Molecular Replacement

(Alexei Vagin's lecture)

Garib N Murshudov
MRC-LMB, Cambridge, UK

Contents

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

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

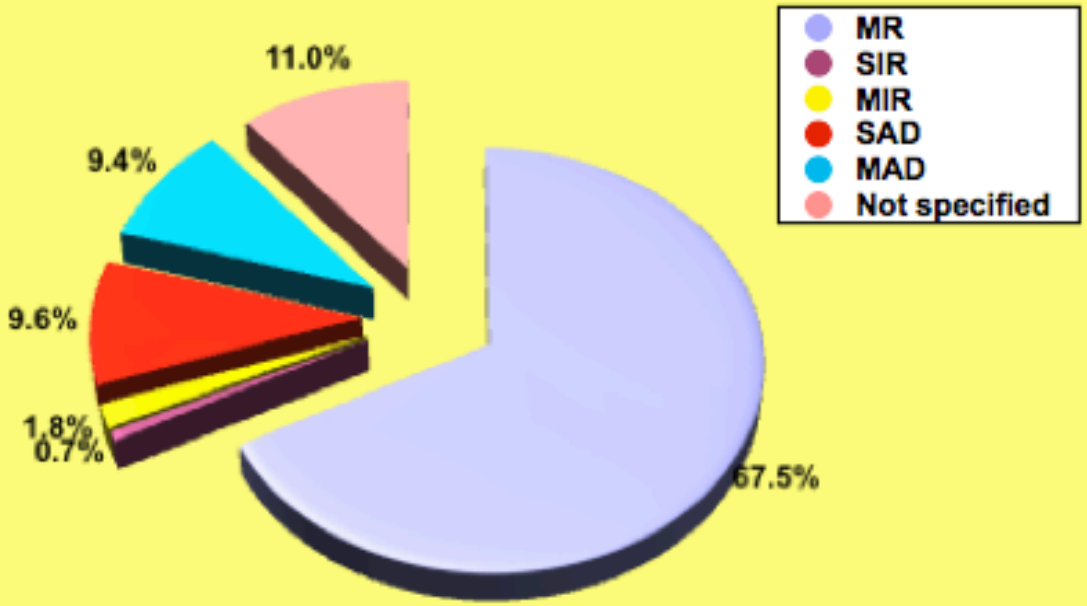


Diagram showing the percentage of structures in the PDB solved by different techniques

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

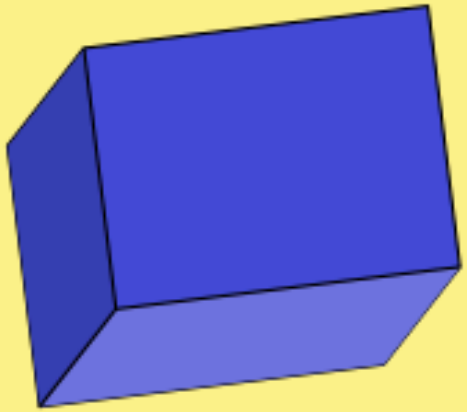
21% of structures are solved by experimental phasing

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
```



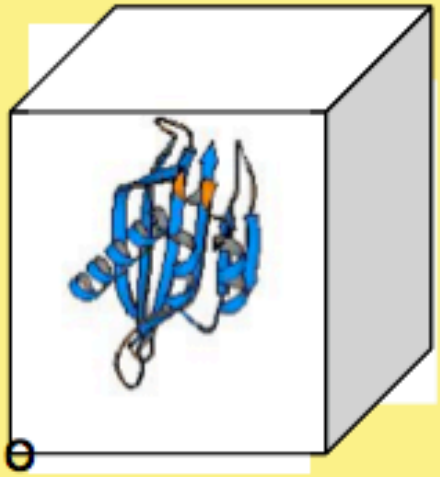
origin o

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

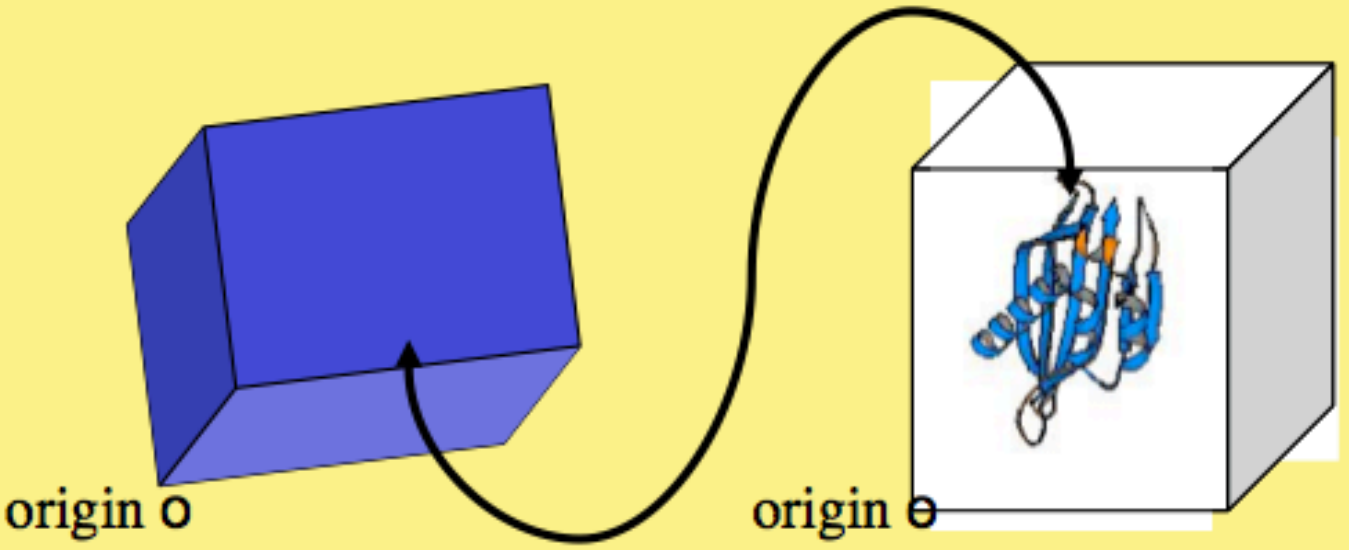
If we can find the rotation and translation that puts the model in the correct position in the crystal cell, THEN we can calculate phases.

unknown structure

```
MGDKPIWEQIGSSFIQHYYQLFDNDRTQLGAIY
IDASCLTWEGQQFQGKAAIVEKLSSLPFQKIQH
SITAQDHQPTPDSCIISMVVGQLKADEDPIMGF
HQMFLLKNINDAWVCTNDMFRLALHNFG
```

known structure

```
PSPLLVGREFVRQYYTLLNKAPEYLHRFYGRNSSY
VHGGVDASGKPQEAVYGQNDIHHKVLSLNFSECHT
KIRHVDAHATLSDGVVVQVMGLLSNSGQPERKFMQ
TFVLAPEGSVPNKFYVHNDMFRYEDE
```



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

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

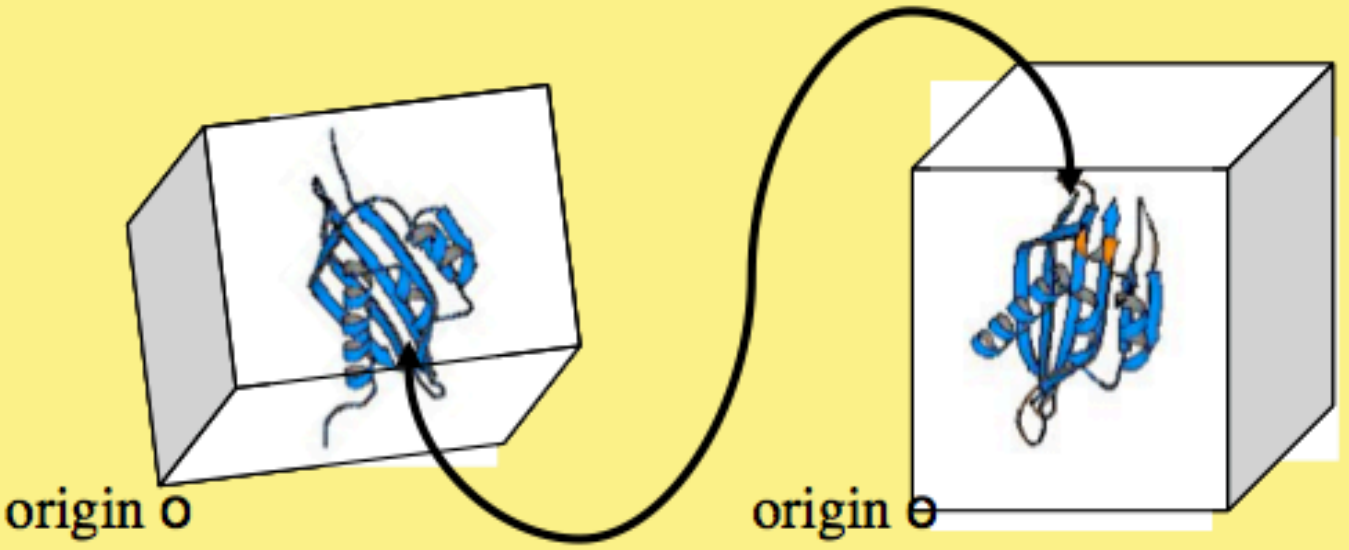
If we can find the rotation and translation that puts the model in the correct position in the crystal cell, THEN we can calculate phases.

unknown structure

```
MGDKPIWEQIGSSFIQHYYQLFDNDRTQLGAIY
IDASCLTWEGQQFQGKAAIVEKLSSLPFQKIQH
SITAQDHQPTPDSCIISMVVGQLKADEDPIMGF
HQMFLLKNINDAWVCTNDMFRLALHNFG
```

known structure

```
PSPLLVGREFVRQYYTLLNKAPEYLHRFYGRNSSY
VHGGVDASGKPQEAVYGQNDIHHKVLSLNFSECHT
KIRHVDAHATLSDGVVVQVMGLLSNSGQPERKFMQ
TFVLAPEGSVPNKFYVHNDMFRYEDE
```



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

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

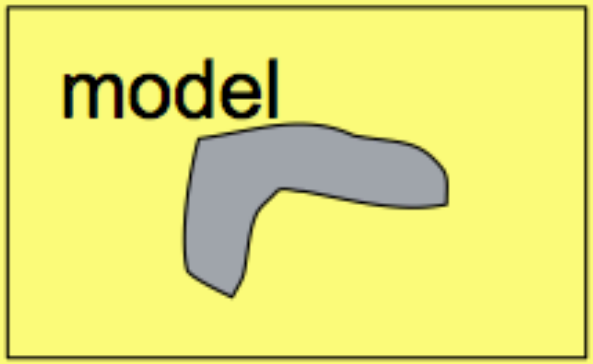
Conventional Molecular Replacement

Functions of molecular replacement

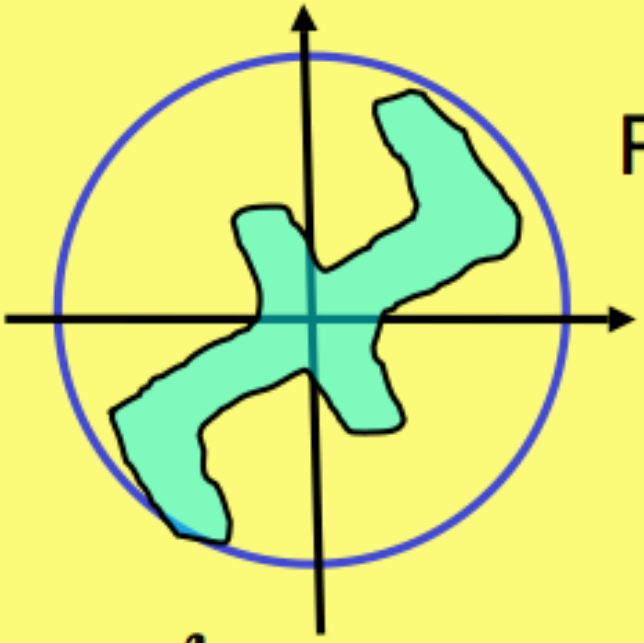
- Cross Rotation function
- Self Rotation function
- Translation function
- Phased Translation function
- Fast Packing function

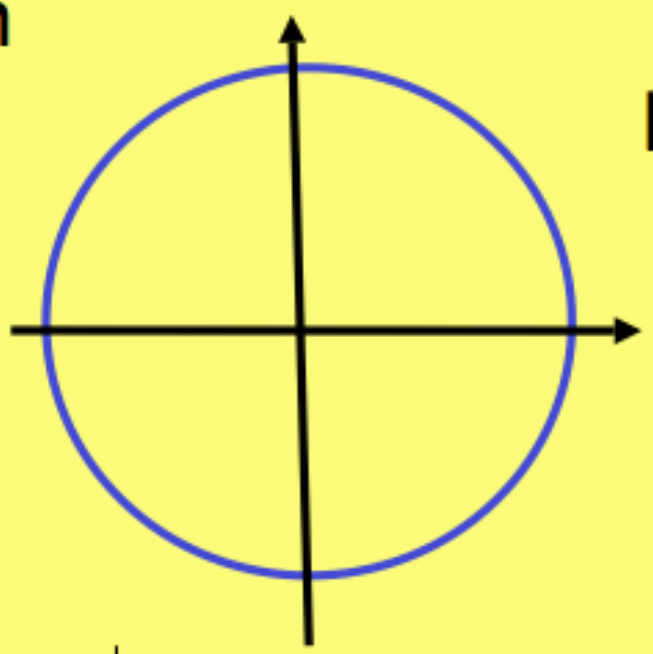
Cross Rotation Function

Cross Rotation Function



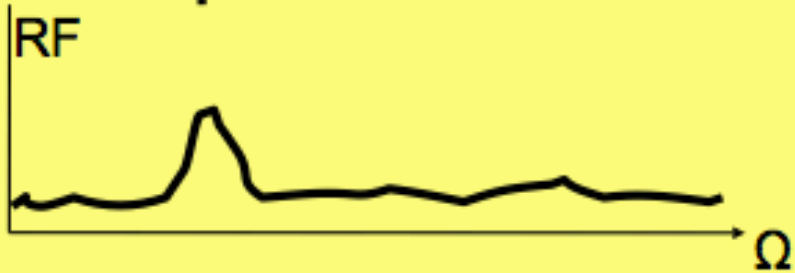
Patterson



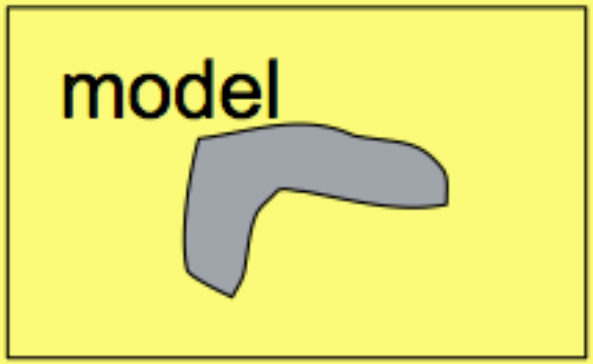


$$RF(\Omega) = \int P_{obs}(r)\ \ \Re_{\Omega}\{ P_{mod}(r)\}\ \ dr$$

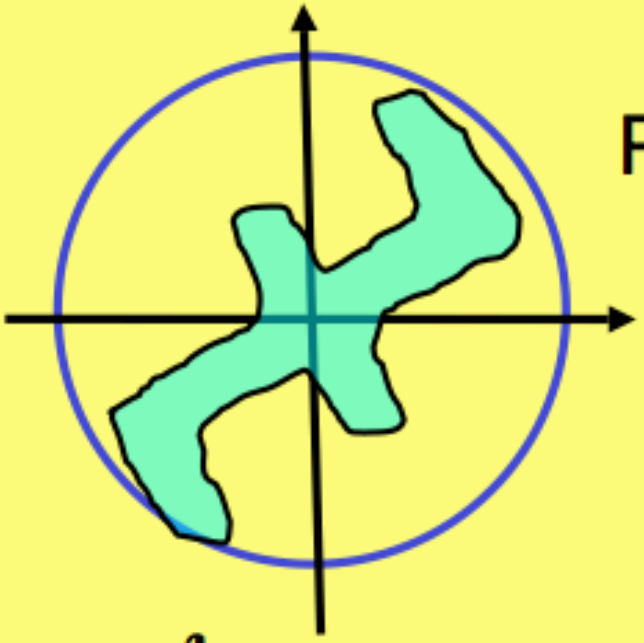
$\Re_{\Omega}$ - rotation operator

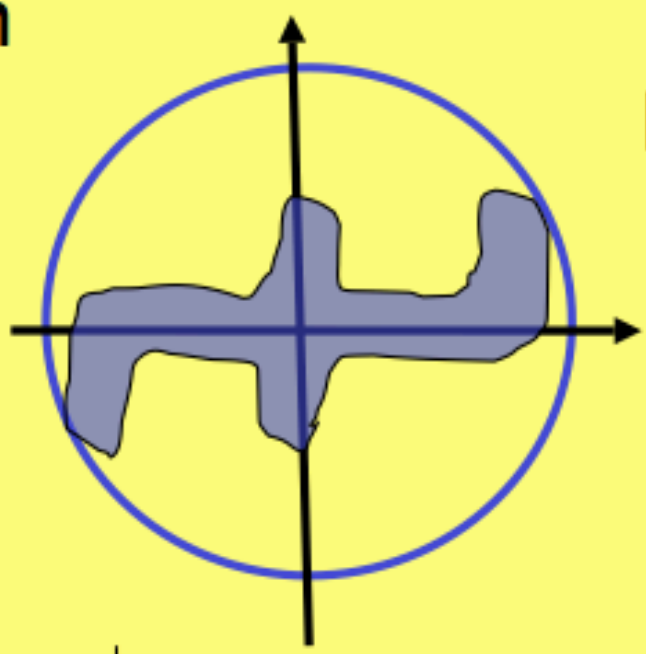


Cross Rotation Function



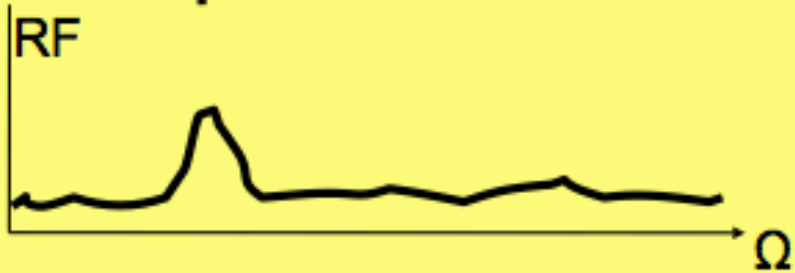
Patterson



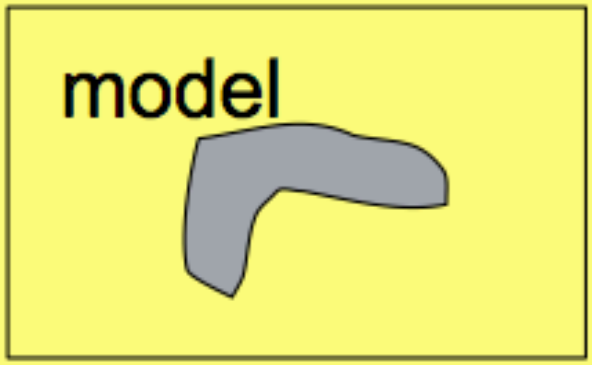


$$RF(\Omega) = \int P_{obs}(r)\ \ \Re_{\Omega}\{ P_{mod}(r)\}\ \ dr$$

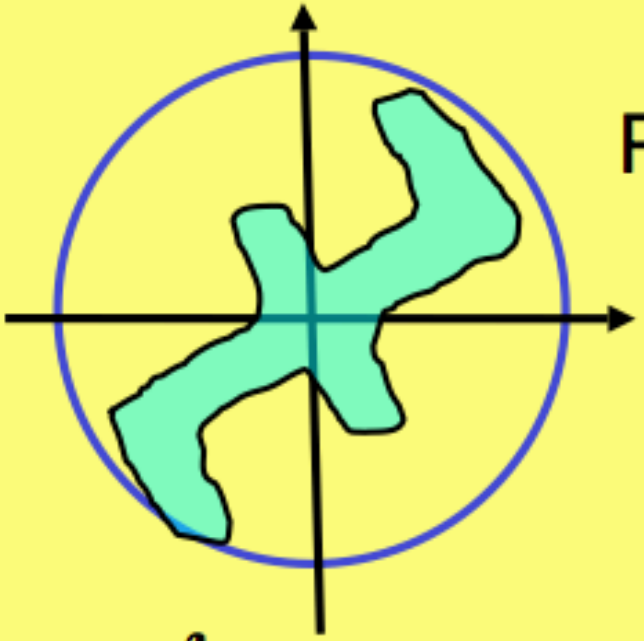
$\Re_{\Omega}$ - rotation operator

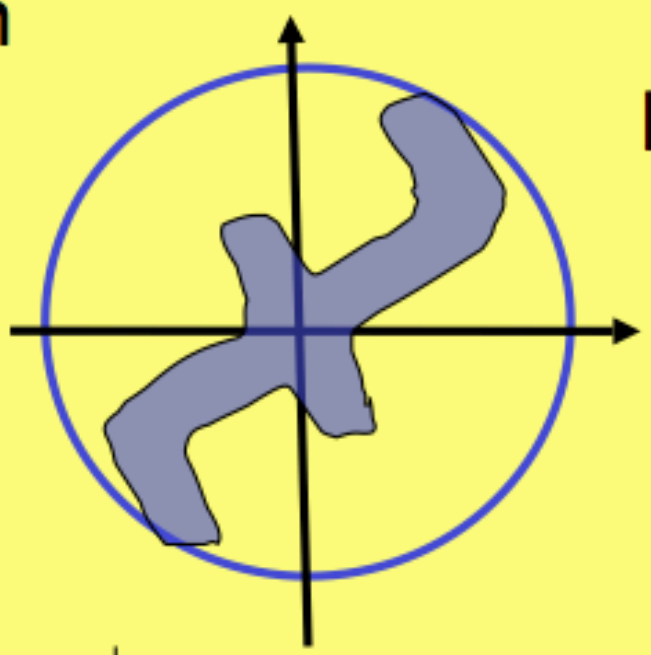


Cross Rotation Function



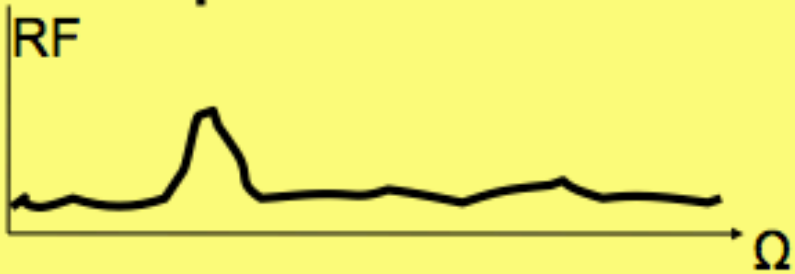
Patterson



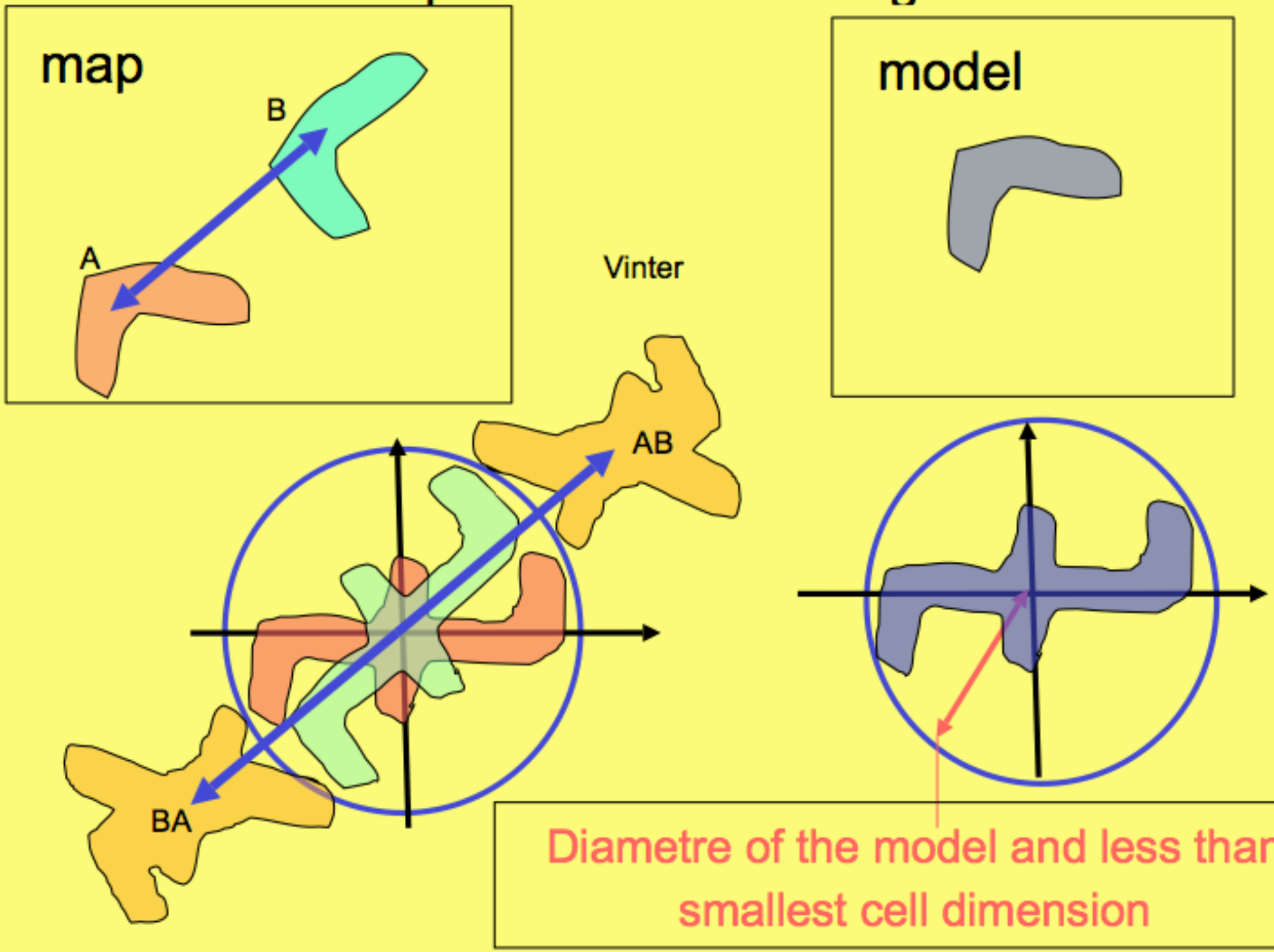


$$RF(\Omega) = \int P_{obs}(r)\ \ \Re_{\Omega}\{ P_{mod}(r)\}\ \ dr$$

$\Re_{\Omega}$ - rotation operator

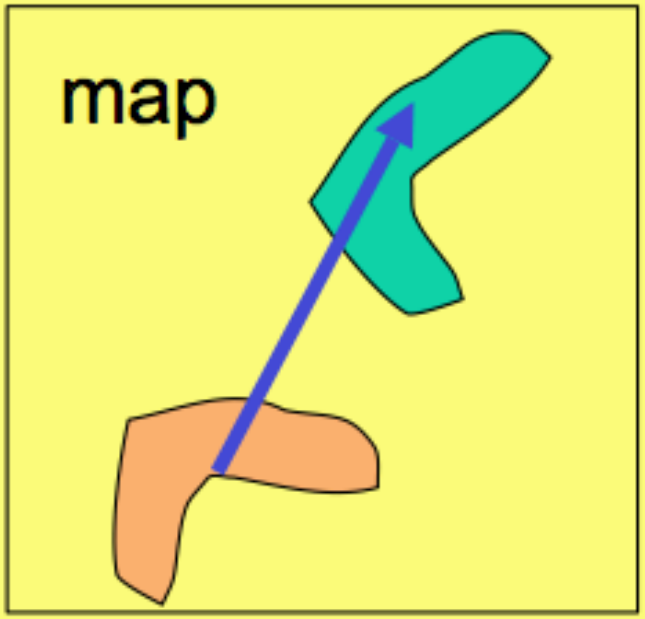


Optimal radius of integration

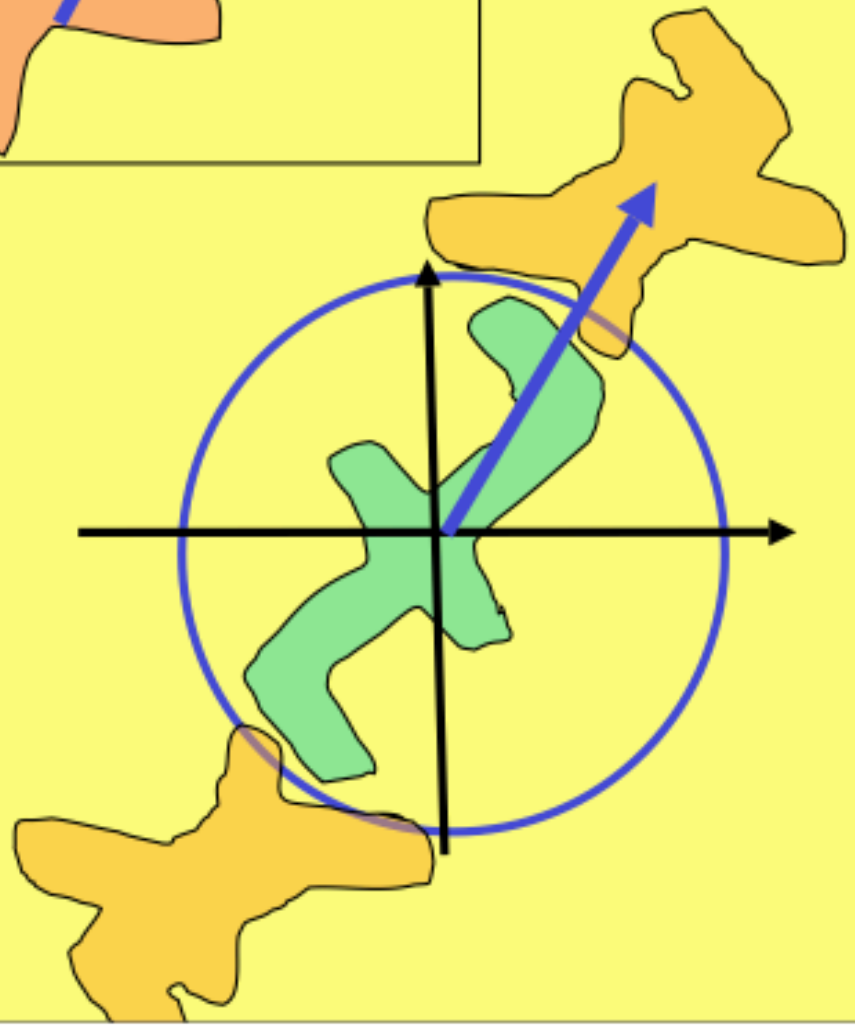


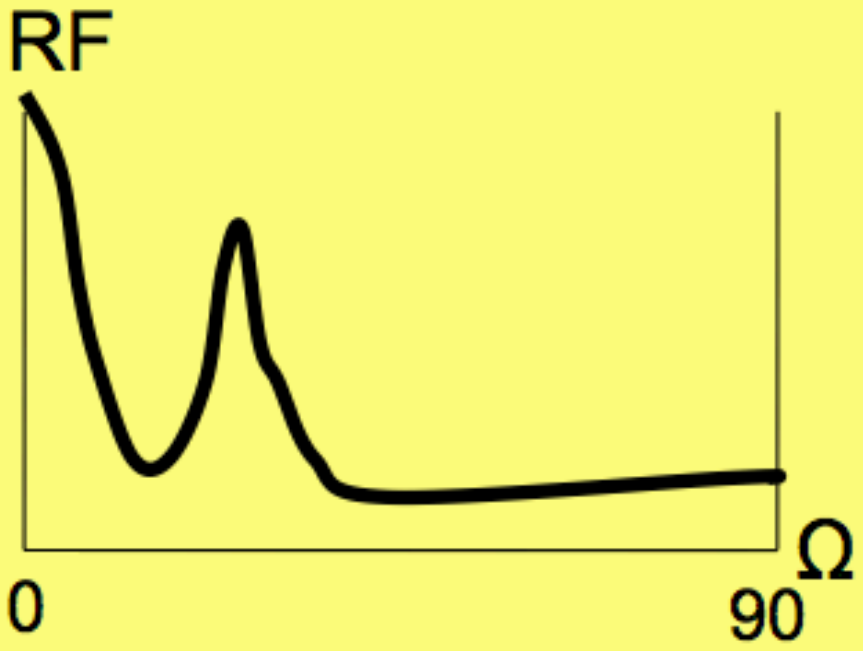
Self Rotation function

Self Rotation Function

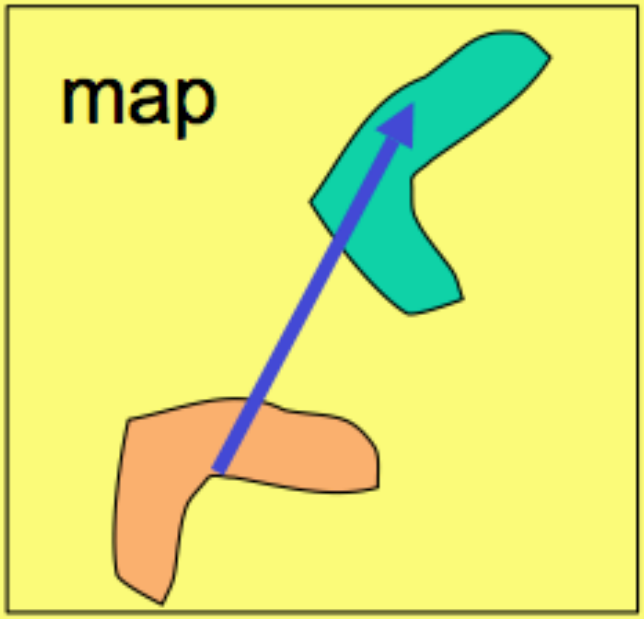


$$RF(\Omega) = \int P_{obs}(r)\ \ \Re_{\Omega}\{ P_{obs}(r)\}\ dr$$

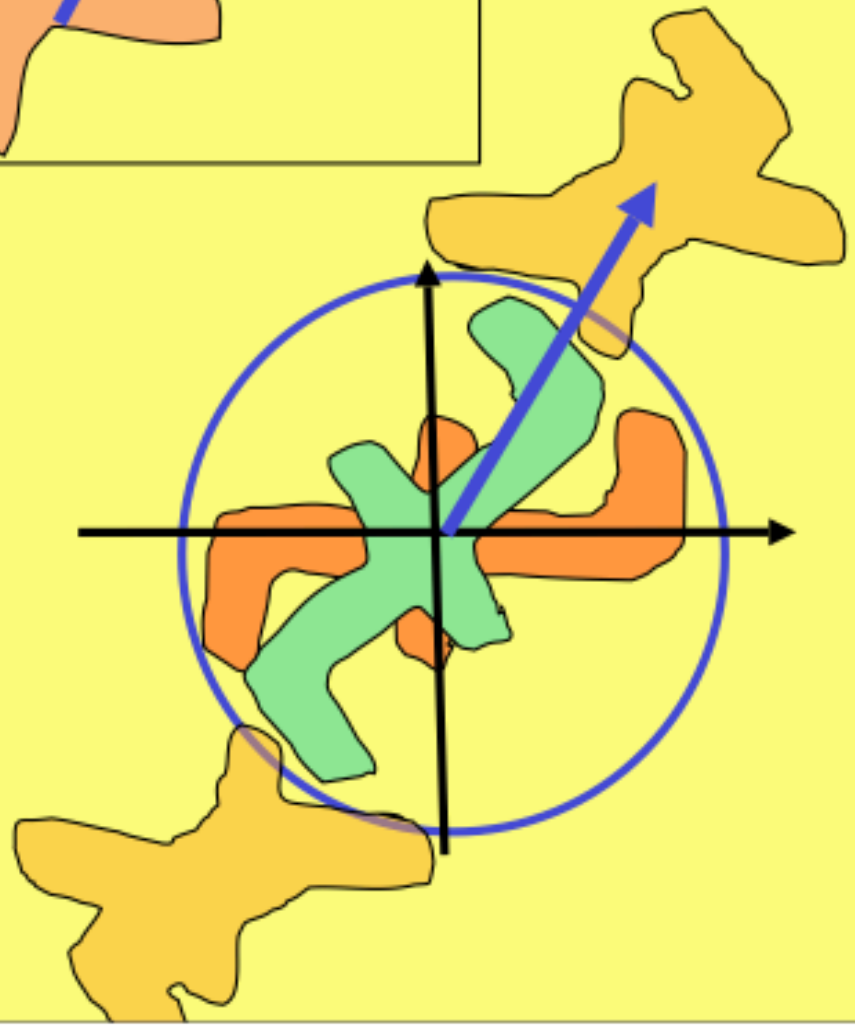


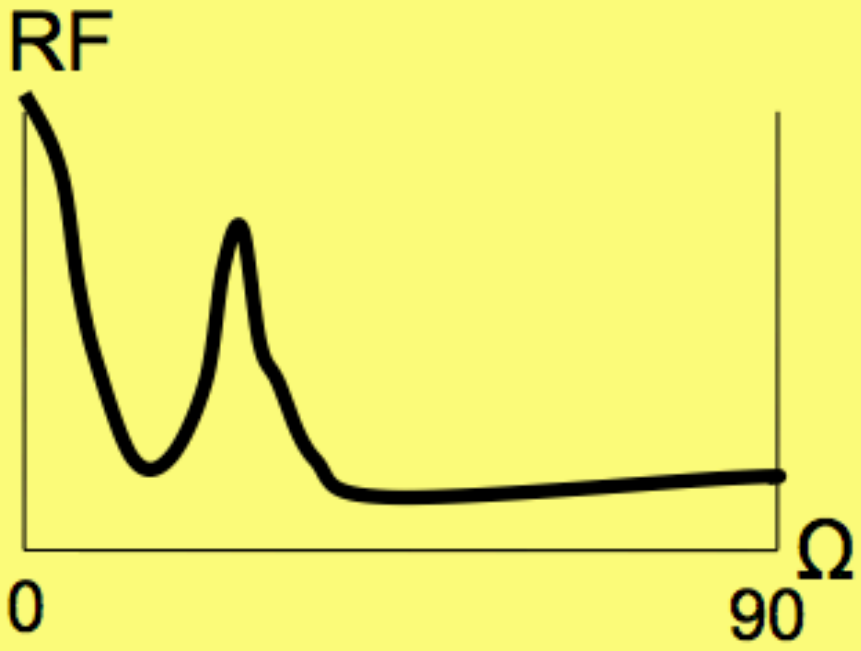


Self Rotation Function

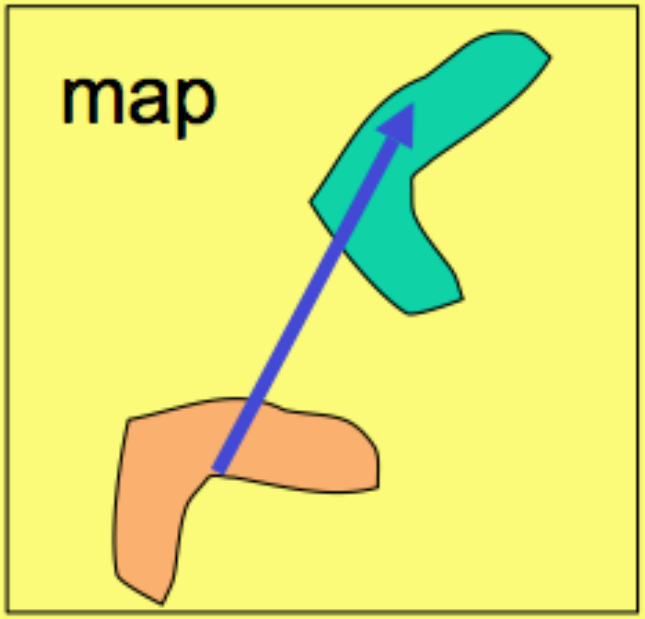


$$RF(\Omega) = \int P_{obs}(r)\ \ \Re_{\Omega}\{ P_{obs}(r)\}\ dr$$

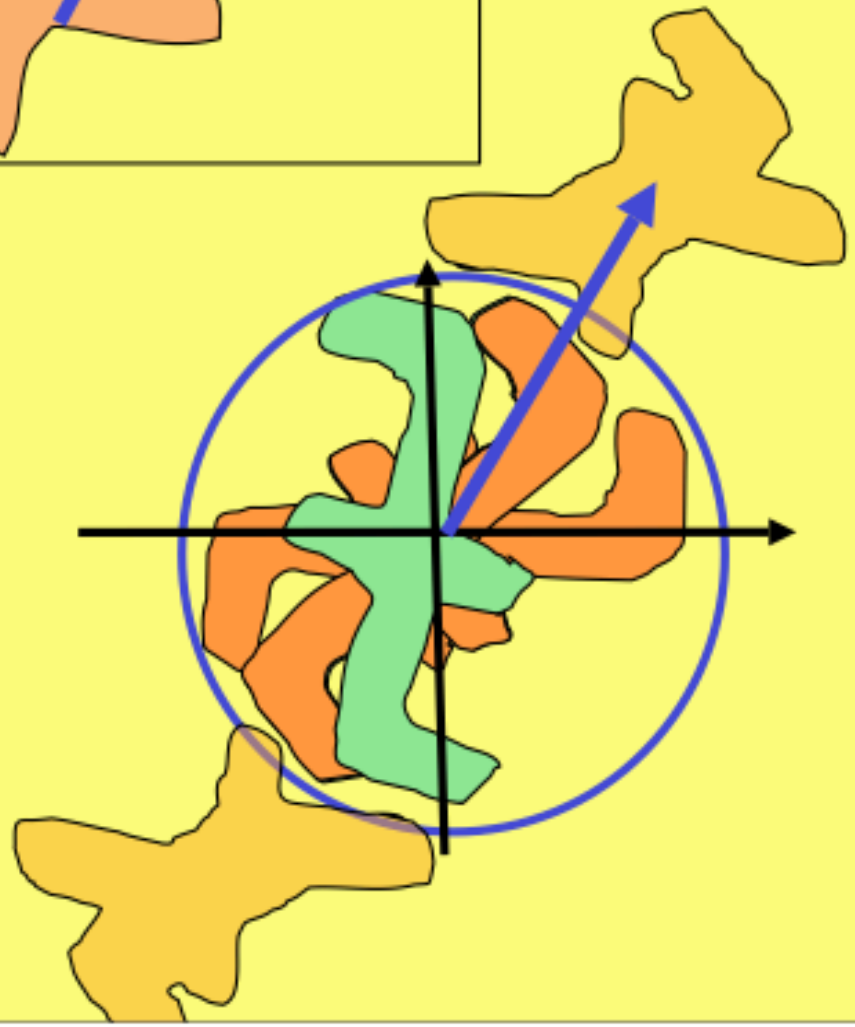


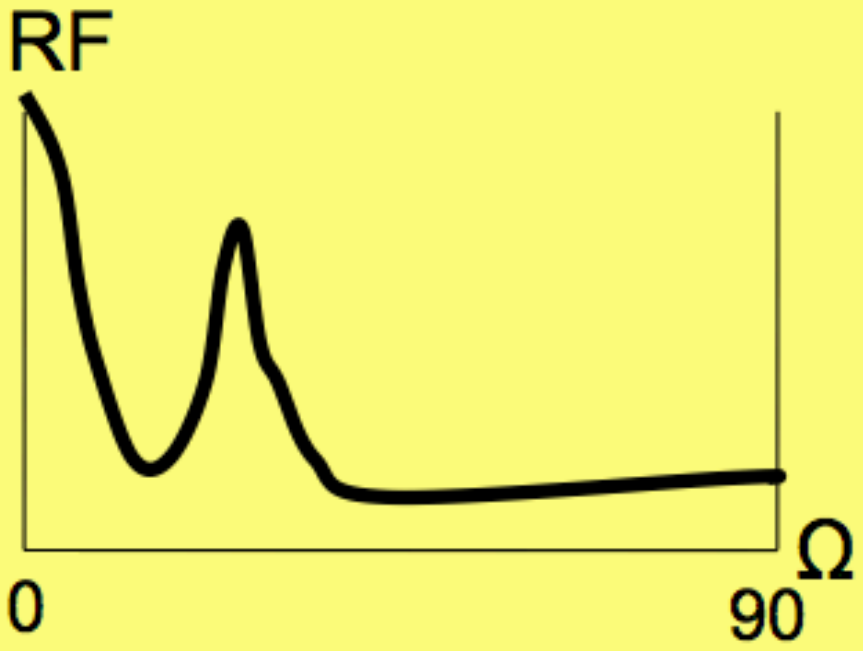


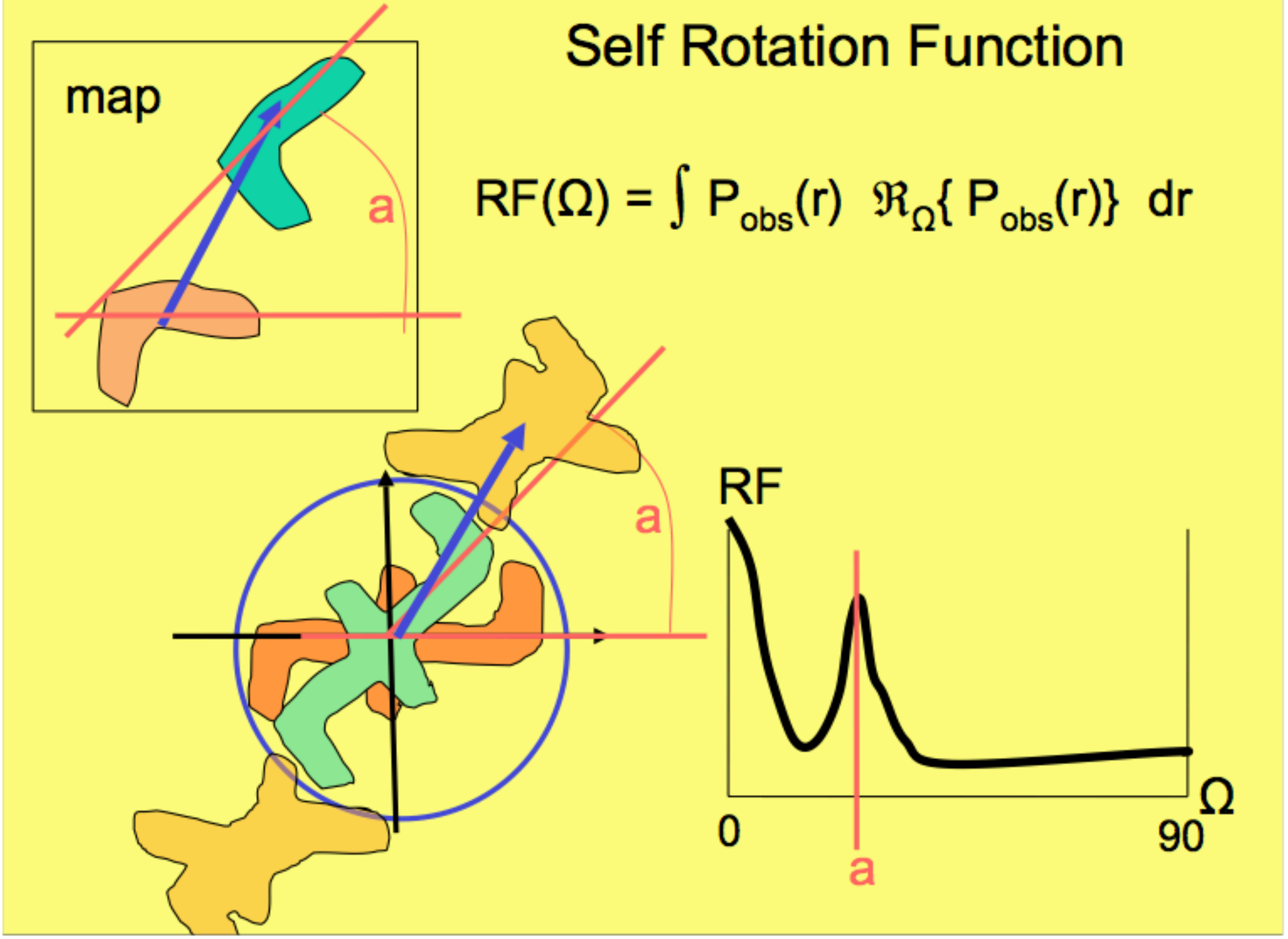
Self Rotation Function



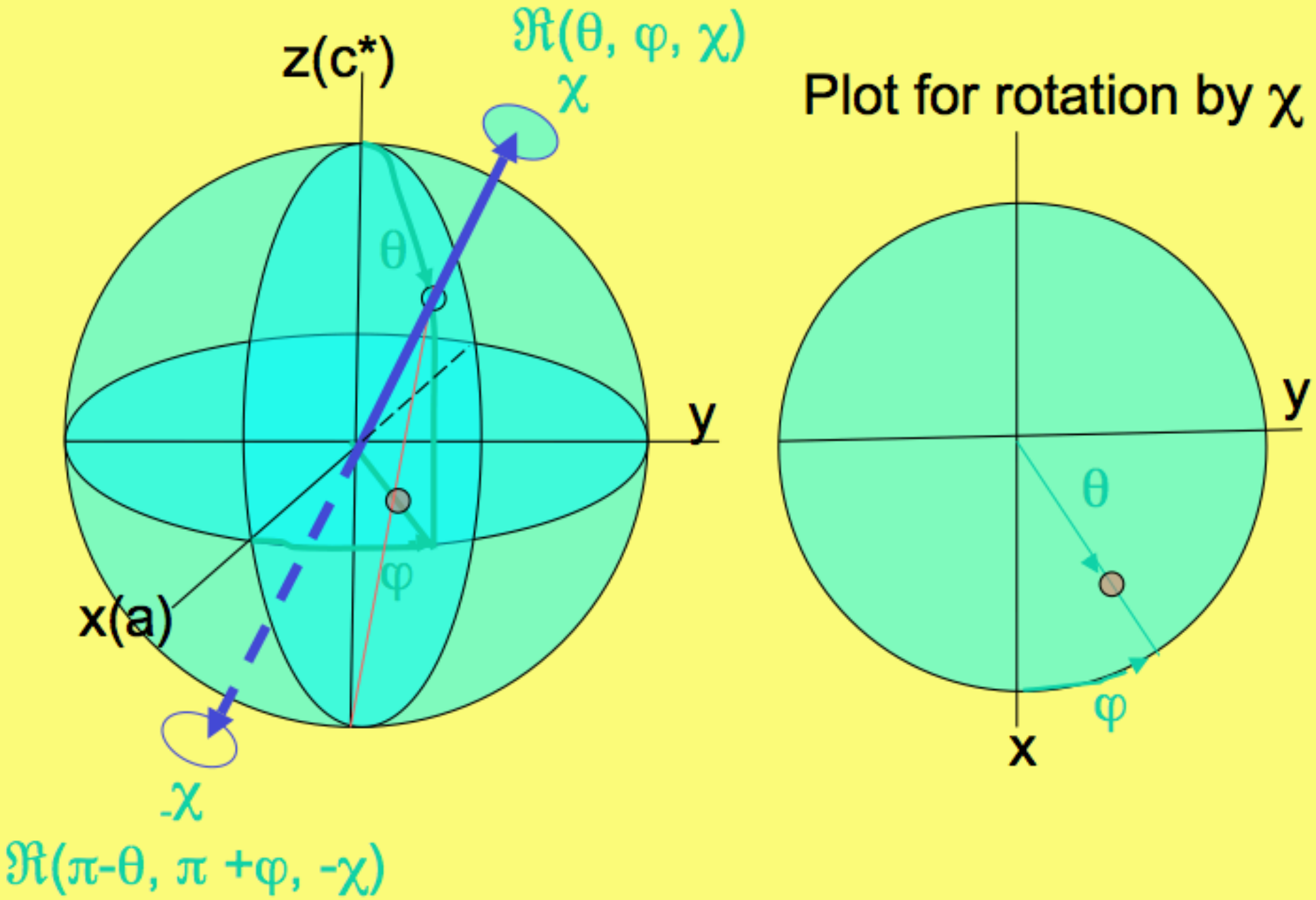
$$RF(\Omega) = \int P_{obs}(r)\ \ \Re_{\Omega}\{ P_{obs}(r)\}\ dr$$





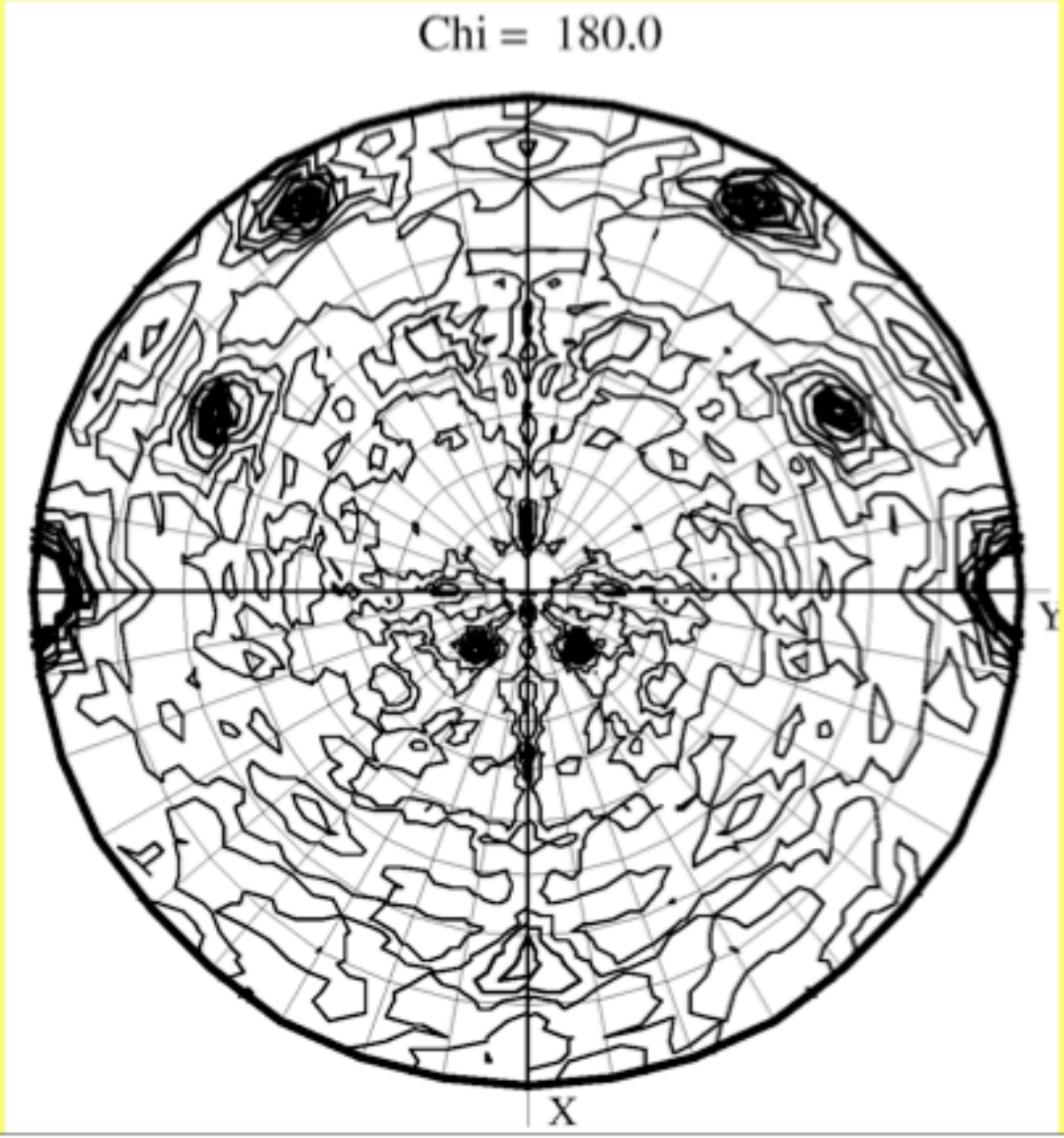

Self Rotation Function
map
a
$RF(\Omega) = \int P_{obs}(r)\ \Re_{\Omega}\{ P_{obs}(r)\}\ dr$
a
RF
0
90
Ω
a

Stereographic projection



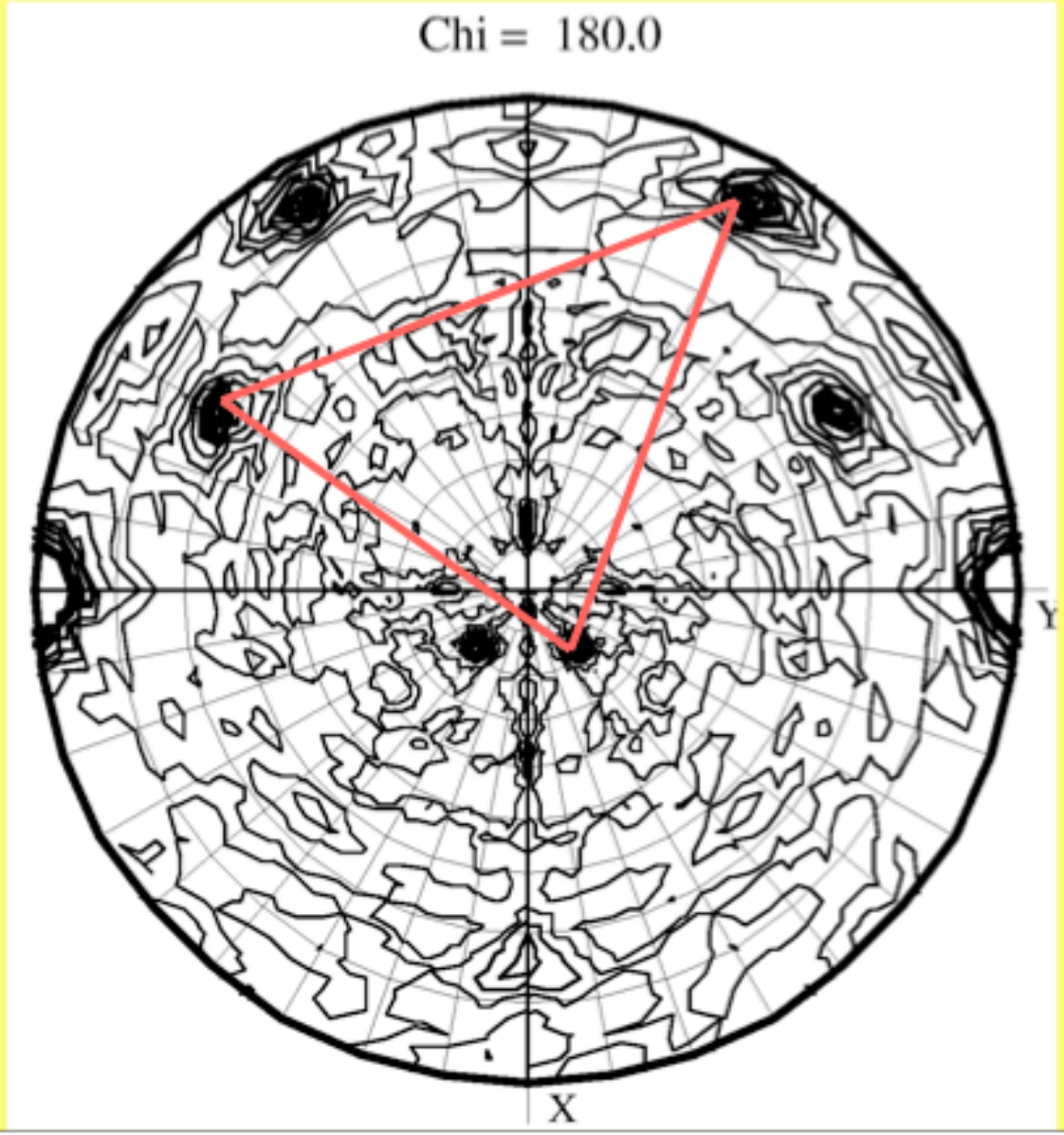
Self Rotation Function

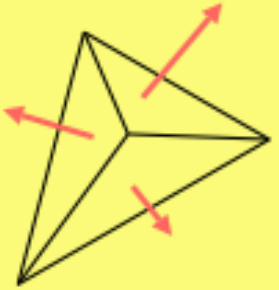
Space group $P2_1$ one tetramer



Self Rotation Function

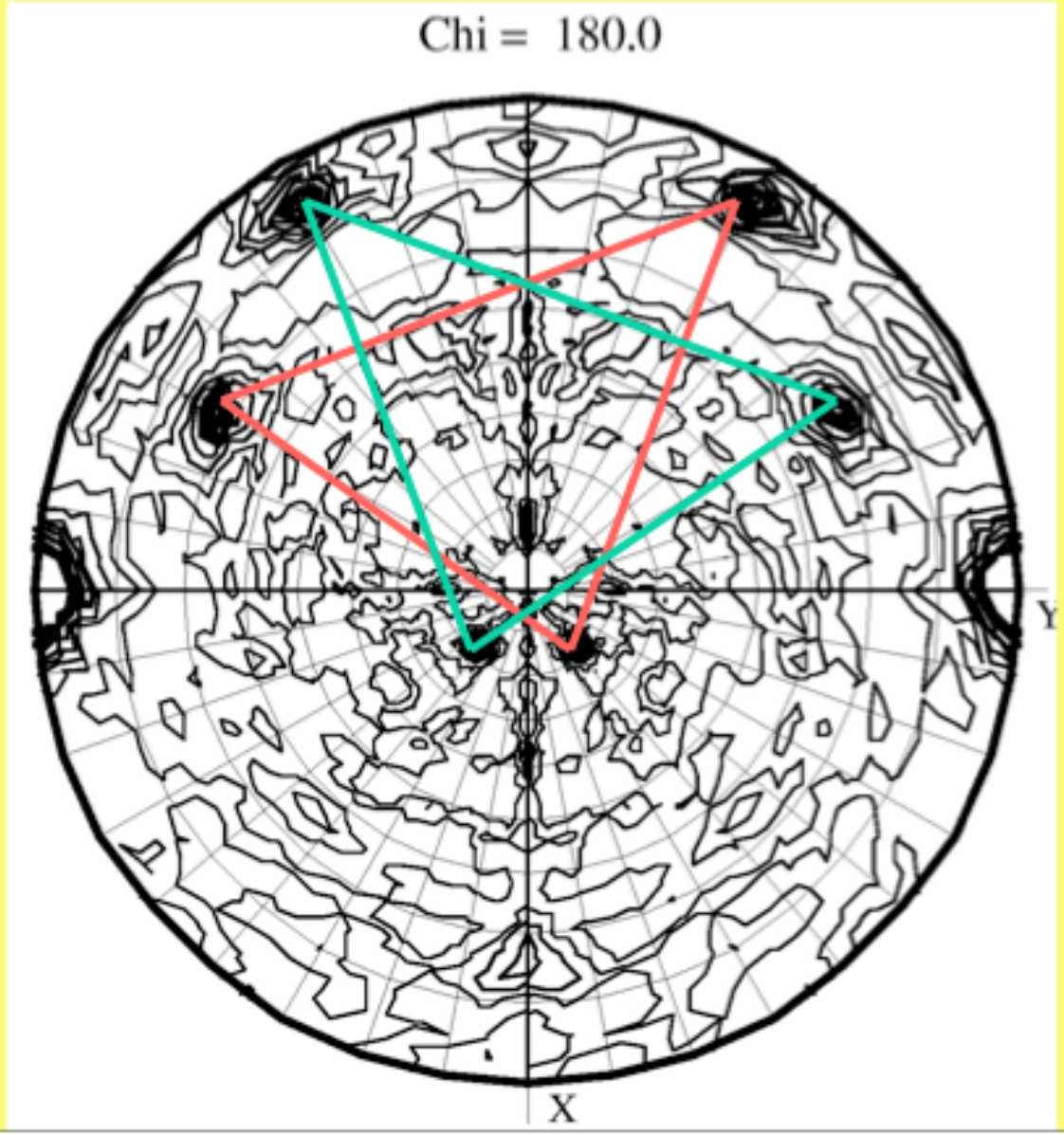
Space group $P2_1$ one tetramer

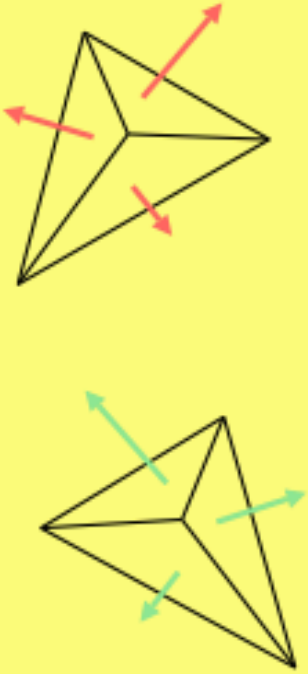




Self Rotation Function

Space group $P2_1$ one tetramer





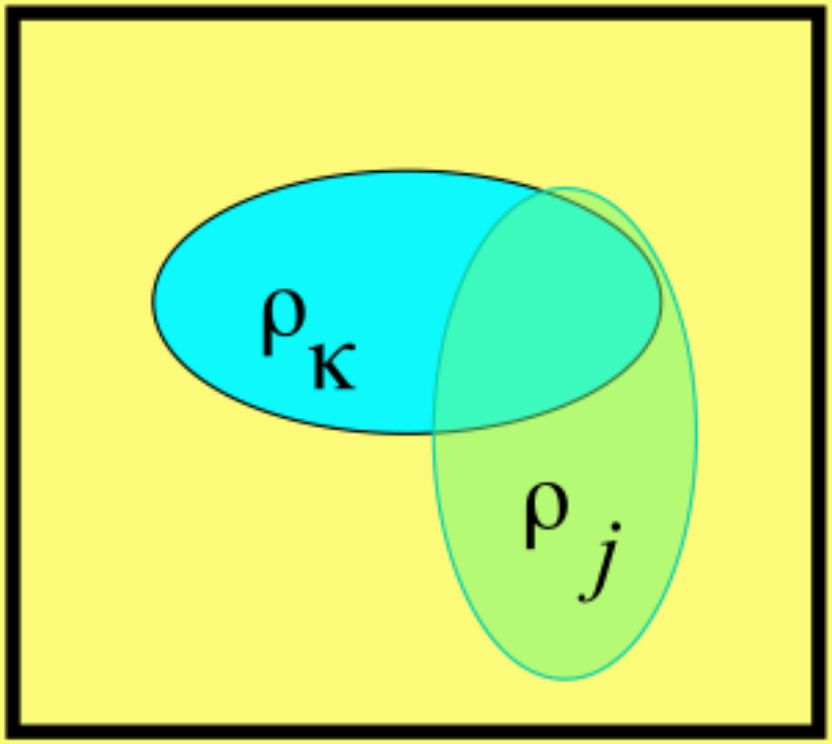
Translation Function

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 ?

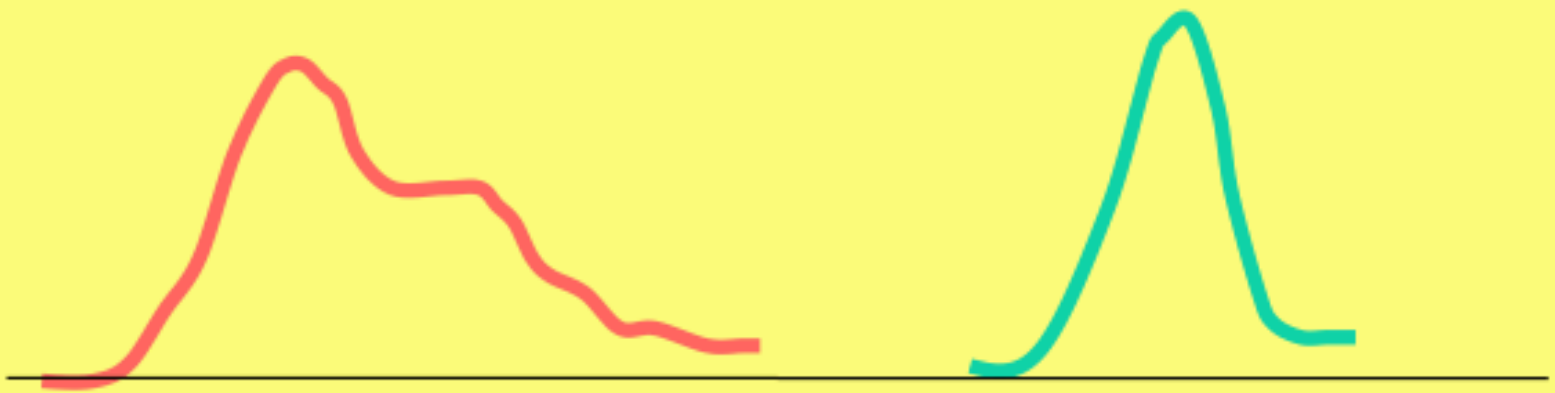
Short introduction to Fourier Transformation

Operators:

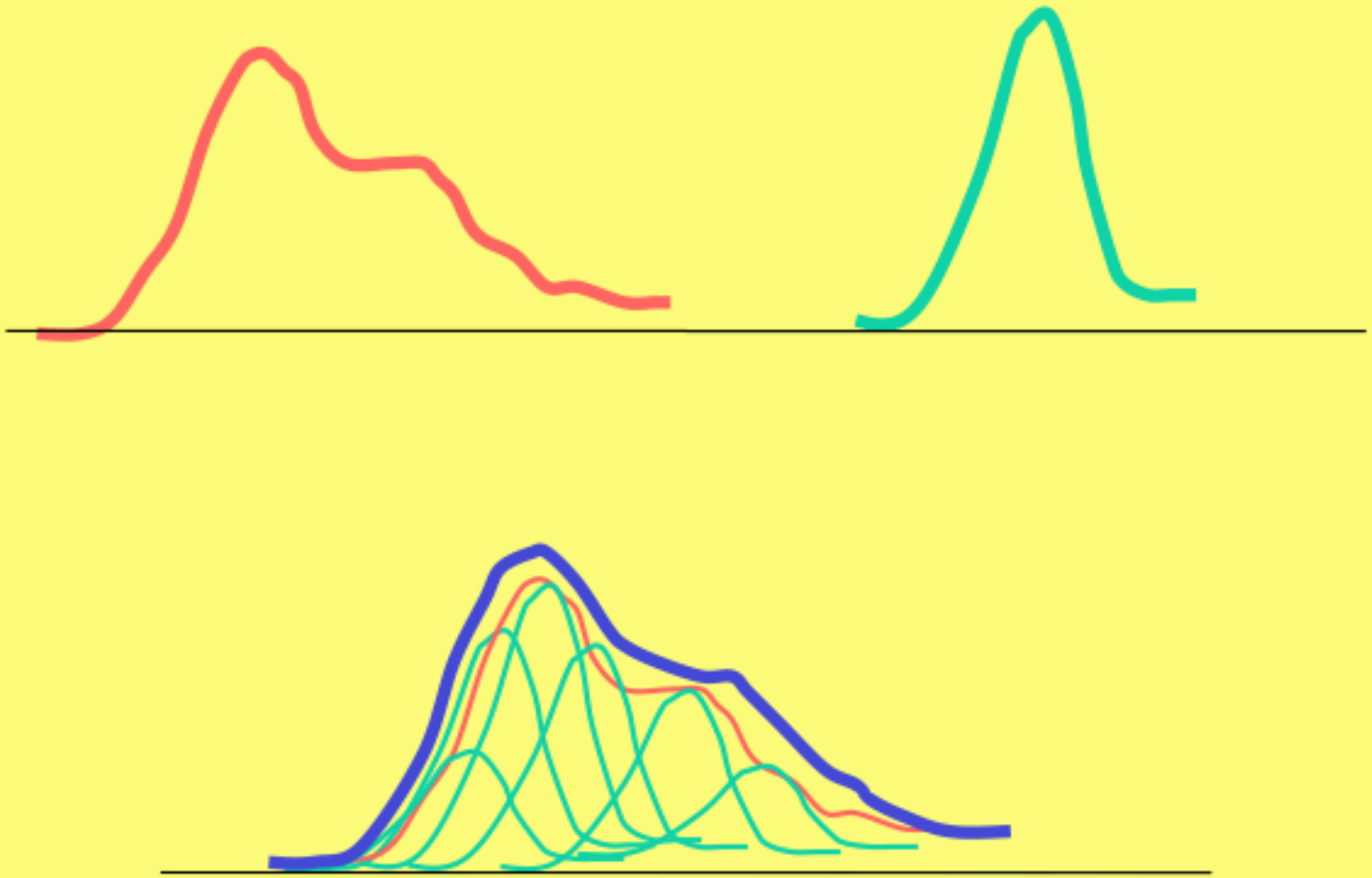
addition $\leftarrow \mathcal{F} \rightarrow$ addition

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

Convolution

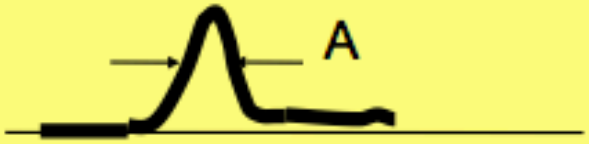


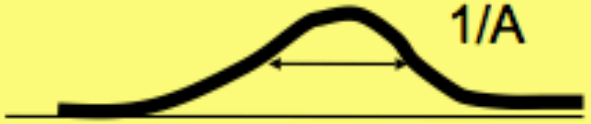
Convolution

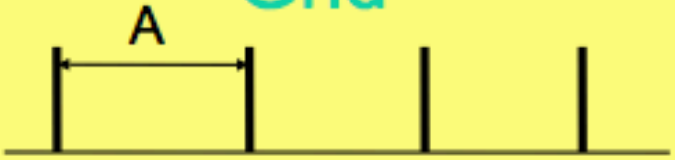


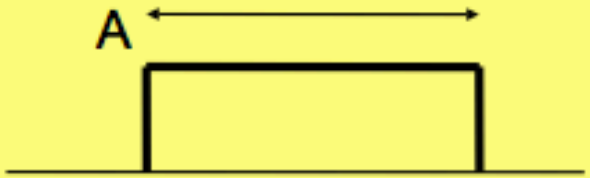
Functions

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

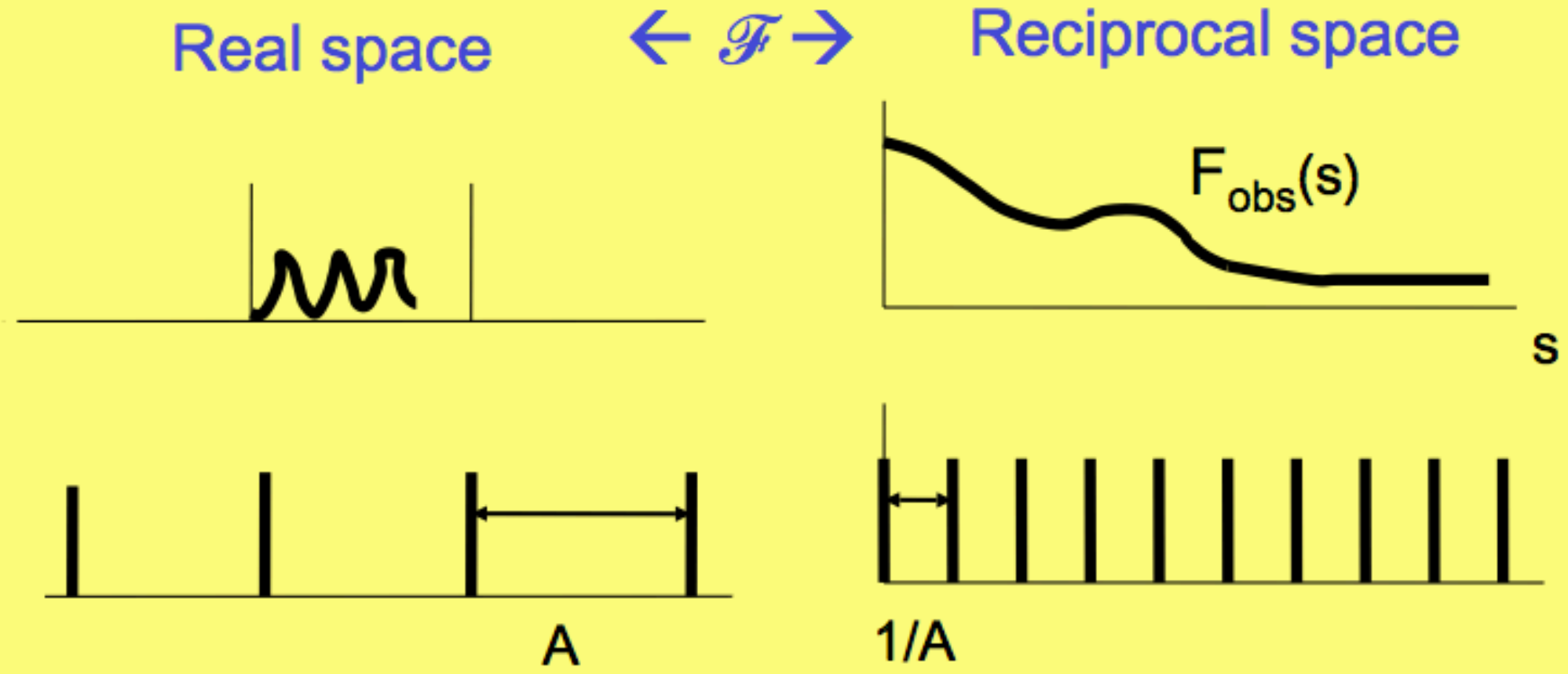


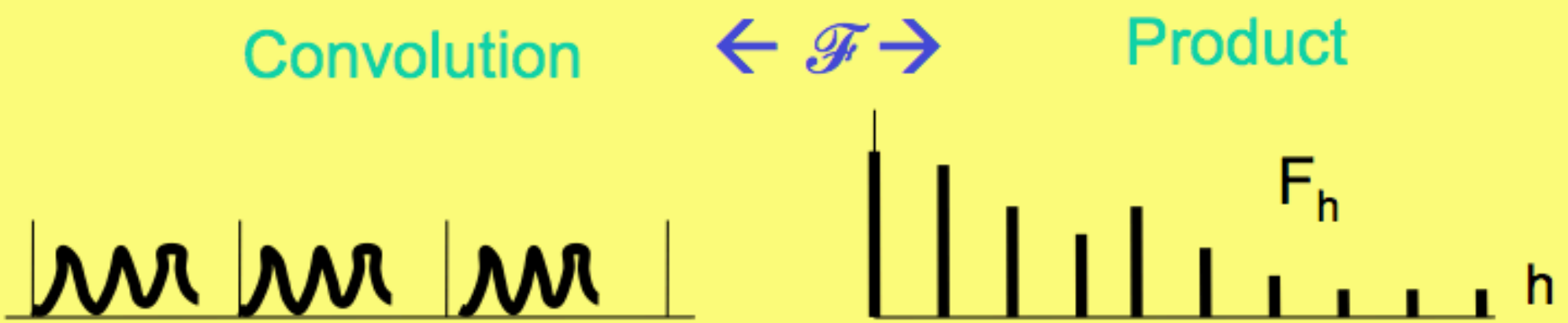


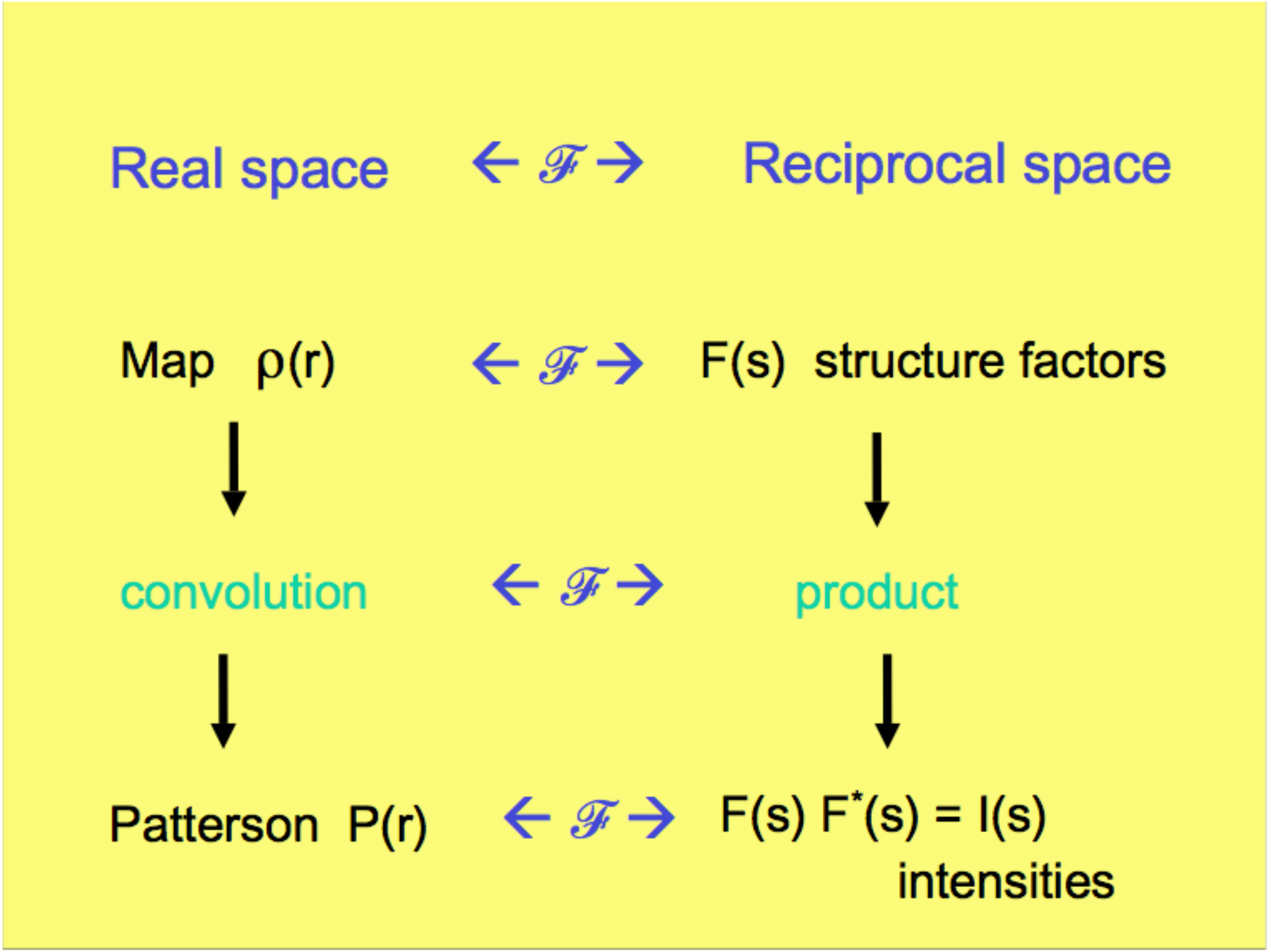




Crystal and Structure Factors





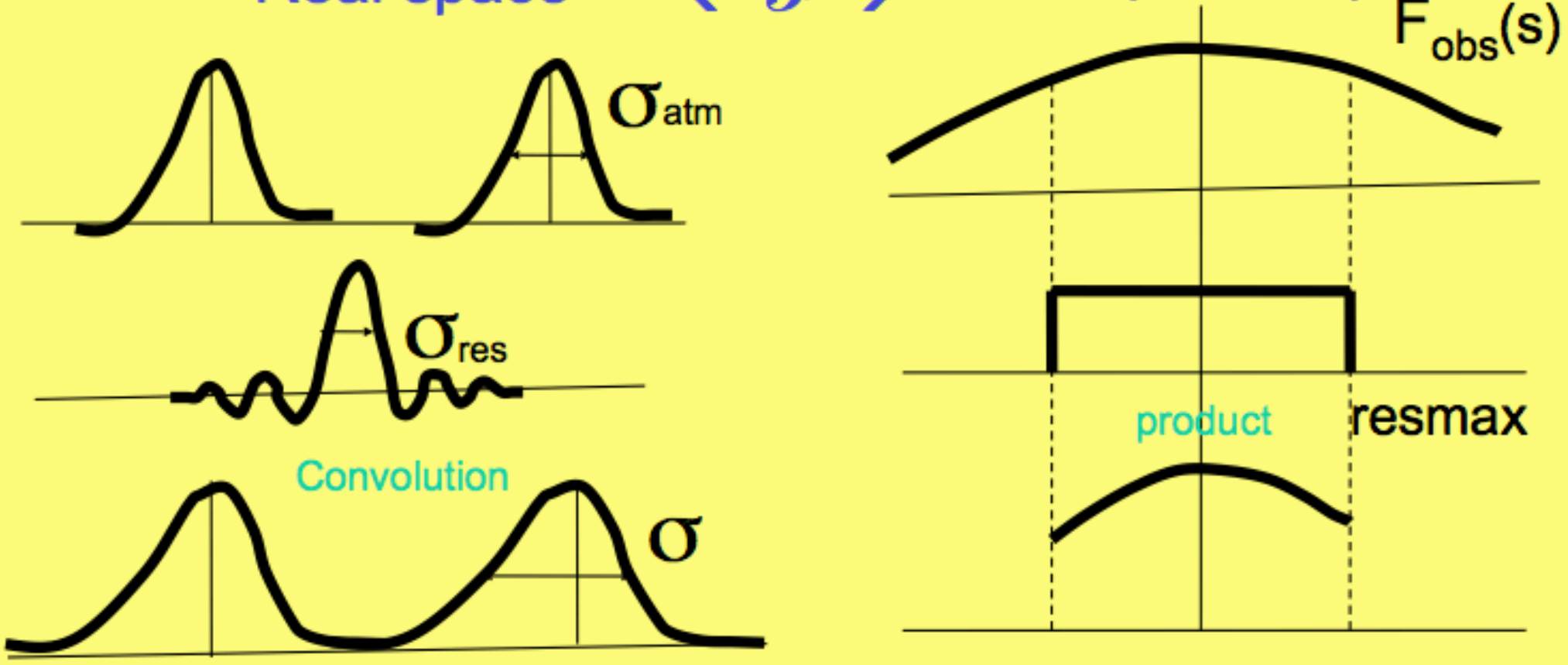

Real space
← 𝓕 →
Reciprocal space
Map ρ(r)
← 𝓕 →
F(s) structure factors
convolution
← 𝓕 →
product
Patterson P(r)
← 𝓕 →
F(s) F*(s) = I(s)
intensities

High resolution data

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

Optical resolution

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

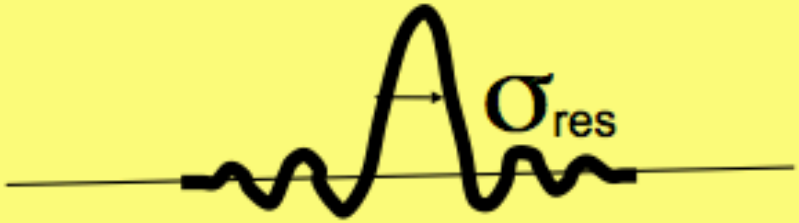


$$\sigma_{res} = 0.356\ resmax$$

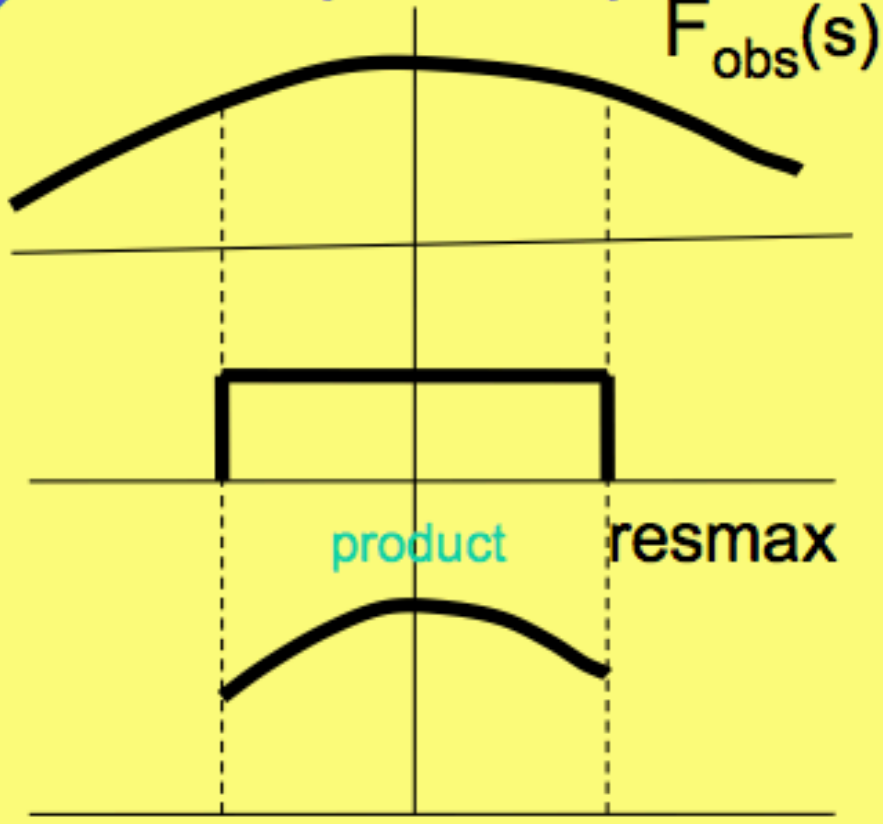
$$\sigma^2 = \sigma_{atm}^2 + \sigma_{res}^2$$

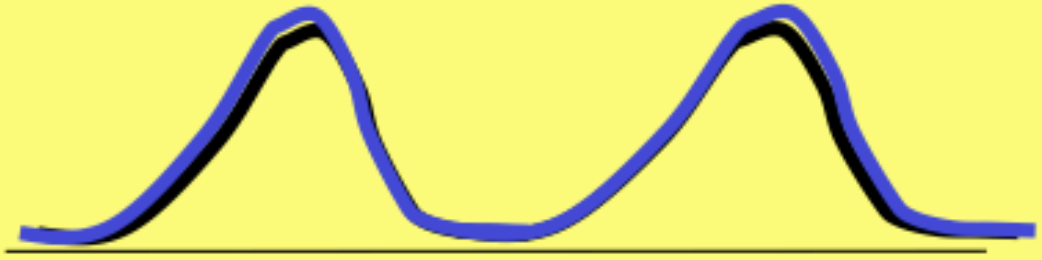
Optical resolution

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



Convolution



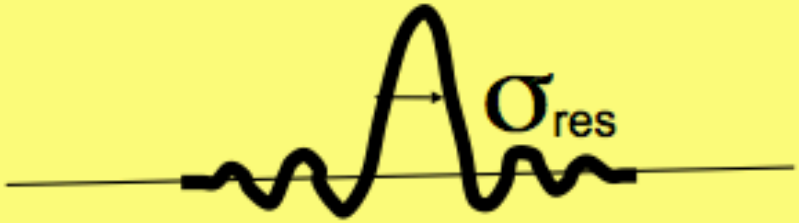


$$\sigma_{res} = 0.356\ resmax$$

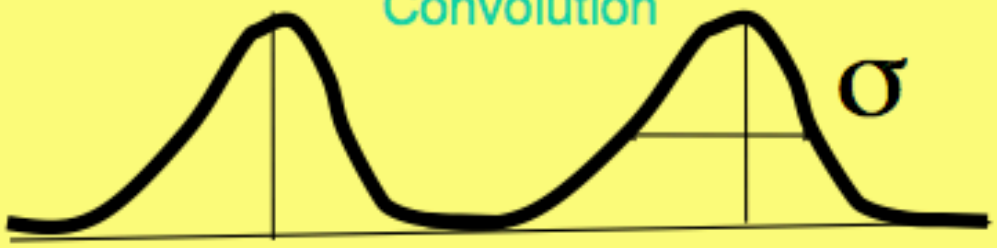
$$\sigma^2 = \sigma_{atm}^2 + \sigma_{res}^2$$

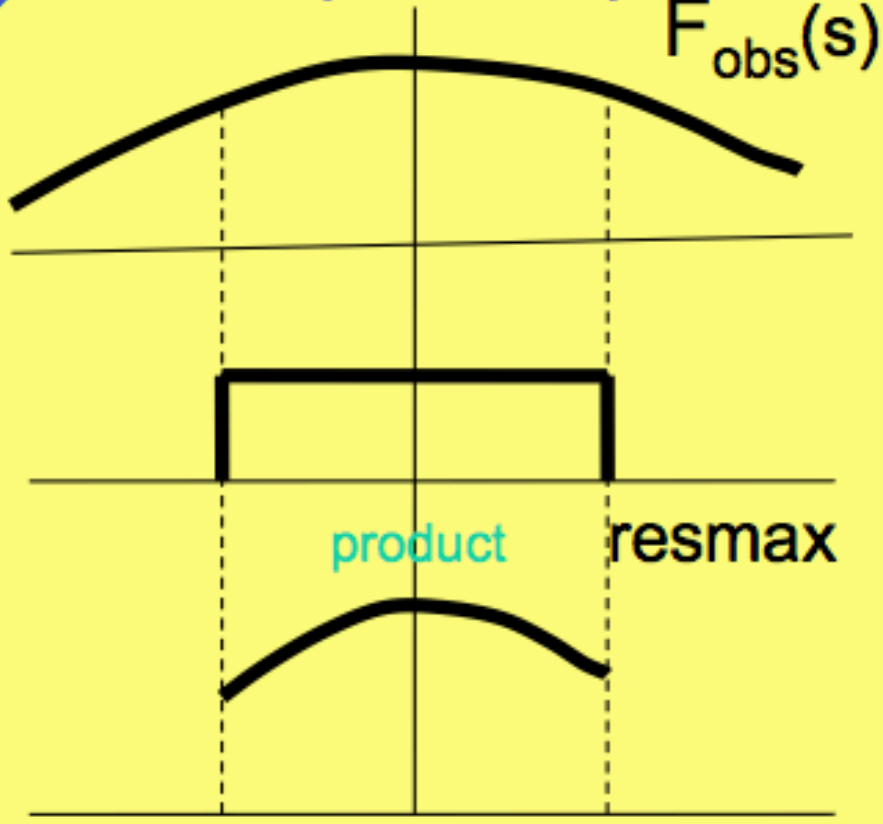
Optical resolution

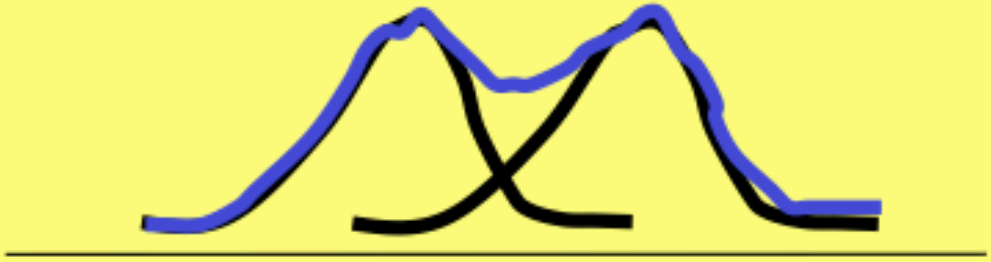
Real space ← $\mathcal{F}$ → Reciprocal space



Convolution





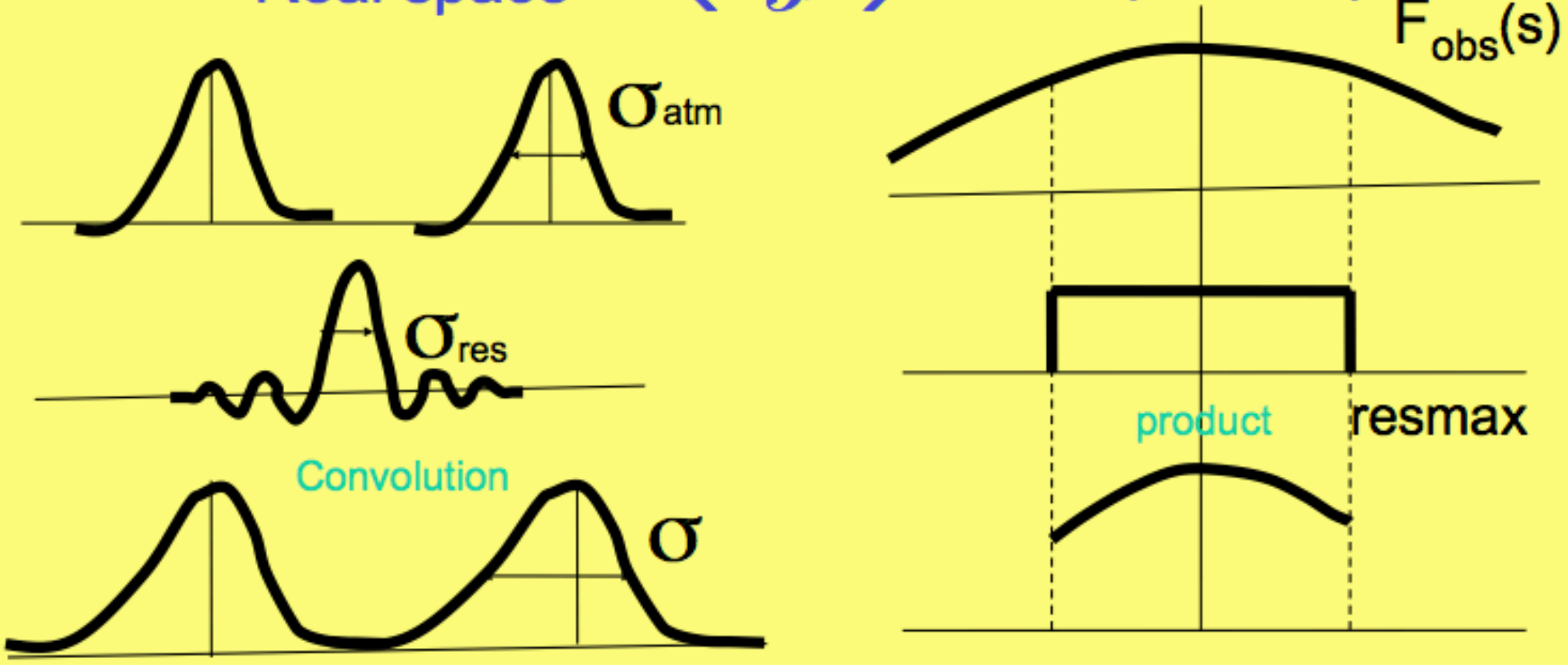


$$\sigma_{res} = 0.356\ resmax$$

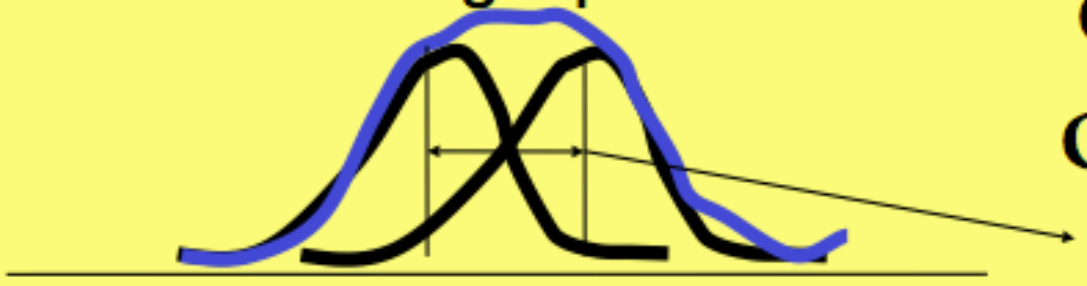
$$\sigma^2 = \sigma_{atm}^2 + \sigma_{res}^2$$

Optical resolution

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



Single peak

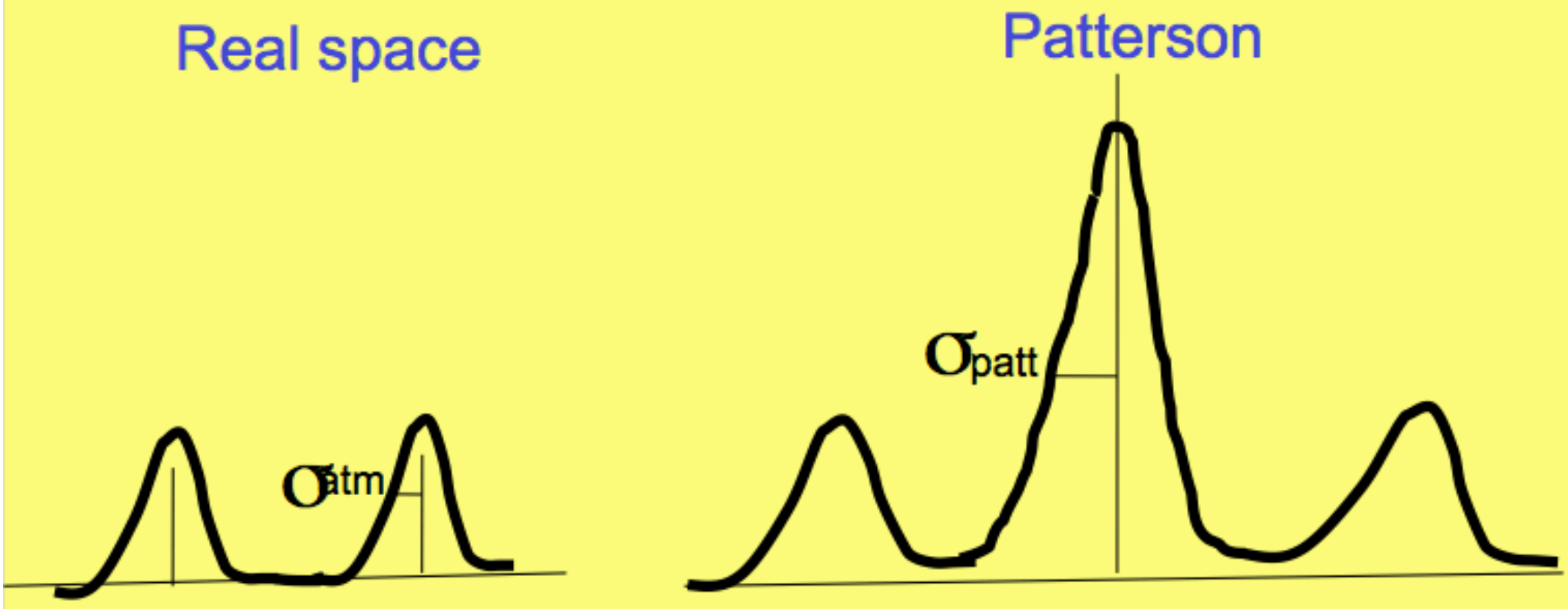


$\sigma_{res} = 0.356\ resmax$

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

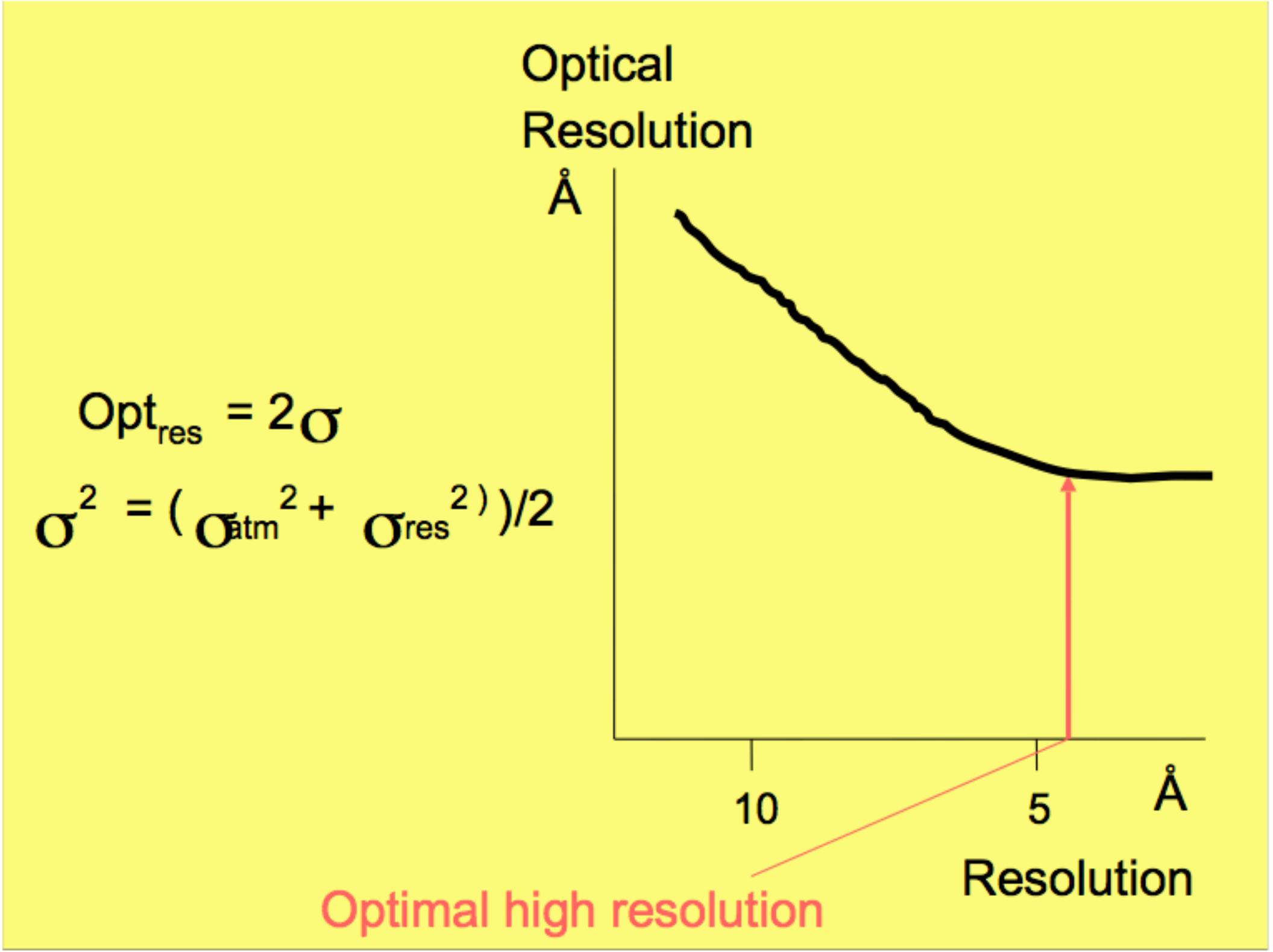
$Opt_{res} = 2\,\sigma$

Optical resolution from origin peak of Patterson

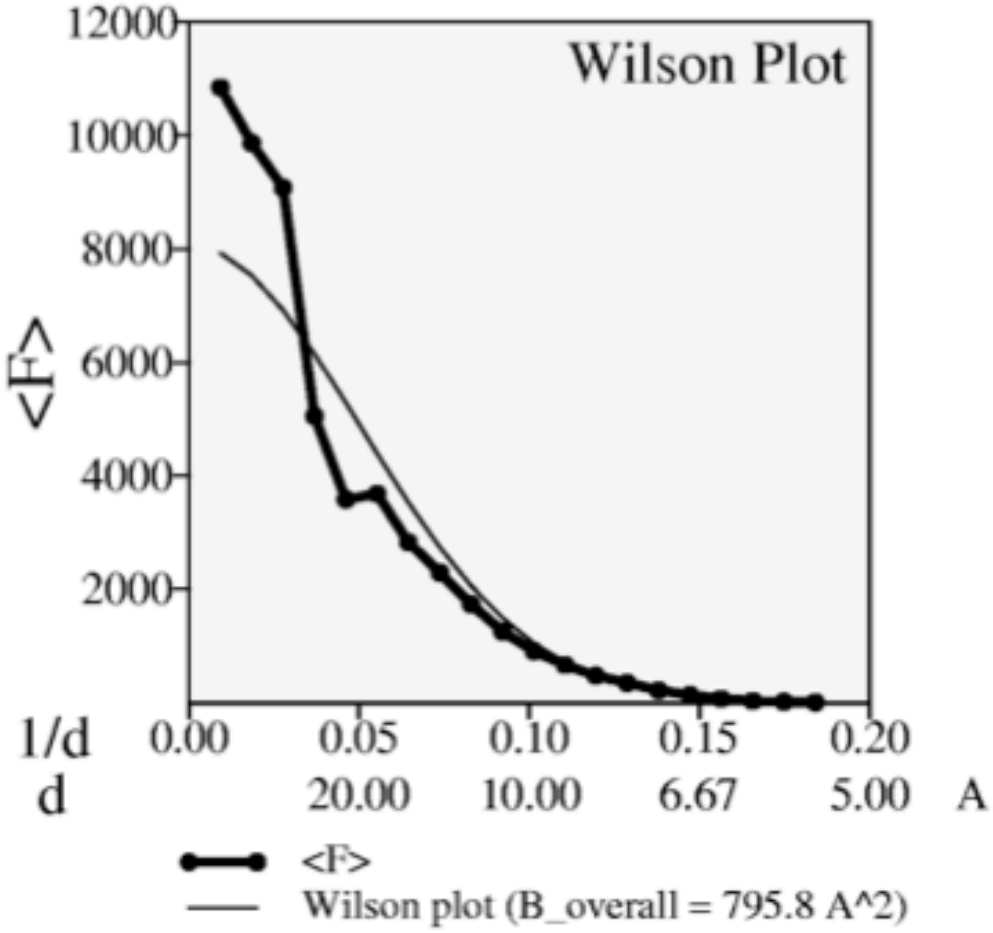


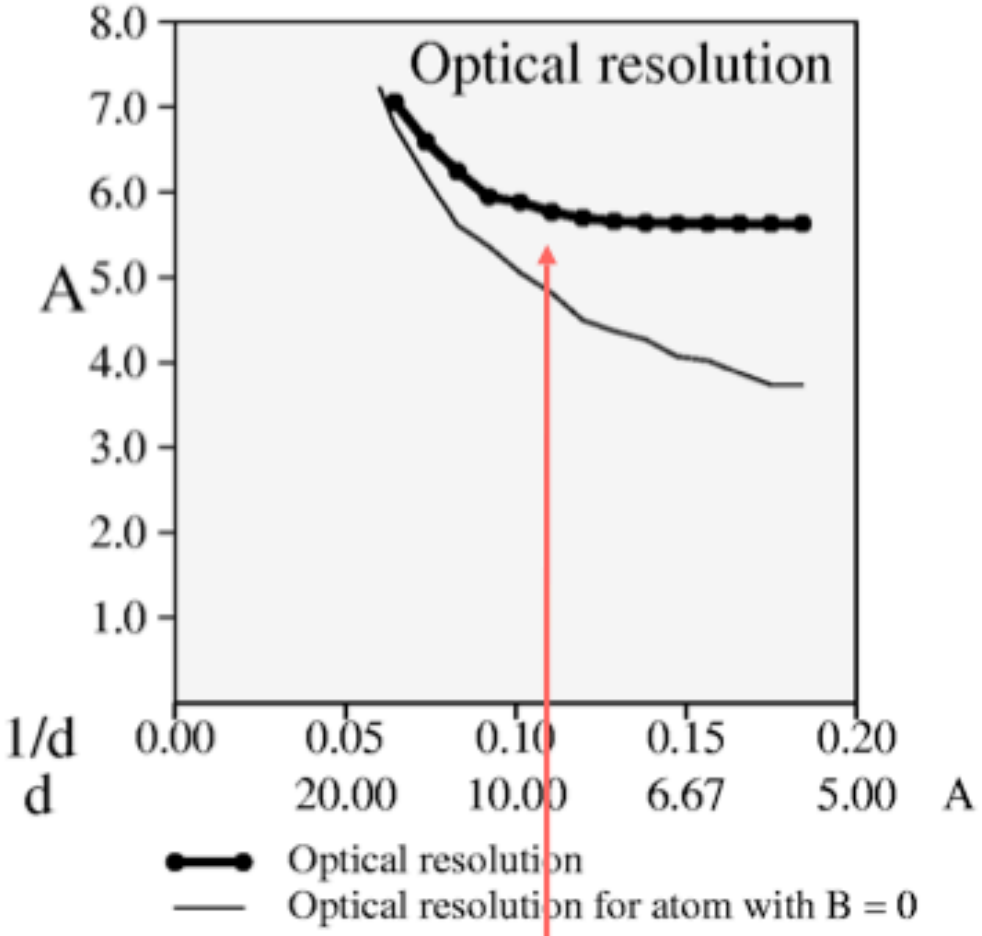
$$Opt_{res} = 2\,\sigma_{atm}$$

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

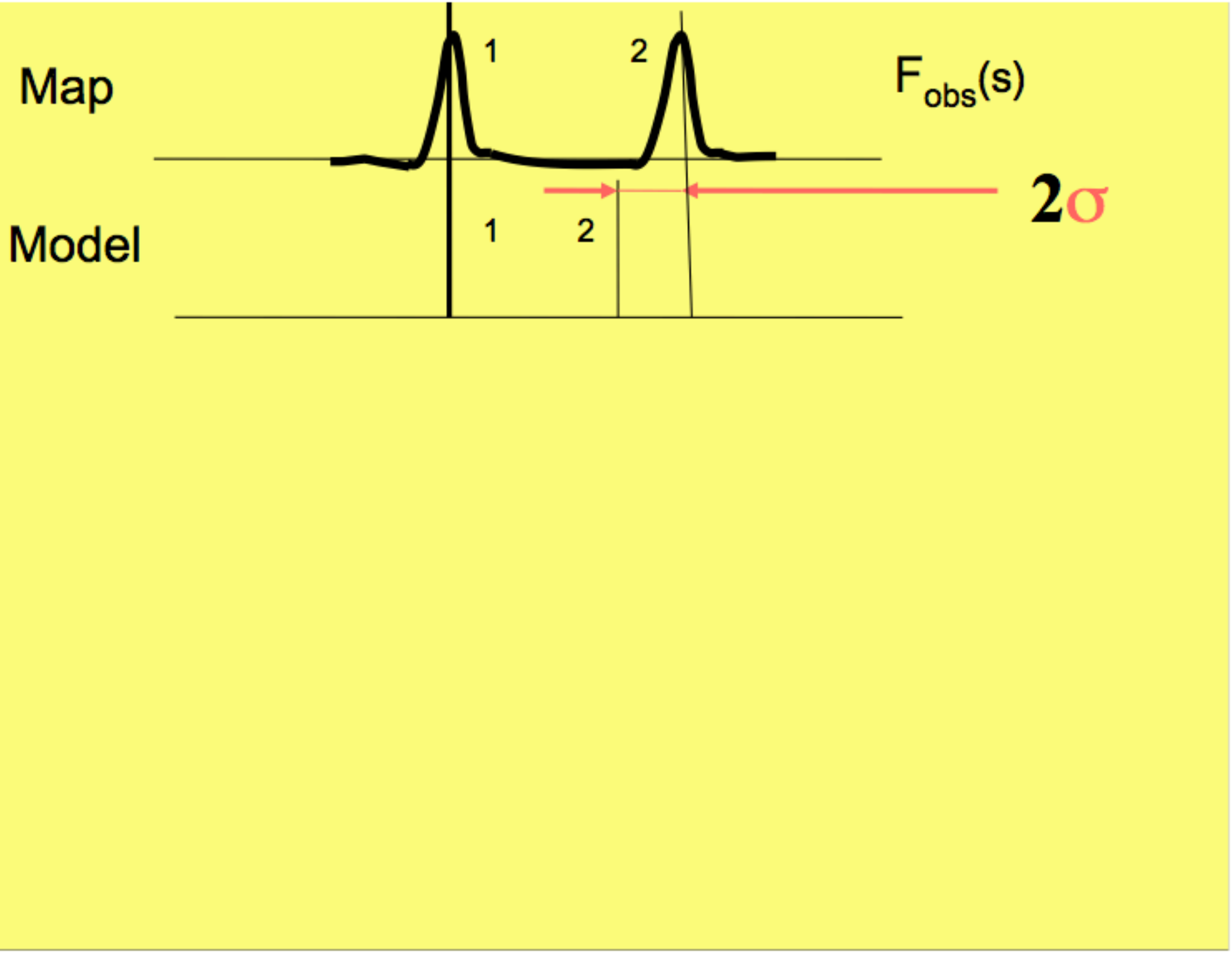

Optical
Resolution
Å
$Opt_{res} = 2\sigma$
$\sigma^2 = (\sigma_{atm}^2 + \sigma_{res}^2)/2$
10
5
Å
Resolution
Optimal high resolution

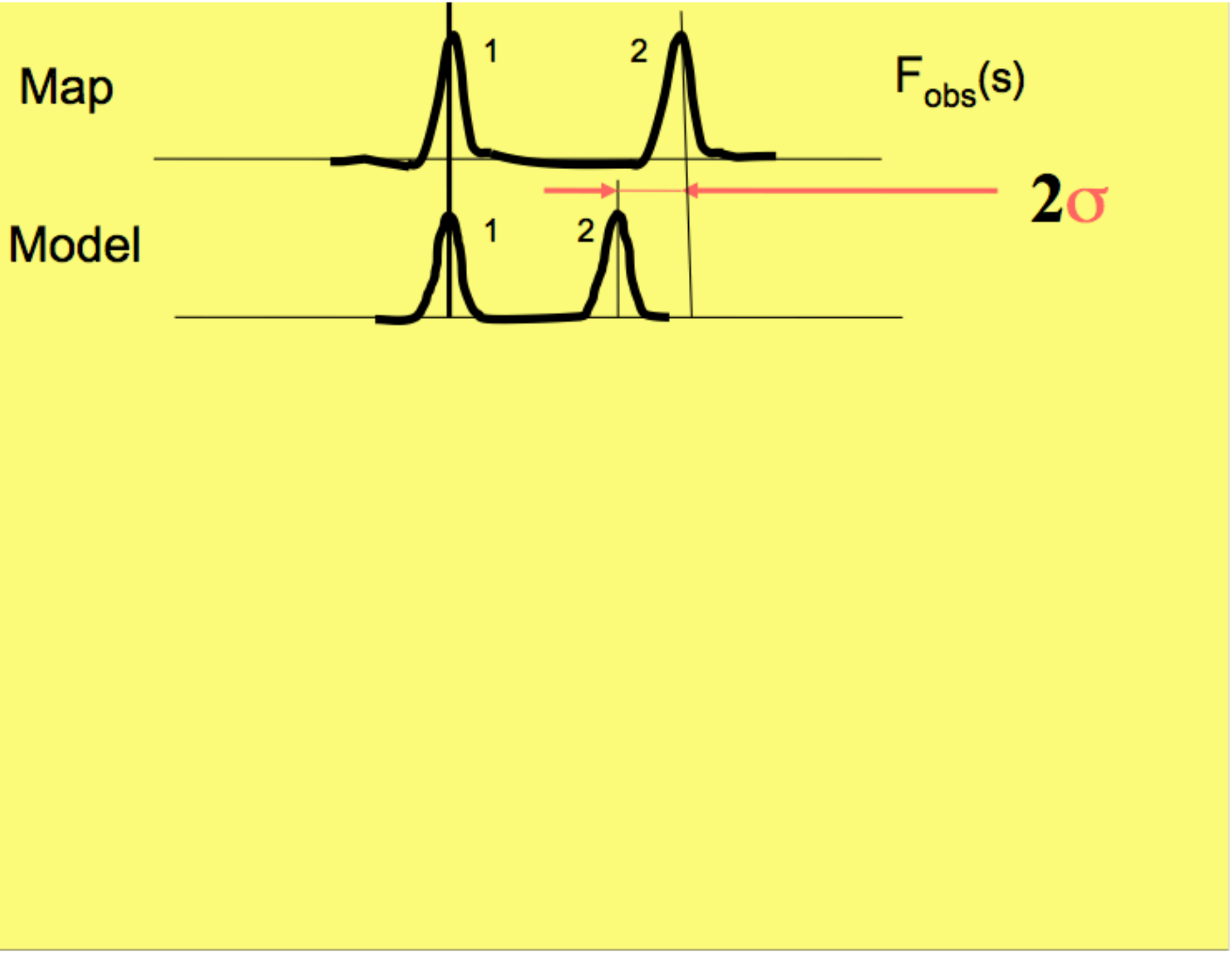
Optical Resolution (by sfcheck)

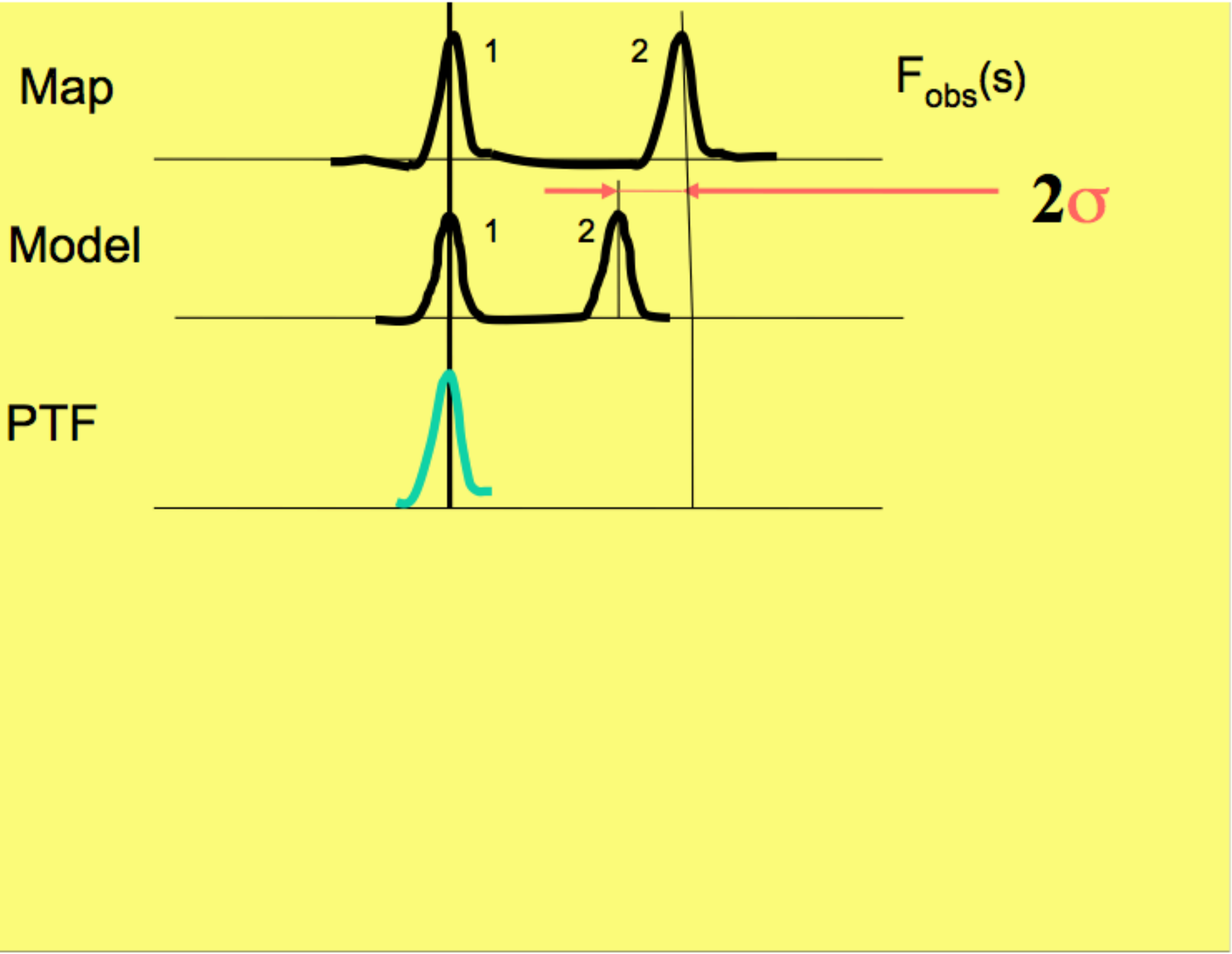


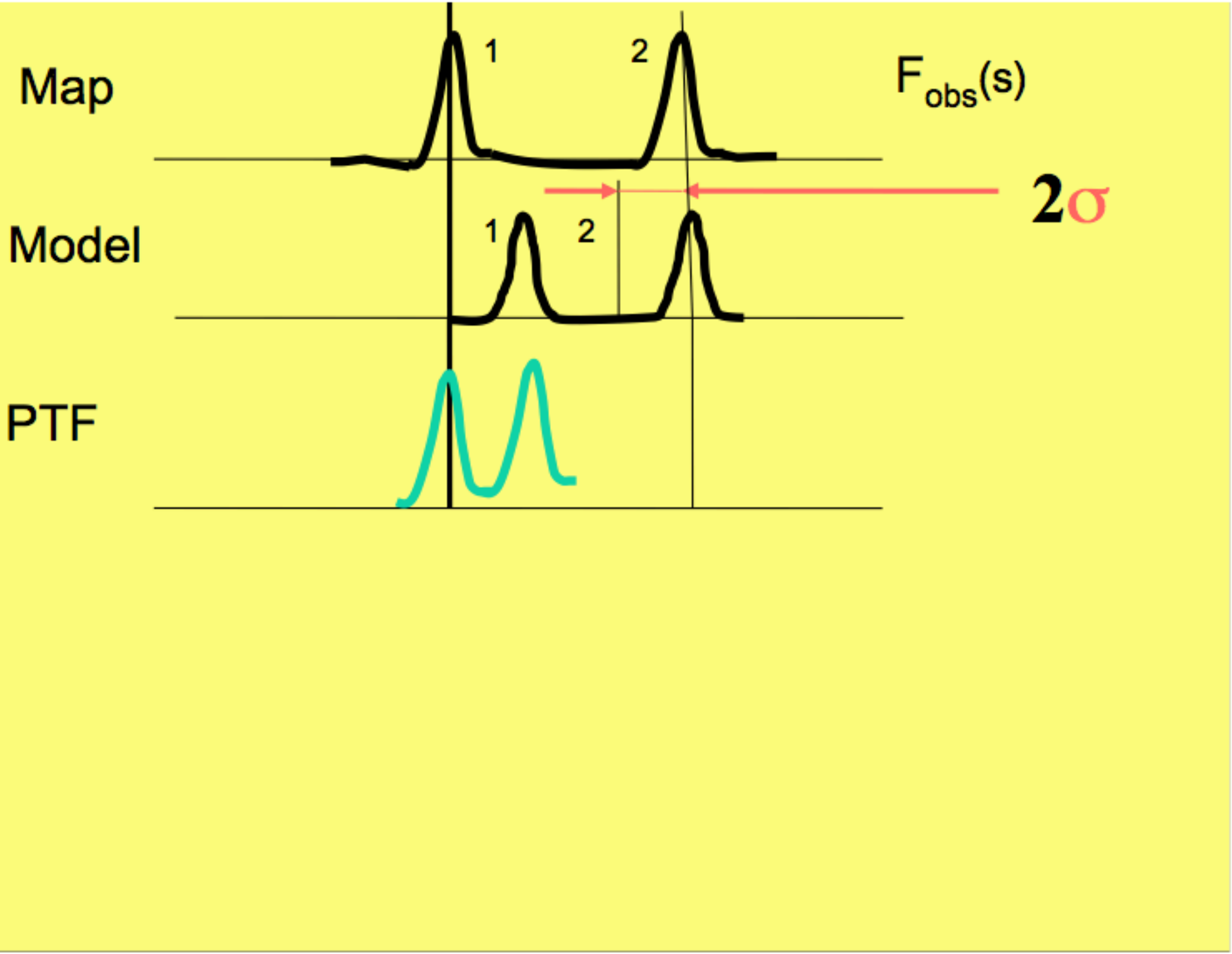


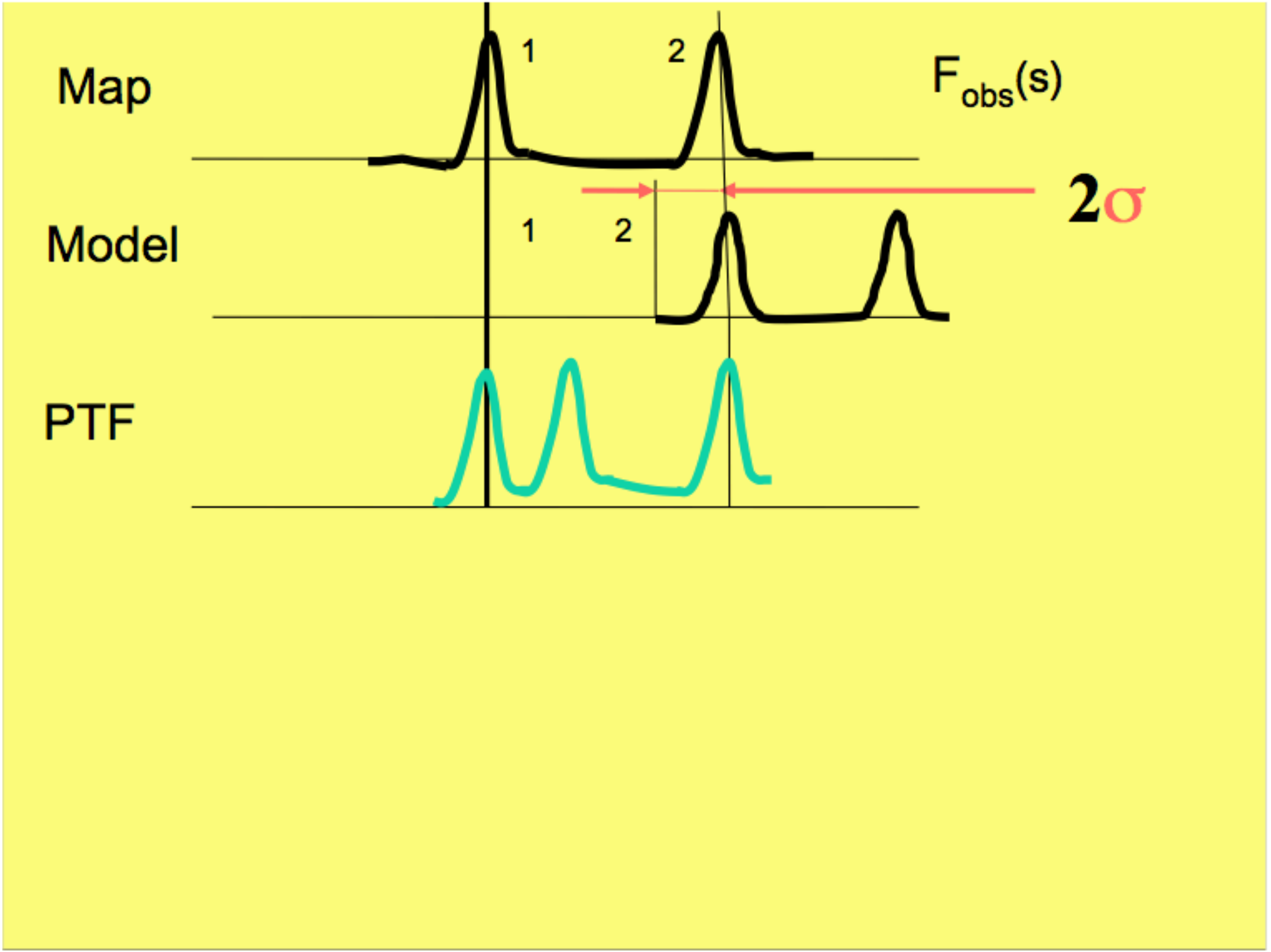
Weights for high resolution data and similarity

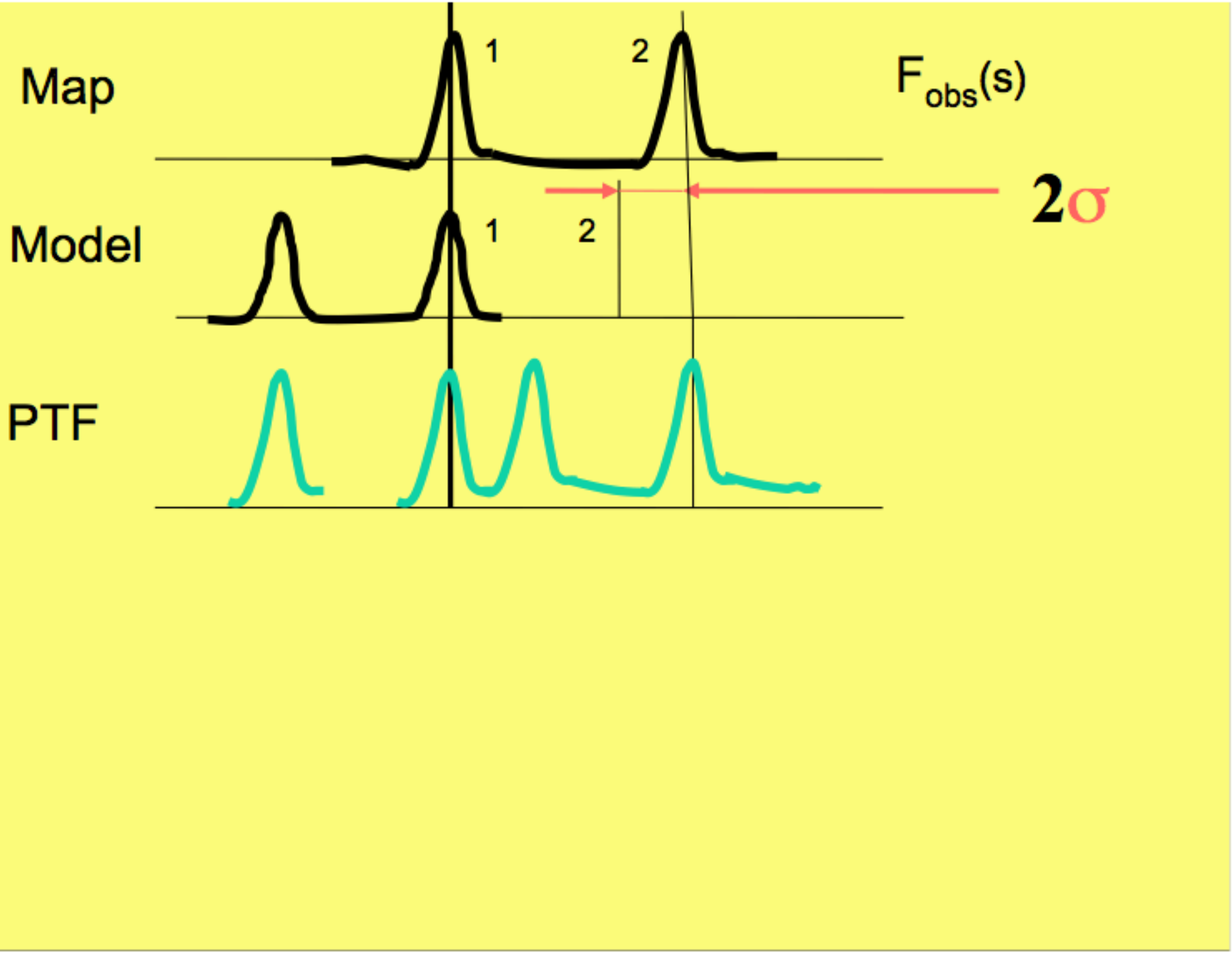

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2

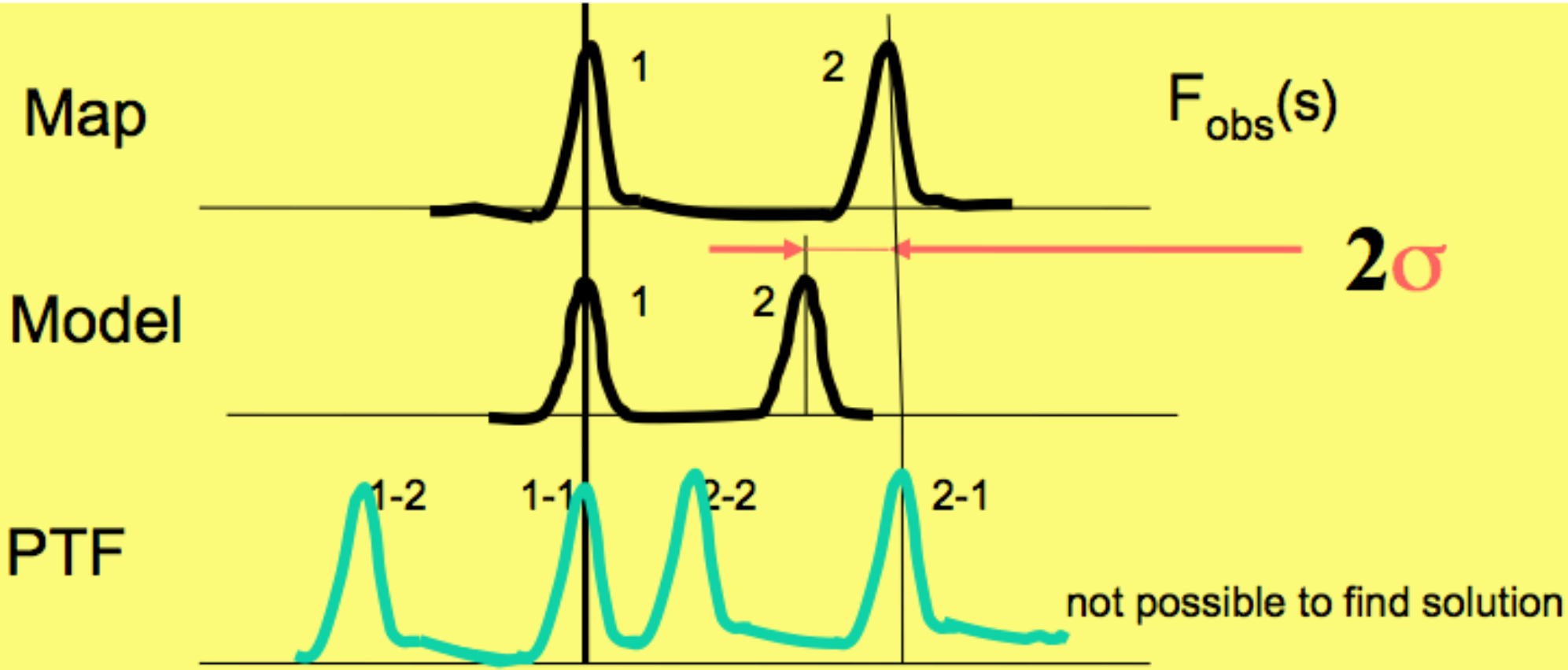

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2

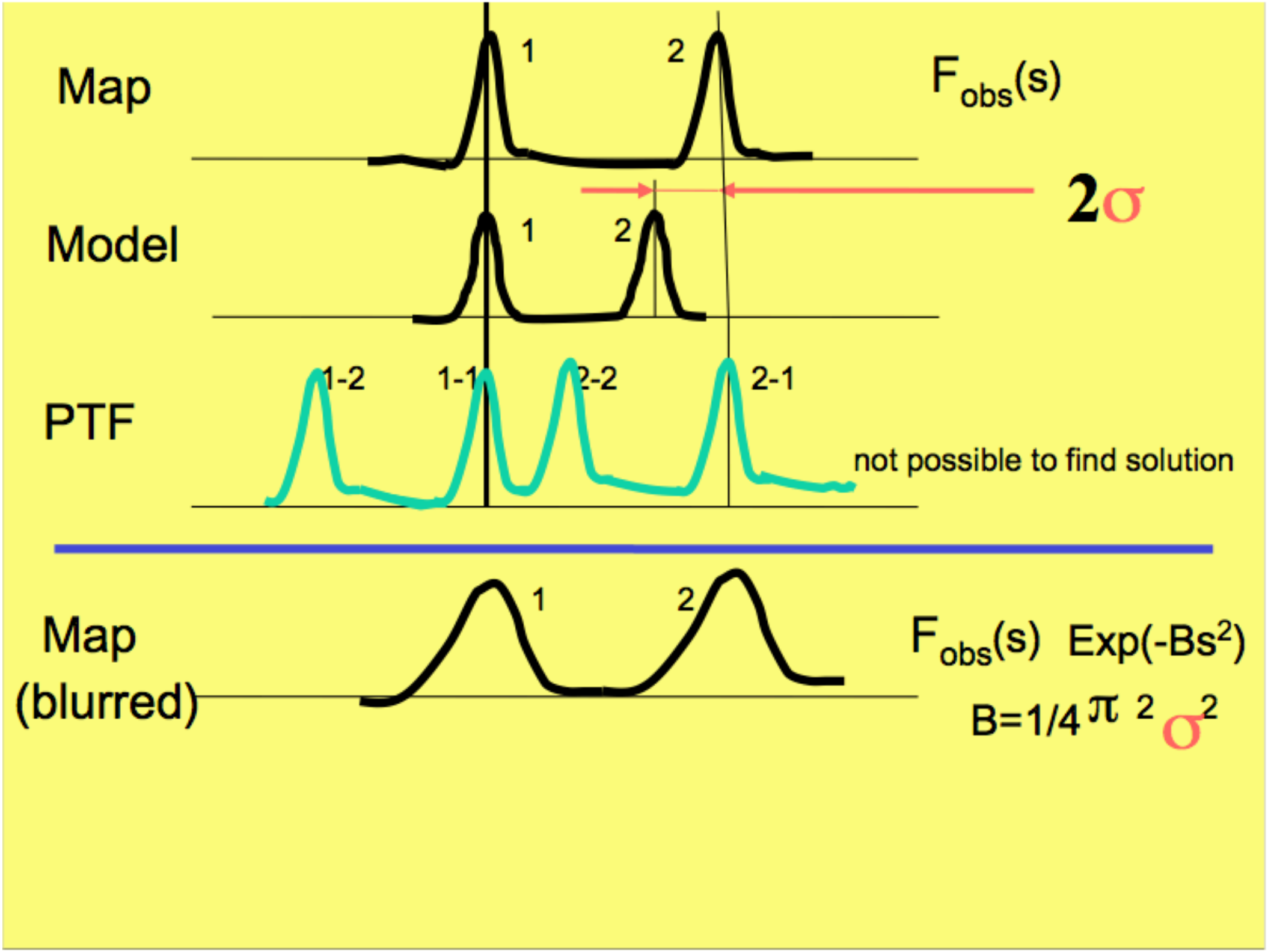

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF

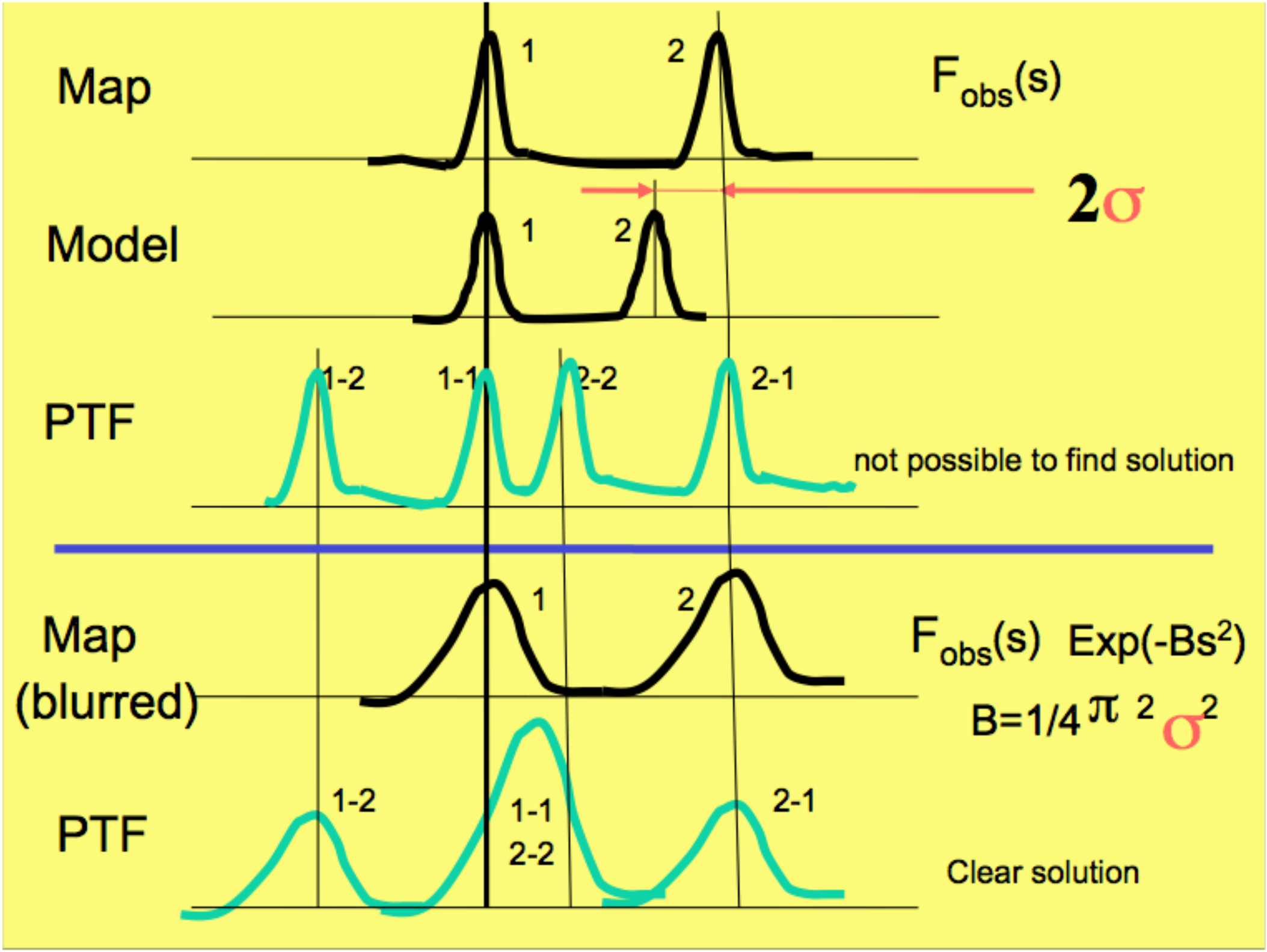

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF


Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF


Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF


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
$F_{obs}(s)\ \ Exp(-Bs^2)$
$B=1/4\,\pi^2\,\sigma^2$

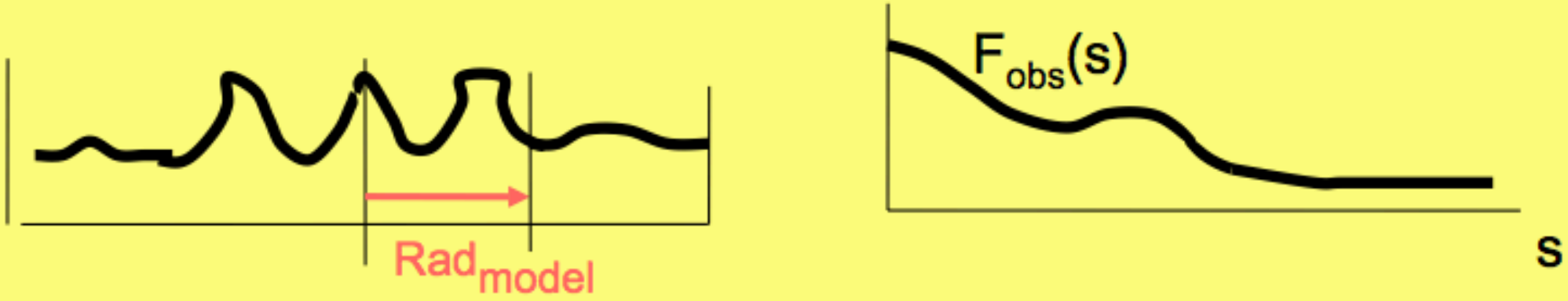

Map
1
2
$F_{obs}(s)$
2σ
Model
1
2
PTF
1-2
1-1
2-2
2-1
not possible to find solution
Map
(blurred)
1
2
$F_{obs}(s)$ Exp($-Bs^2$)
$B=1/4\,\pi^2\sigma^2$
PTF
1-2
1-1
2-2
2-1
Clear solution

Low resolution data

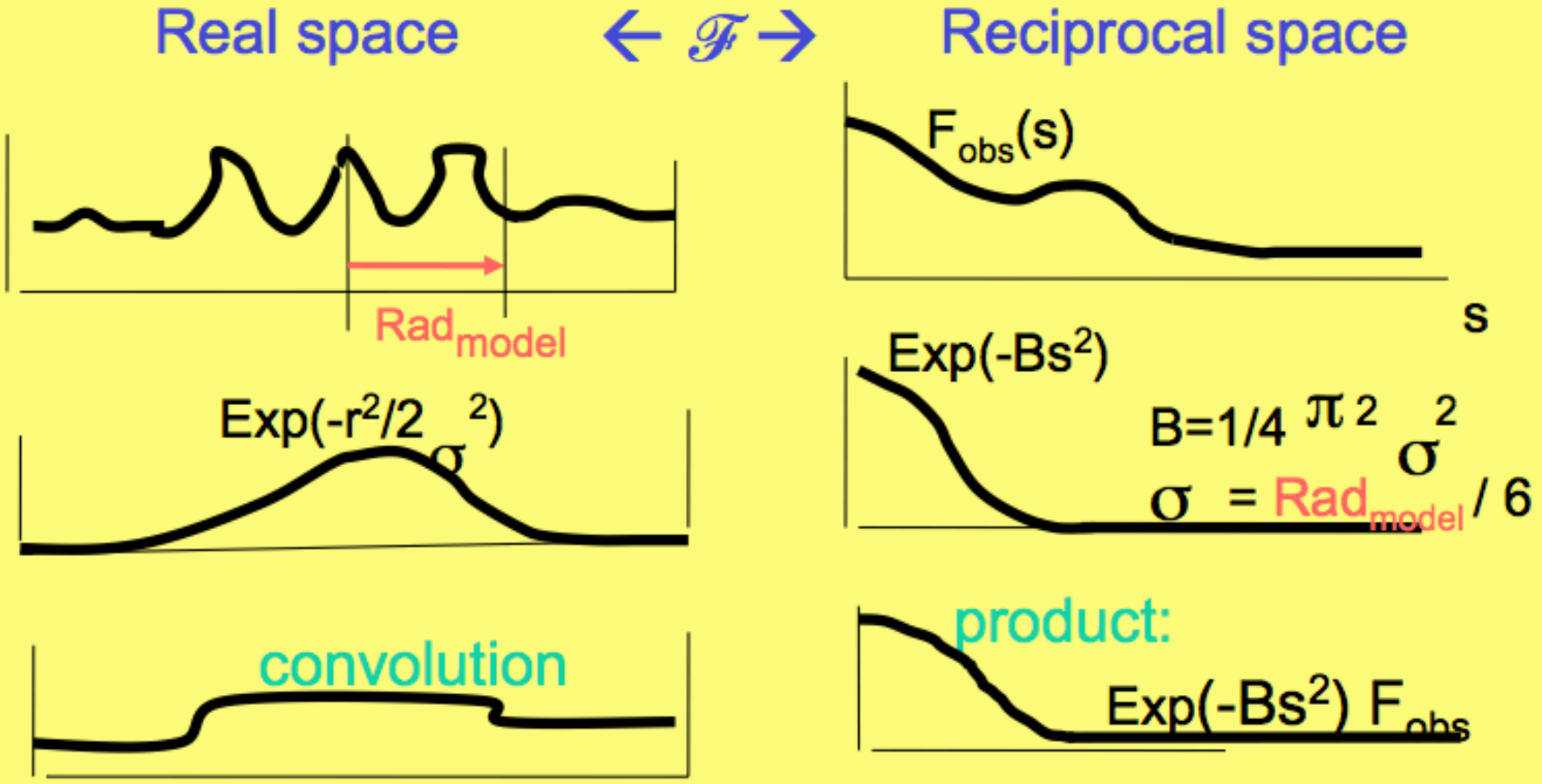
Weights for low resolution data and size of model

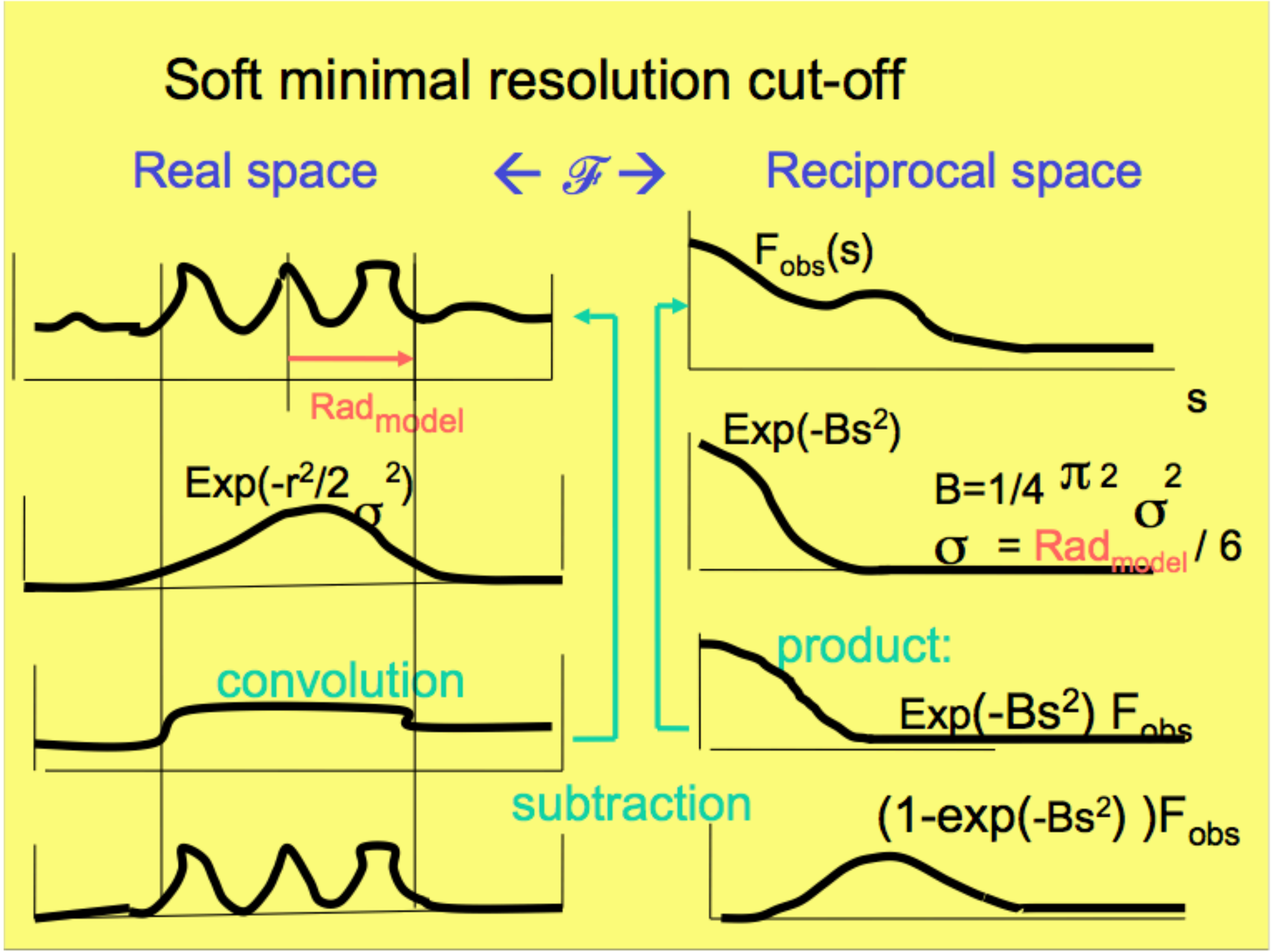
Soft minimal resolution cut-off

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



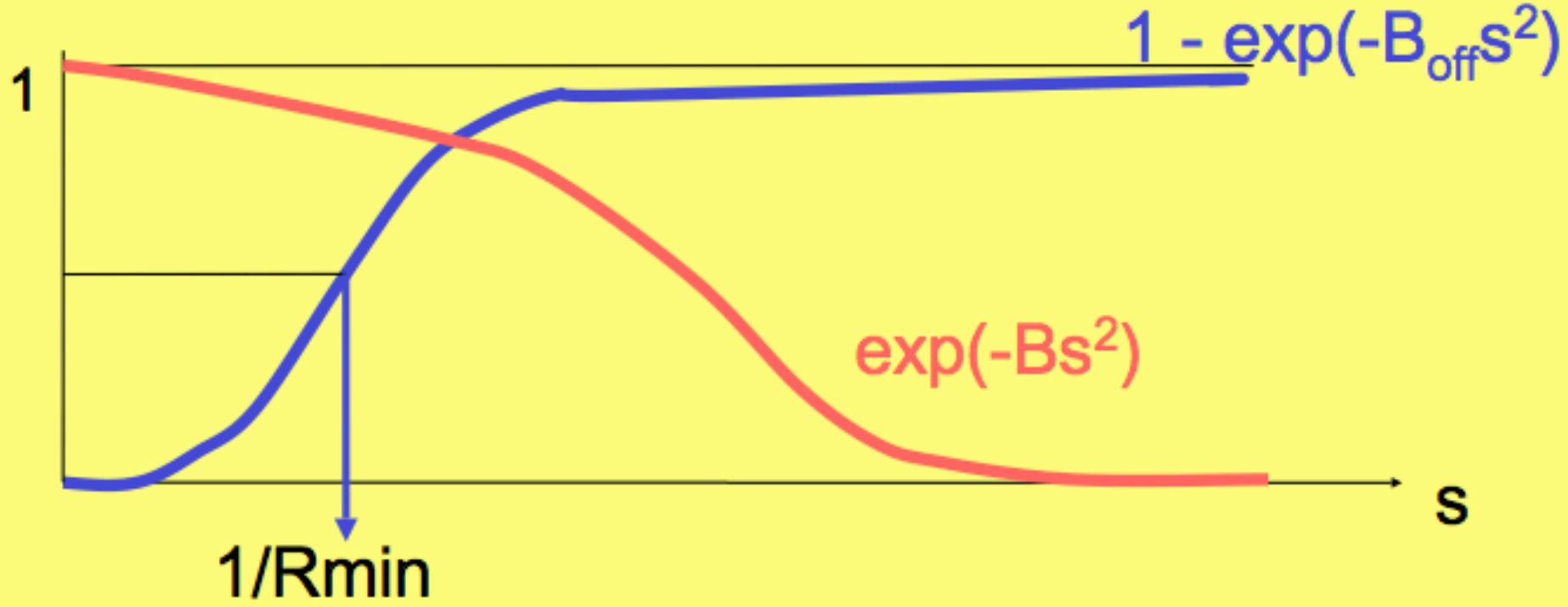
Soft minimal resolution cut-off




Soft minimal resolution cut-off
Real space
← 𝓕 →
Reciprocal space
$F_{obs}(s)$
s
Rad_{model}
$Exp(-r^2/2\sigma^2)$
$Exp(-Bs^2)$
$B=1/4\ \pi^2\ \sigma^2$
$\sigma = Rad_{model} / 6$
convolution
product:
$Exp(-Bs^2)\ F_{obs}$
subtraction
$(1-exp(-Bs^2))F_{obs}$

Weighting scheme

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



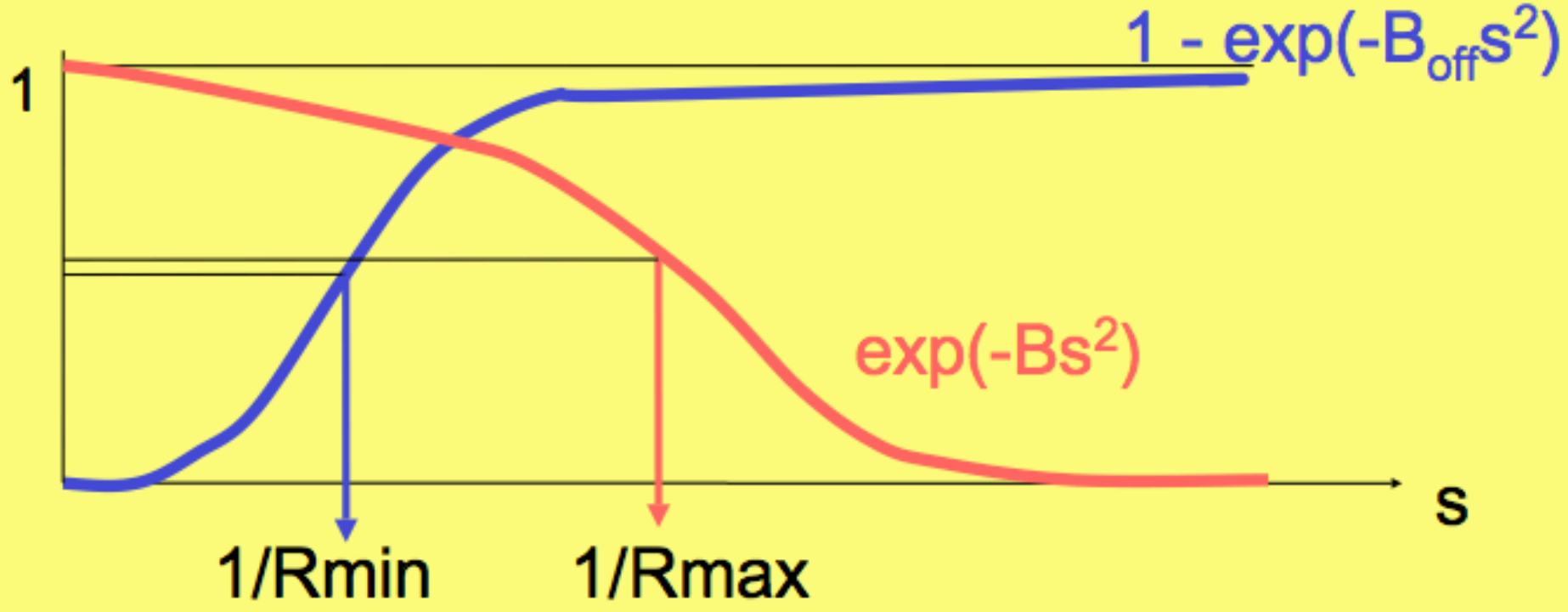
$Rmin = \sqrt{2B_{off}} = 2 \; \pi \sigma_{model}$

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

$Rmin \approx Rad_{model}$

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

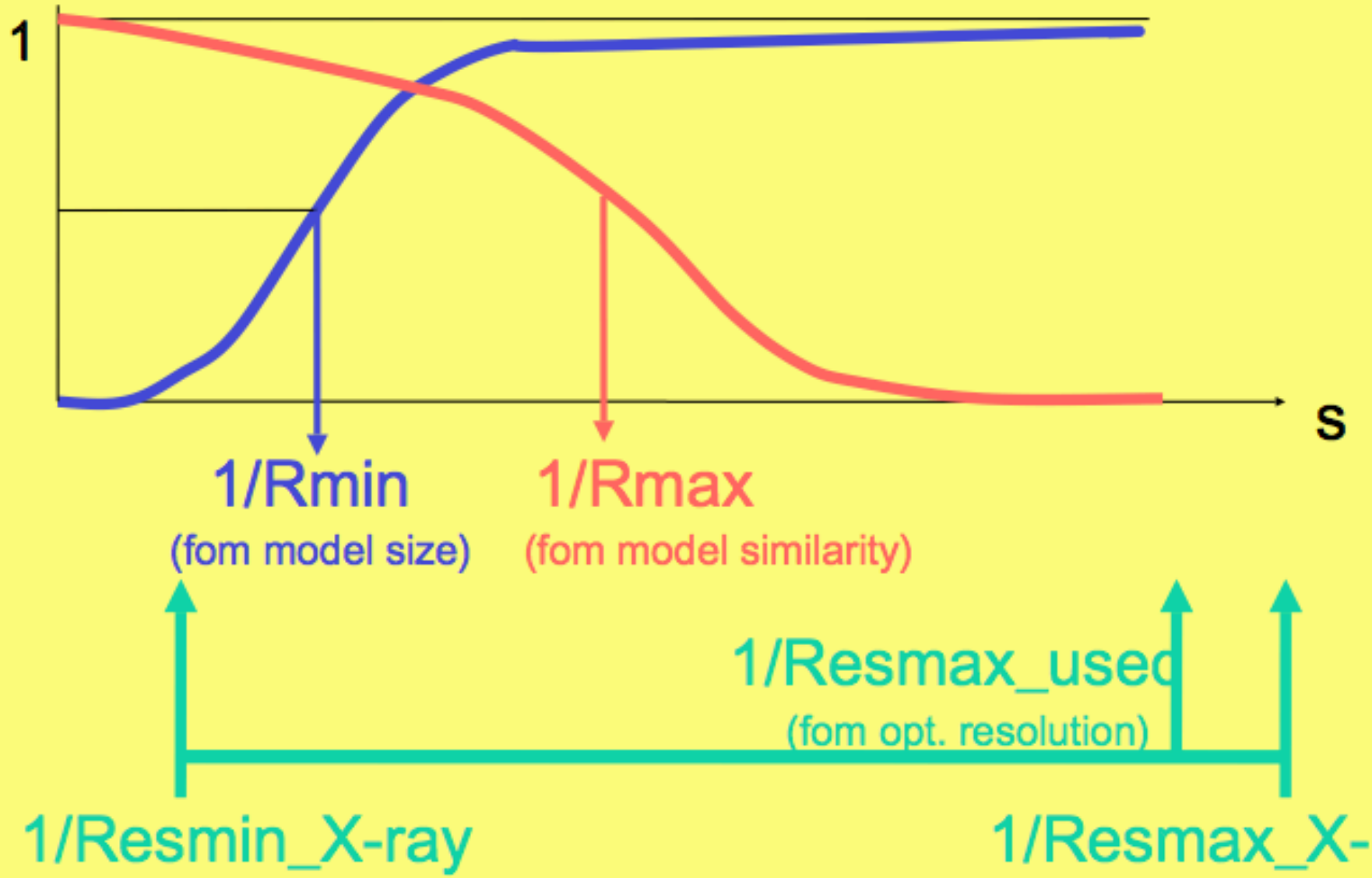
Two filters in Image processing:

Gaussian highpass filter Gaussian lowpass filter

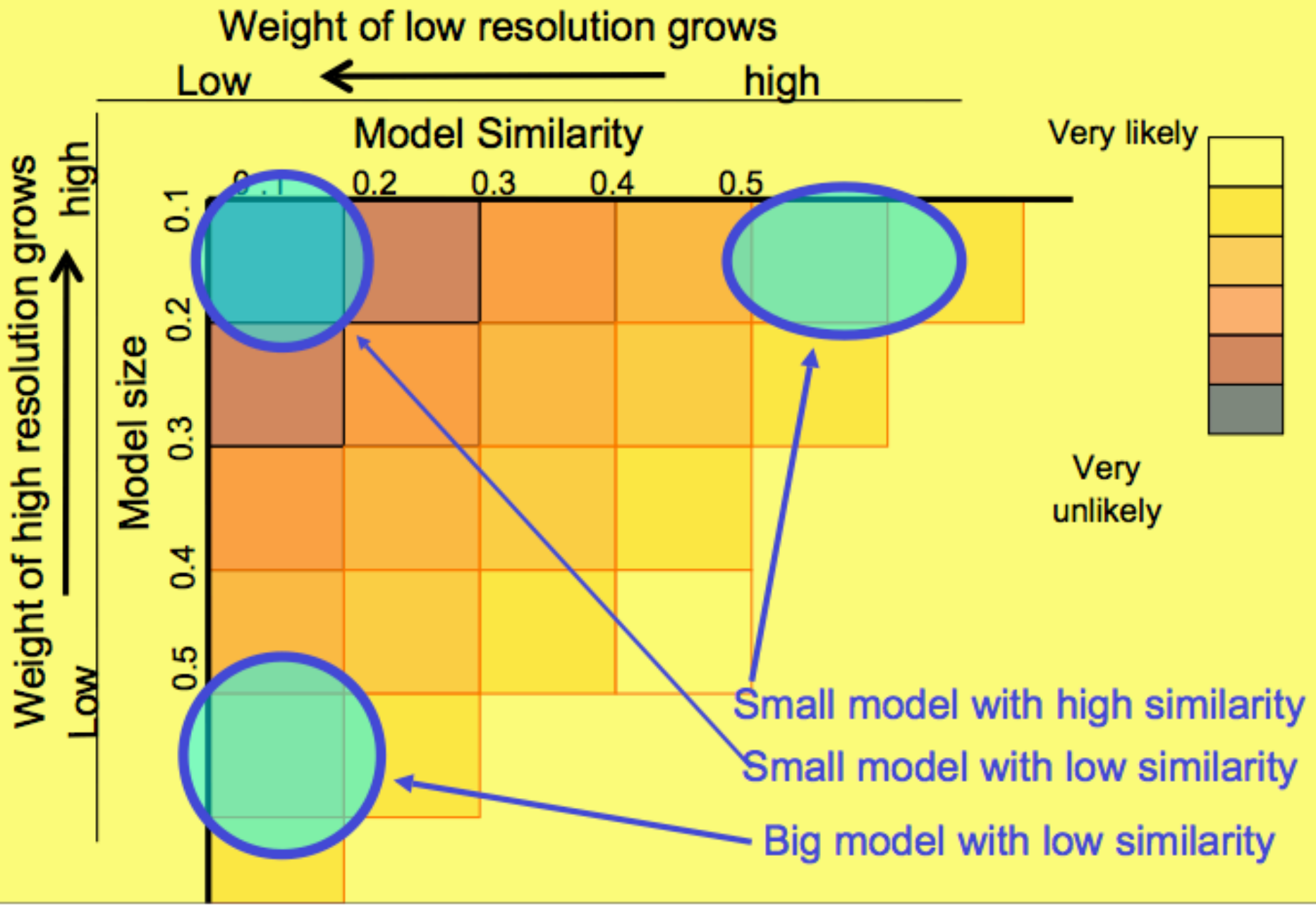
$$F_{used} = (1 - \exp(-B_{off}s^2)) \; F_{obs} \; \exp(-Bs^2)$$

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?



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

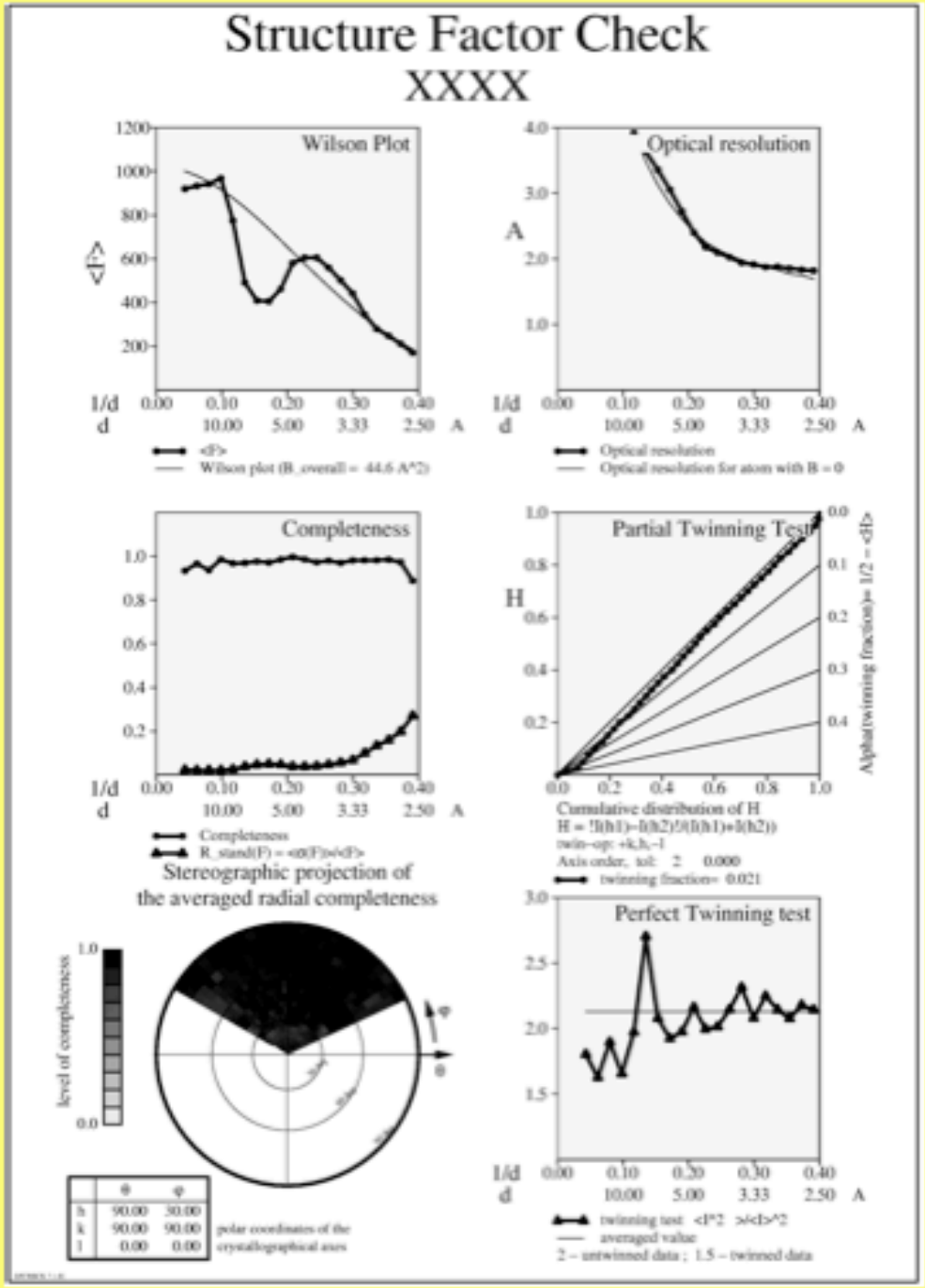
Nominal resolution range:		29.1 – 2.50 A
Reflections in file:		7974
Unique reflections	above 0:	7974
	above 1σ:	7973
	above 3σ:	5026

SFCHECK

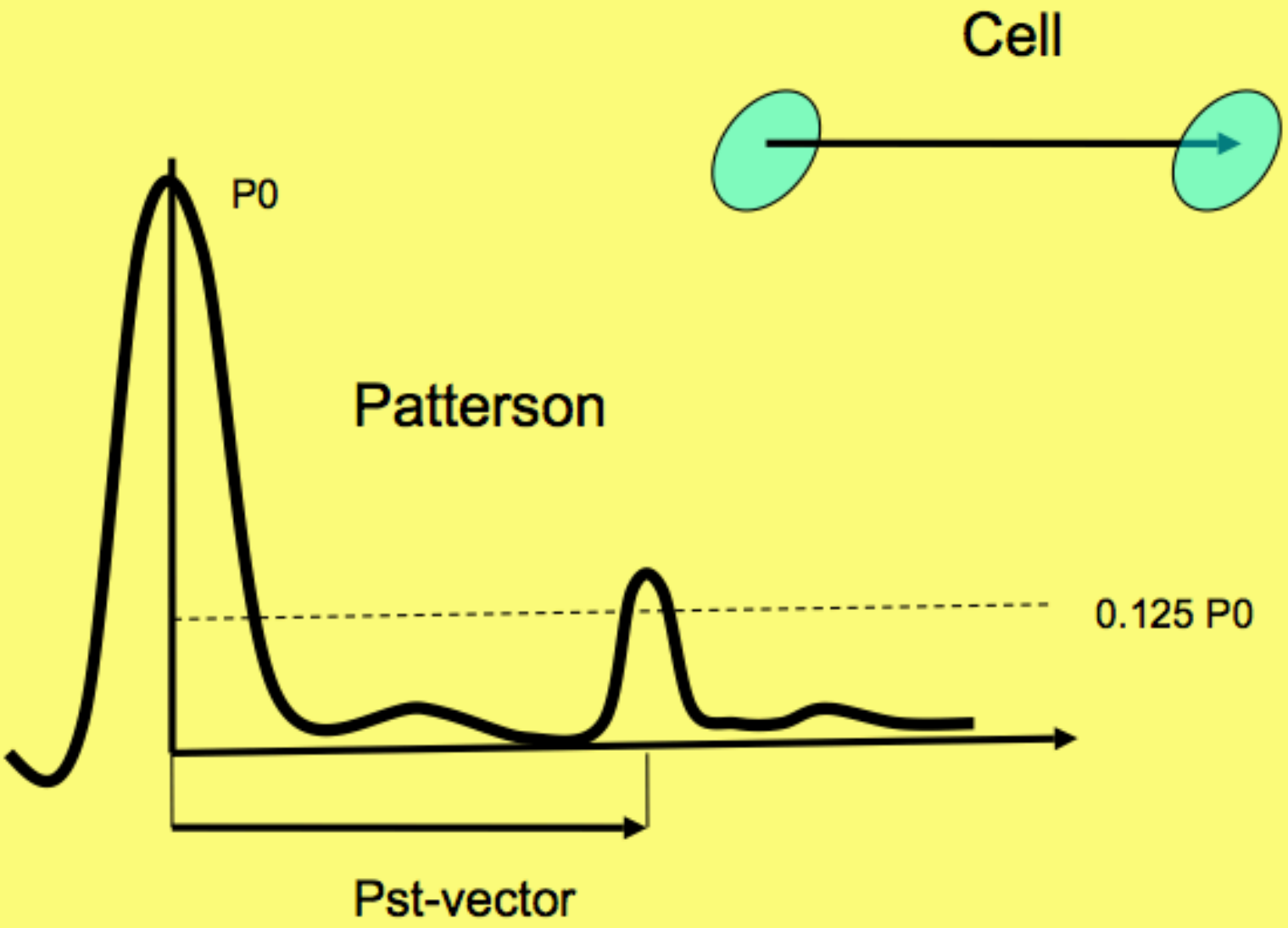
Nominal resolution range: 29.1 – 2.50 A
\05max. from input data, min. from author\05

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
<I>
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
Completeness
R_stand(F) = <n(F)>/<F>
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
Stereographic projection of
the averaged radial completeness
level of completeness
polar coordinates of the
crystallographical axes
Perfect Twinning test
twinning test <I^2 >/<I>^2
averaged value
2 – untwinned data ; 1.5 – twinned data

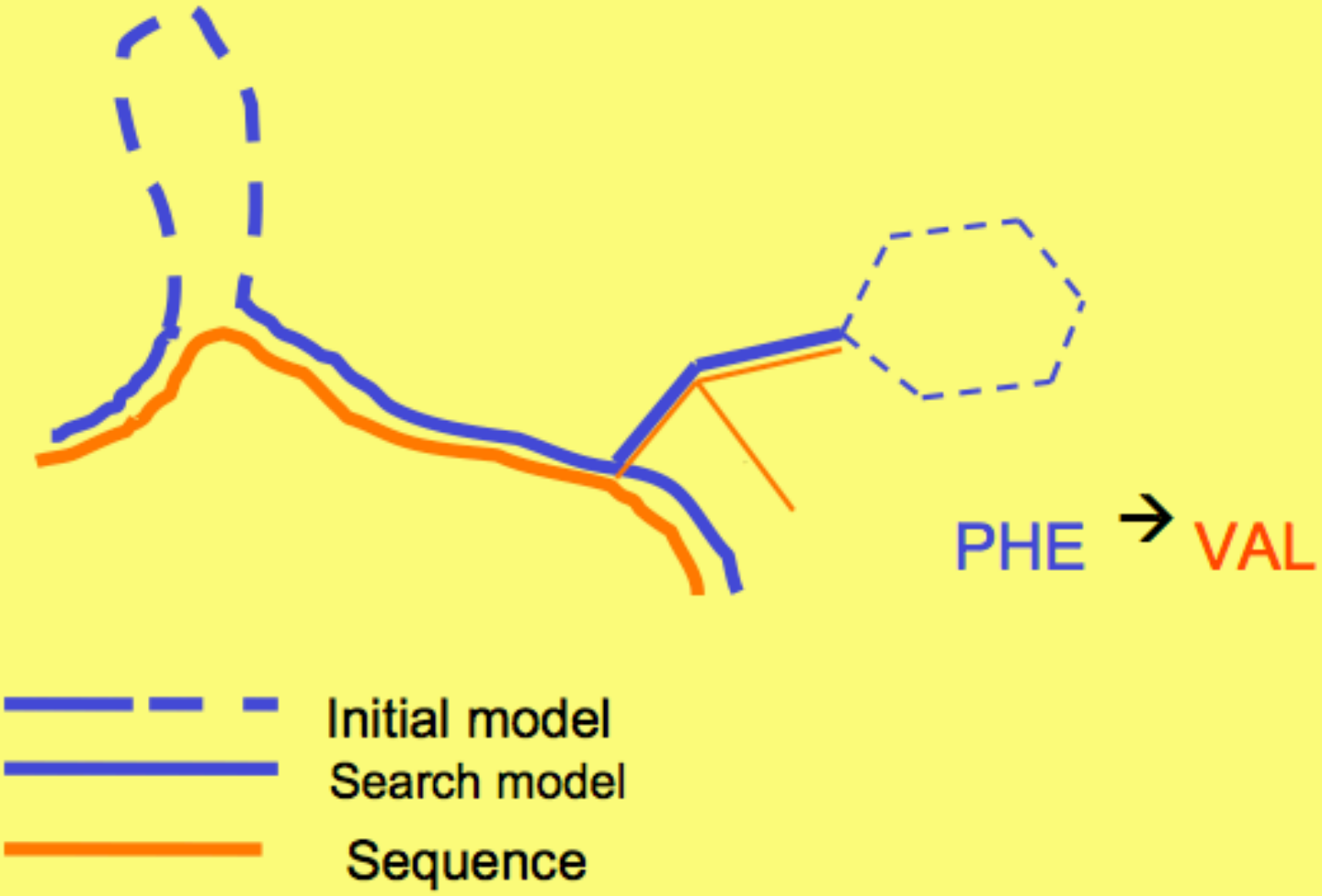
Pseudo-translation



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



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
```

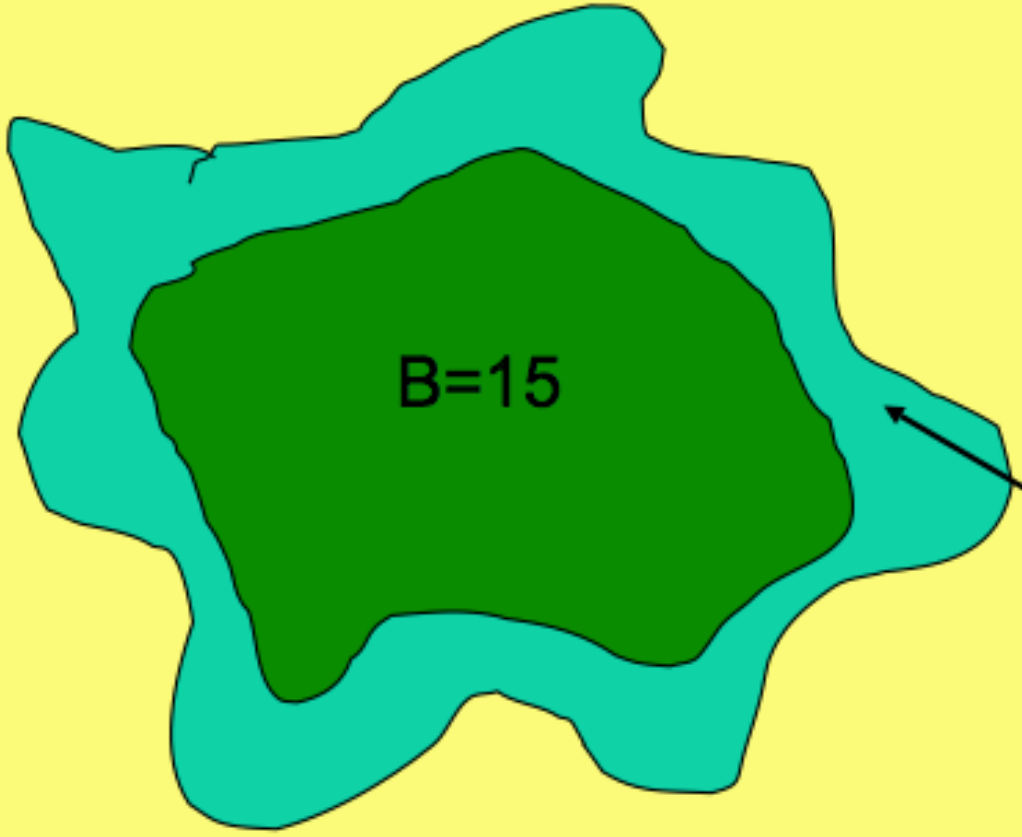
with alignment correction

```
                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 *

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