for atom with B = 0
Completeness
R_stand(F) = <σ(F)>/<F>
Stereographic projection of
the averaged radial completeness
level of completeness
φ
θ
polar coordinates of the
crystallographical axes
Partial Twinning Test
H
Alpha(twinning fraction)= 1/2 − <H>
Cumulative distribution of H
H = !I(h1)−I(h2)!/(I(h1)+I(h2))
twin−op: +k,h,−l
Axis order, tol: 2 0.000
twinning fraction= 0.021
Perfect Twinning test
twinning test <I^2 >/<I>^2
averaged value
2 − untwinned data ; 1.5 − twinned data
SFCHECK 7.1.01

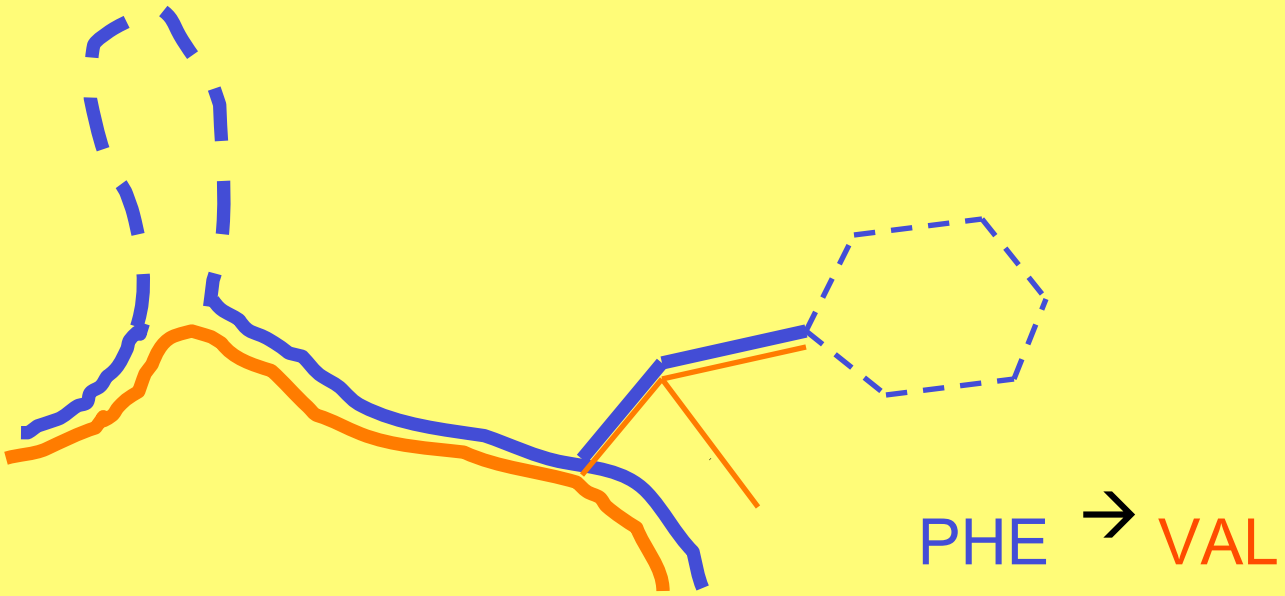
Pseudo-translation



Examine the model

- Look at the molecular shape and flexibility
- Check the sequence similarity
- Estimate the model size
- Choose the method of the model correction
- Estimate number of copies

Automatic correction of the model using sequence alignment



Initial model

Search model

Sequence

without alignment correction

P 21 21 2 2 models in a.u.c. Identity 27%

--- Rotation function ---

	Rf	Rf/sigma
RF 1	252.9	4.99
RF 2	230.5	4.55 *
RF 3	220.3	4.34
RF 4	206.1	4.06
RF 5	200.3	3.95
. . . .		

--- Translation function ---

RF	TF	Rfac	Score
1	3	0.554	0.206
2	3	0.554	0.205
6	1	0.556	0.199
3	4	0.556	0.199
. . .			

can not find solution

with alignment correction

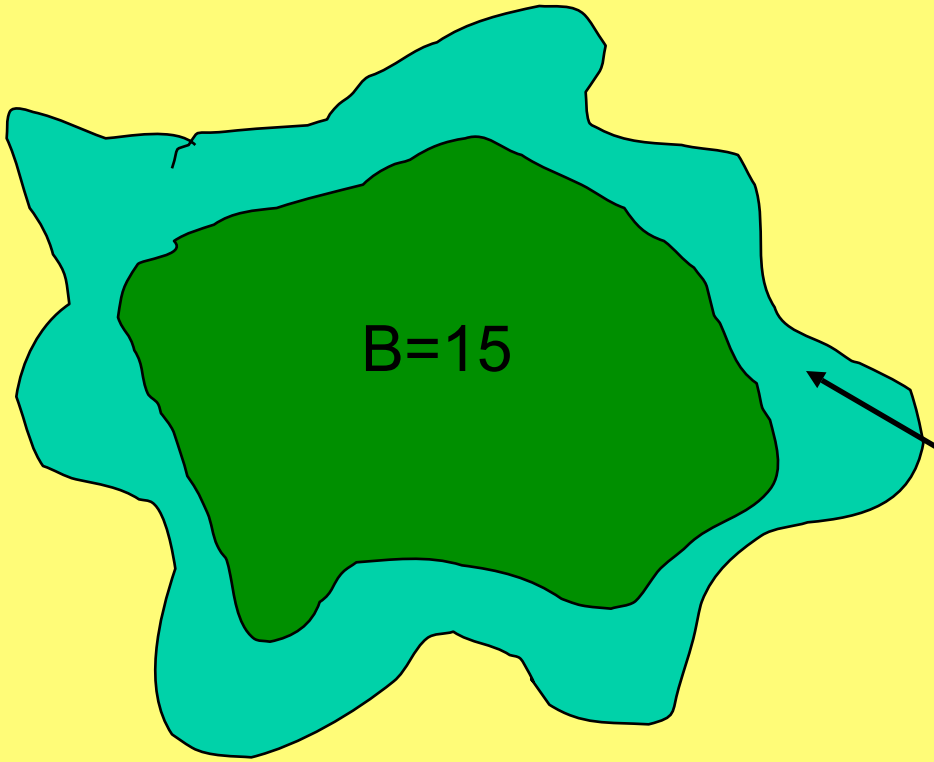
	Rf	Rf/sig
RF 1	329.2	5.27
RF 2	304.9	4.88
RF 3	282.6	4.52 *
RF 4	249.6	3.99
. . . .		
RF 18	205.7	3.29 *

--- Translation function ---

RF	TF	Rfac	Score
3	2	0.556	0.197
1	4	0.559	0.194
18	2	0.560	0.194
2	4	0.562	0.186
with fixed model			
18	1	0.547	0.233
20	4	0.558	0.200
2	4	0.557	0.200

Model improvement

Set atomic B values according to accessible surface area



$B\ (A^2) = 15\ (A^2) + 10\ SA\ (A^2)$

SA – surface area

Expected number of copies

$$\text{No of copies} = 0.8 \frac{\text{Volume of the au}}{\text{Volume of the molecule}}$$

Time to have a break

NMR model

- Rotation function

  Use as single model or

  Averaged individual RF $\Leftrightarrow$ Averaged intensities

- Translation function

  Use as single model or

  Averaged individual TF

Special techniques of molecular replacement

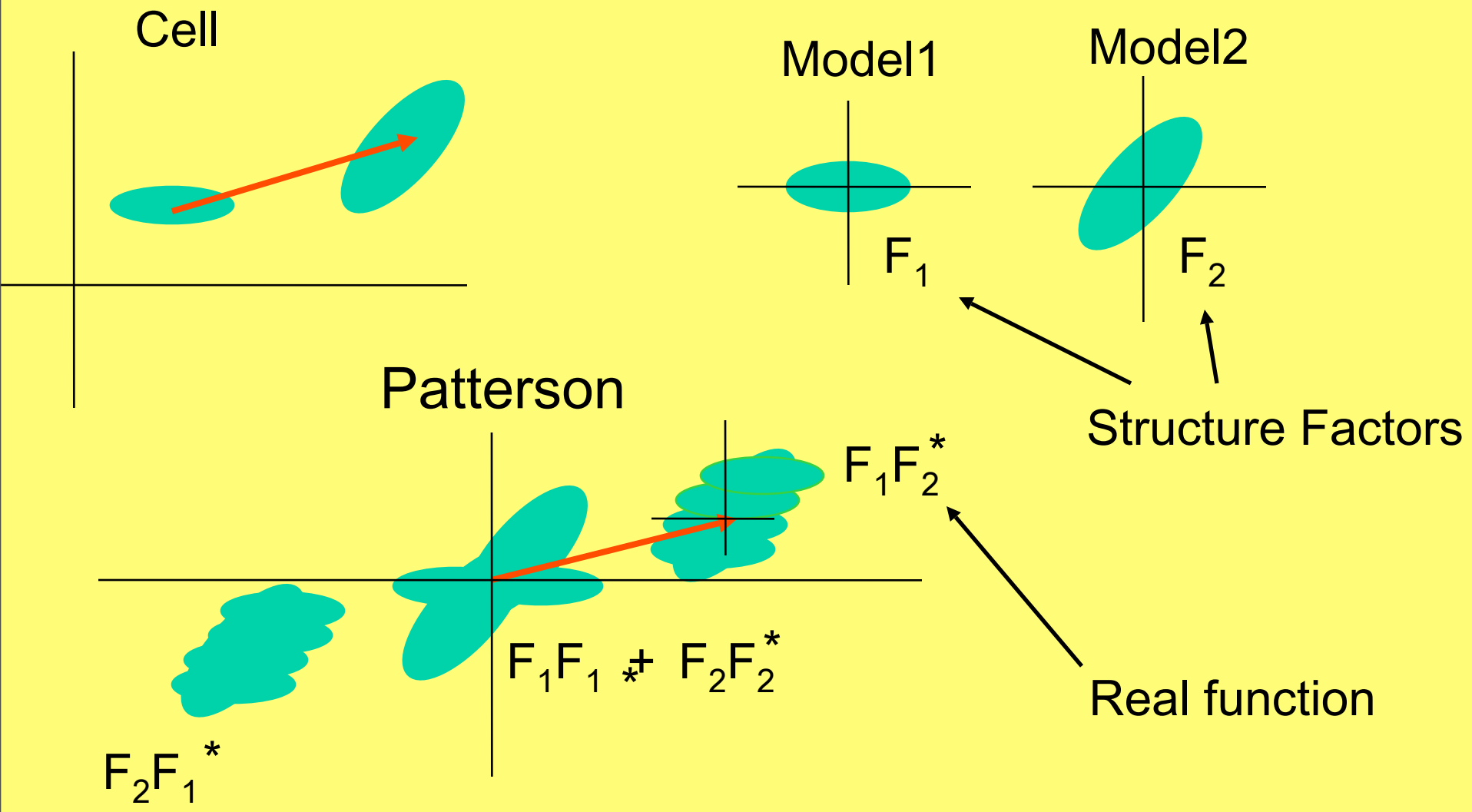
- Locked Rotation function
- Multi-copy search
- Use phases after Refinement
- Spherically Average Phased Translation function

Self rotation and locked RF

Peaks selected from the self rotation function can be used for locked cross rotation function. Locked rotation function is averaged RF according to NCS

```
Sol_ Space group : H 3
Sol_--- Rotation function ---
      theta     phi     chi        Rf   Rf/sig
RF 1   47.90    67.54 158.59     1190    6.37
RF 2   79.14 -166.90  89.47     1050    5.05
RF 3   97.26 -139.11 145.11      848    4.55
RF 4 137.75 -156.31  94.80      843    4.44
` ` ` `
Sol_--- Locked Rotation function ---
        theta    phi     chi        Rf   Rf/sig
RF 1 127.99   139.59 122.00     2034    6.90
RF 2 123.49   -52.42 122.11     1979    6.80
RF 3  71.51 -171.88 105.08     1541    5.16
RF 4  44.71 -107.06 154.01     1500    4.45
```

Multi-copy search



Difficult case

Space group H3
Resolution 1.8A

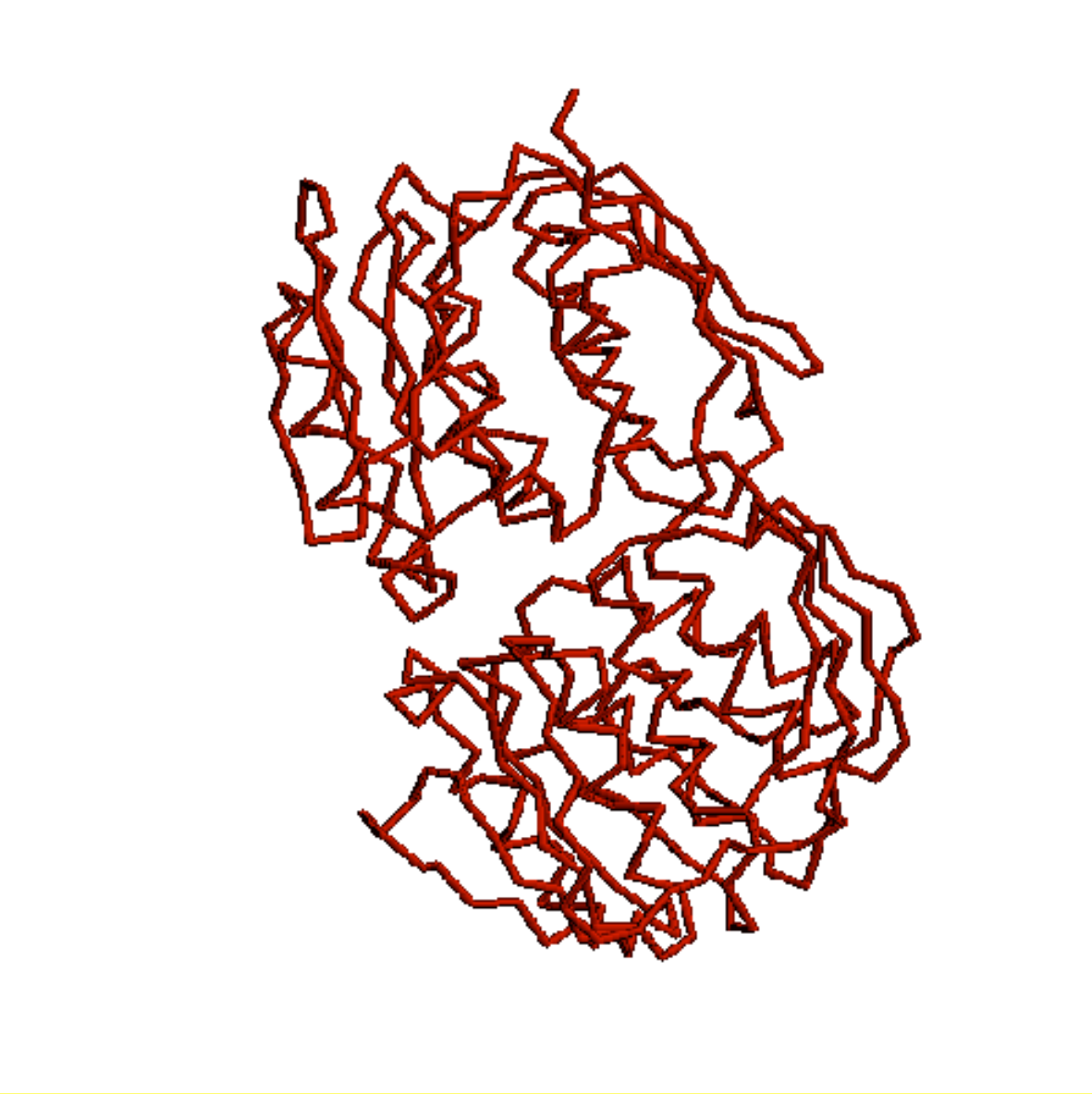
One molecule in a.u.c
Identity 35%

1. Using complete model - failed

2. Using domains separately - failed

3. Multi-copy search - success

Initial model



RF & TF

	complete model			domain 1			domain 2		
	RF	TF	Score	RF	TF	Score	RF	TF	Score
1	17	7	0.208	10	2	0.211	14	1	0.205
2	1	1	0.204	15	7	0.209	8	1	0.204
3	8	8	0.203	12	1	0.206	26	8	0.203
4	14	7	0.199	5	3	0.202	5	6	0.203
5	5	6	0.196	3	1	0.202	10	1	0.202
6	11	2	0.195	6	7	0.201	1	2	0.200
7	18	5	0.191	7	5	0.199	.	.	.
8	16	6	0.191	4	2	0.198	21	24	1	0.192
9	12	1	0.191	1	2	0.197	.	.	.
.	.	.		.	.	.			

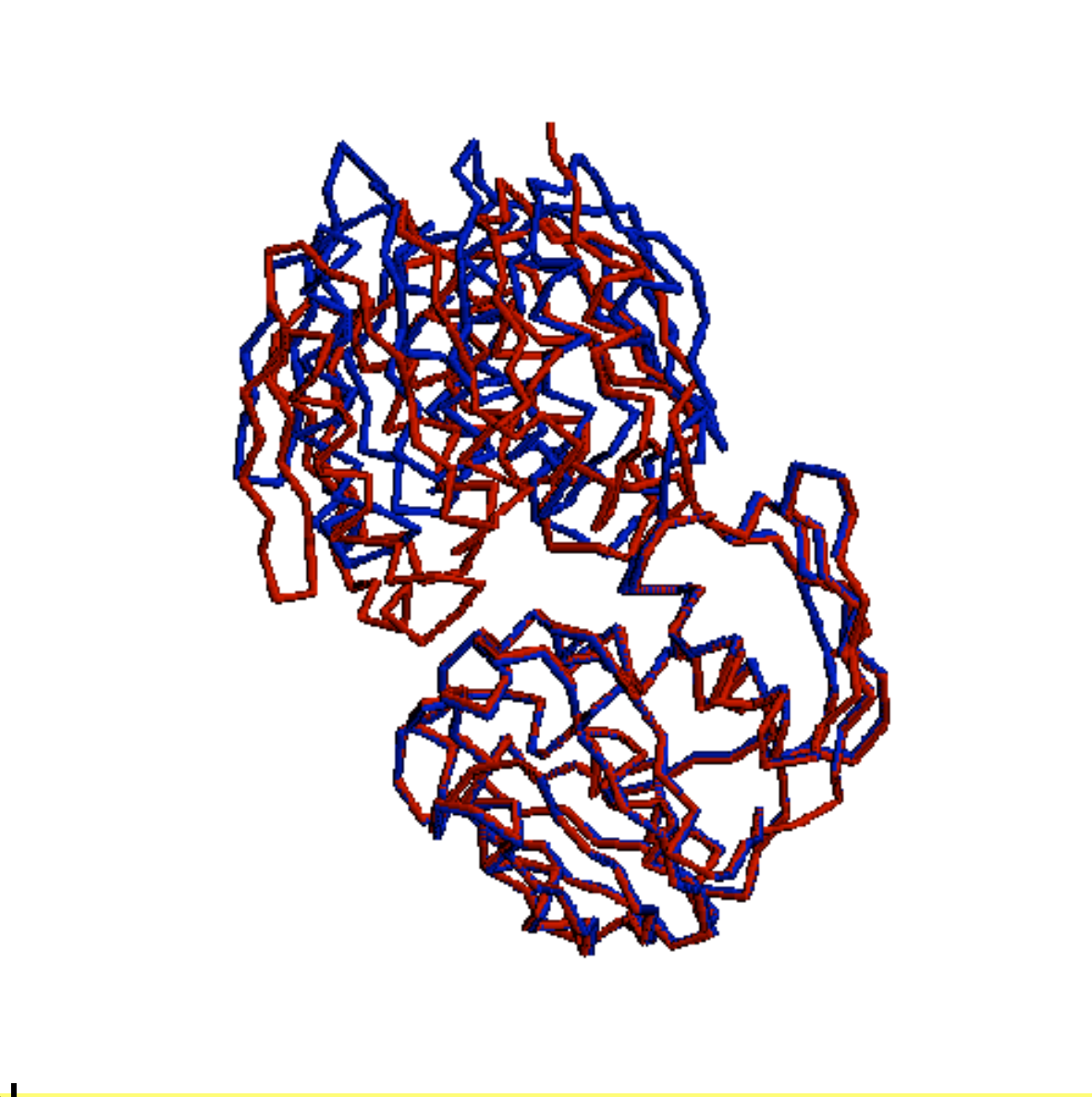
Domain 1 + 2 : Multi-copy search

multi-copy Search

R1	R2	STF	TF	PFmax	PFmin	Score
1	1	2	2	0.65	-11.55	0.209
1	2	5	1	0.98	-15.90	0.212
1	3	1	1	0.99	-12.73	0.223
.	.	.	.			
7	24	3	1	0.99	-13.59	0.248
.	.	.				

domain1 (rf7) , domain2(rf24)

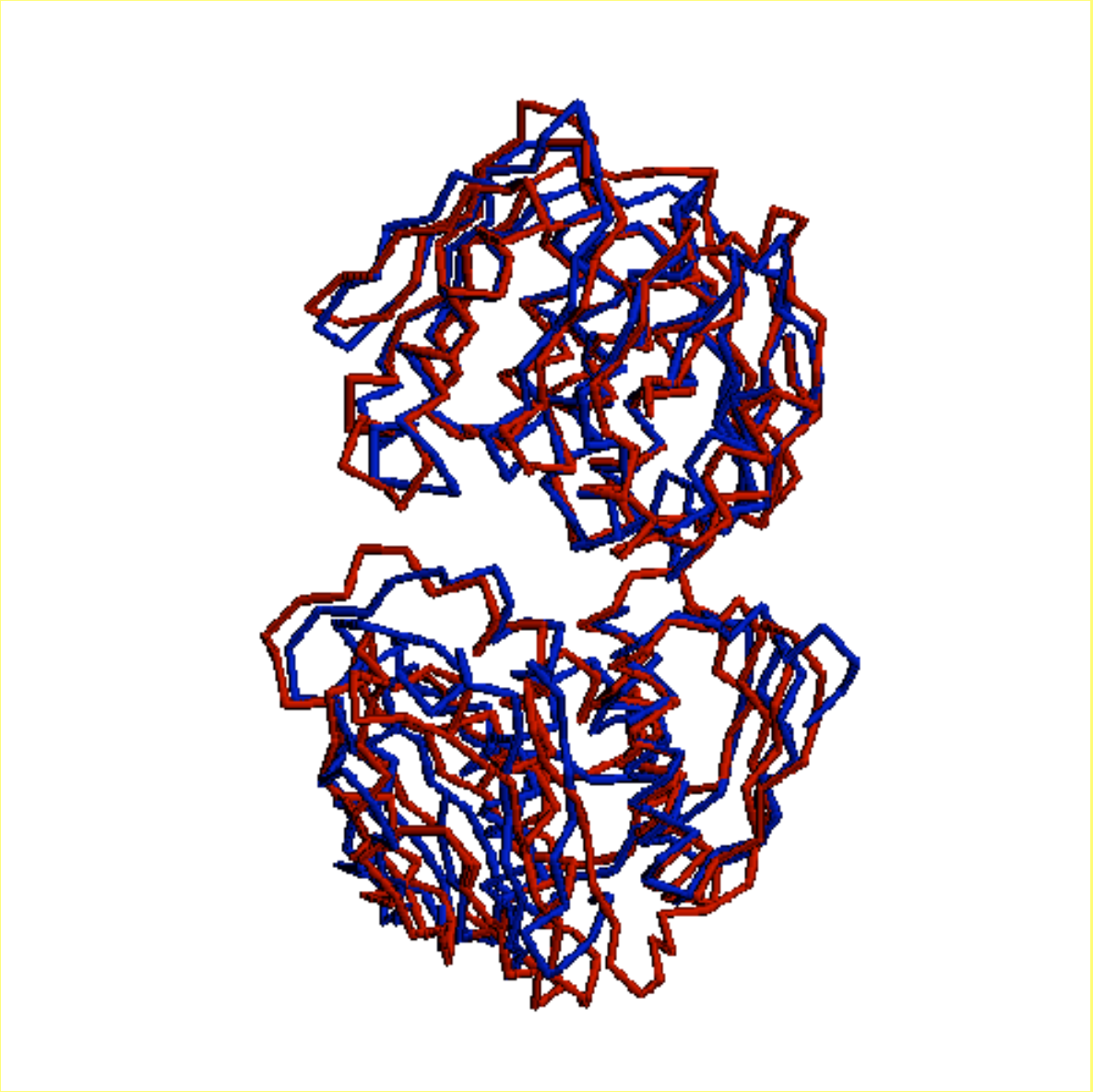
Initial model and MR solution



MR

initial

MR solution and final structure



MR

final

Use Phases after Refinement

Example: Domain motions - 1tj3

"unknown" structure (1tj3)

search model with sequence identity 100%

Search for the whole molecule using standard MR protocol failed because of domain flexibility.

Search by domains using standard MR protocol failed because of small size of the second domain.

Structure was then solved manually in three steps:

1) standard MR search for larger domain;

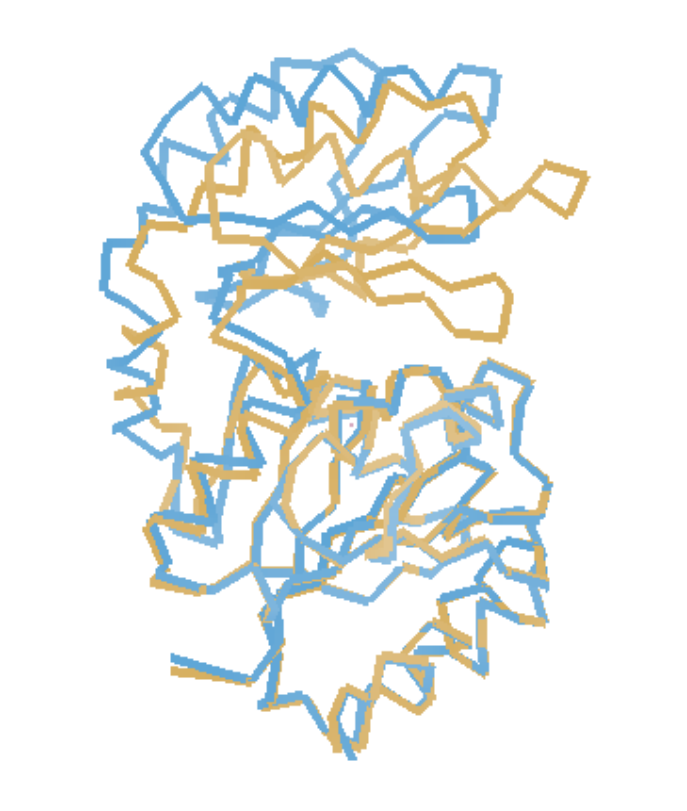
2) refinement of the partial model;

3) search for smaller domain in the masked map (generated from REFMAC's FWT and PHIWT)

Example: Domain motions - 1tj3

"unknown" structure (1tj3)

search model with sequence identity 100%



Search for the whole molecule using standard MR protocol failed because of domain flexibility.

Search by domains using standard MR protocol failed because of small size of the second domain.

Structure was then solved manually in three steps:

1) standard MR search for larger domain;

2) refinement of the partial model;

3) search for smaller domain in the masked map (generated from REFMAC's FWT and PHIWT)

Fitting model into X-ray or EM map

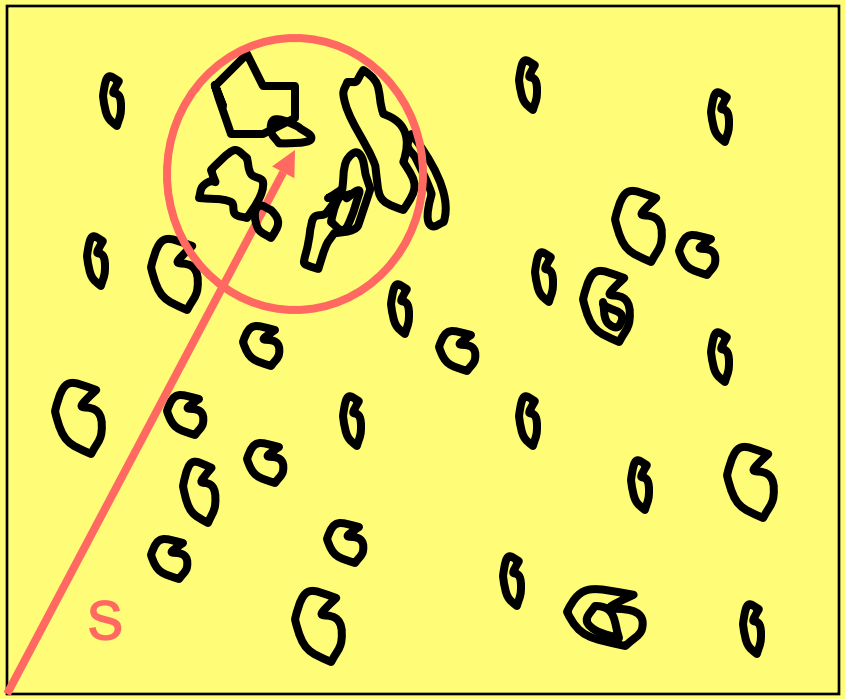
1. find orientation (RF)
2. find position (PTF)

Alternative approach:

1. find position
2. find orientation

Spherically Averaged Phased Translation Function (SAPTF)

Map $\rho(r)$



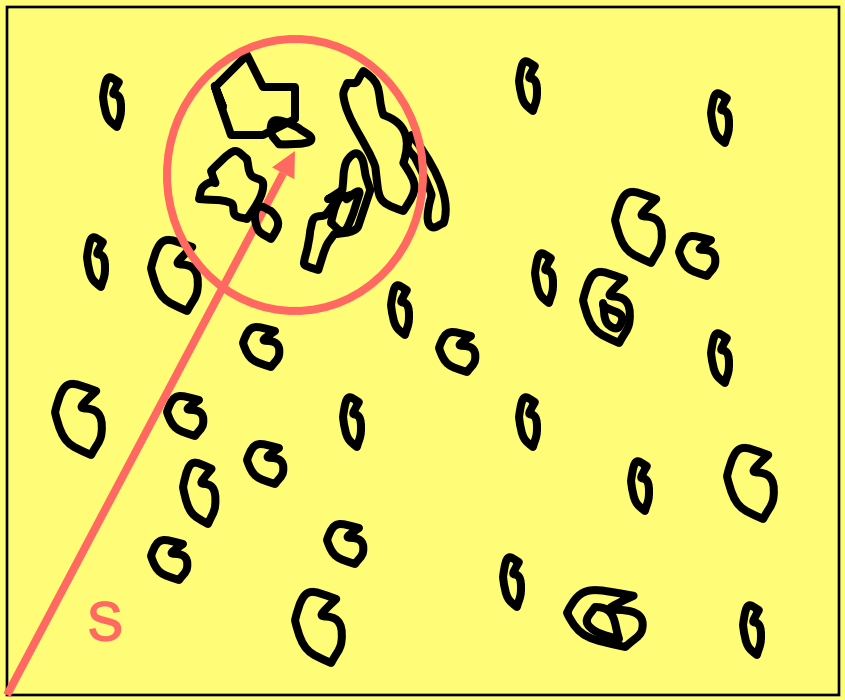
radial distribution

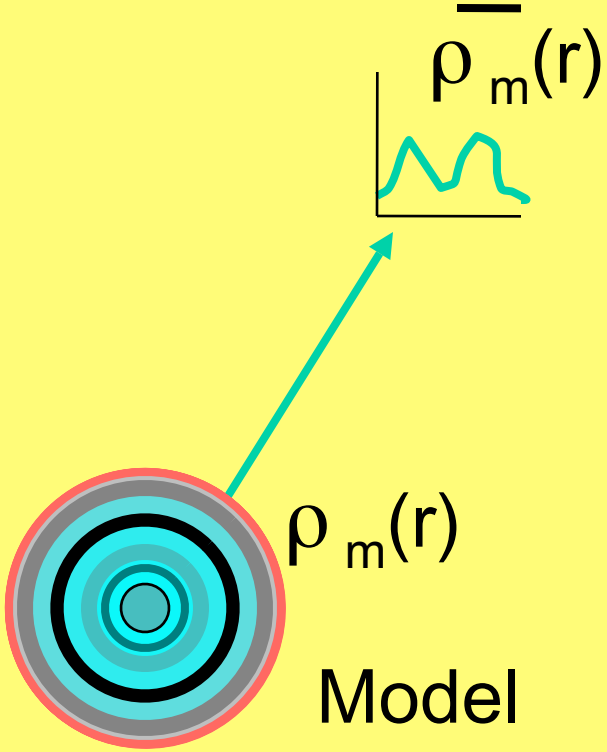
$\rho_m(r)$

Model

$$SAPTF(s) = \int \overline{\rho_s}(r)\ \overline{\rho_m}(r)\ r^2\ dr$$

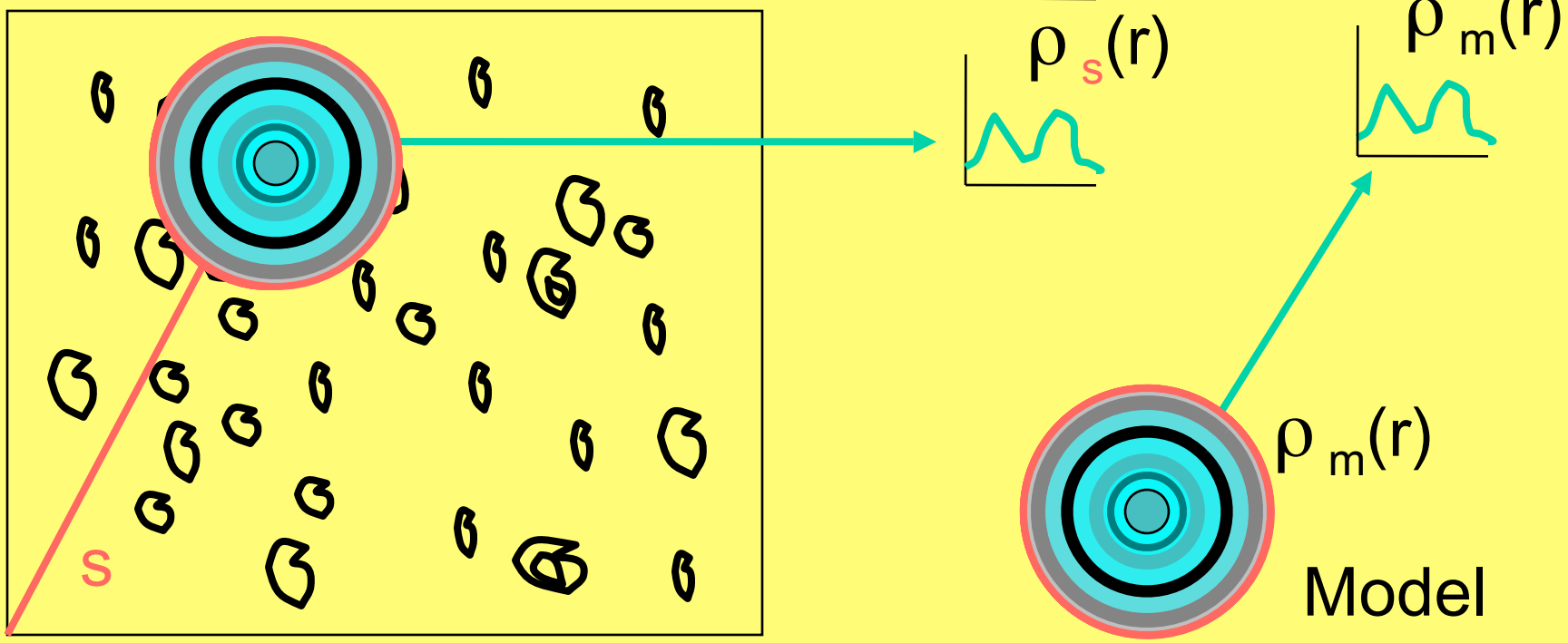
Spherically Averaged Phased Translation Function (SAPTF)





$$SAPTF(s) = \int \overline{\rho_s}(r)\, \overline{\rho_m}(r)\ r^2\, dr$$

Spherically Averaged Phased Translation Function (SAPTF)



$$SAPTF(s) = \int \overline{\rho_s}(r)\, \overline{\rho_m}(r)\; r^2\, dr$$

SAPTF as Fourier series

By expanding SAPTF into spherical harmonics it is possible to represent it as a Fourier series

$$SAPTF(s) = \int \overline{\rho_s}(r)\ \overline{\rho_m}(r)\ r^2\, dr =$$

$$= \sum_h A_h \exp(2\pi i\, hs)$$

$$A_h = \sum_n F_h c_{00n}(R)\, j_0(2\pi Ra)\, b_{00n}$$

Algorithm

1. Find position:
   Spherically averaged phased translation function

2. Find orientation:
   Local phased rotation function

3. Check and refine position :
   Phased translation function