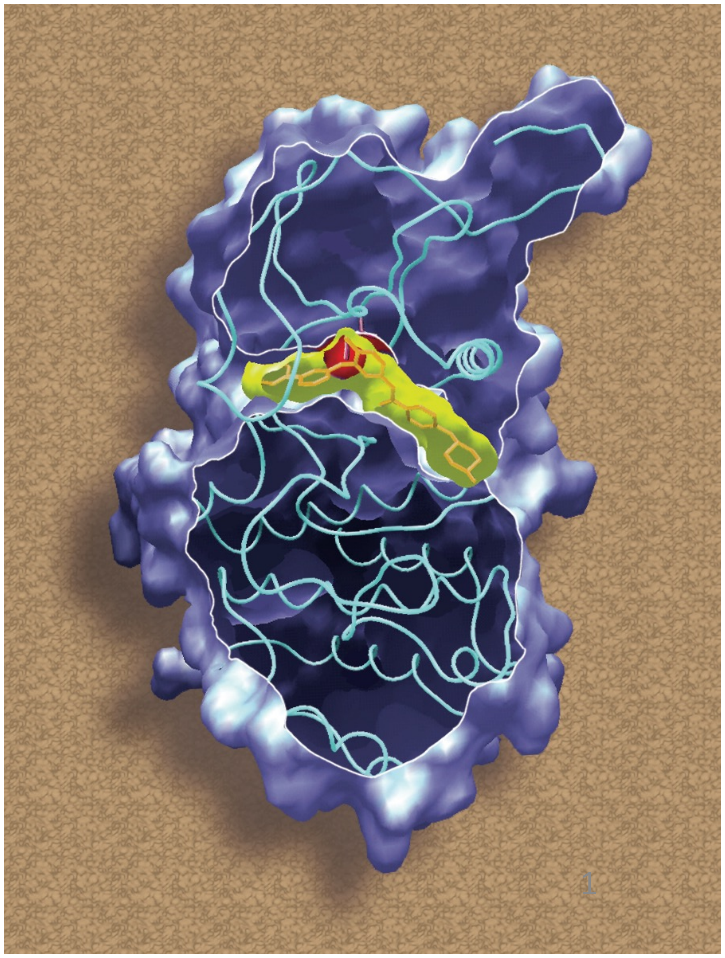
Bi 204 Methods:

Seeing atomic Structure:

Calibrating Molecular Interactions

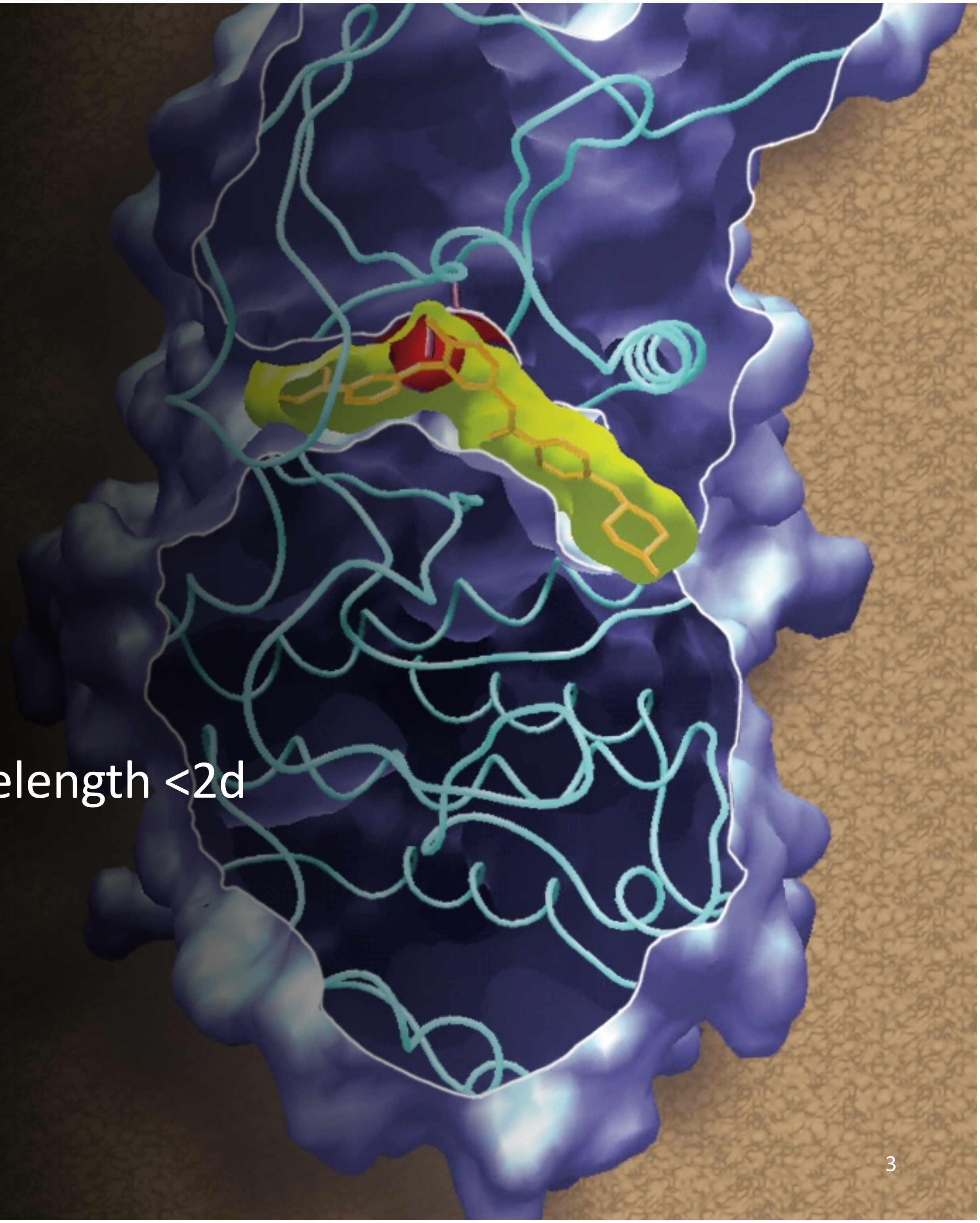
Bob Stroud
2022

stroud@msg.ucsf.edu



A ‘Ligand’ the cancer drug imatinib (Gleevec) bound to the tyrosine kinase Abl.

PRINCIPLE #1

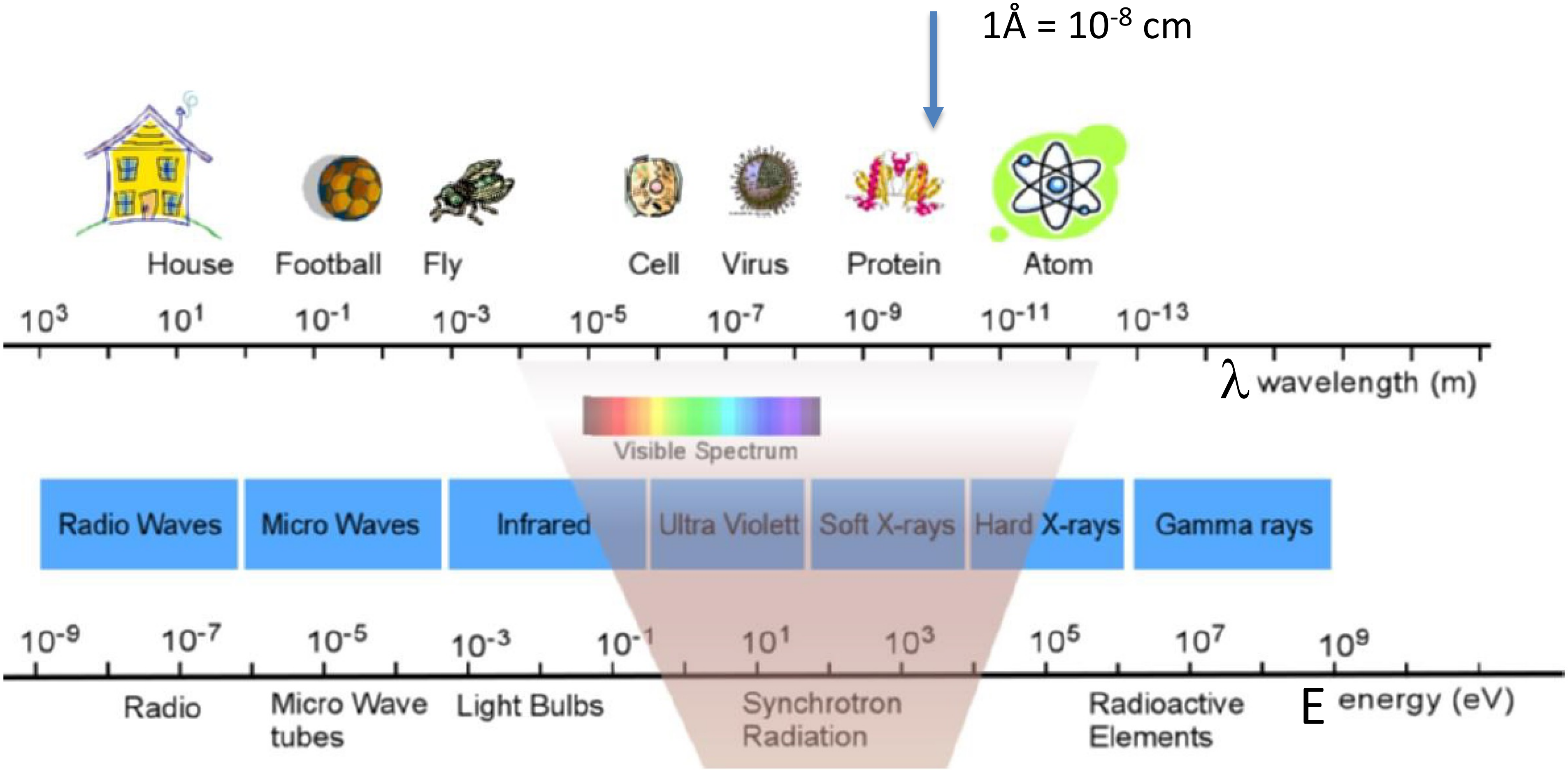


RESOLUTION?

To See things of size d

Need Wave-particles Wavelength <2d

The Electromagnetic Spectrum

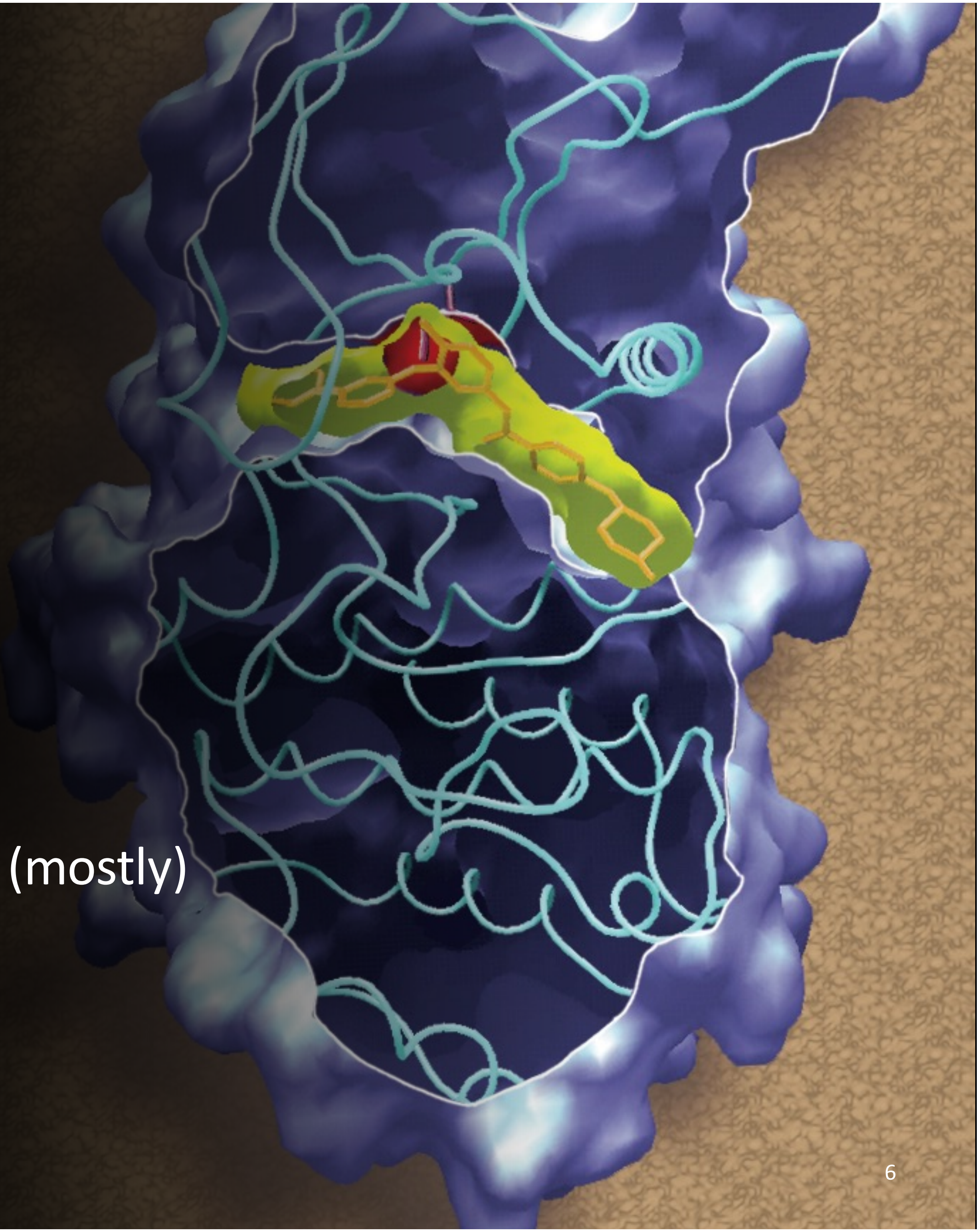

1Å = 10^{-8} cm

$E = h\nu = hc/\lambda$ C = 3.10^{8} m/sec. h = 6.6×10^{-34} J.s

Wavelength of Particles: electrons (neutrons etc…)

- λ=h/p = h/mv   Louis de Broglie
- $eV=1/2\ mv^2$
- Electron mass $m = 9.11 \times 10^{-31}$ kg
- $h = 6.6 \times 10^{-34}$ J.s
- $e= 1.6\ 10^{-19}$ coulombs
- $\lambda= h/ (2meV)^{½} = 12.25 \times 10^{-10}$ cm.

- λ= 0.039 Å at V=100 keV,
-       0.027 Å at   200 keV,
-       0.022 Å at   300 keV



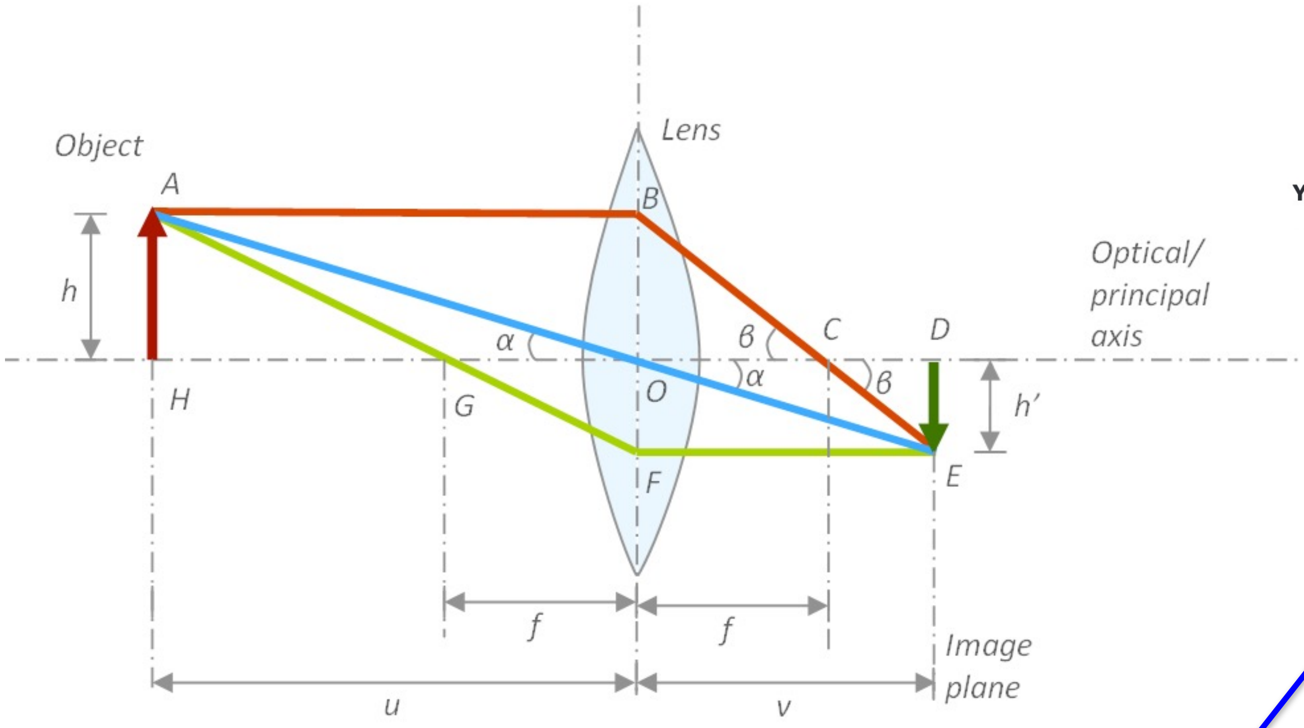
PRINCIPLE #2

SEE WHAT ?

ONLY See what scatters λ

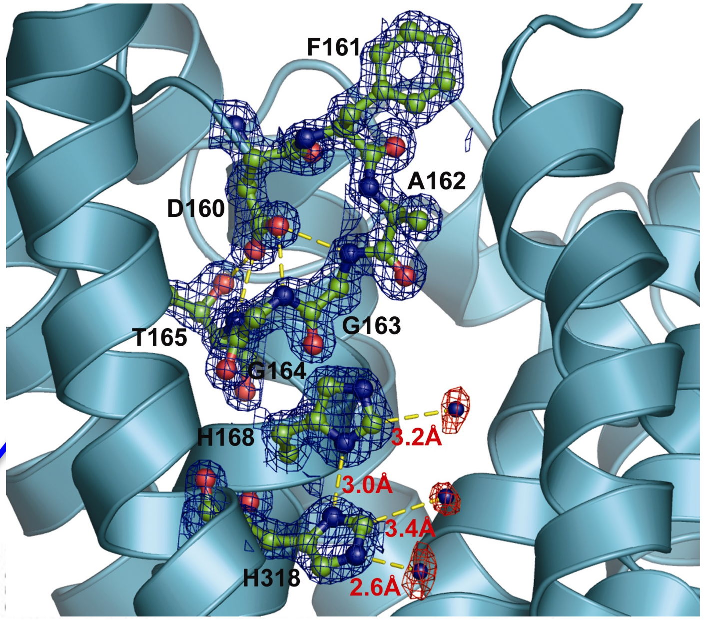
Experiments: SINGLE λ (mostly)

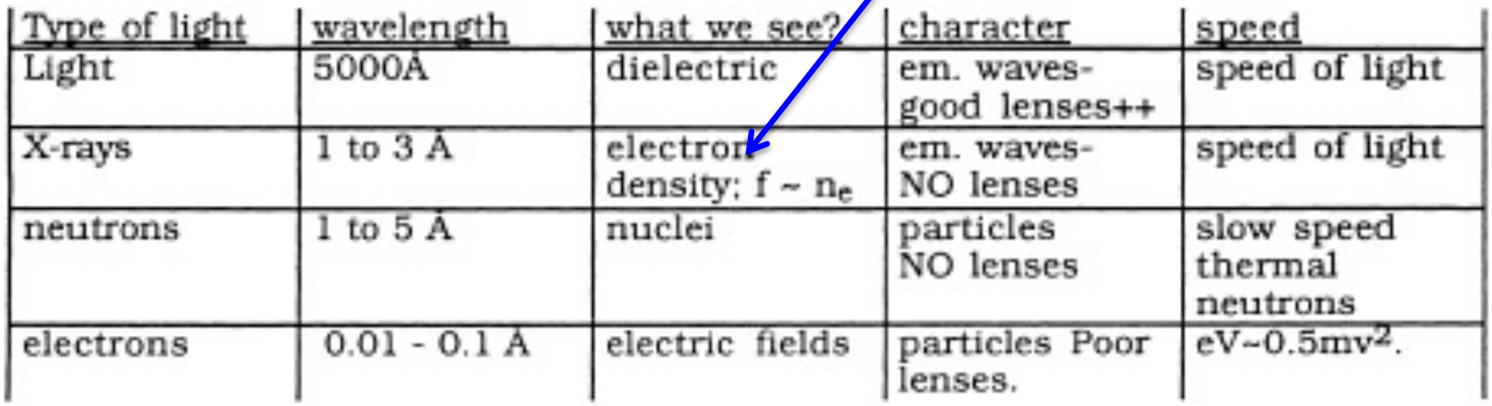
Optical image formation, - with/without lenses



Mechanism of Ammonia Transport by Amt/MEP/Rh: Structure of AmtB at 1.35 Å

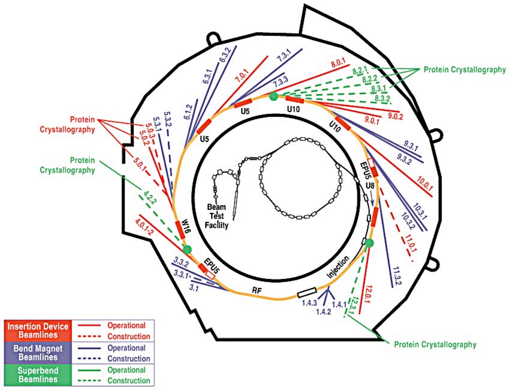
Shahram Khademi, Joseph O'Connell III, Jonathan Remis,
Yaneth Robles-Colmenares, Larry J. W. Miercke, Robert M. Stroud*

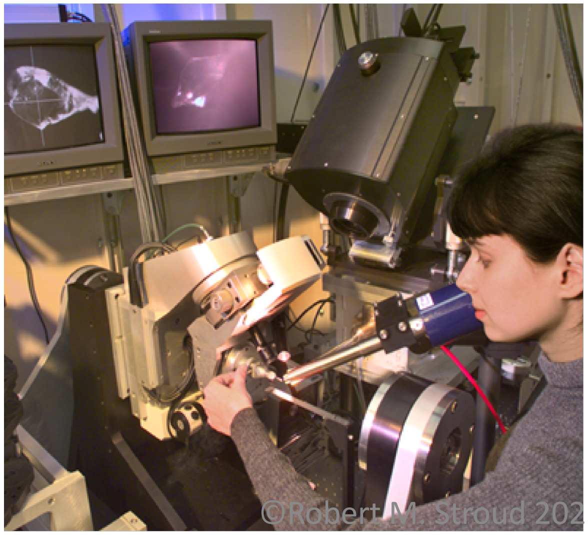




Type of light	wavelength	what we see?	character	speed
Light	5000A	dielectric	em. waves- good lenses++	speed of light
X-rays	1 to 3 Å	electron density: f ~ n_e	em. waves- NO lenses	speed of light
neutrons	1 to 5 Å	nuclei	particles NO lenses	slow speed thermal neutrons
electrons	0.01 - 0.1 Å	electric fields	particles Poor lenses.	eV~$0.5mv^2$.

The UCSF beamline 8.3.1





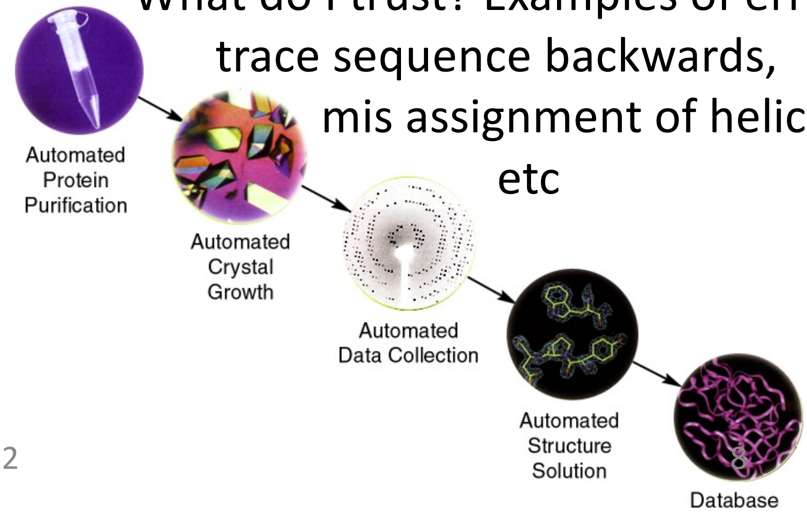
If automated- why are there errors?

What do I trust? Examples of errors

trace sequence backwards,

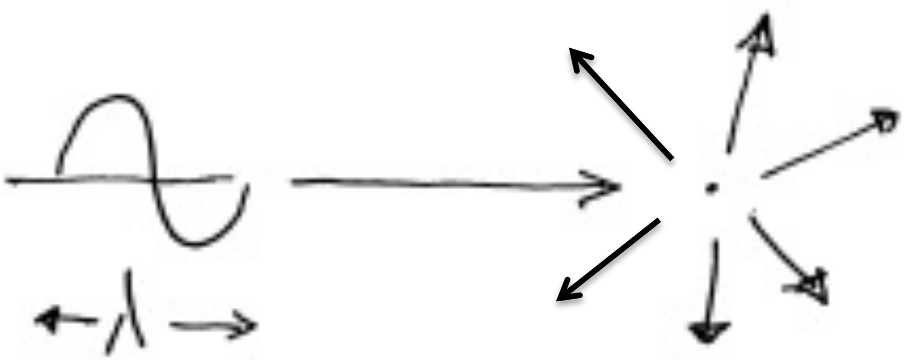
mis assignment of helices

etc



THE CENTRAL AXIOM

Elastic Scattering from a point is equal in every direction



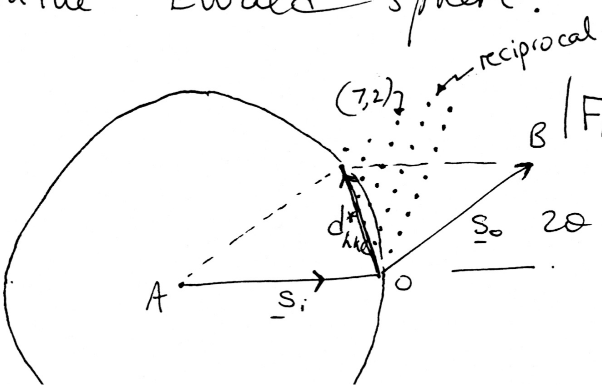
Scattering from a point is equal in all directions.

John Desmond Bernal

describe any reciprocal lattice point as

$$\underline{s} = \underline{d}^* = ha^* + kb^* + lc^*$$

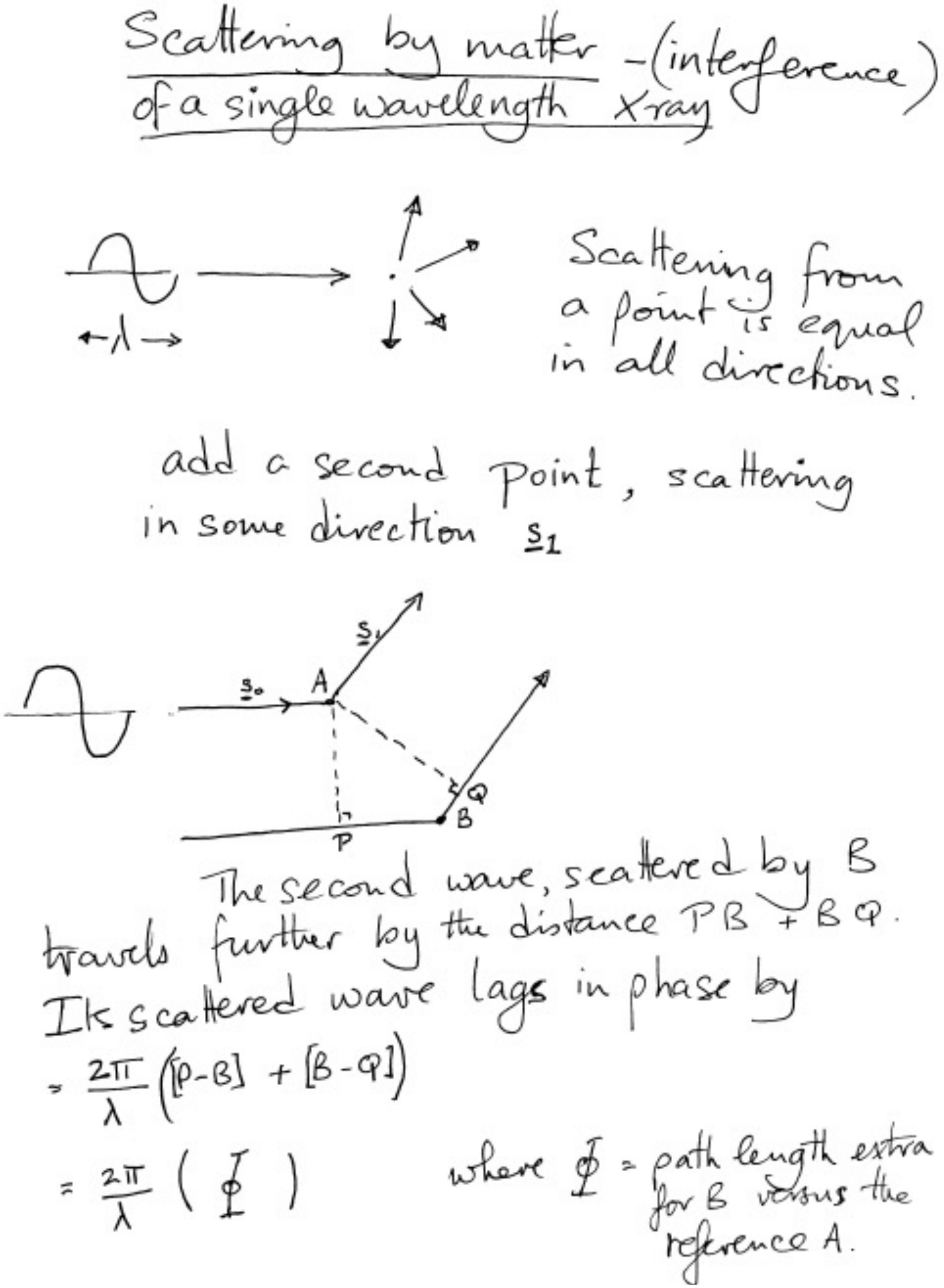
This (hkl) lattice point will only be observed when the crystal is turned so ~~that~~ the (hkl) point $\underline{s} = d^*_{hkl}$ ~~lies~~ lands in the Ewald sphere.



$$|F_{hkl}|^2 = G(\underline{s})\,G^*(\underline{s}) = I_{hkl}$$

Scattering from multiple points? Add wave amplitudes with phase change

Adding up the scattering of Atoms:
Amplitudes, ‘interference’ of waves



Scattering by matter -(interference) of a single wavelength X-ray

Scattering from a point is equal in all directions.

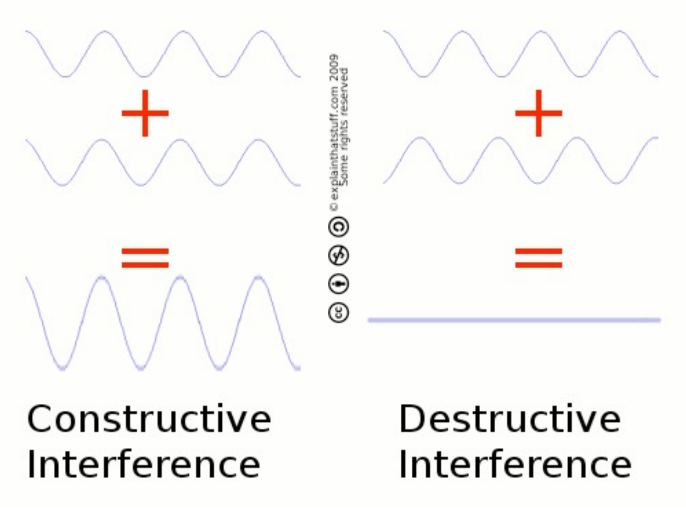
add a second point, scattering in some direction $\underline{s}_1$

The second wave, scattered by B travels further by the distance PB + BQ. Its scattered wave lags in phase by

$$= \frac{2\pi}{\lambda}\left([P-B] + [B-Q]\right)$$

$$= \frac{2\pi}{\lambda}\left(\Phi\right)$$

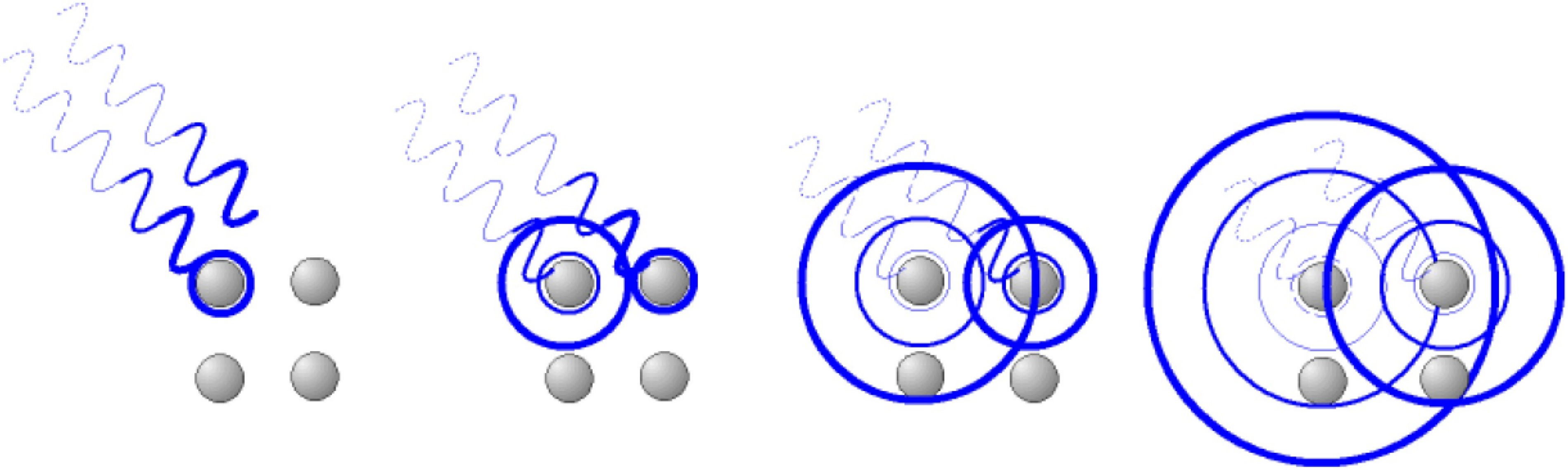
where Φ = path length extra for B versus the reference A.



Waves add out of phase
by 2π[extra path/λ]

Elastic scattering for structure determination

X-ray scattering



X-rays are scattered at the electrons of the atomic shell. During the scattering process the electron is starts oscilating. It becomes a dipol and a spherical wave is sent out. The wavelength and energy of the scattered wave does not change (elastic scattering).

We only ever observe INTENSITY:

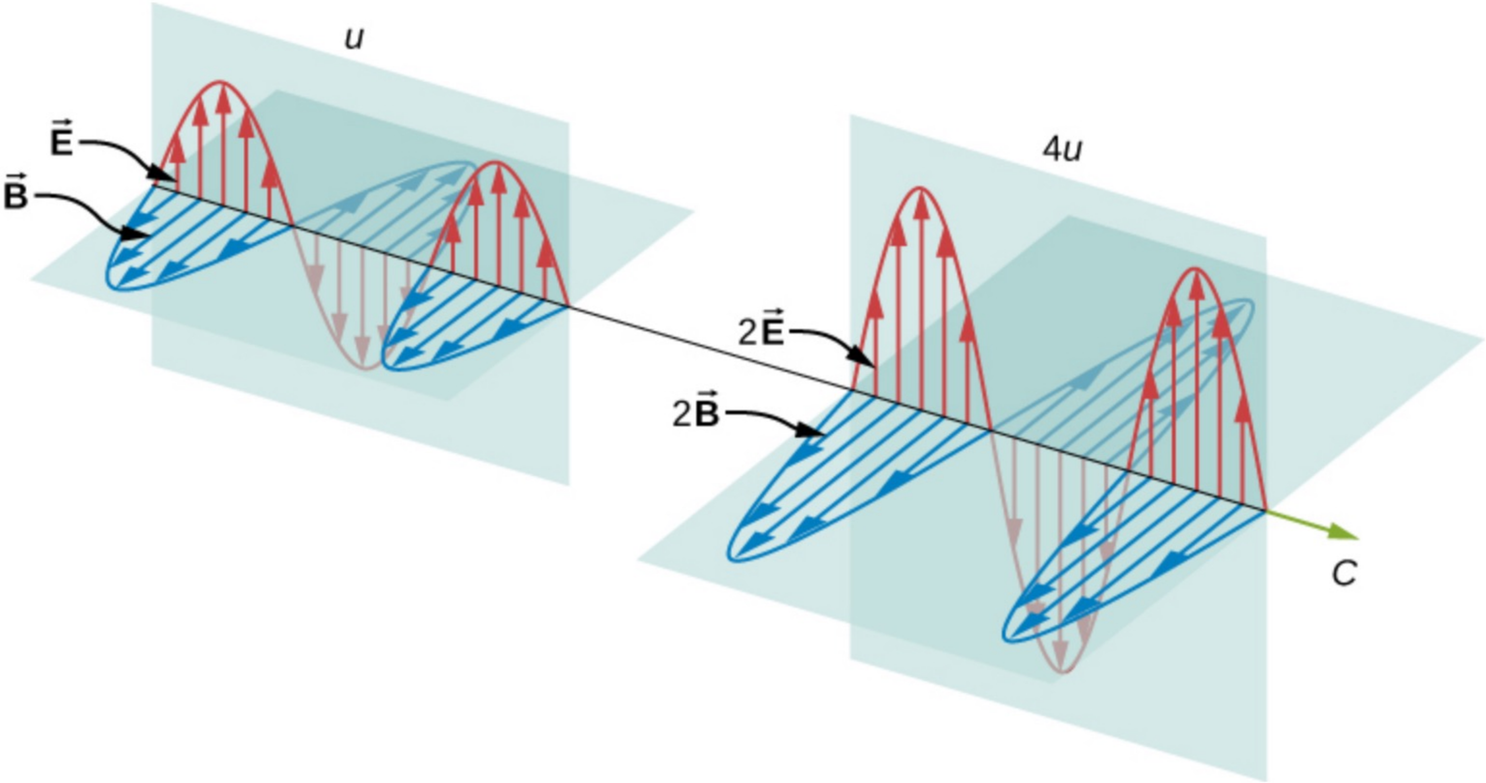
eg. For X-rays, No Lenses……

Intensity (we can observe) = |Amplitude|^2

calculate Amplitude = $\sqrt{\textbf{Intensity}}$

I= ***E***x***B*** in the c direction

I = $E_0\ B_0\ /2\mu_0$ ~ Amplitude^2



μ_0 = Permeability of free space

We observe Intensity, (can’t observe Amplitude directly)

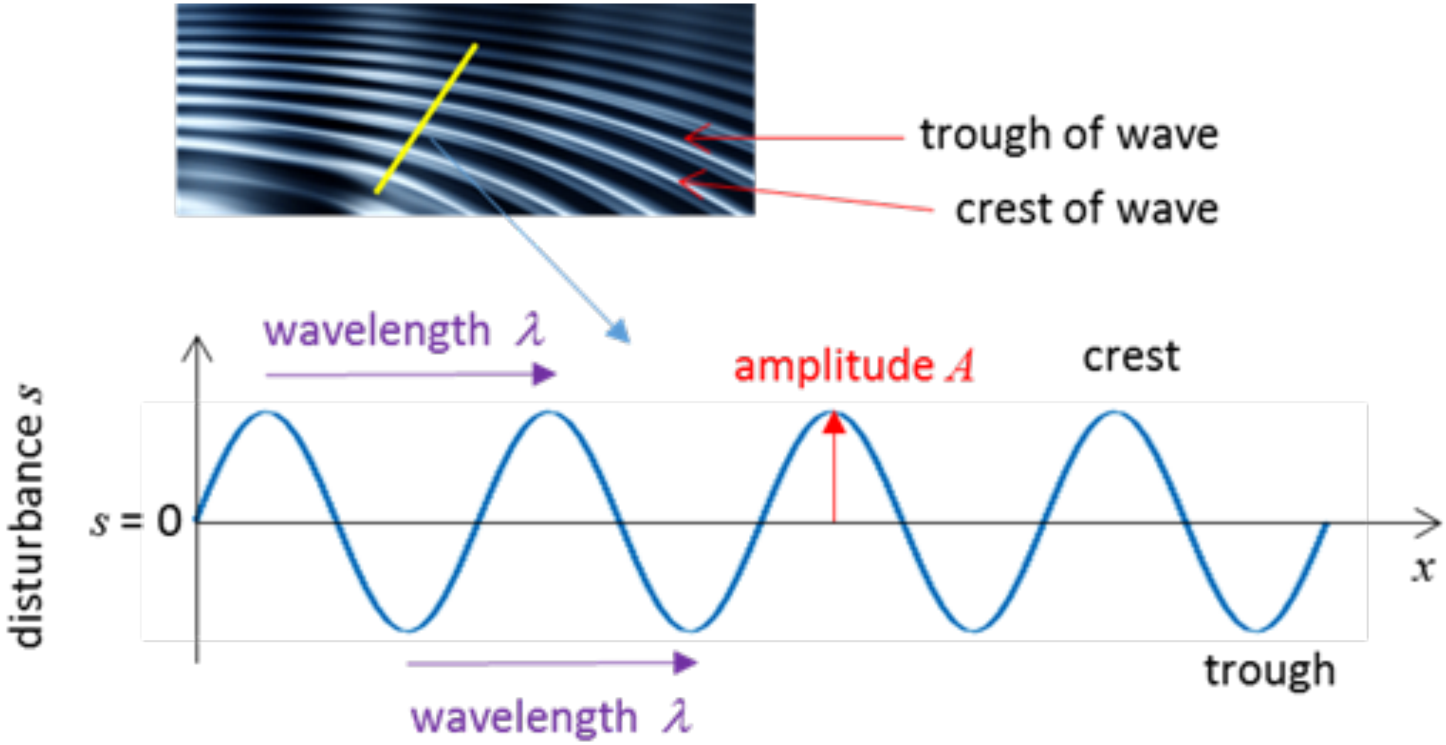
Intensity = Power/ unit area = Energy/sec . unit area

For a vibrating particle, Energy = $\frac{1}{2} mv^2 \sim \frac{1}{2} m\,(ds/dt)^2$

Eg. If $s = a_0 \sin(\omega t)$ so $ds/dt = a_0 \cos(\omega t)$

Integral $(ds/dt)^2$ over time $= a_0{}^2/2$

Energy $\sim \frac{1}{2} m\,(a_0)^2$ per $1/\omega$ time **Intensity ~ $(a_0)^2$**



water wave at one instance of time

Hence Amplitude = $|F|$

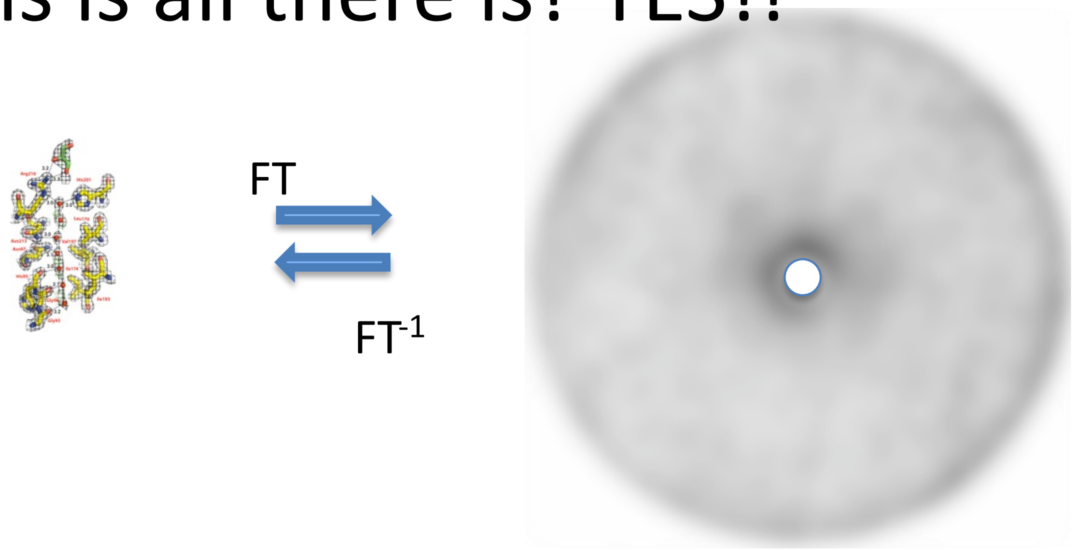
$|F|^2 = \underline{\mathbf{F}}.\underline{\mathbf{F}}^* =$ Intensity

Measure Intensity, each spot position (h,k,l)

Take $\sqrt{\text{Intensity}} = |F|_{h,k,l}$

Then Need relative phase of each….

This is all there is? YES!!

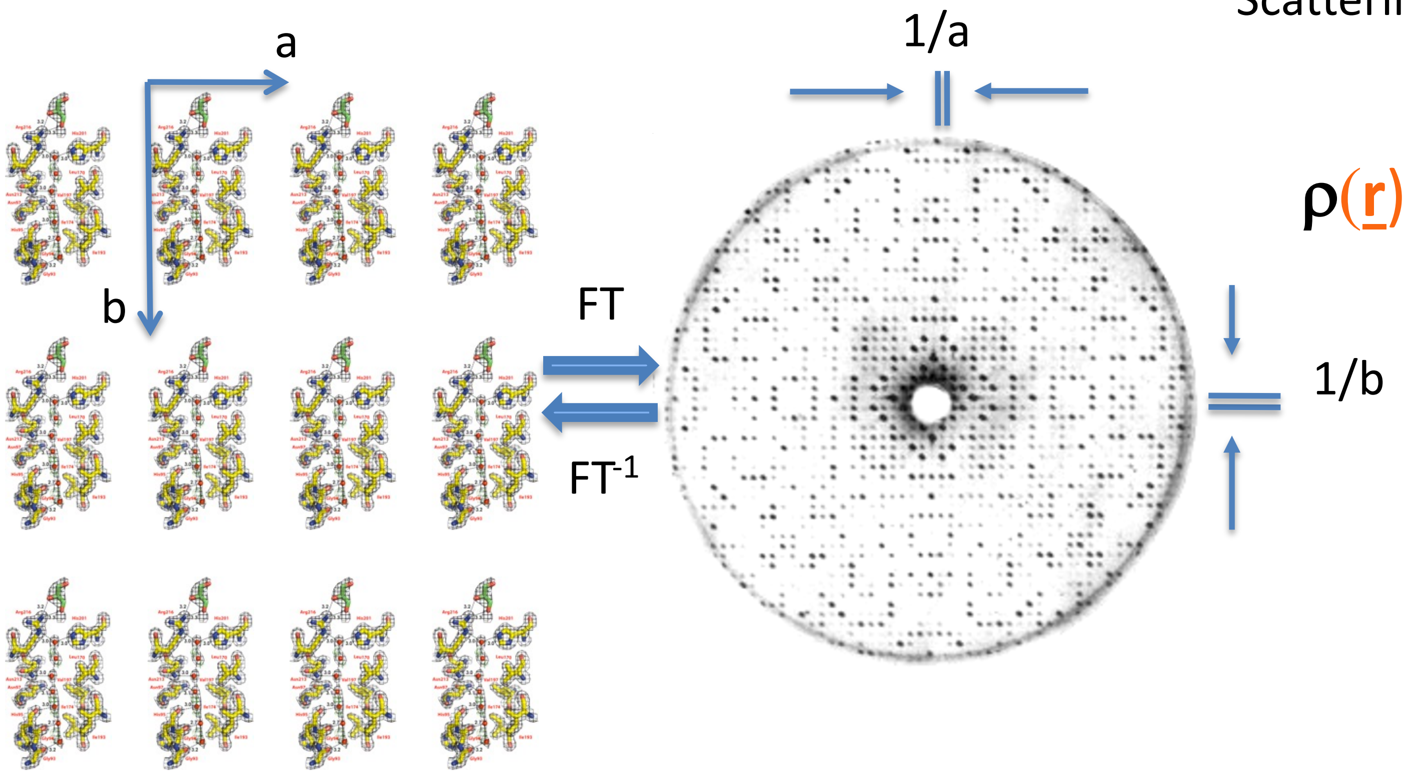


Scattering pattern is the
Fourier transform (FT) of the structure:
Amplitude and phase of waves is a
sum of waves from each atom **j**

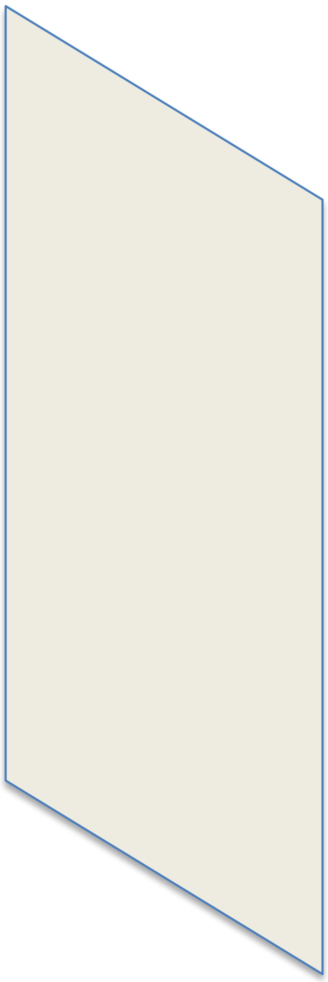
$$\underline{F}(\underline{s}) = \Sigma_j f_j e^{(2\pi i \underline{r}_j . \underline{s})}$$

$$\textbf{Observe } I(\underline{S}) = \underline{F}(\underline{s}).\underline{F}^*(\underline{s})$$

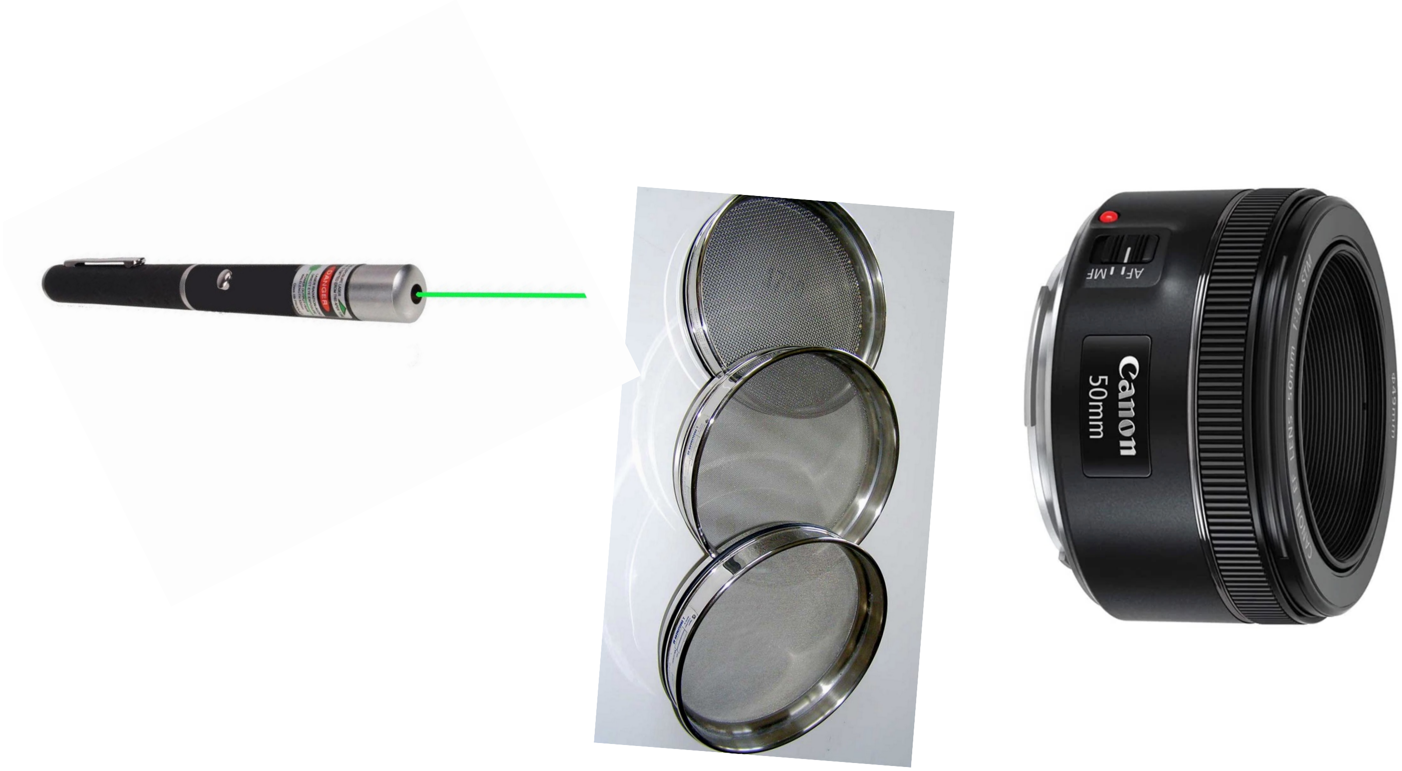
Structure is the ‘inverse’
Fourier transform of the
Scattering pattern $\underline{F}(\underline{s})$

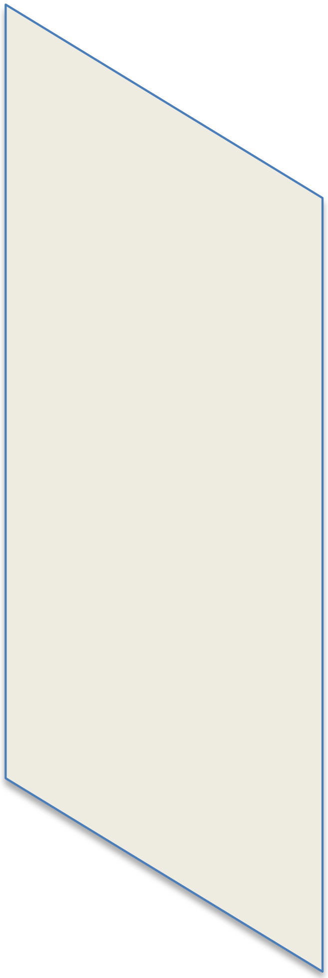


$$\rho(\underline{r}) = \Sigma_s \underline{F}(\underline{s})\ e^{(-2\pi i \underline{r}.\underline{s})}$$



Optical Diffraction.
Source of photons a single wavelength
A ‘2D crystal’ (repeating square shapes)
A screen 10 feet in front of the crystal

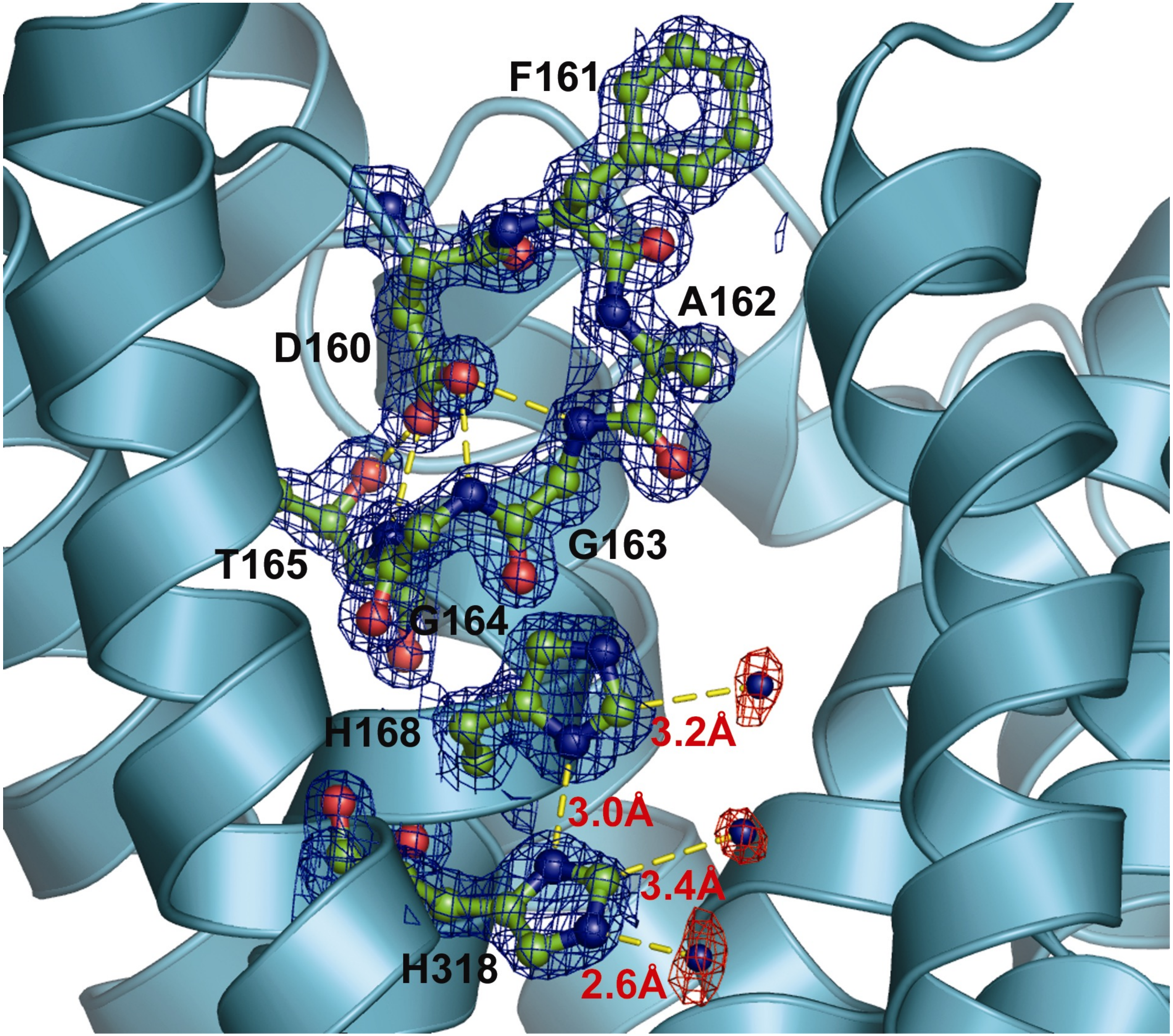




Optical Diffraction.
Source of photons a single wavelength
A '2D crystal' (repeating square shapes)
A screen 10 feet in front of the crystal

And a lens..

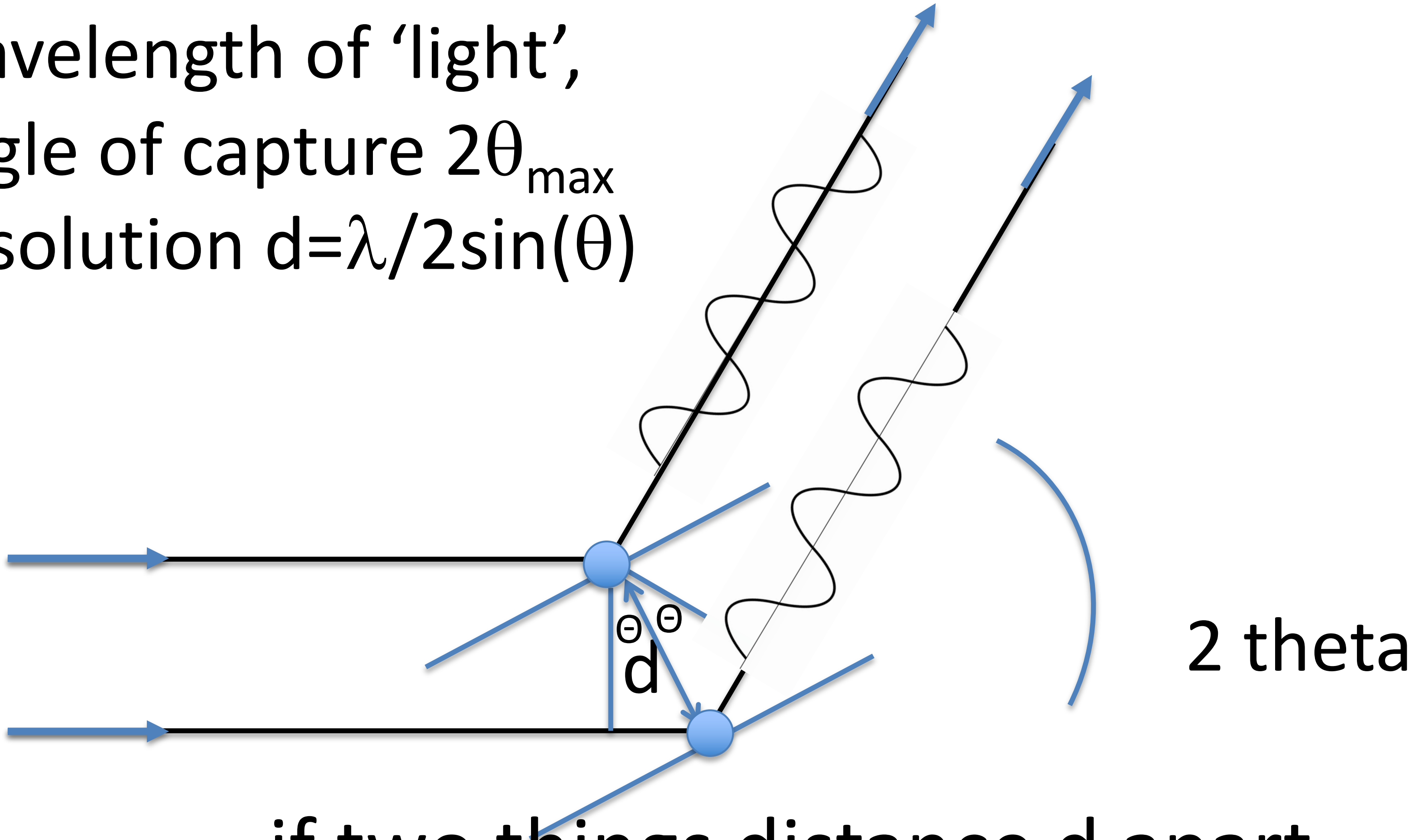
RESOLUTION ? NH3 sites and role of D160 at 1.35Å Resolution



©Robert M. Stroud 2022

Resolution depends on

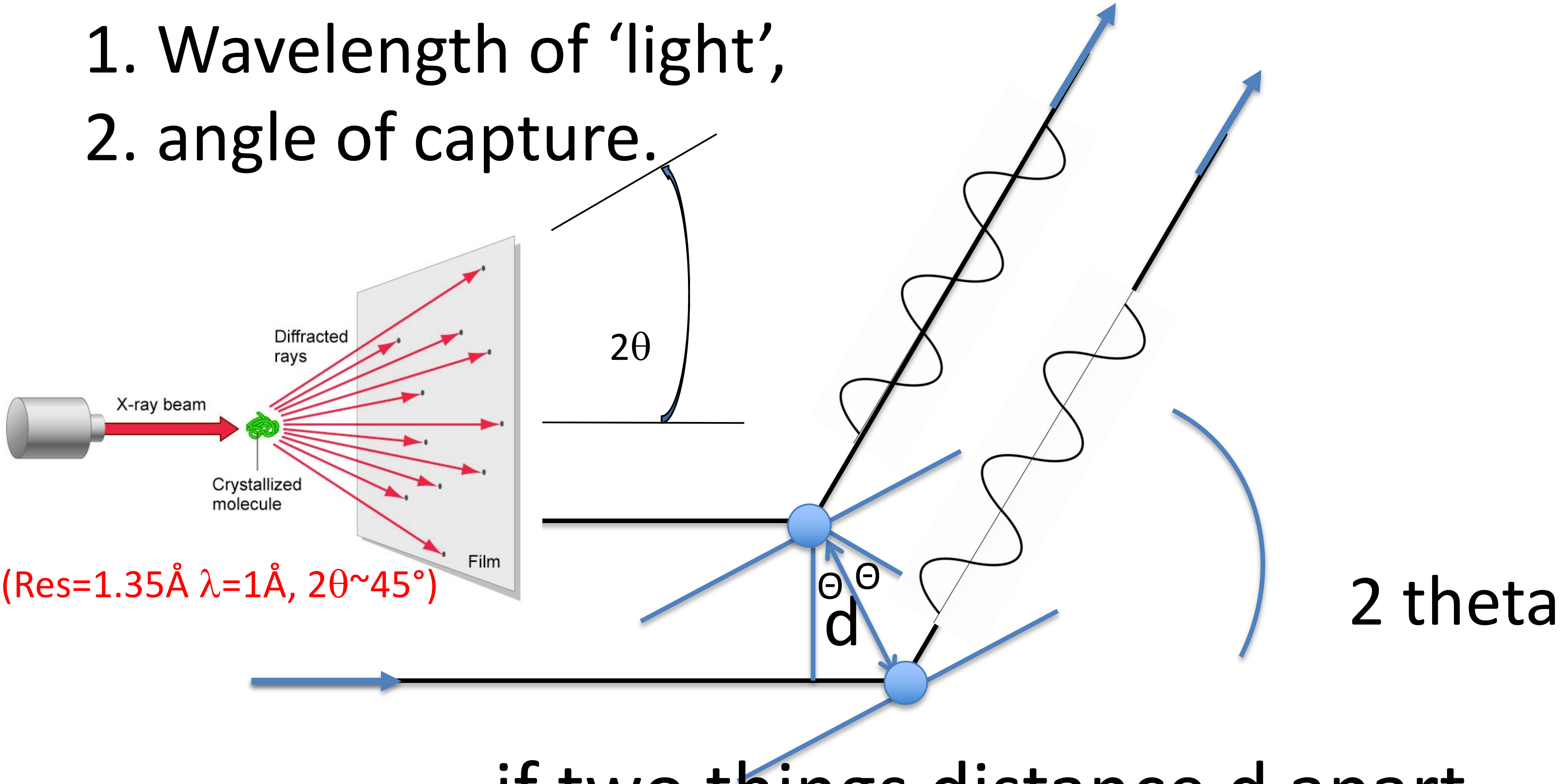
1. Wavelength of ‘light’,
2. angle of capture $2\theta_{max}$
3. Resolution $d=\lambda/2\sin(\theta)$



if two things distance d apart.
scattered waves reinforce when
$2\ d \sin(\Theta) = \lambda$

Resolution depends on

1. Wavelength of ‘light’,
2. angle of capture.

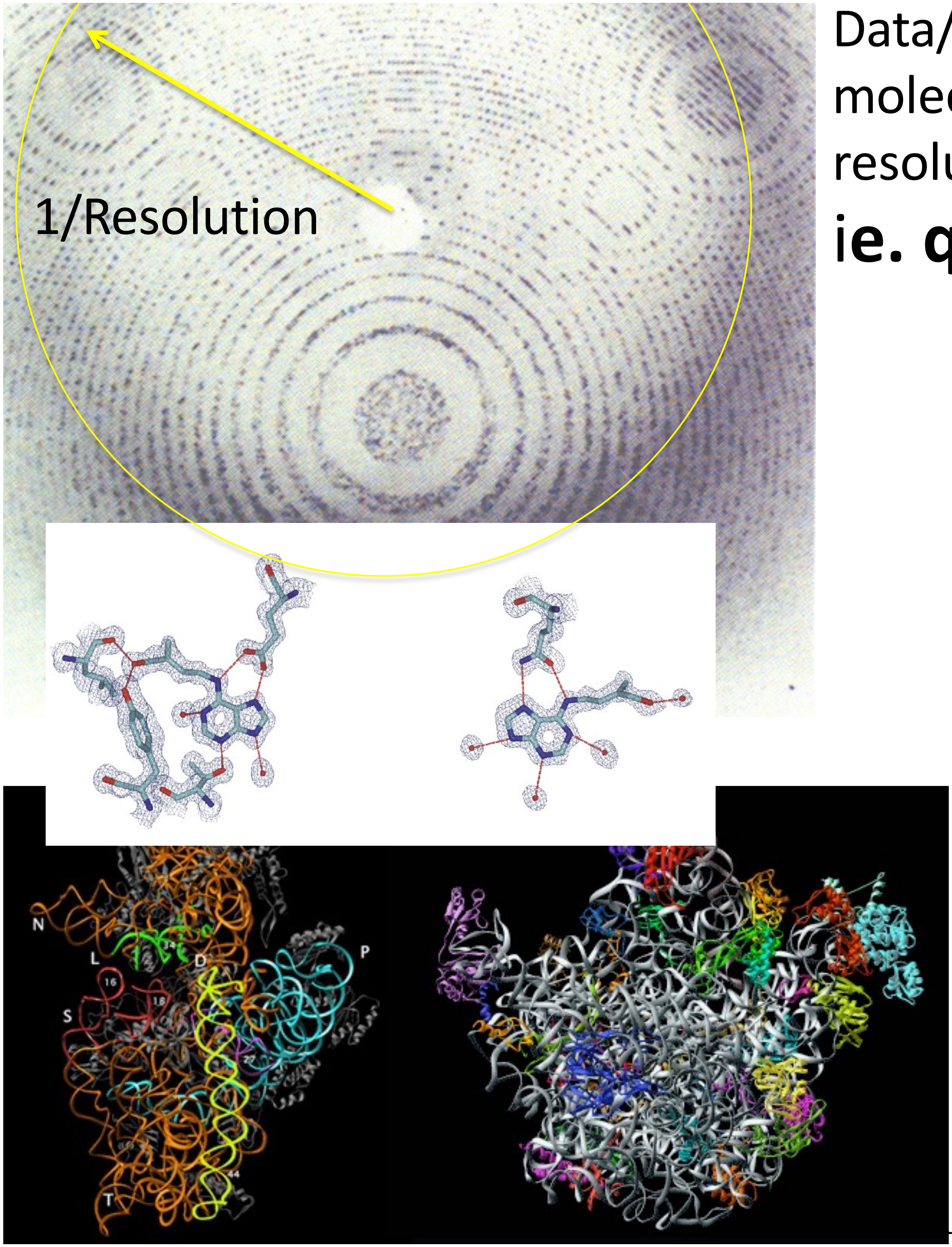


(Res=1.35Å λ=1Å, 2θ~45°)

if two things distance d apart.

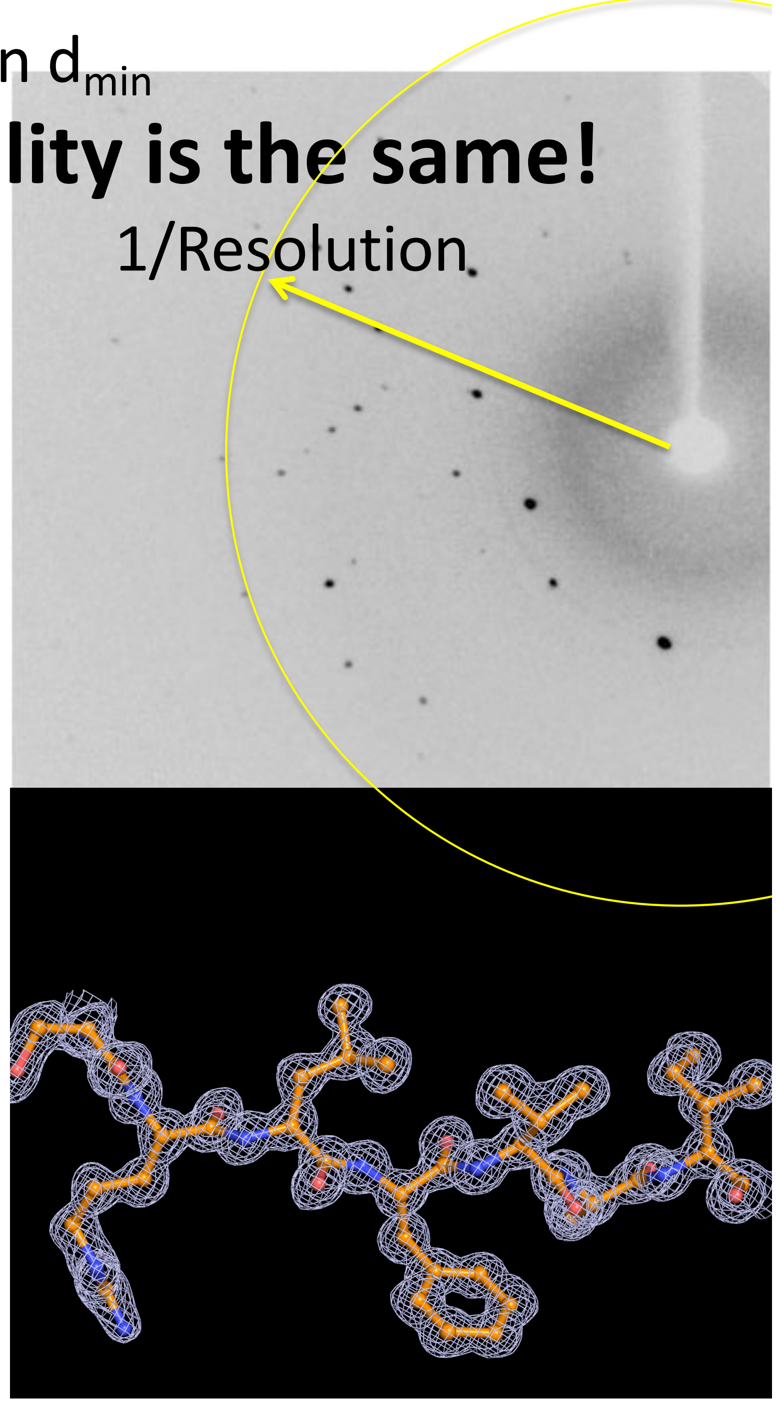
scattered waves reinforce when

$2\ d \sin(\Theta) = \lambda$



Data/Parameter is the same for all molecular sizes at the same resolution d_{min}

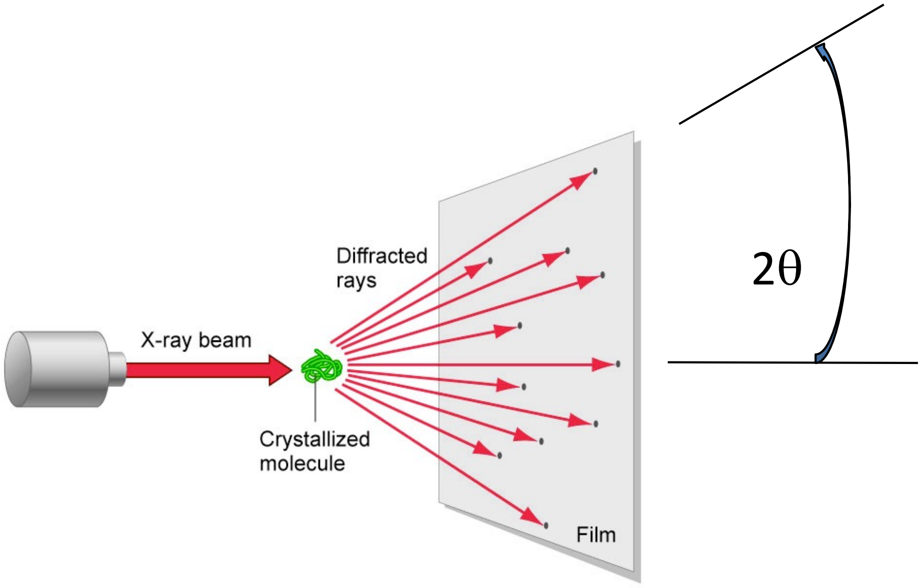
ie. quality is the same!

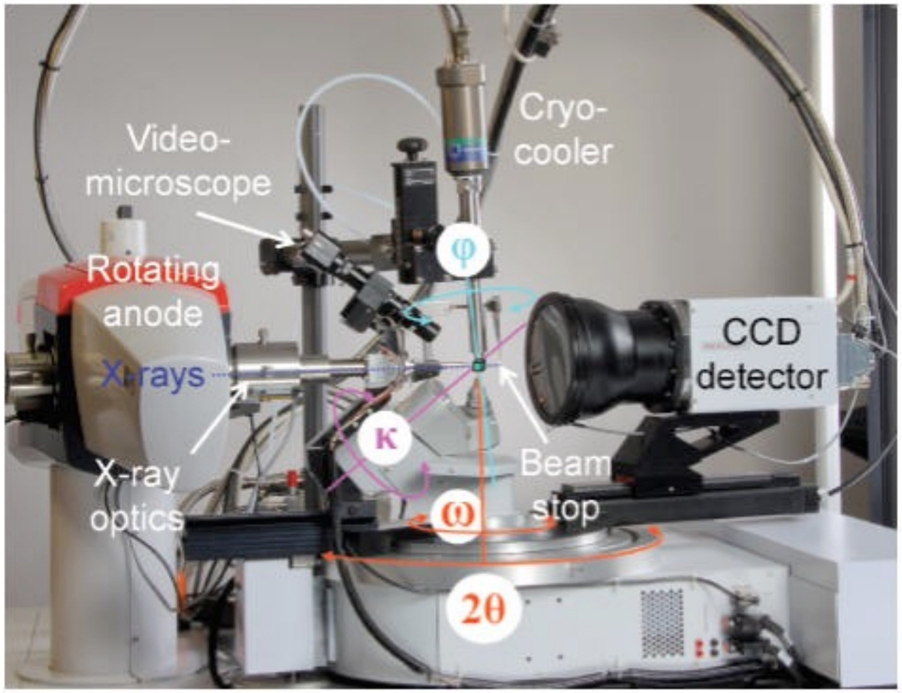


X-Ray Structure Determination

- X-ray crystallography = optics λ ~ 1.5Å (no lenses)
- Bond lengths ~1.4Å
- Electrons scatter X-rays; X-rays ‘see electrons’
- Resolution –Best is $\lambda/2$ Typical is 1 to 3 Å
- Accuracy of atom center positions ±1/10 Resolution

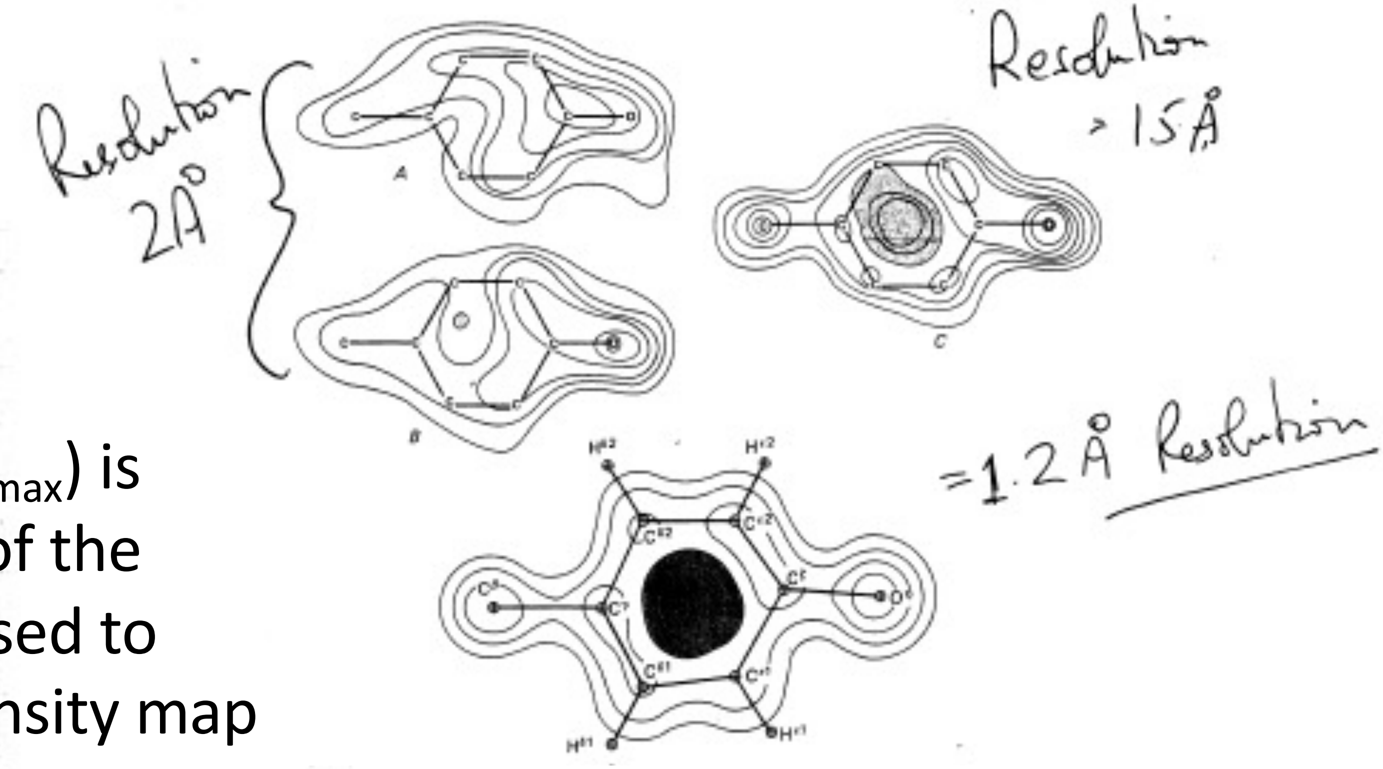
$$d_{\min} = \frac{\lambda}{2 \sin \theta_{\max}}.$$





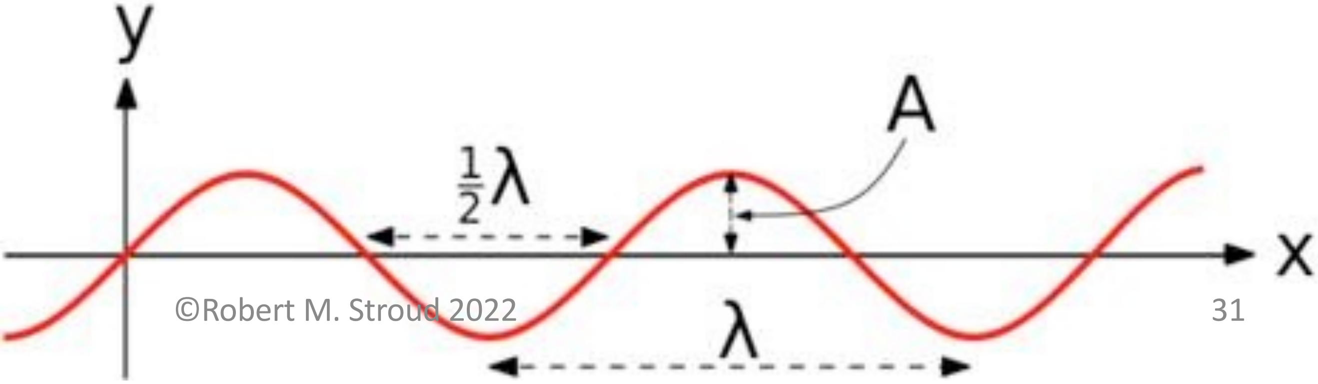
Resolution $d_{min} = \lambda / 2 \sin(\theta_{max})$

differs from Rayleigh criterion



$d_{min} = \lambda / 2 \sin(\theta_{max})$ is the wavelength of the shortest wave used to construct the density map

FIGURE (



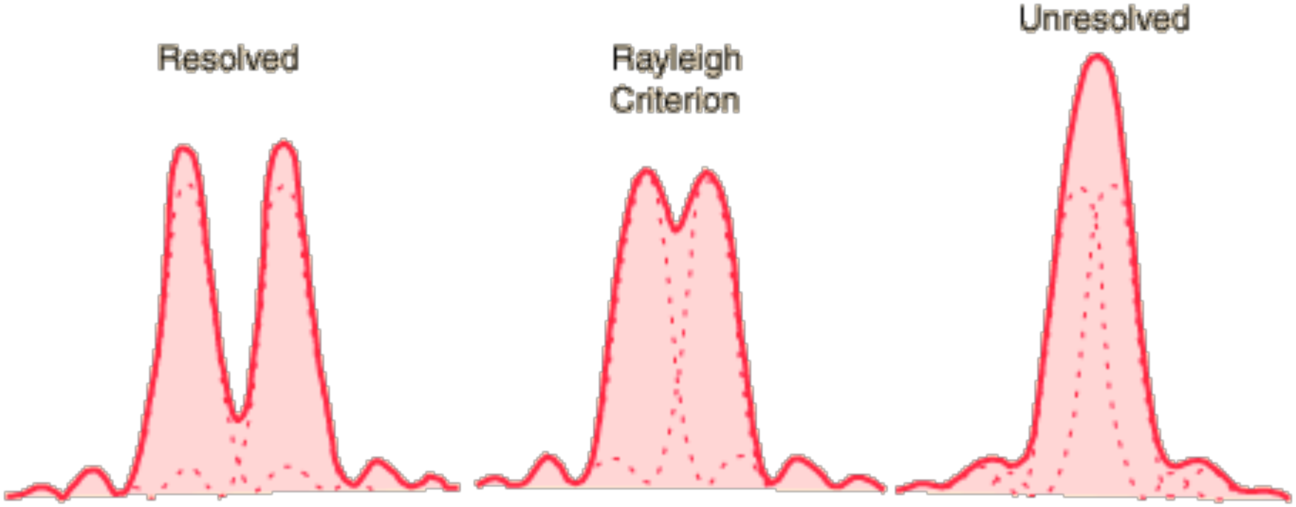
The Rayleigh Criterion

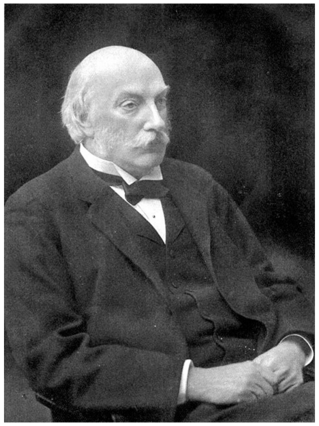
- The Rayleigh criterion is the generally accepted criterion for the minimum resolvable detail - the imaging process is said to be diffraction-limited when the first diffraction minimum of the image of one source point coincides with the maximum of another.

Single slit perpendicular to beam
$d_{min} = \lambda /\sin (2\theta_{max})$

Circular hole
$d_{min} = 1.22\, \lambda /\sin (2\theta_{max})$

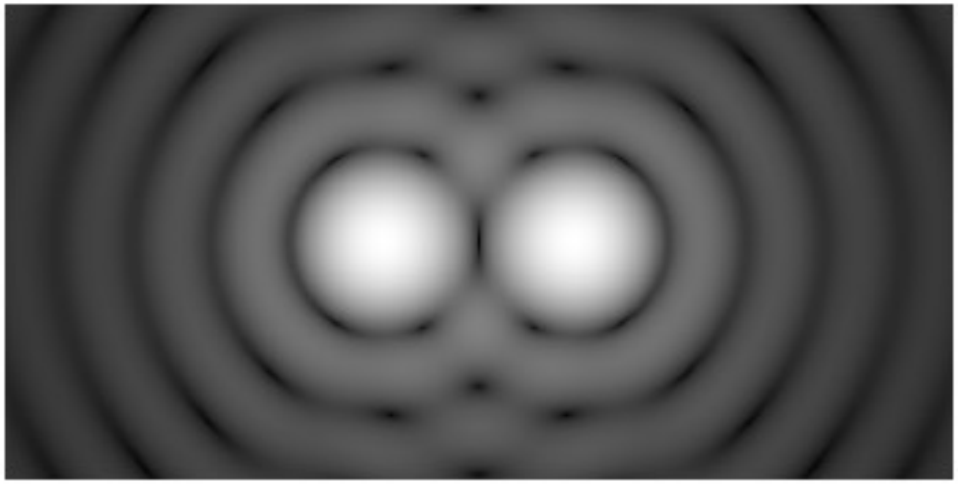
X-ray/EM/Neutrons
In 3 Dimensions:
$d_{min} = \lambda /2 \sin (\theta_{max})$



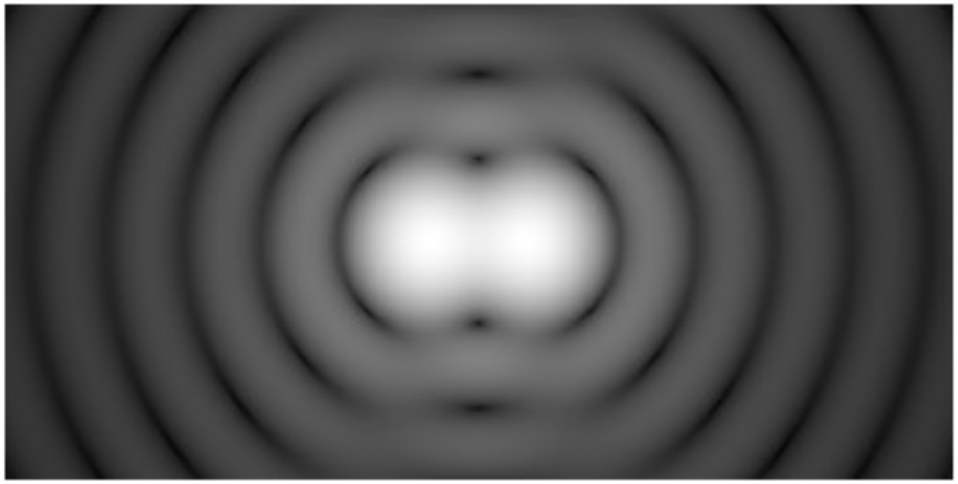


Lord Rayleigh
U.Cambridge
Nobel 1904

Crystal

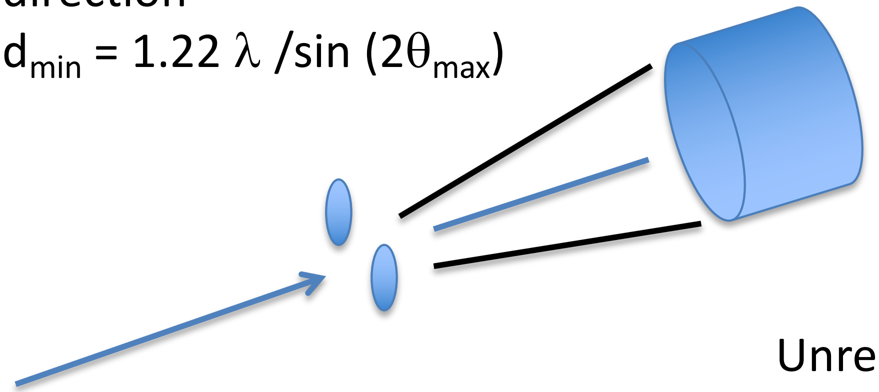


Rayleigh

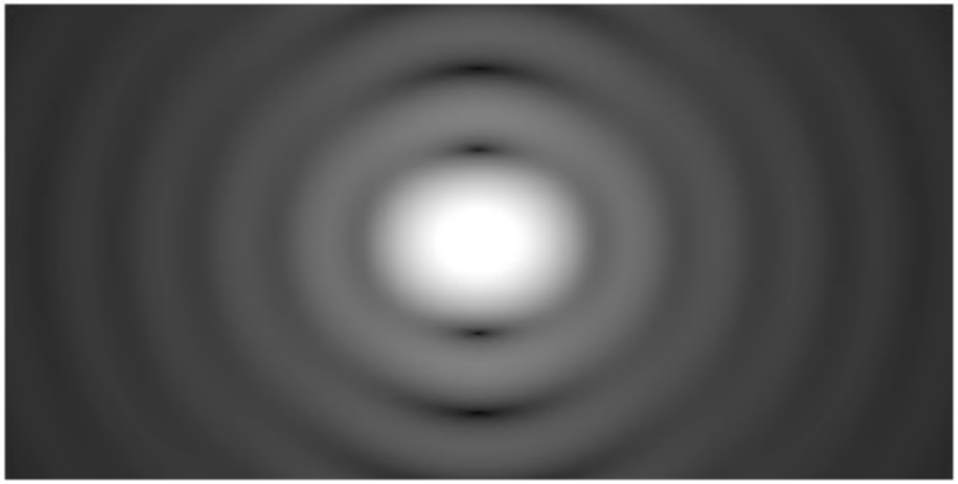


Diffraction from Circular hole, forward direction

$d_{min} = 1.22\ \lambda\ /\sin(2\theta_{max})$



Unresolved



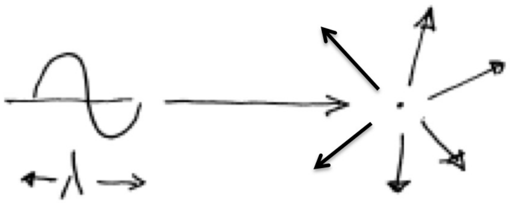
Stroud 2022

#2   Dot Product,

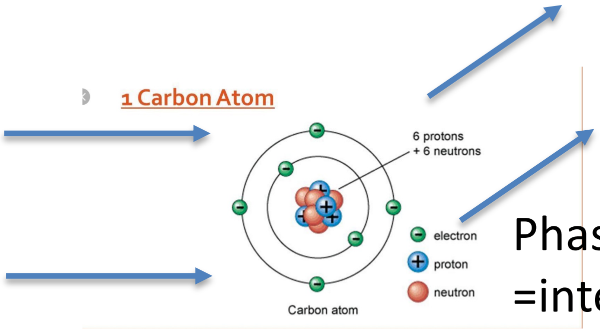
Whole Protein Scattering

Lattice of Proteins

Form factor: finite size means fall off of scattering $f(\mathbf{s})$

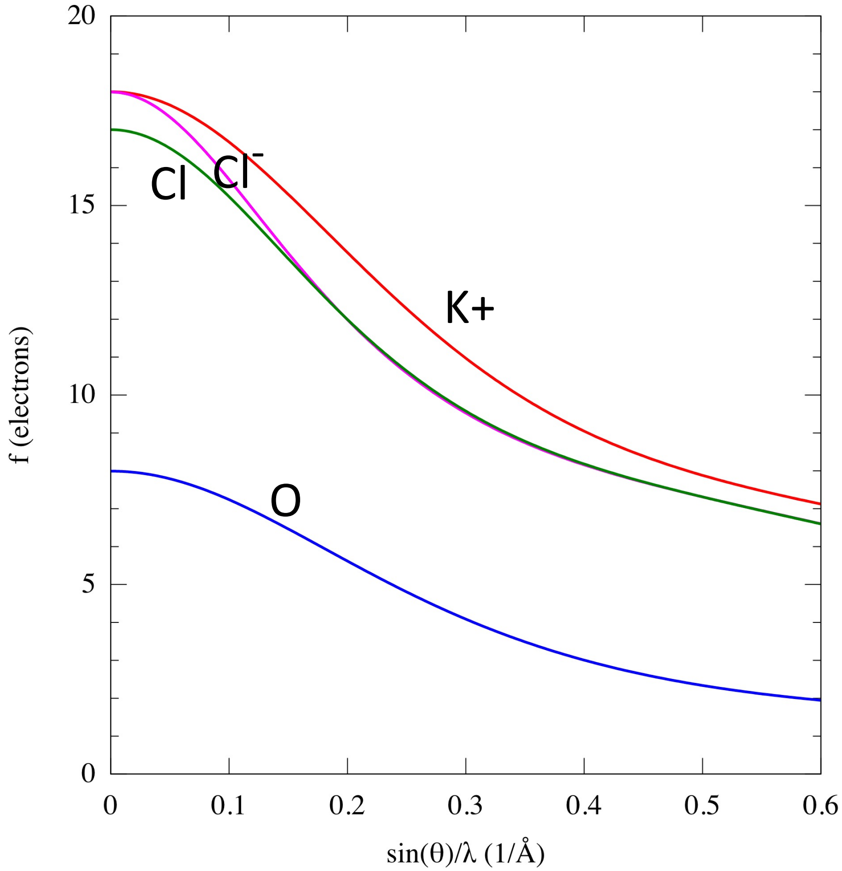


Scattering from a point is equal in all directions.



Phase lag
=interference

Atoms finite size of
Electron density,
Scatters waves out of phase with
Density on the other side. These
Cancel each other out at higher
Angles (or $\sin(\theta)/\lambda$)



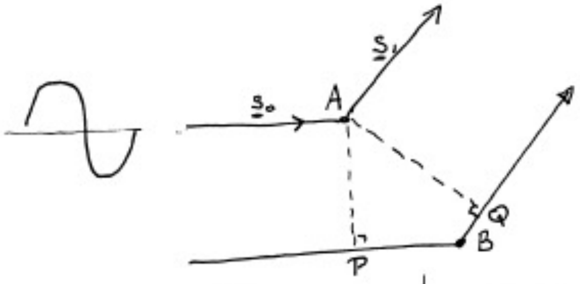
X-ray atomic ‘form factors’ f of oxygen (blue), chlorine (green), Cl^- (magenta), and K^+ (red); smaller charge distributions have a wider form factor.

Scattering from multiple points? Add wave amplitudes with phase change

Scattering by matter -(interference) of a single wavelength X-ray

Scattering from a point is equal in all directions.

add a second point, scattering in some direction $\underline{s}_1$



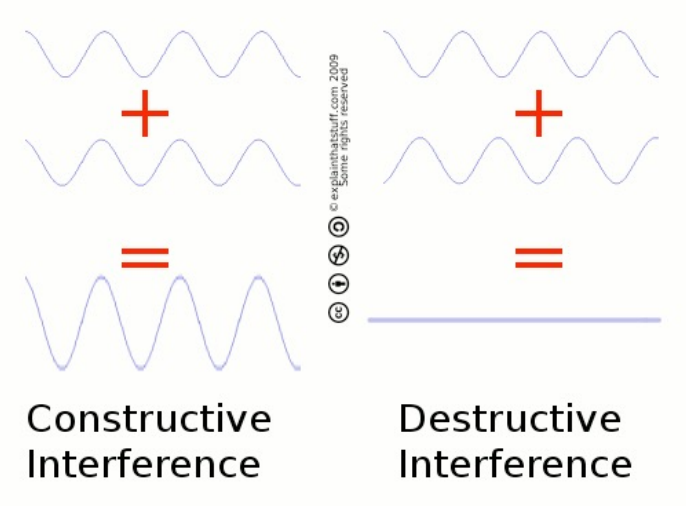
The second wave, scattered by B travels further by the distance PB + BQ.
The scattered wave lags in phase by

$$= \frac{2\pi}{\lambda}\left([P-B] + [B-Q]\right)$$

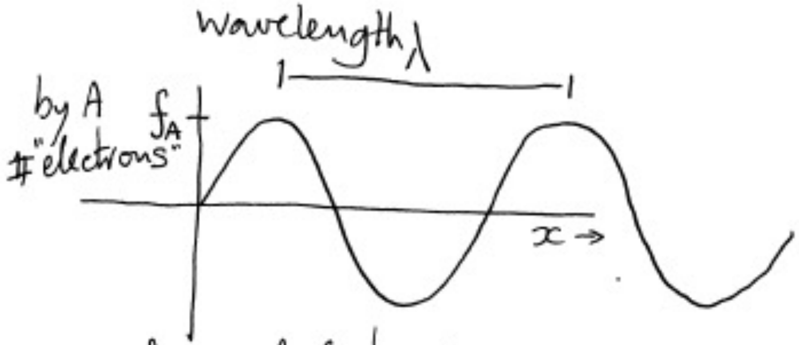
$$= \frac{2\pi}{\lambda}\left(\Phi \right)$$

where Φ = path length extra for B versus the reference A.

Adding up the scattering of Atoms:
Amplitudes, ‘interference’ of waves

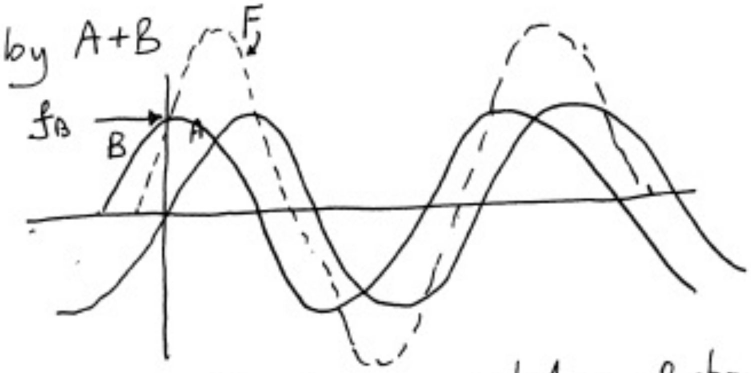


Waves add out of phase by 2π[extra path/λ]



f_A = # of electrons

$$F = A \sin 2\pi\left(\frac{x}{\lambda} - \omega t\right)$$



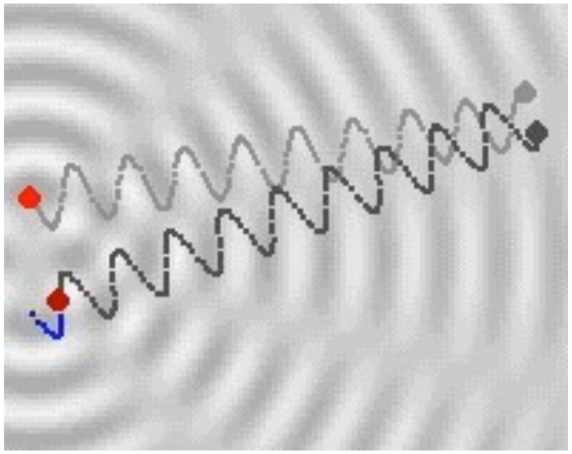
For $f_A = f_B = 1$ say, total $\underline{2}$ electrons scatter

$$F = A \sin 2\pi\left(\frac{x}{\lambda} - \omega t\right) + A \sin 2\pi\left(\frac{x+\Phi}{\lambda} - \omega t\right)$$

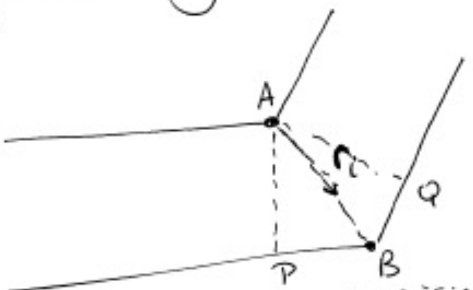
$$F = \underbrace{2A \cos 2\pi\left(\frac{\Phi}{2\lambda}\right)}_{\substack{\text{Amplitude of F,} \\ \text{Always less than 2} \\ \text{(less than the sum of all} \\ \text{electrons in object)}}} \underbrace{\sin 2\pi\left(\frac{x+\Phi/2}{\lambda} - \omega t\right)}_{\substack{\text{the wave is} \\ \text{'phase shifted} \\ \text{by } \left(\frac{2\pi\Phi}{2\lambda}\right)}}$$

In general they add up to something amplitude In between -2f and +2f.
For n atoms

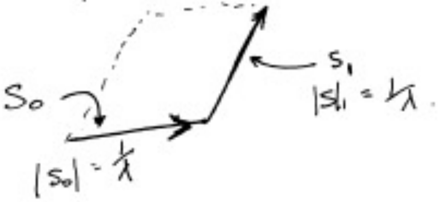
$F^2 = f \times \sqrt{n}$



The phase lag can also be simplified



We define s_0

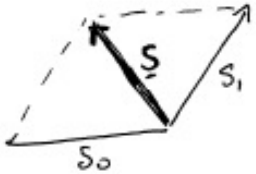


the PB = projection of r_i on s_0
BQ = projection of r_i on s_1

$$\text{so } \frac{\Phi}{\lambda} = \frac{PB + BQ}{\lambda} = -\underline{r_i} \cdot \underline{s_0} + \underline{r_i} \cdot \underline{s_1}$$

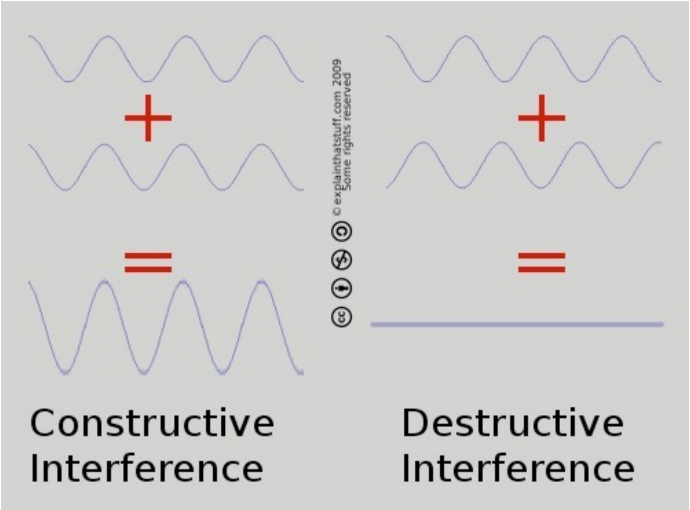
$$= \underline{r_i} \cdot (\underline{s_1} - \underline{s_0})$$

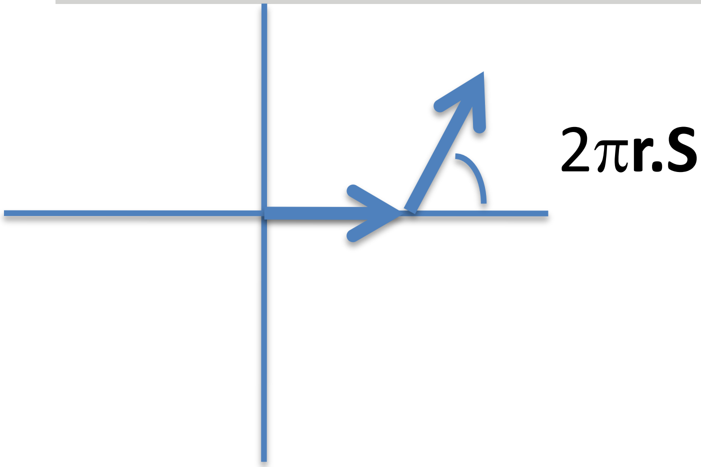
So we define $\underline{S} = s_1 - s_0$



$$\text{Then } \frac{\Phi}{\lambda} = \underline{r} \cdot \underline{S}$$

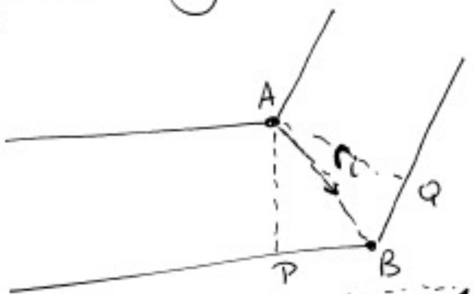
Adding up the scattering of Atoms: 'interference' of waves



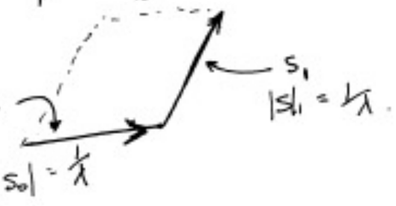


Adding up the scattering of Atoms: ‘interference’ of waves

The phase lag can also be simplified



We define S_0

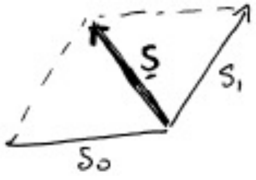


the PB = projection of r_i on s_0
BQ = projection of r_i on s_1

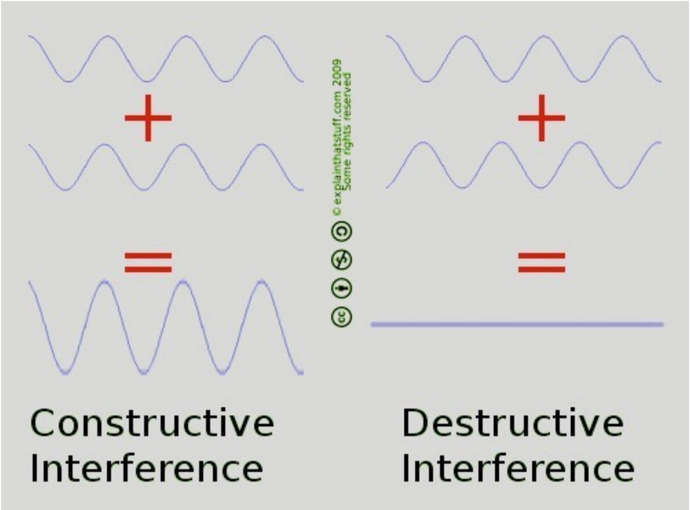
$$\text{so } \frac{\Phi}{\lambda} = \frac{PB + BQ}{\lambda} = -\underline{r}_i \cdot \underline{s}_0 + \underline{r}_i \cdot \underline{s}_1$$

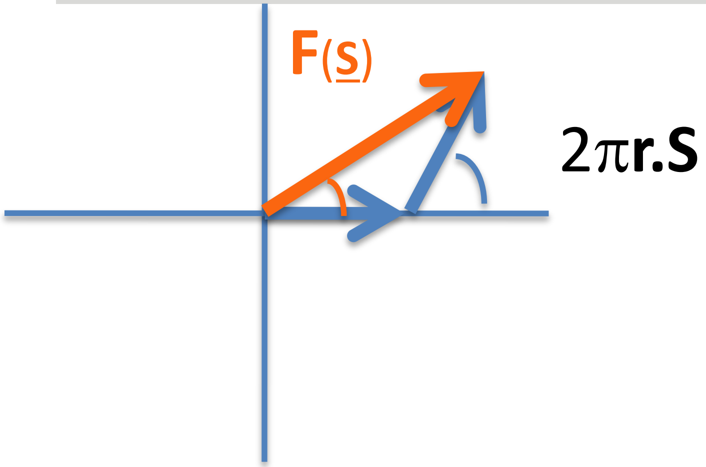
$$= \underline{r}_i \cdot (\underline{s}_1 - \underline{s}_0)$$

So we define $\underline{S} = \underline{s}_1 - \underline{s}_0$



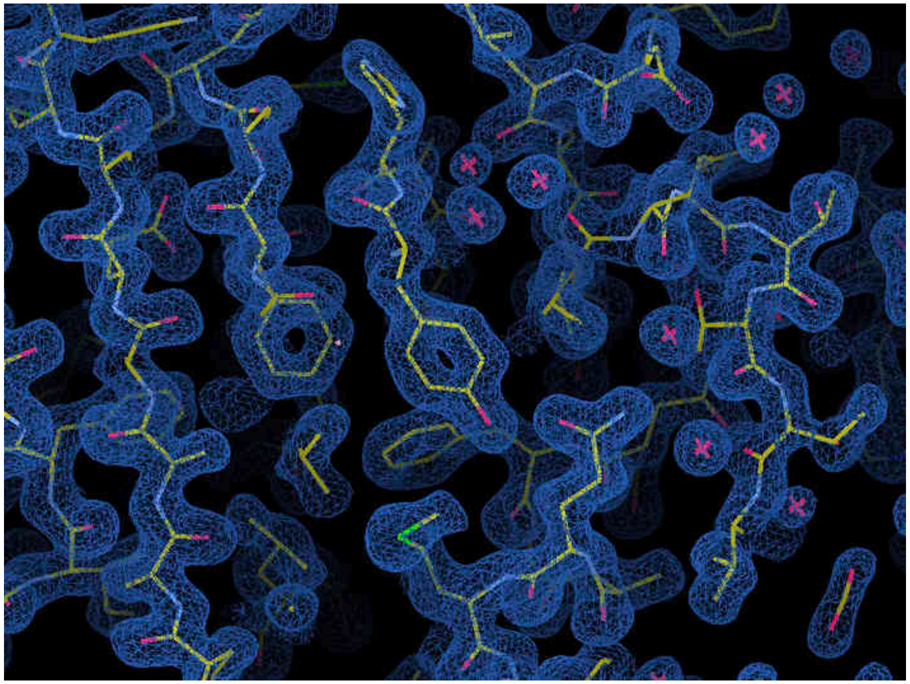
$$\text{Then } \frac{\Phi}{\lambda} = \underline{r} \cdot \underline{S}$$



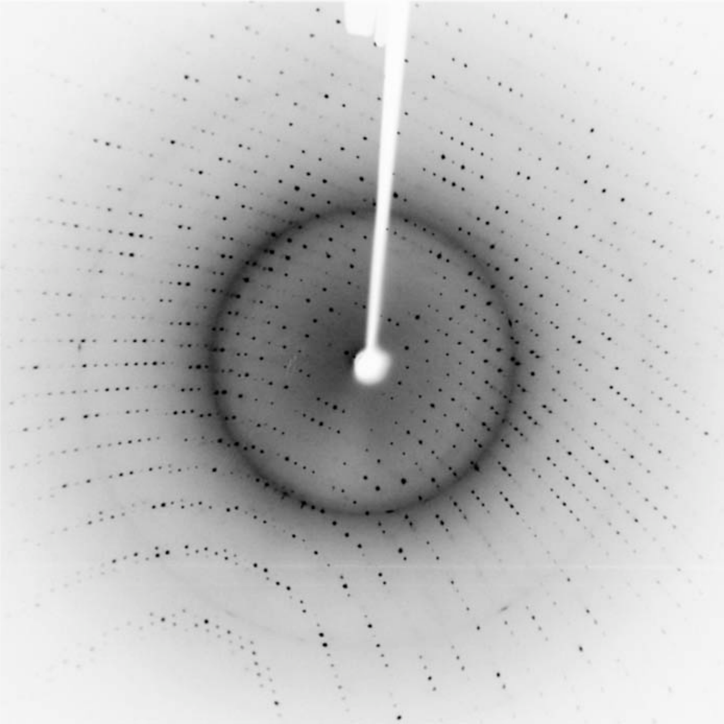


What do we want?

“Real” space (x,y,z)

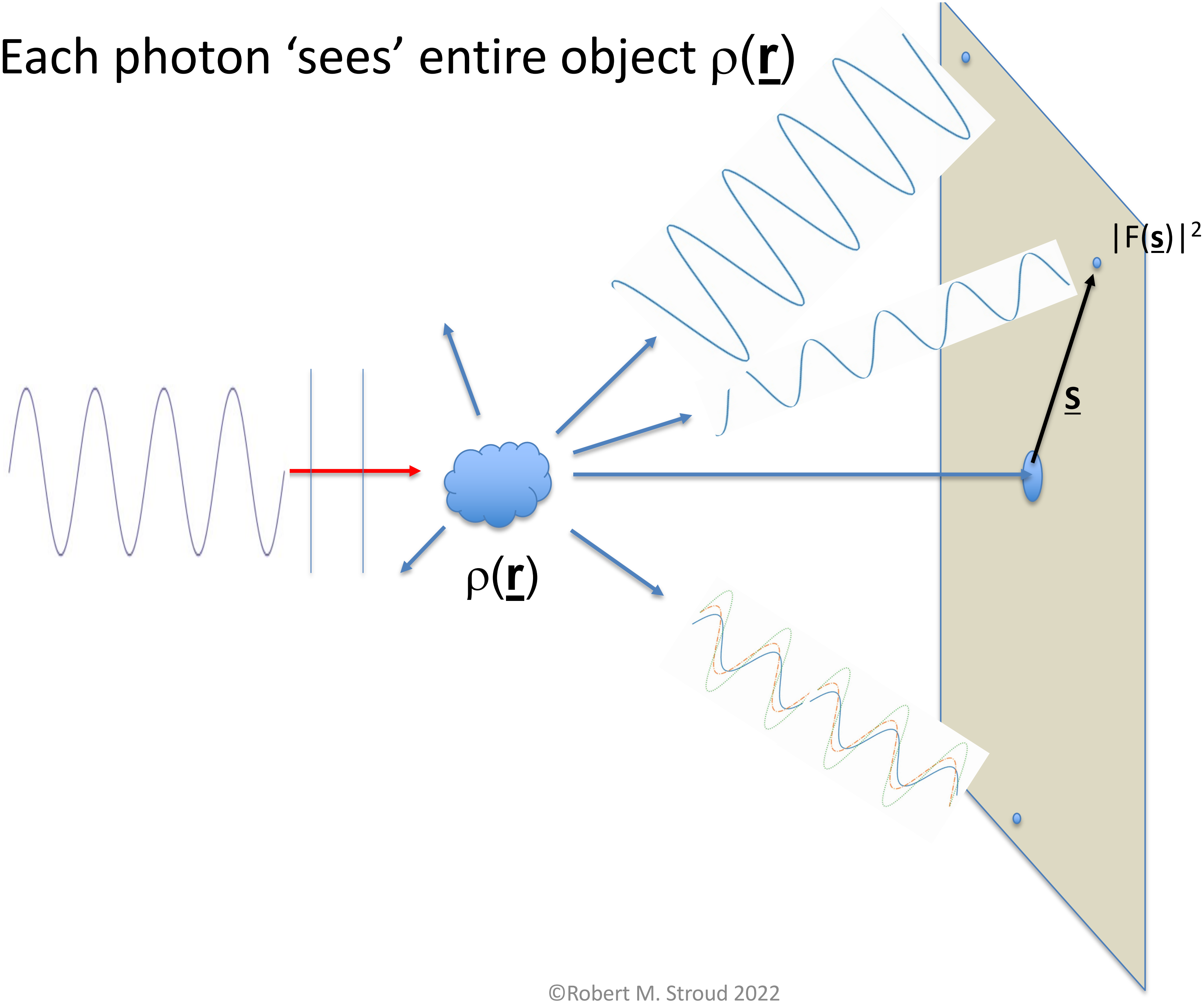


Scattering / Diffraction $F^2_{(h,k,l)}$



$$\rho(x,y,z) = const \cdot \int_{hkl} F(h,k,l)e^{-2\pi i(hx+ky+lz)+i\varphi(h,k,l)}dhdkdl$$

Each photon ‘sees’ entire object ρ(r̲)



A proof: of Euler's relation (De Moivre's theorem).

Using McClaurin's theorem:

that $f(x) = f(0) + x f'(x) + \frac{x^2}{2!} f''(0) \dots$

So $e^{x} = 1 + x + \frac{x^2}{2!} \dots$

$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!}$

$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!}$

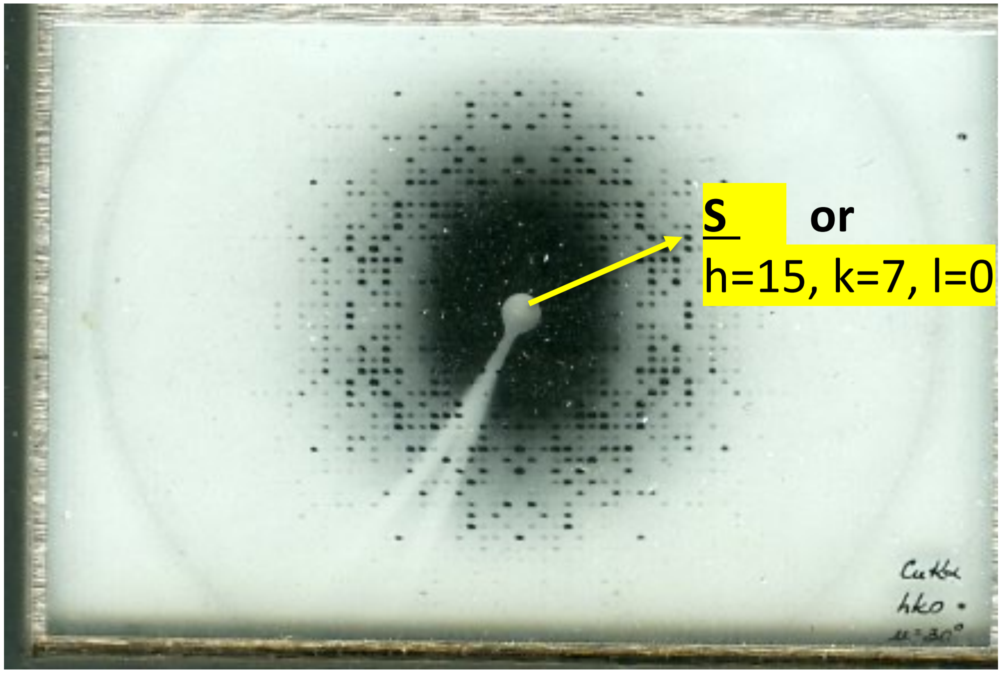
$\therefore e^{ix} = 1 + ix - \frac{x^2}{2!} - \frac{ix^3}{3!} + \frac{x^4}{4!} + \frac{ix^5}{5!} - \frac{x^6}{6!}$

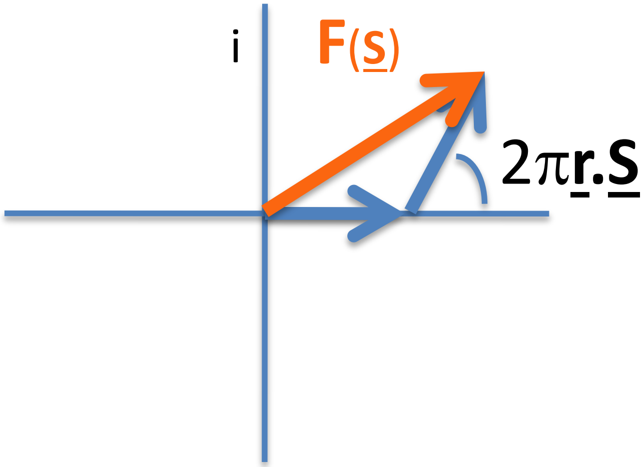
$= \cos x + i \sin x.$

Scattering from a molecule is described by

$$\underline{\mathbf{F}}(\underline{\mathbf{s}}) = \Sigma_j \, f_j \, e^{(2\pi i \mathbf{r}_j . \mathbf{s})}$$

The Scattering from one molecule of j=1....N atoms is sampled at the diffraction positions. How so??



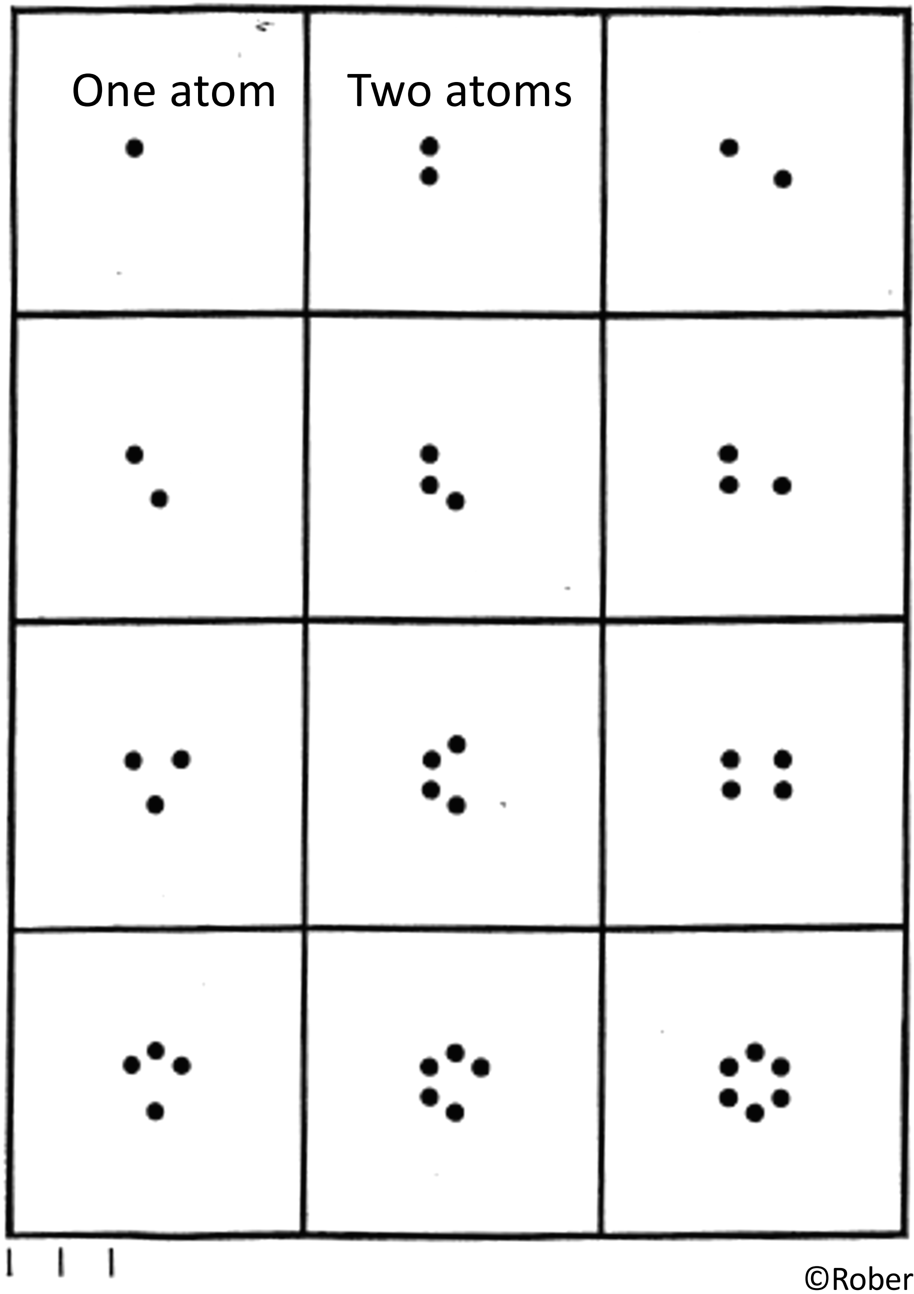


Repetition in crystal==sampling in diffraction

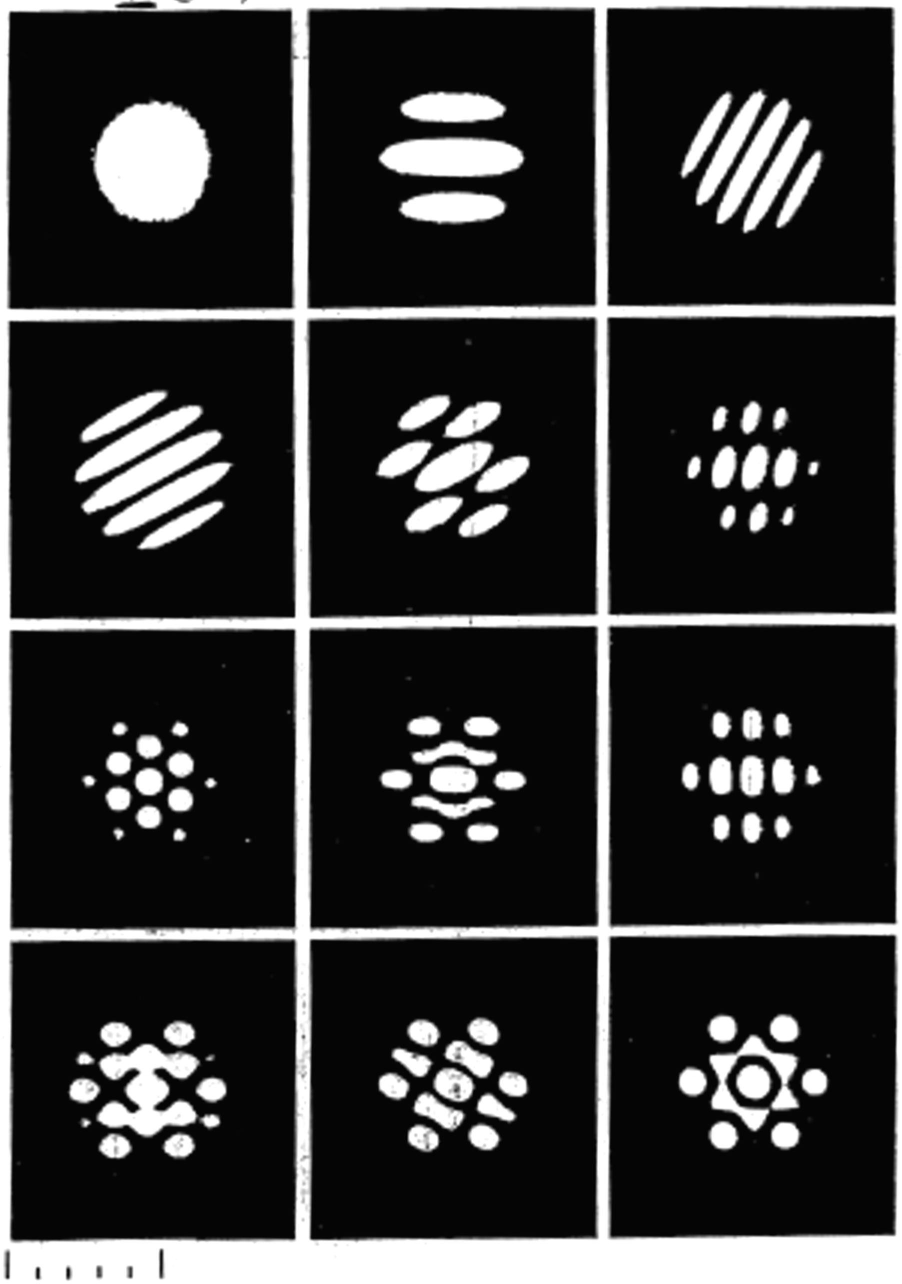
Plate 2 Superimposed fringes

The sequence shows how the diffraction pattern of a simple object is built up by superposition of sets of fringes.

Plate 2 $\rho(\underline{r})$



$\underline{F}(\underline{s})$ Plate 2



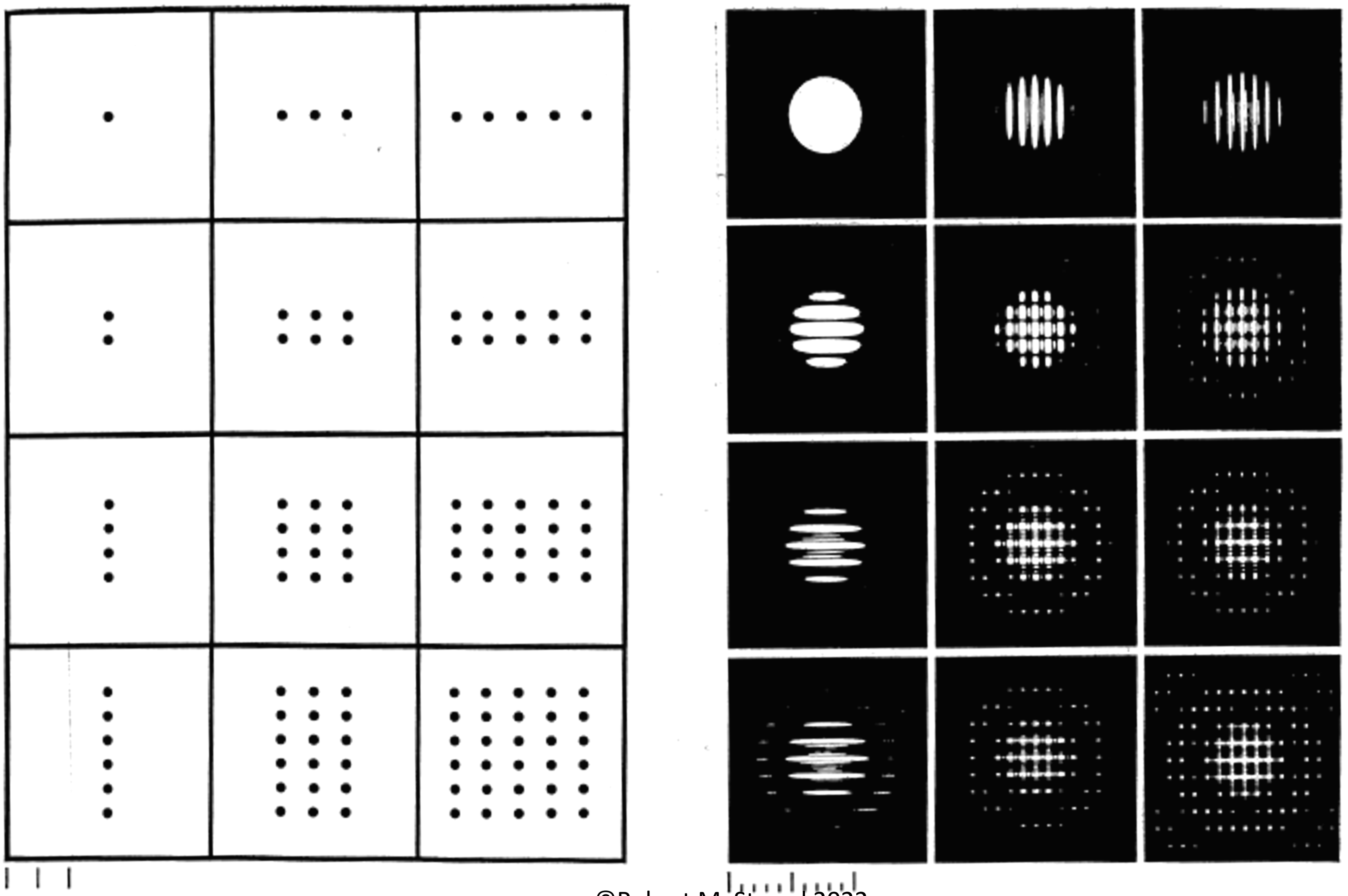
Diffraction due to repeats, is the same as the object, sampled by (1/repeat)

Plate 10 Development of a lattice

A single aperture is repeated to build up rows and columns which are combined to produce the lattice of 12. Subsidiary diffraction maxima due to the small number of apertures can be seen clearly.

Plate 10

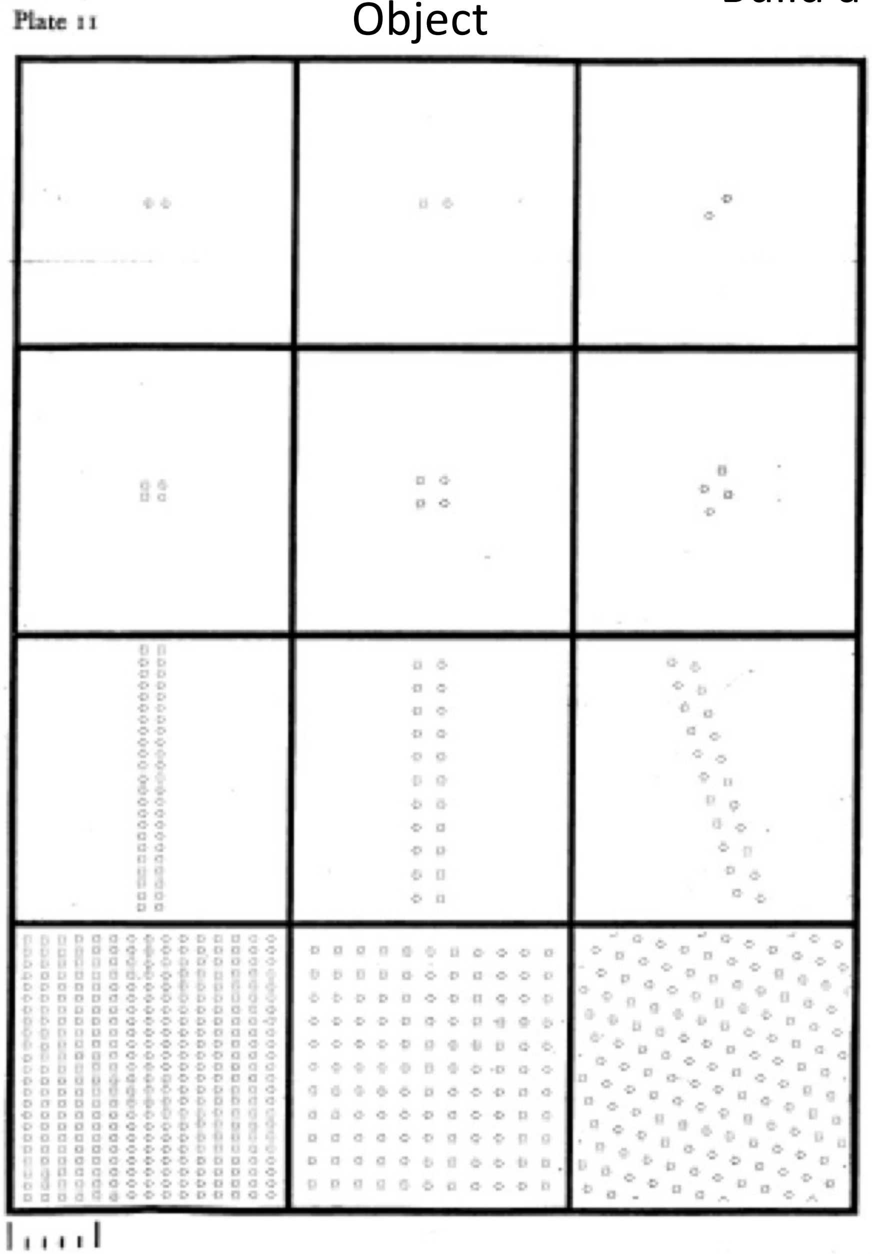
Plate 10



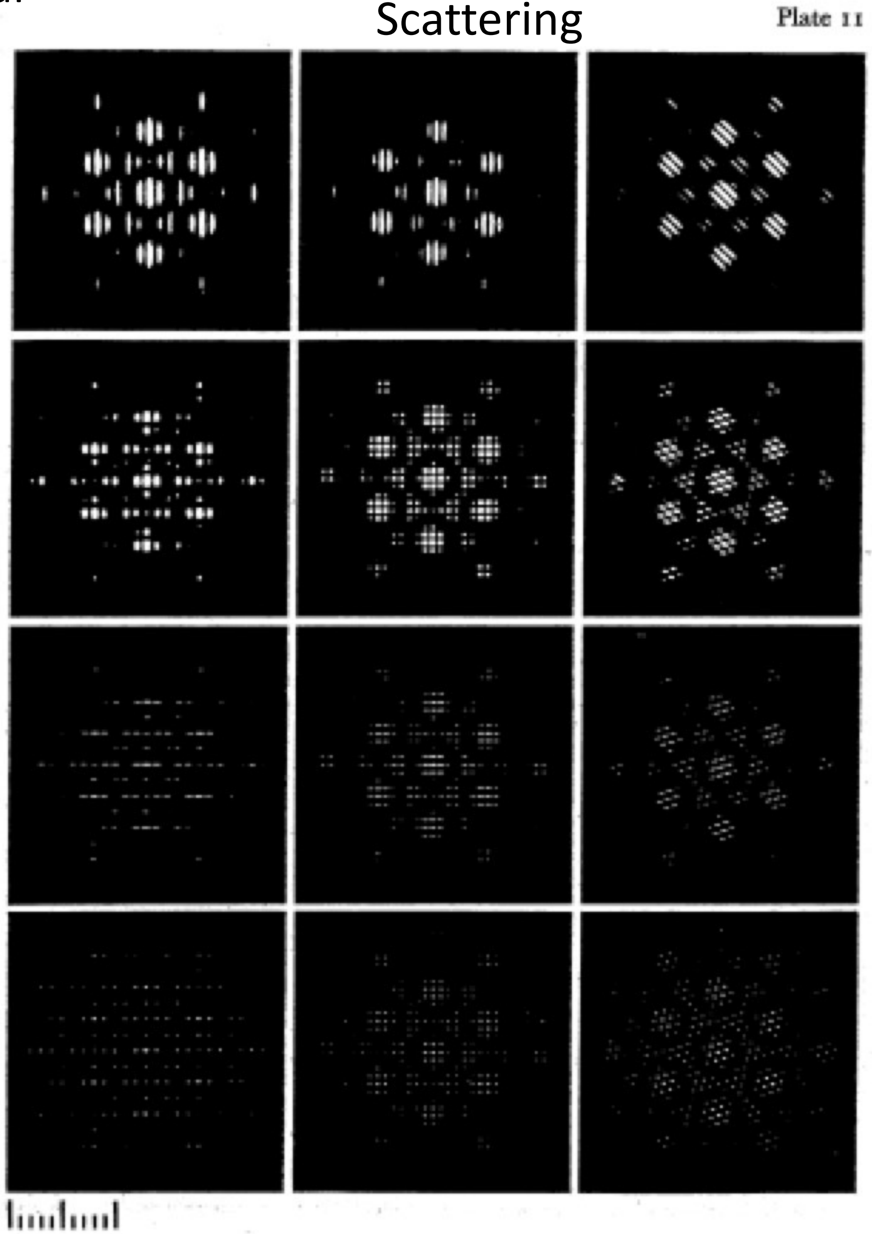
©Robert M. Stroud 2022

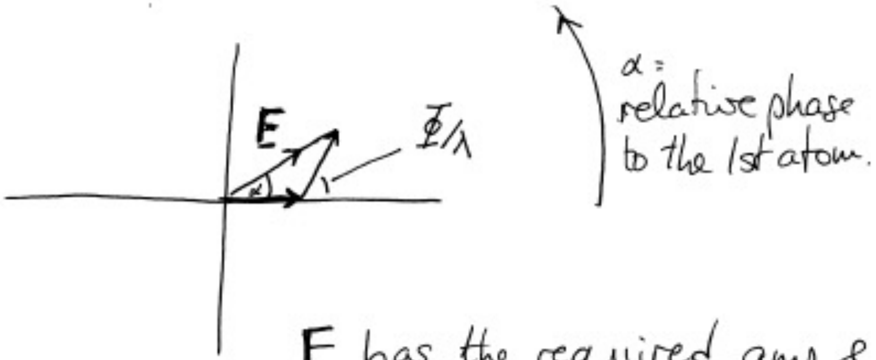
Build a crystal

Object



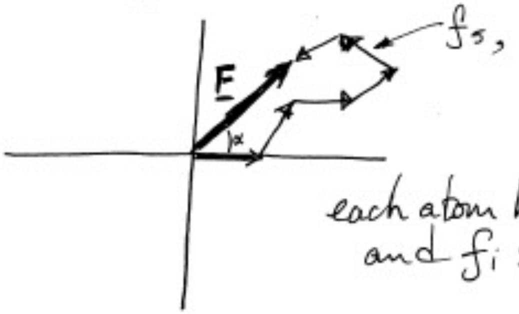
Scattering





F has the required amp. & phase.

If we use this method we can add $i = 1$ to n different atoms; each amplitude f_i



each atom has $f_i \cos\alpha_i$ along x and $f_i \sin\alpha_i$ along y

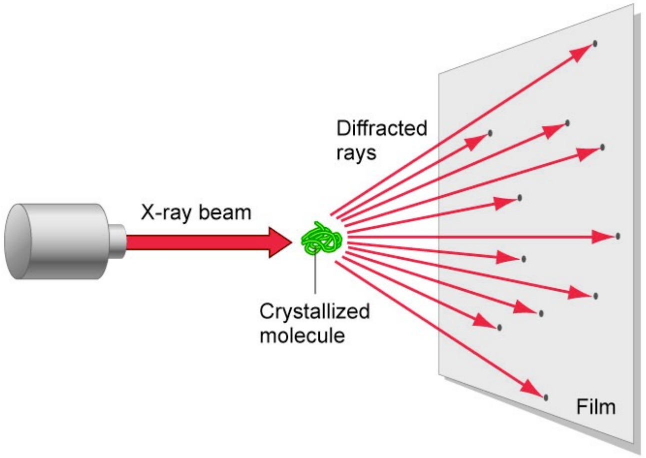
If we put 'units' on the axes, we can add up the 'x' and 'y' components to write the sum over "x"; the sum over "y", — hence calculate F as a wave of amplitude

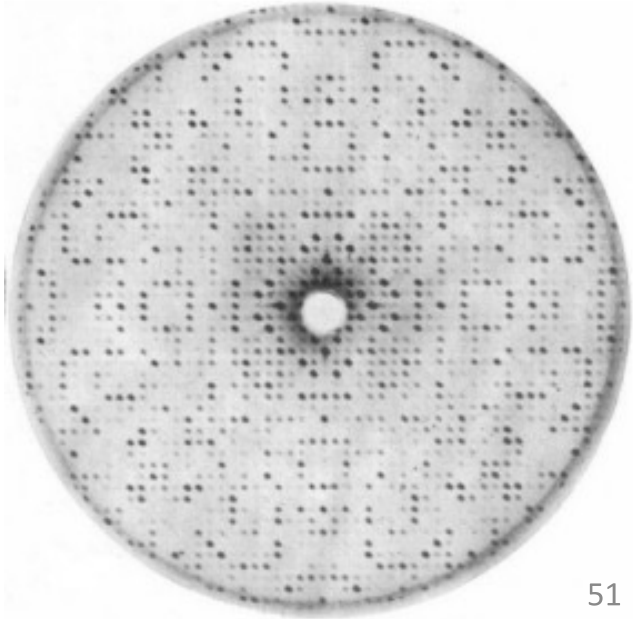
$$|F| = \sqrt{\left(\sum_x f_i \cos\alpha_i\right)^2 + \left(\sum_y f_i \sin\alpha_i\right)^2}$$

and $$\alpha = \tan^{-1}\left(\frac{\sum_y f_i \sin\alpha_i}{\sum_x f_i \cos\alpha_i}\right)$$

Many atoms add by the same rules.

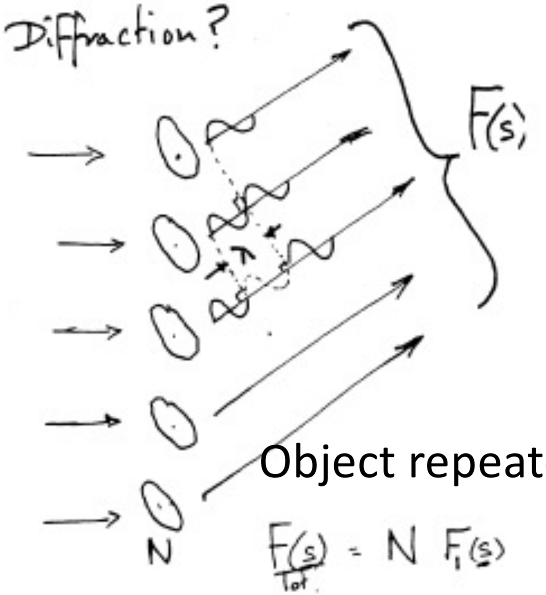
Different in every direction.





⑧

Why Diffraction?

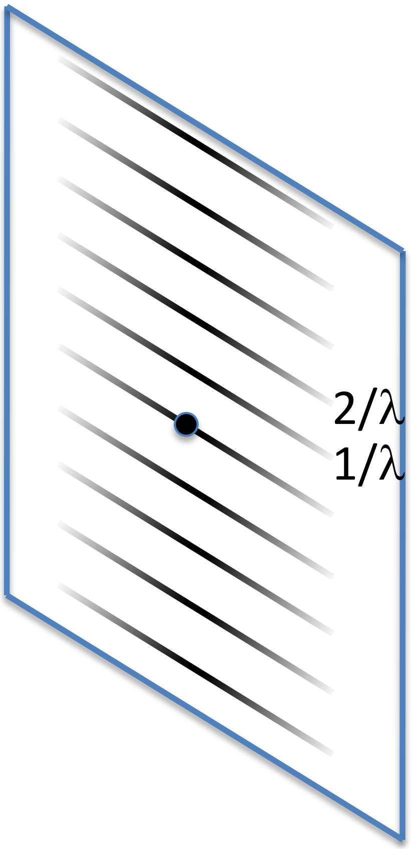


Object repeated

$F_{Tot}(\underline{s}) = N\ F_1(\underline{s})$

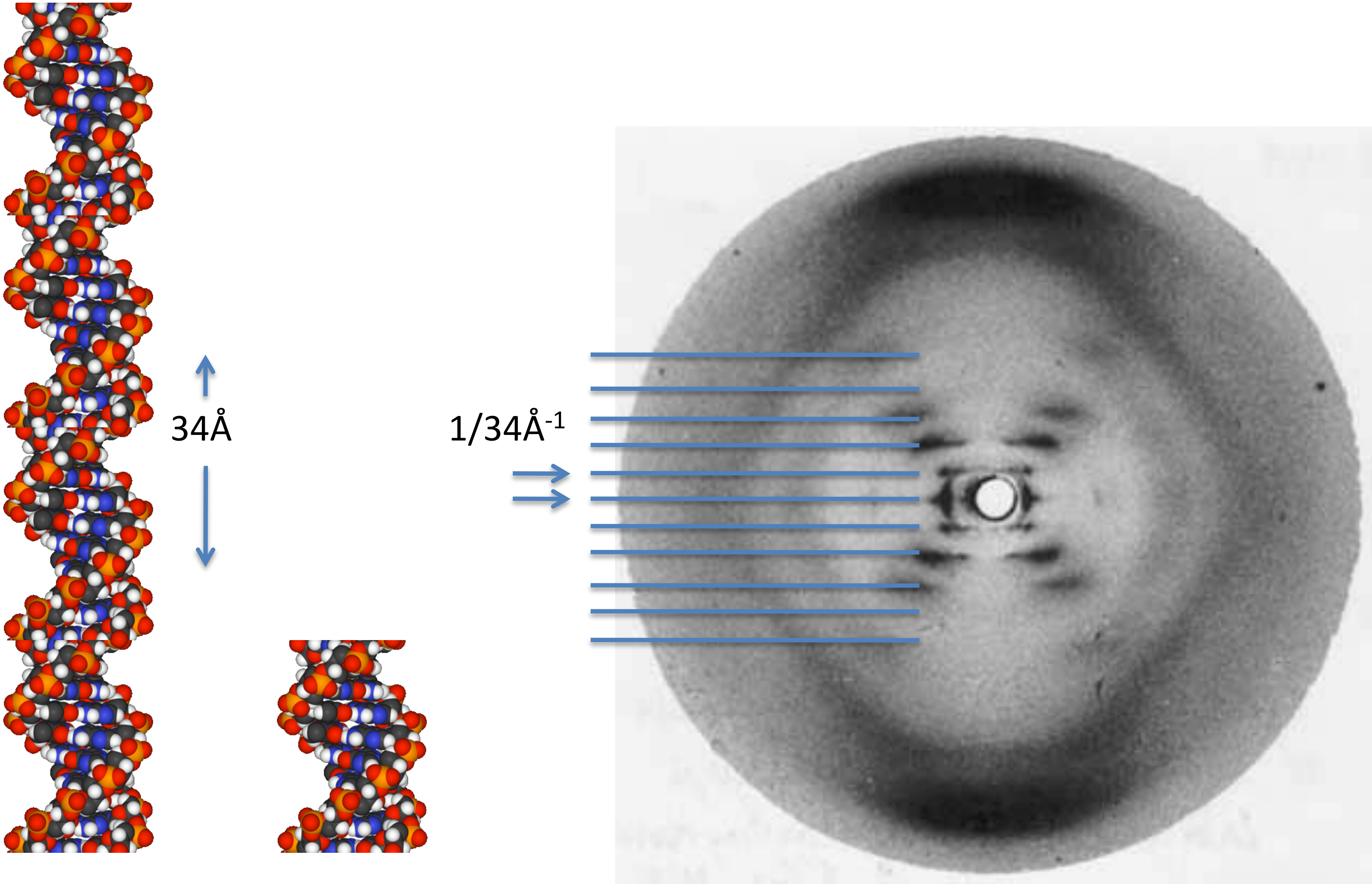
$\underline{s}$ refers to a scattering direction

so $|\underline{s}| = \frac{2\sin\theta}{\lambda}$



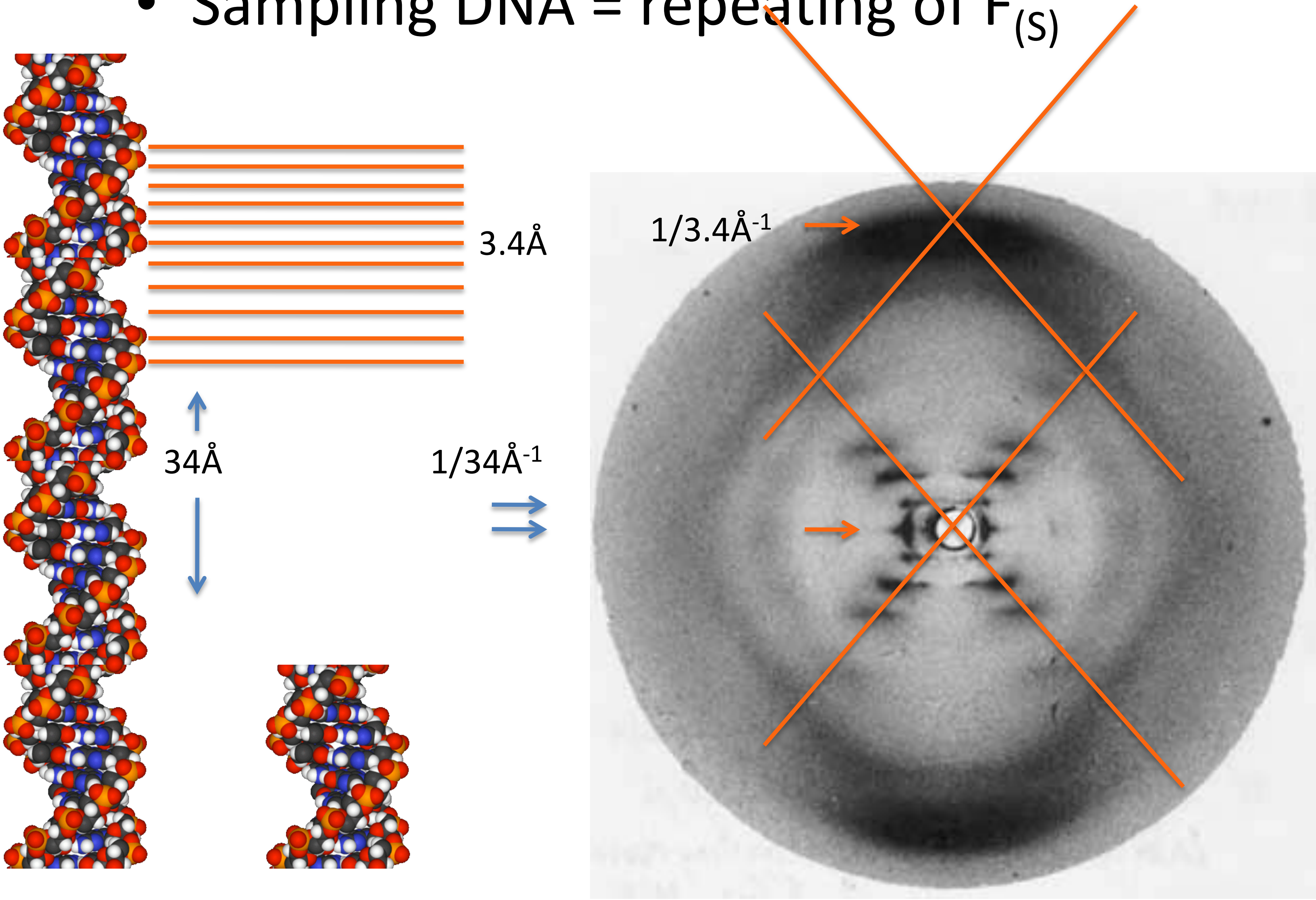
Consequences of being a crystal?

- Repetition = sampling of $F_{(S)}$



Consequences of being a crystal?

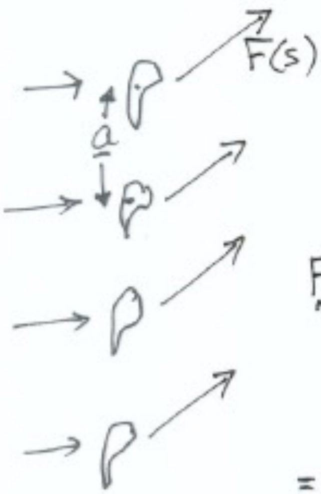
- Sampling DNA = repeating of $F_{(S)}$



Build up a crystal from Molecules…

First 1 dimension, **a** direction

When Phase shift is 2π they will add Amplitude F(**s**)

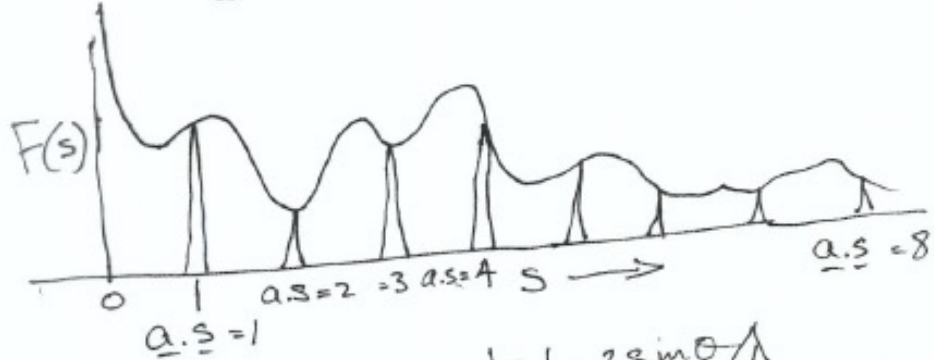


$$F_m(s)_{total} = F(s)\left[1 + e^{2\pi i \underline{a}.\underline{s}} + e^{2\pi i\, 2a.s} + e^{2\pi i\, 3\underline{a}.\underline{s}} \dots \dots e^{2\pi i (M-1)a.s}\right]$$

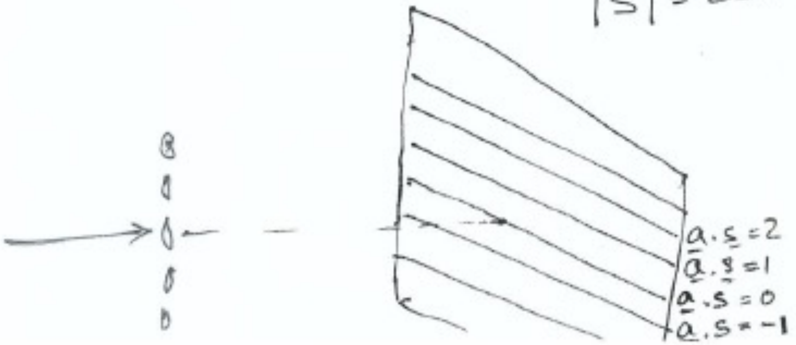
= A Geometric series

$$= F(s)\left(\frac{1-e^{2\pi i M\underline{a}.\underline{s}}}{1-e^{2\pi i \underline{a}.\underline{s}}}\right) = F(s)\frac{e^{\pi i M a.s}}{e^{\pi a.s}}\left[\frac{e^{-\pi i M\underline{a}.\underline{s}} - e^{\pi i M\underline{a}.\underline{s}}}{e^{-\pi i \underline{a}.\underline{s}} - e^{\pi i \underline{a}.\underline{s}}}\right]$$

$$= F(s)\frac{e^{\pi i M\underline{a}.\underline{s}}}{e^{\pi \underline{a}.\underline{s}}}\left[\frac{\sin\pi M\underline{a}.\underline{s}}{\sin\pi\,\underline{a}.\underline{s}}\right] = e^{\pi i (M-1)\underline{a}.\underline{s}}\left[\frac{\sin\pi M a.s}{\sin\pi a.s}\right]$$

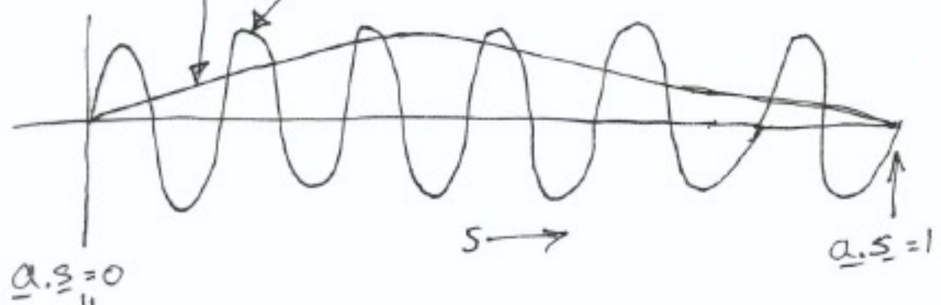


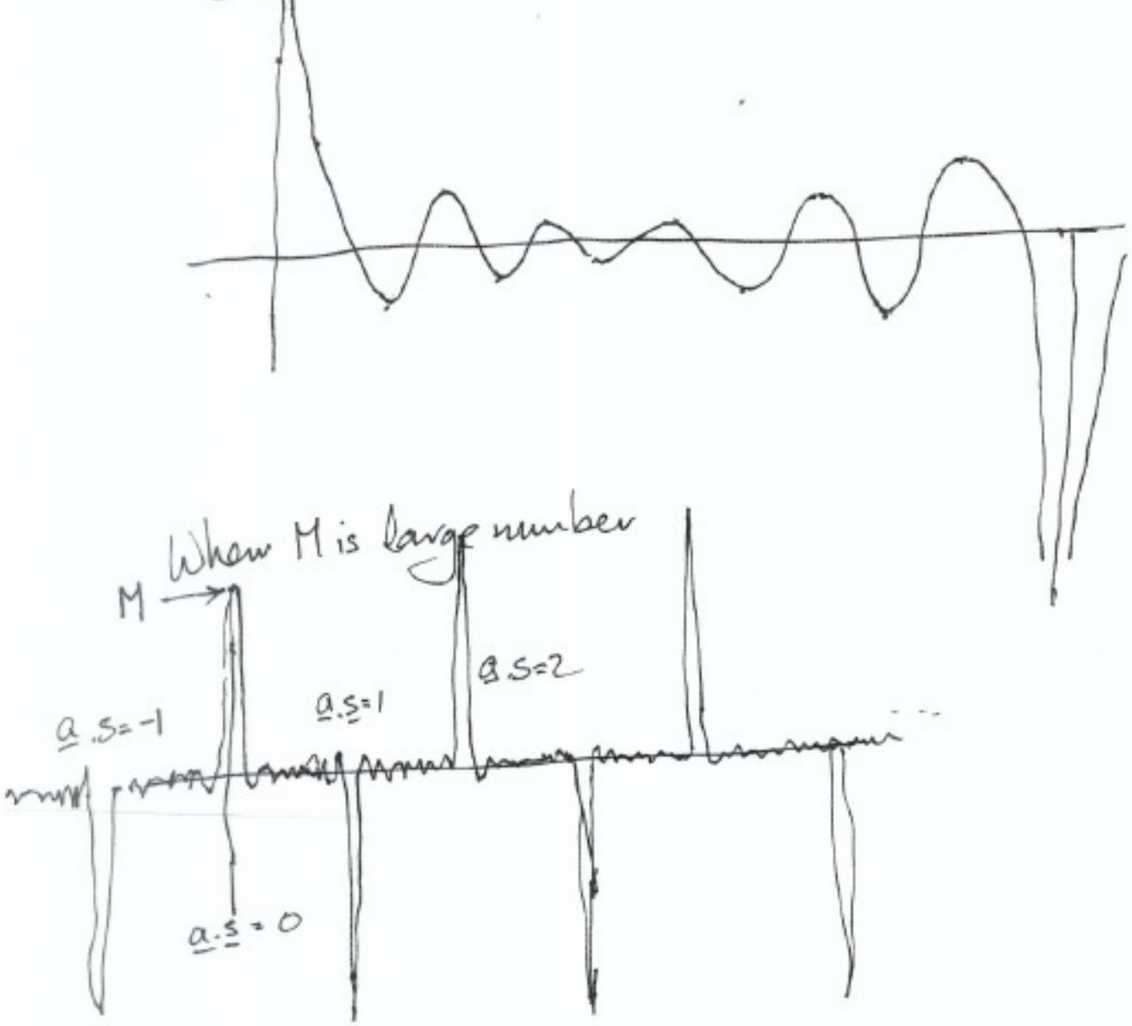
$$|s| = 2\sin\theta/\lambda$$



$$\frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}}$$?

eg. for M=6





When M is large number

6/ The reciprocal lattice

For a crystal, the function $G(\underline{s})$ can only be observed at all, ie for any rotation of the scattering object; - $G(s)$ only exists for

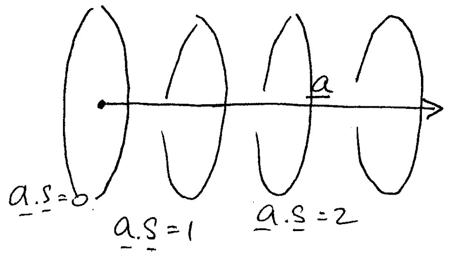
$$\underline{a}\cdot\underline{s} = h$$
$$\underline{b}\cdot\underline{s} = k$$
$$\underline{c}\cdot\underline{s} = l$$

as a consequence of summing all unit cells. $\left(\frac{\sin \pi M a.s}{\sin \pi a.s}\right)$ etc

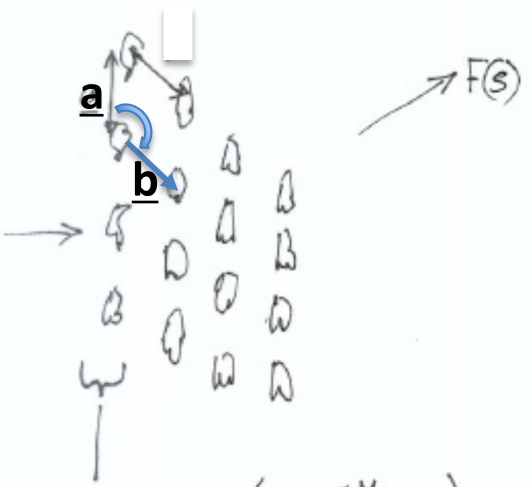
These 3 intersecting sets of planes describe a lattice of points: The first two planesets

$a.s = h$
$b.s = k$ describe a set of lines and the third set of planes cut these lines. at positions where $c.s = l$.

The planes $\underline{a}.\underline{s}$ are perpendicular to $\underline{a}$



add a second repeat axis $\underline{b}$



$$F(s)_{tot} = F(s)\left(\frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}}\right)$$

Next column, is just $\underline{b}$ away

$$= F(s)\left(\frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}}\right) e^{2\pi i \underline{b}.\underline{s}}$$

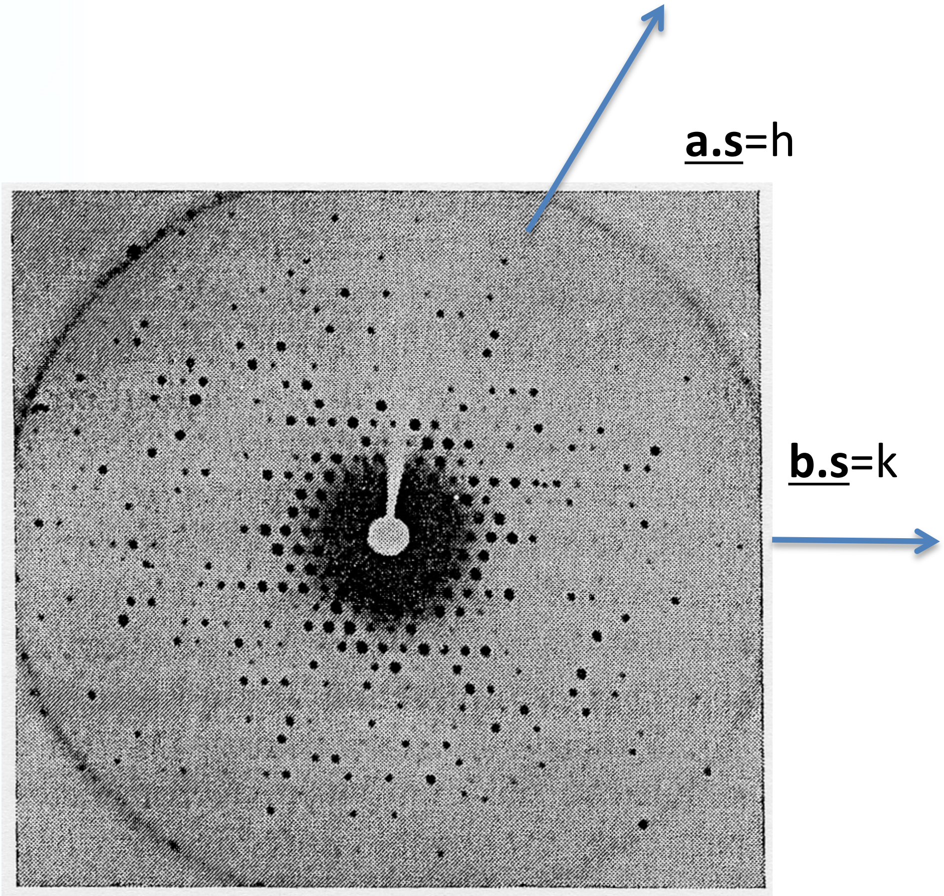
etc.

so for the 2D array:-

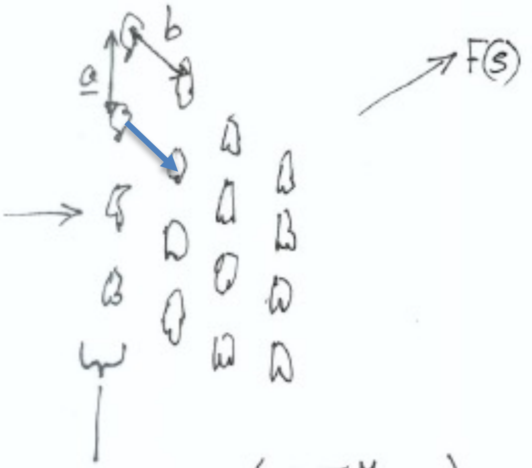
$$= F(s)\left(\frac{\sin \pi M a.s}{\sin \pi a.s}\right)\left(\frac{\sin \pi N \underline{b}.\underline{s}}{\sin \pi \underline{b}.\underline{s}}\right)$$

for 3D array

$$= F(s)\,\frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}}\;\frac{\sin \pi N \underline{b}.\underline{s}}{\sin \pi \underline{b}.\underline{s}}\;\frac{\sin \pi P \underline{c}.\underline{s}}{\sin \pi c.s}$$



add a second repeat axis $\underline{b}$



$$F(s)_{tot} = F(s)\left(\frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}}\right)$$

Next column, is just $\underline{b}$ away

$$= F(s)\left(\frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi \underline{a}.\underline{s}}\right) e^{2\pi i \underline{b}.\underline{s}}$$

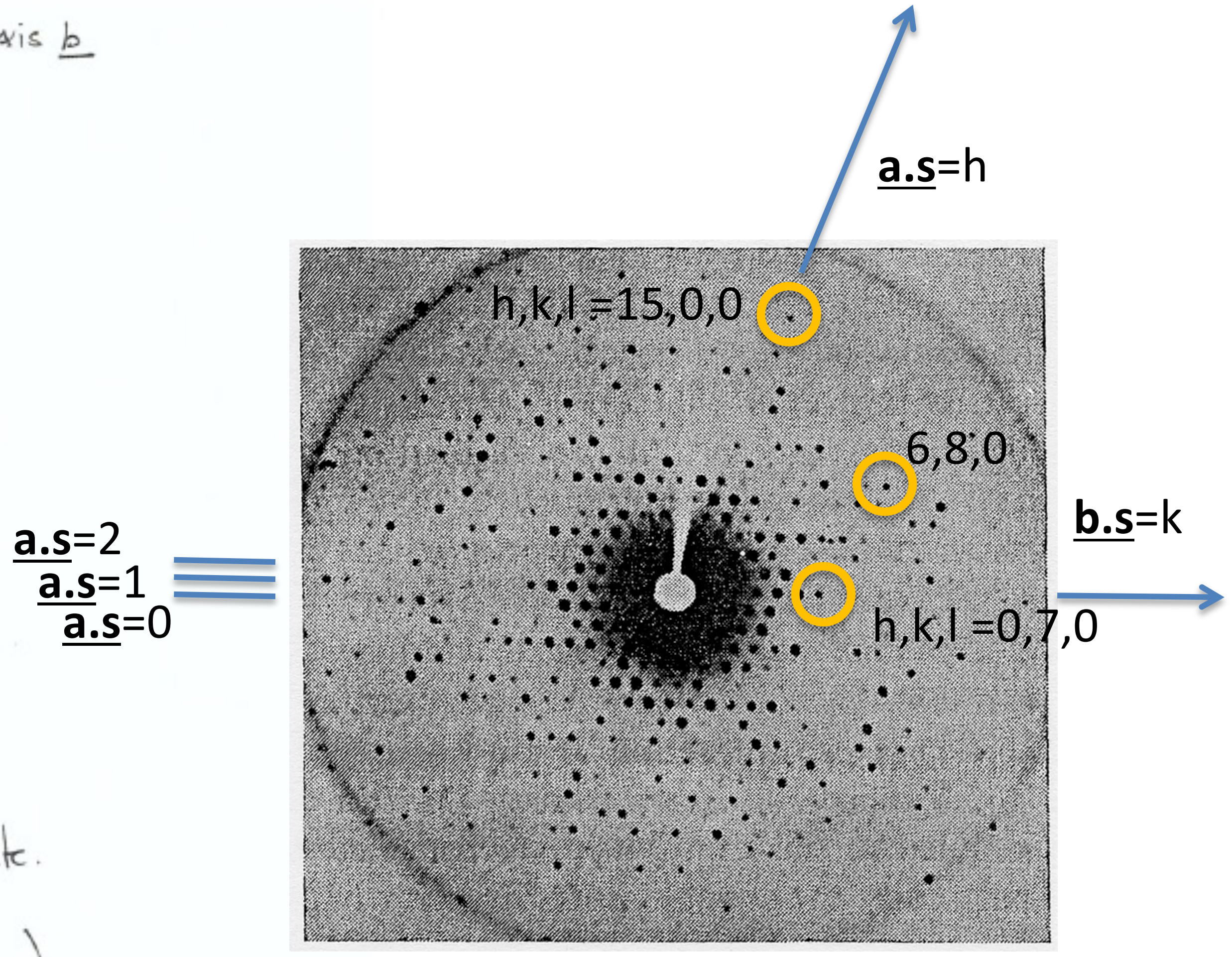
etc.

so for the 2D array:-

$$= F(s)\left(\frac{\sin \pi M a.s}{\sin \pi a.s}\right)\left(\frac{\sin \pi N \underline{b}.\underline{s}}{\sin \pi \underline{b}.\underline{s}}\right)$$

for 3D array

$$= F(s) \frac{\sin \pi M \underline{a}.\underline{s}}{\sin \pi a.\underline{s}} \frac{\sin \pi N \underline{b}.\underline{s}}{\sin \pi \underline{b}.\underline{s}} \frac{\sin \pi P \underline{c}.\underline{s}}{\sin \pi c.s}$$



a.s=h

h,k,l =15,0,0

6,8,0

b.s=k

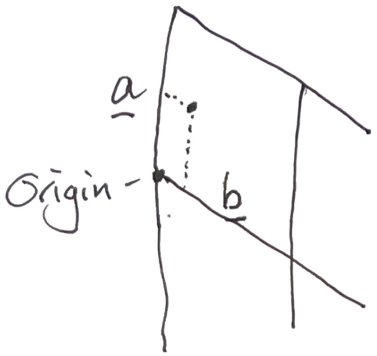
a.s=2

a.s=1

a.s=0

h,k,l =0,7,0

Positions can be described in a "unit cell"



atom$_j$

$$\underline{\mathbf{r}} = x_j \underline{\mathbf{a}} + y_j \underline{\mathbf{b}} + z_j \underline{\mathbf{c}}$$

X-ray "reflections" only when

$$\underline{a} . \underline{s} = h$$
$$\underline{b} . \underline{s} = k$$
$$c . \underline{s} = l$$

(perpendicular to $\underline{a}$)

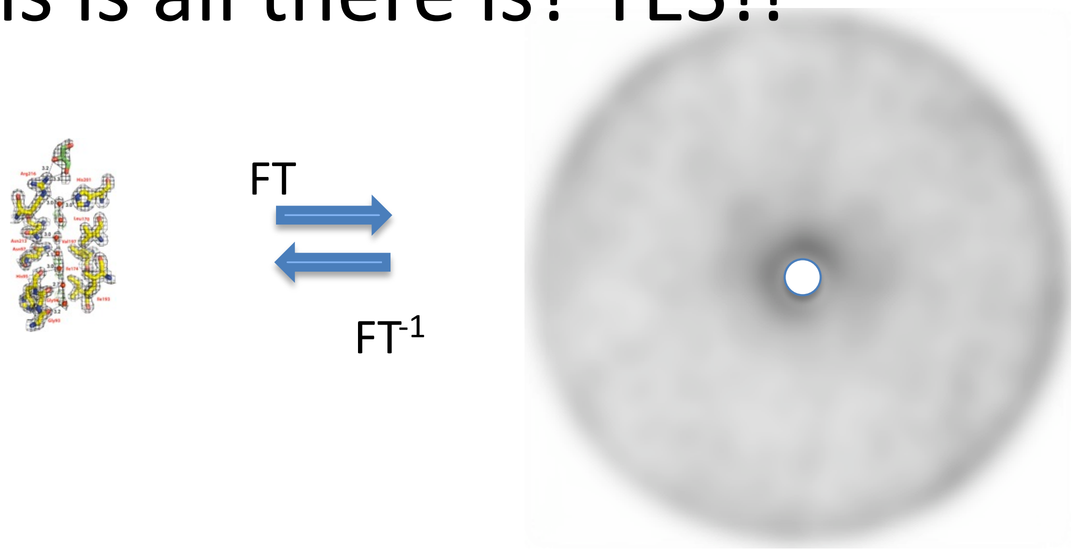
Thus

$$\underline{\mathbf{r}}_j . \underline{\mathbf{s}} = (h x_j + k y_j + l z_j)$$

so $$\underline{F}(\underline{s}) = \sum_j f_j \, e^{2\pi i (h x_j + k y_j + l z_j)}$$

and $$\rho(\underline{r}) = \rho(x,y,z) = \sum_{hkl} |\mathbf{F}| e^{2\pi i \varphi_{hkl}} \cdot e^{2\pi i (hx + ky + lz)}$$

This is all there is? YES!!

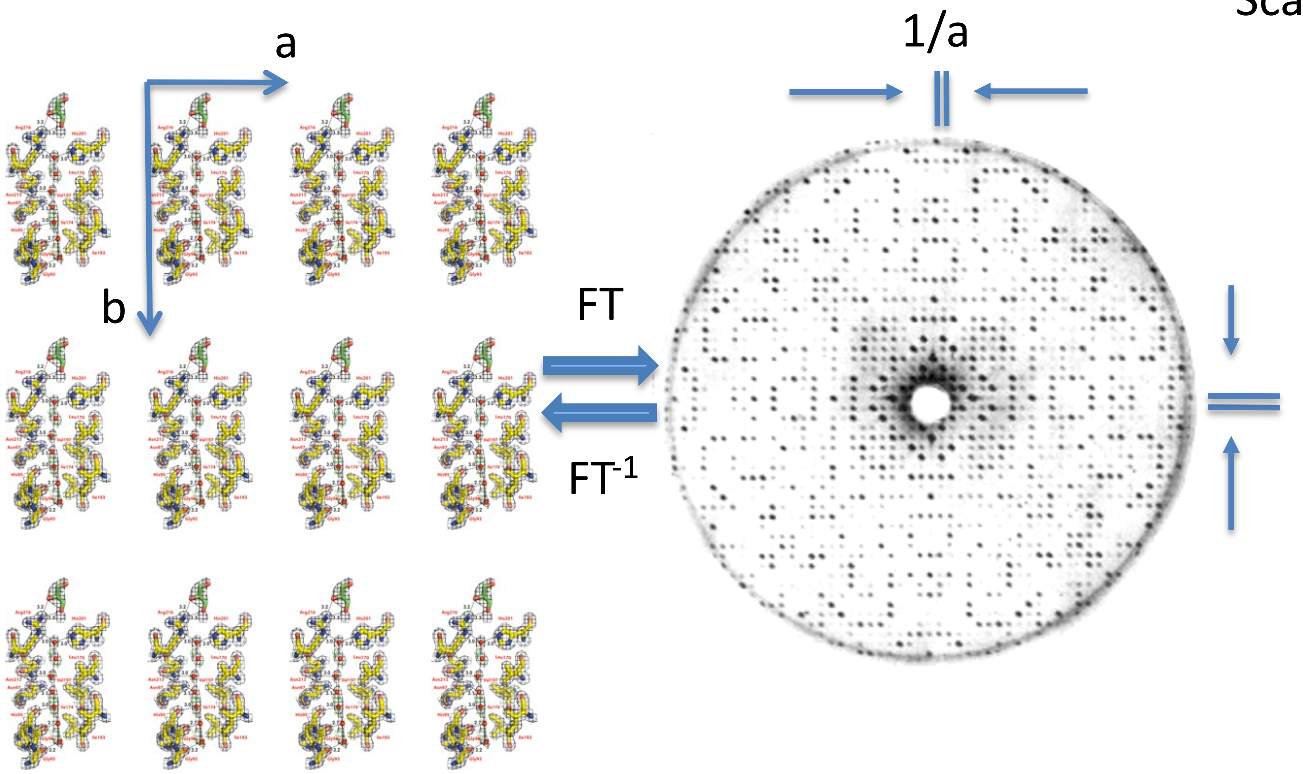


Scattering pattern is the Fourier transform (FT) of the structure: Amplitude and phase of waves is a sum of waves from each atom **j**

$$\underline{F}(\underline{s}) = \Sigma_j f_j e^{(2\pi i \underline{r}_j . \underline{s})}$$

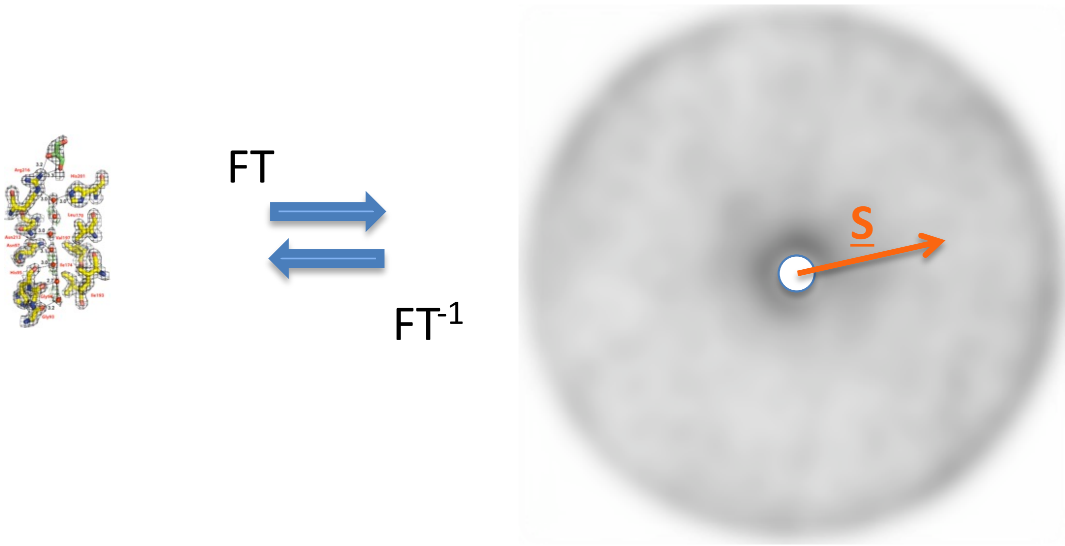
$$\textbf{Observe } \mathbf{I(\underline{S})} = \underline{F}(\underline{s}).\underline{F}^*(\underline{s})$$

Structure is the 'inverse' Fourier transform of the Scattering pattern $\underline{F}(\underline{s})$



$$\rho(\underline{r}) = \Sigma_s \underline{F}(\underline{s})\ e^{(-2\pi i \underline{r}.\underline{s})}$$

This is all there is?

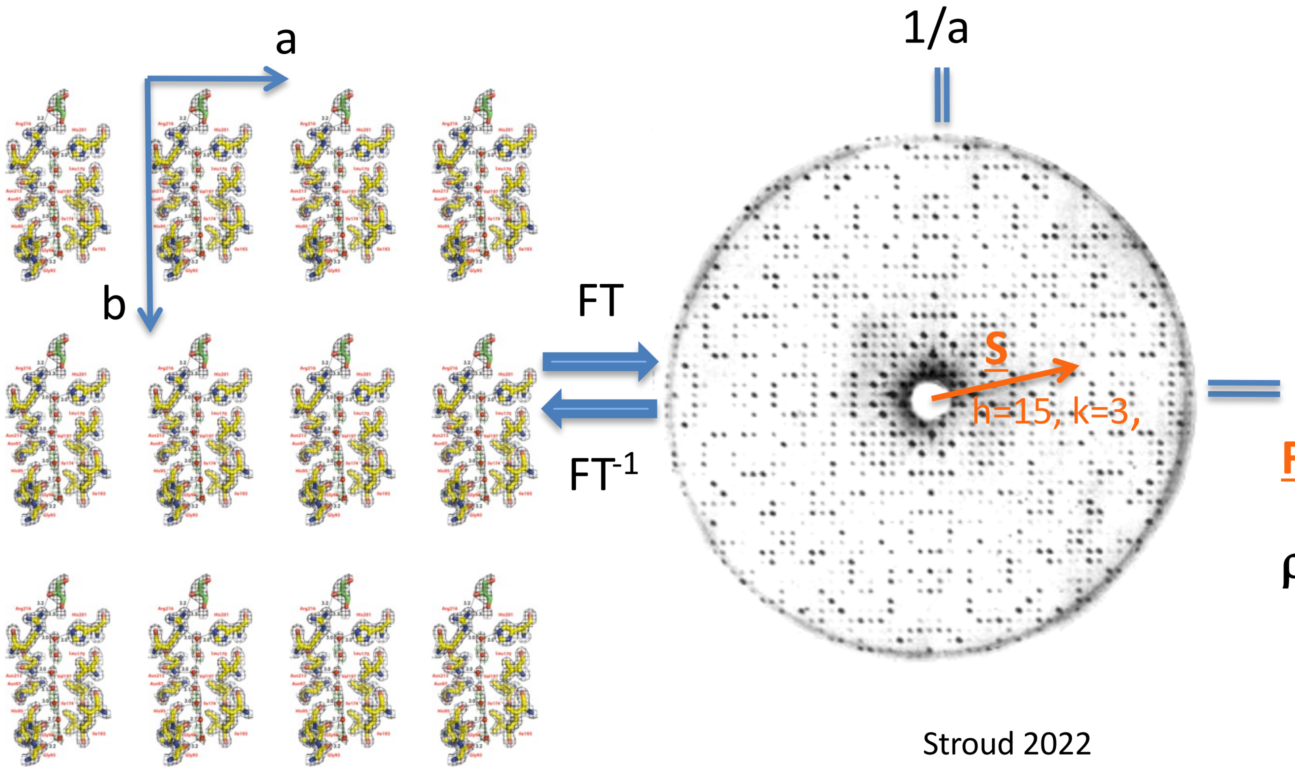


Scattering pattern is the Fourier transform of the structure

$$\underline{F}(\underline{S}) = \Sigma_j \, f_j \, e^{(2\pi i \underline{r_j} \cdot \underline{s})}$$

Structure is the ‘inverse’ Fourier transform of the Scattering pattern

$$\rho(\underline{r}) = \Sigma_s \, \underline{F}(\underline{s}) \, e^{(-2\pi i \underline{r}.\underline{s})}$$



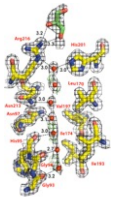
$$\underline{F}(h,k,l) = \Sigma_j \, f_j \, e^{(2\pi i \, (hx_j+ky_j+lz_j))}$$

$$\rho(x,y,z) = \Sigma_{h,k,l} \, \underline{F}_{(\underline{h,k,l})} \, e^{(-2\pi i \, (hx+ky+lz))}$$

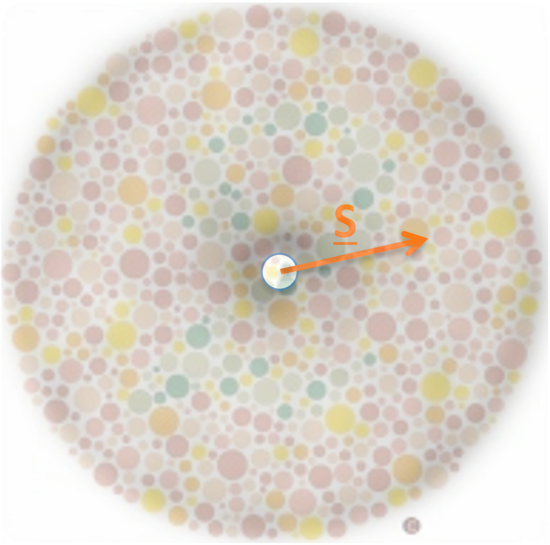
Stroud 2022

This is all there is? **PHASES-as colors !**

$\rho(\underline{r})$



FT ⇔ FT^{-1}

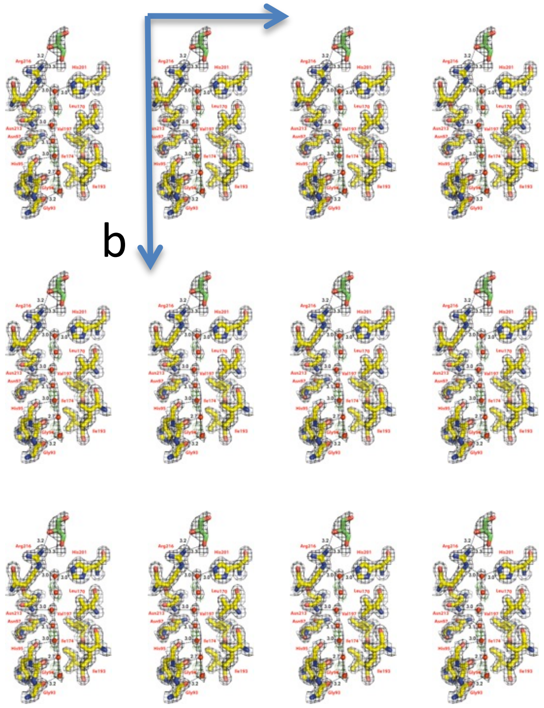


Scattering pattern is the Fourier transform of the structure

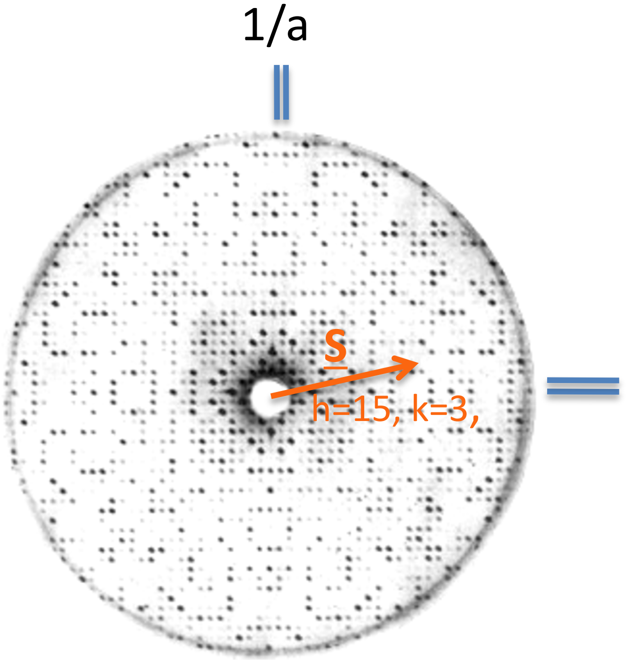
$$\underline{F}(\underline{S}) = \Sigma_j \, f_j \, e^{(2\pi i \underline{r}_j \cdot \underline{s})}$$

Structure is the 'inverse' Fourier transform of the Scattering pattern

$\rho(\underline{r})$



FT ⇔ FT^{-1}



$$\rho(\underline{r}) = \Sigma_s \, \underline{F}(\underline{S}) \, e^{(-2\pi i \mathbf{r.S})}$$

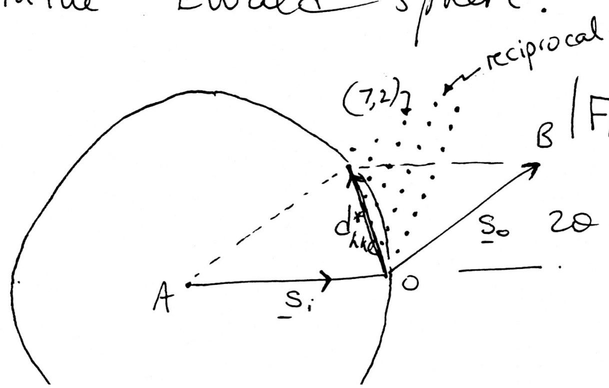
$$\underline{F}(h,k,l) = \Sigma_j \, f_j \, e^{(2\pi i \, (hx+ky+lz))}$$

$$\rho(x,y,z) = \Sigma_{h,k,l} \, \underline{F(h,k,l)} \, e^{(-2\pi i \mathbf{r.S})}$$

describe any reciprocal lattice point as

$$\underline{s} = \underline{d}^* = h a^* + k b^* + l c^*$$

This (hkl) lattice point will only be observed when the crystal is turned so ~~that~~ the (hkl) point $\underline{s} = d^*_{hkl}$ ~~lies~~ lands in the Ewald sphere.



$$|F_{hkl}|^2 = G(\underline{s})\,G^*(\underline{s}) = I_{hkl}$$

Relative Information in Intensities versus phases

$\rho(\underline{r})$ duck

$\underline{F}_{(\underline{S})}$ duck →

$|F_{(\underline{S})}|$

$\phi_{(\underline{s})}$

$|F_{(\underline{S})}|$ duck

$\phi_{(\underline{s})}$ cat

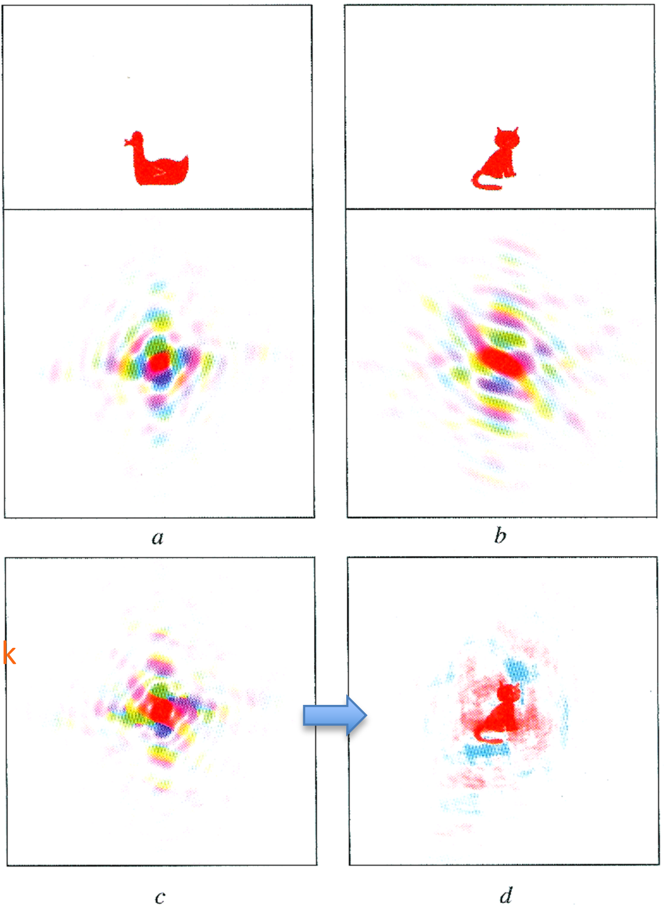


Figure 6.1 ▶ Relative amounts of information contained in reflection intensities and phases. (*a*) and (*b*) Duck and cat, along with their Fourier transforms. (*c*) Intensity (shading) of the duck transform, combined with the phases (colors) of the cat transform. (*d*) Back-transform of (*c*) produces recognizable image of cat, but not duck. Phases contain more information than intensities. Figure generously provided by Dr. Kevin Cowtan.

$\rho(\underline{r})$ cat

$\underline{F}_{(\underline{S})}$ cat

$$\underline{F}_{(\underline{S})} = \Sigma_r \rho_{(r)} e^{(2\pi i \mathbf{r}.\mathbf{S})}$$

$$\rho(\underline{r}) = \Sigma \underline{F}_{(\underline{S})} e^{(-2\pi i \mathbf{r}.\mathbf{S})}$$

Looks like a …..???

Relative Information in Intensities versus phases

$\rho(\underline{\mathbf{r}})$ duck

$\rho(\underline{\mathbf{r}})$ cat

$\underline{\mathbf{F}}_{(\underline{S})}$ duck →

$|F_{(\underline{S})}|$

$\phi_{(\underline{S})}$

$\underline{\mathbf{F}}_{(\underline{S})}$ cat

$$\underline{\mathbf{F}}_{(\underline{S})} = \Sigma_j f_j e^{(2\pi i\mathbf{r.S})}$$

$|F_{(\underline{S})}|$ duck

$\phi_{(\underline{S})}$ cat

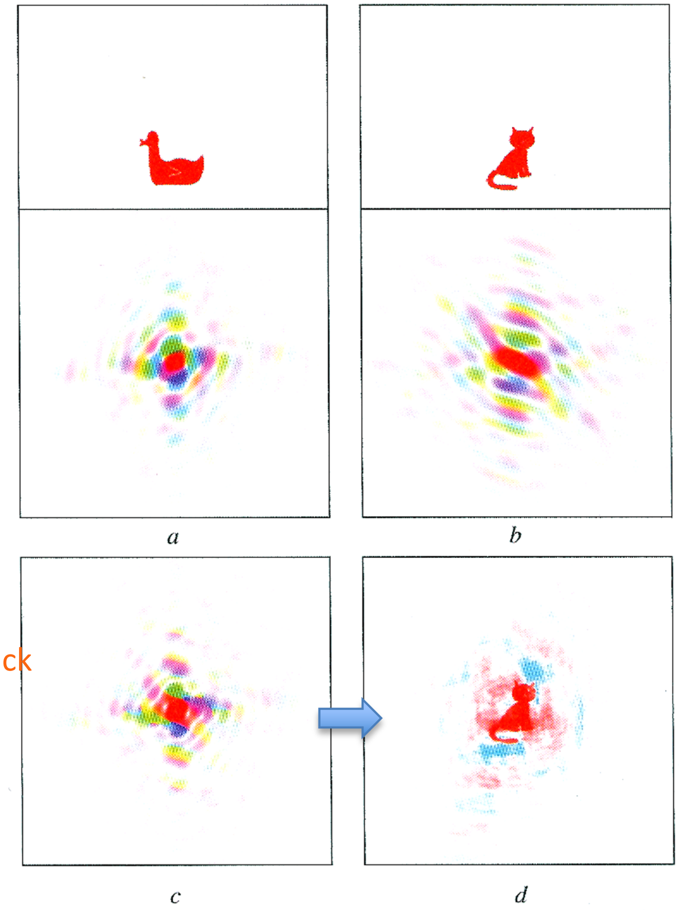


Figure 6.1 ▶ Relative amounts of information contained in reflection intensities and phases. (*a*) and (*b*) Duck and cat, along with their Fourier transforms. (*c*) Intensity (shading) of the duck transform, combined with the phases (colors) of the cat transform. (*d*) Back-transform of (*c*) produces recognizable image of cat, but not duck. Phases contain more information than intensities. Figure generously provided by Dr. Kevin Cowtan.

$$\rho(\underline{\mathbf{r}}) = \sum_{\underline{S}} \mathbf{F}_{(\underline{S})} e^{(-2\pi i\mathbf{r.S})}$$

Looks like a CAT

PHASES DOMINATE:

-Incorrect phases = incorrect structure

-incorrect model = incorrect structure

-incorrect assumption = incorrect structure

2. Molecular Replacement.

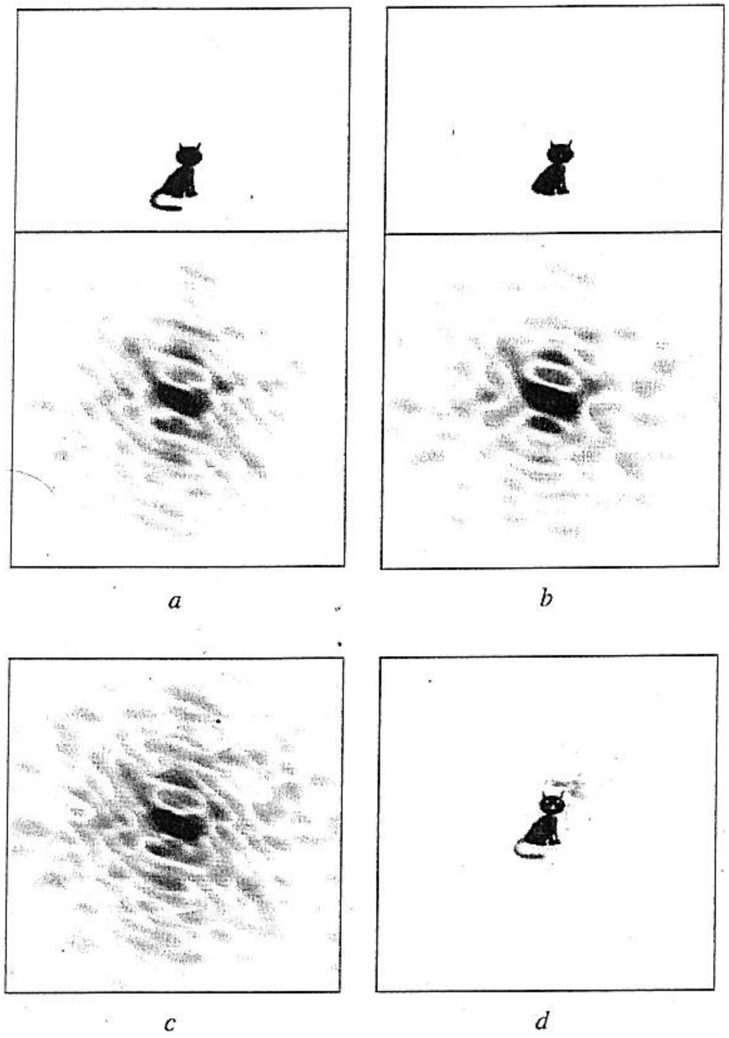


Figure 6.17 ▶ Structure determination by molecular replacement. (*a*) Unknown structure, cat, and its diffraction pattern (not colored, because phases are unknown). (*b*) Known structure and phasing model, Manx cat, and transform computed from the model (colored, because calculation of transform from a model tells us phases). (*c*) Manx-cat phases combined with unknown-cat intensities. (*d*) Back-transform of (*c*). Intensities contain enough information to reveal differences (the tail) between phasing model and unknown structure.

Molecular Replacement

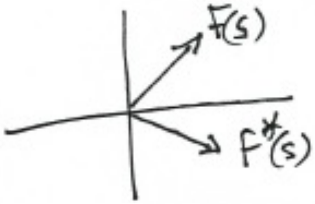
$$\underline{F}(\underline{s}) = \sum f_i \, e^{2\pi i \, \underline{r}_i \cdot \underline{s}}$$

What happens if we transform the observed intensities $I(s)$?

ie $$P(\underline{r}) = \sum_{\underline{s}} I(\underline{s}) \, e^{2\pi i \, \underline{r} \cdot \underline{s}} = \sum_{hkl} I_{hkl} \, e^{2\pi i (hx+ky+l}$$

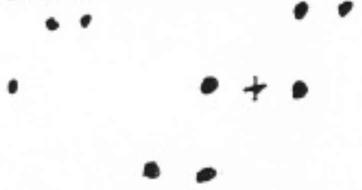
$$P(\underline{r}) = \left(\underline{F}(\underline{s}) \times \underline{F}^{*}(\underline{s}) \right) e^{2\pi i \, \underline{r} \cdot \underline{s}}$$

$$= I(\underline{s}) \, e^{2\pi i \, \underline{r} \cdot \underline{s}}$$

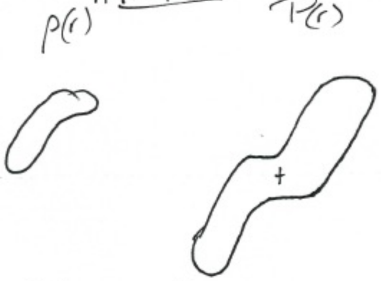


$P(\underline{r})$ = Patterson function
≡ "All vectors in the crystal, weighted by electron, brought to a common origin"

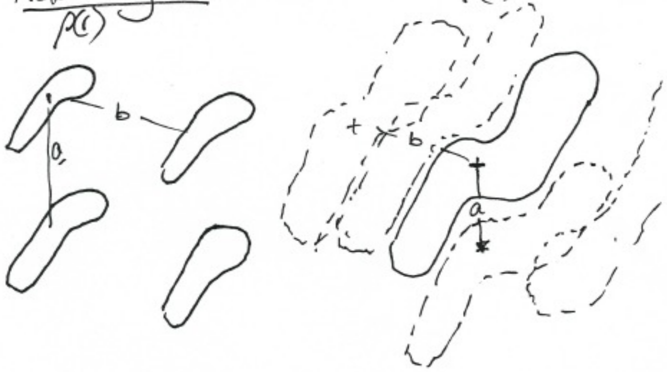
$\rho(r)$ Molecule $\qquad P(r) = \rho(r) \, \rho(-r)$



A Protein



Protein Crystal

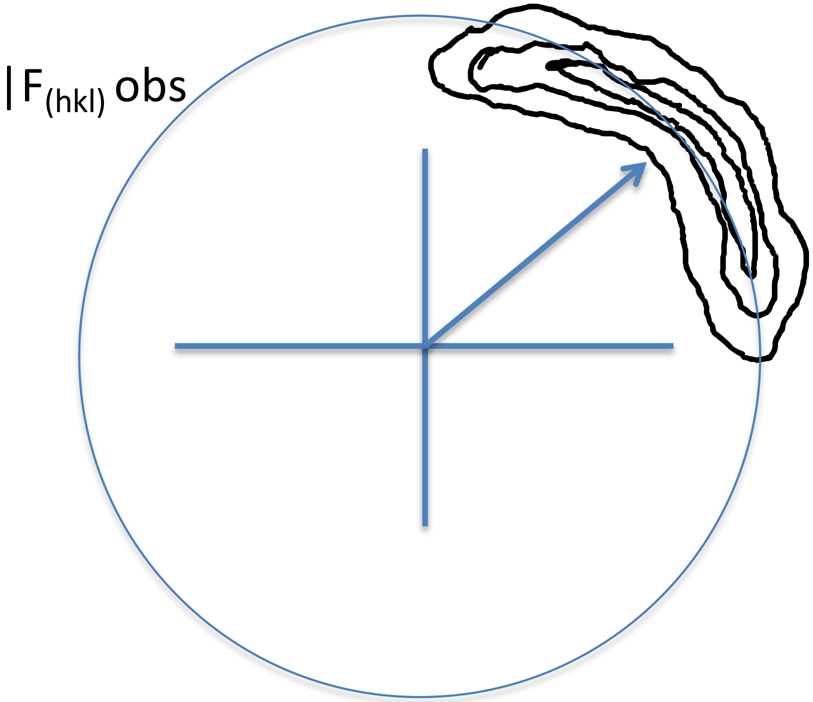


A protein Molecule

Protein Crystal

Molecular Replacement:
Search P(r) (calculated from observed I(s) only)
with a P(r) calculated for a similar molecule
⇒ find orientation
⇒ search all translations.

Phase determination by any means, ends up as a probabilty distribution. So $F_{h,k,l}$, $\phi_{(h,k,l)}$



Then what to use for the best map?

$\rho(\underline{r}) = \Sigma\, \underline{\mathbf{F}}_{(\underline{S})}\, e^{(-2\pi i\mathbf{r.S})}$?

the signal towards some F true will be

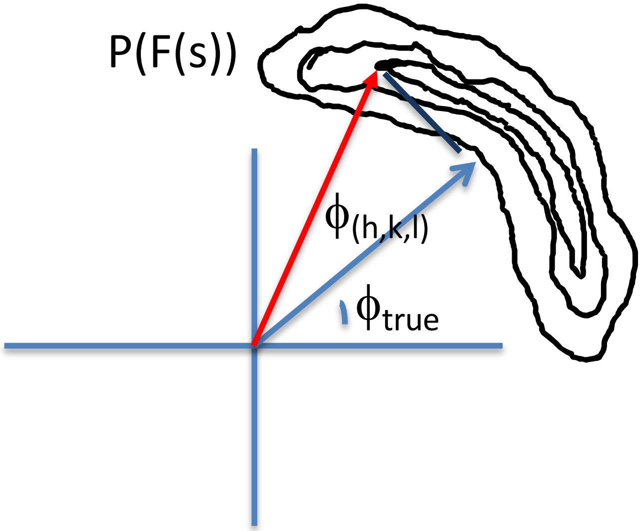
$\int_\phi P(F(S)) \cos(\phi_{(h,k,l)} - \phi_{true})$

and the ‘noise ‘ will be

$\int_\phi P(F(S)) \sin(\phi_{(h,k,l)} - \phi_{true})$

The map with the least noise will have F(s) = center of mass of P(F(S))

Phase determination by any means, ends up as a probabilty distribution. So $F_{h,k,l}$, $\phi_{(h,k,l)}$



Then what to use for the best map?

$\rho(\underline{\mathbf{r}}) = \Sigma\, \underline{\mathbf{F}}_{(\underline{\mathbf{s}})}\, e^{(-2\pi i\mathbf{r.S})}$?

the signal towards some F true will be

$\int_{\phi} P(|F(\underline{\mathbf{s}})|\, \cos\, (\phi_{(h,k,l)} - \phi_{true})$

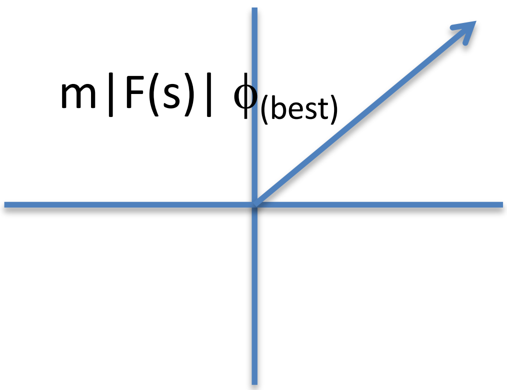
and the 'noise ' will be

$\int_{\phi} P(|F(\underline{\mathbf{s}}))\, \sin\, (\phi_{(h,k,l)} - \phi_{true})$

The map with the least noise will have F(s) = center of mass of P(F(S))

Figure of merit weights to ‘minimum error’

Phase determination by any means, ends up as a probabilty distribution. So $F_{h,k,l}$, $\phi_{(h,k,l)}$



Then what to use for the best map?

$\rho(\underline{\mathbf{r}})_{best} = \underline{\mathbf{F}}_{(\underline{\mathbf{S}})}\, e^{(-2\pi i\mathbf{r.S})}$?

The map with the least noise will have

F(s) = center of mass of P(F(S))= $\int_\phi P_\phi\, F \sin(\phi-\phi_{(best)})$

is a minimum. Then m = $\int_\phi P_\phi\, F \cos(\phi-\phi_{(best)})/F$

Signal = $m|F(s)|\ \phi_{(best)} = \int_\phi P.F$

where m = figure of merit = $\int_\phi P_\phi(F)\ F(s)$

m = $\langle\cos\Delta\phi\rangle$

noise = $\int_\phi F(s) \sin\Delta\phi$

If a map is produced with some $\phi_{(hkl)}$
The probability of it being correct is $\Pi_{(hkl)} P_{(hkl)} (\phi_{(hkl)})$

Maximum value of $P_{(hkl)} (\phi_{(hkl)})$ gives the 'Most probable' map

Map with the least mean square error, is when noise is minimum,
Int find $\phi_{(best)}$ such that
$Q = \int_{\phi} [|F|\ P_{(hkl)} (\phi_{(hkl)}) \exp (i\phi_{(hkl)}) - \mathbf{F}_{best}\, \phi_{(best)})]^2 d\phi$ is minimum.

is minimum when $dQ/d\mathbf{F}_{best} = 0$
so $\mathbf{F}_{best}\, \phi_{(best)} = \int_{\phi} |F|\ P_{(hkl)} (\phi_{(hkl)}) \exp (i\phi_{(hkl)})\ d\phi$

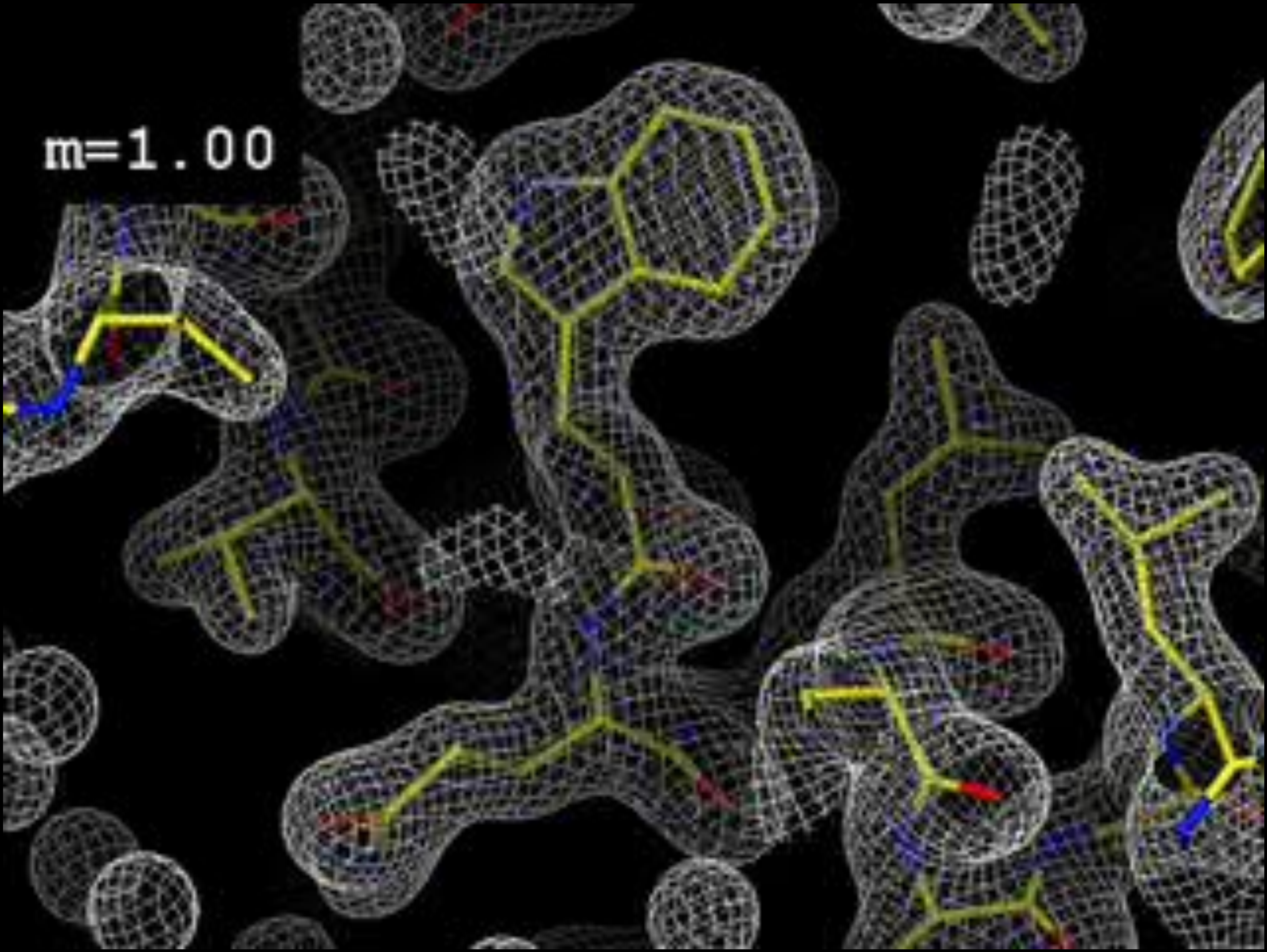
$\mathbf{F}_{best}\, \phi_{(best)} = m|F|$ center of 'mass' of the Probability distribution

where $m = \int_{\phi} P_{(hkl)} (\phi_{(hkl)}) \cos(\phi - \phi_{(best)})$
consider rms errors from one reflection, and its complex conjugate
$<(\Delta\rho)^2> = 2/V^2 \int_{\phi} P_{(hkl)} (\phi_{(hkl)}) (\sin(\phi - \phi_{(best)}))^2$

Then $|F|_{best} = \int_{\phi} F \cos(\phi - \phi_{(best)}) / F$
Noise $<(\Delta\rho)^2> = 2/V^2 \int_{\phi} F (\sin (\phi - \phi_{(best)}))^2/F = F(1-m^2)$

mF where $m = \int_{\phi} F \cos(\phi - \phi_{(best)})$

Figure of Merit



http://bl831.als.lbl.gov/~jamesh/movies/dephase.mpeg

‘Difference maps’

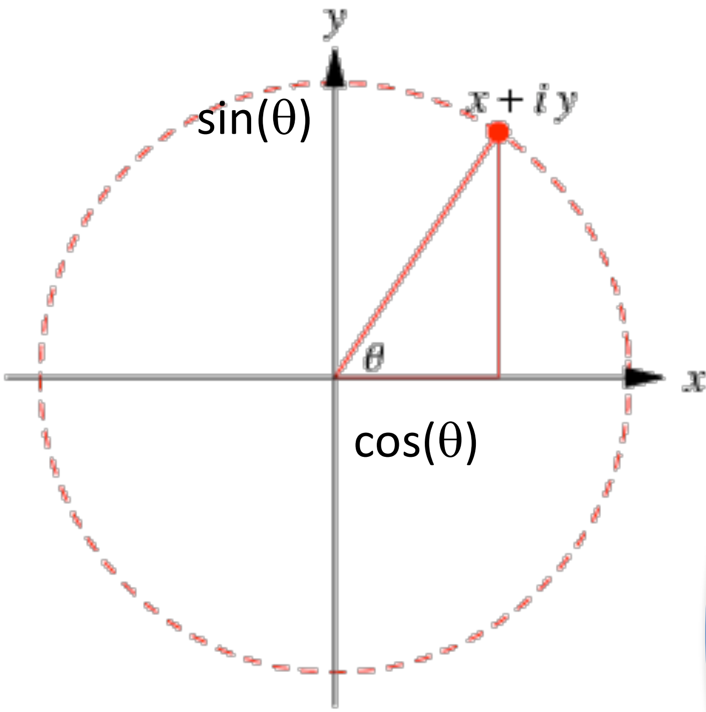
-Define bound ligands
-to find any missing atoms during refinement,
-to find ligands
-define movements of protein or water
-determine ion positions
-determine changes in dynamic motion

Suppose we interpret 7 atoms; but 3 remain to be found in density

Result is a wave of
amplitude $|\mathbf{F}(\underline{\mathbf{S}})|$
phase $\Phi(\underline{\mathbf{S}})$

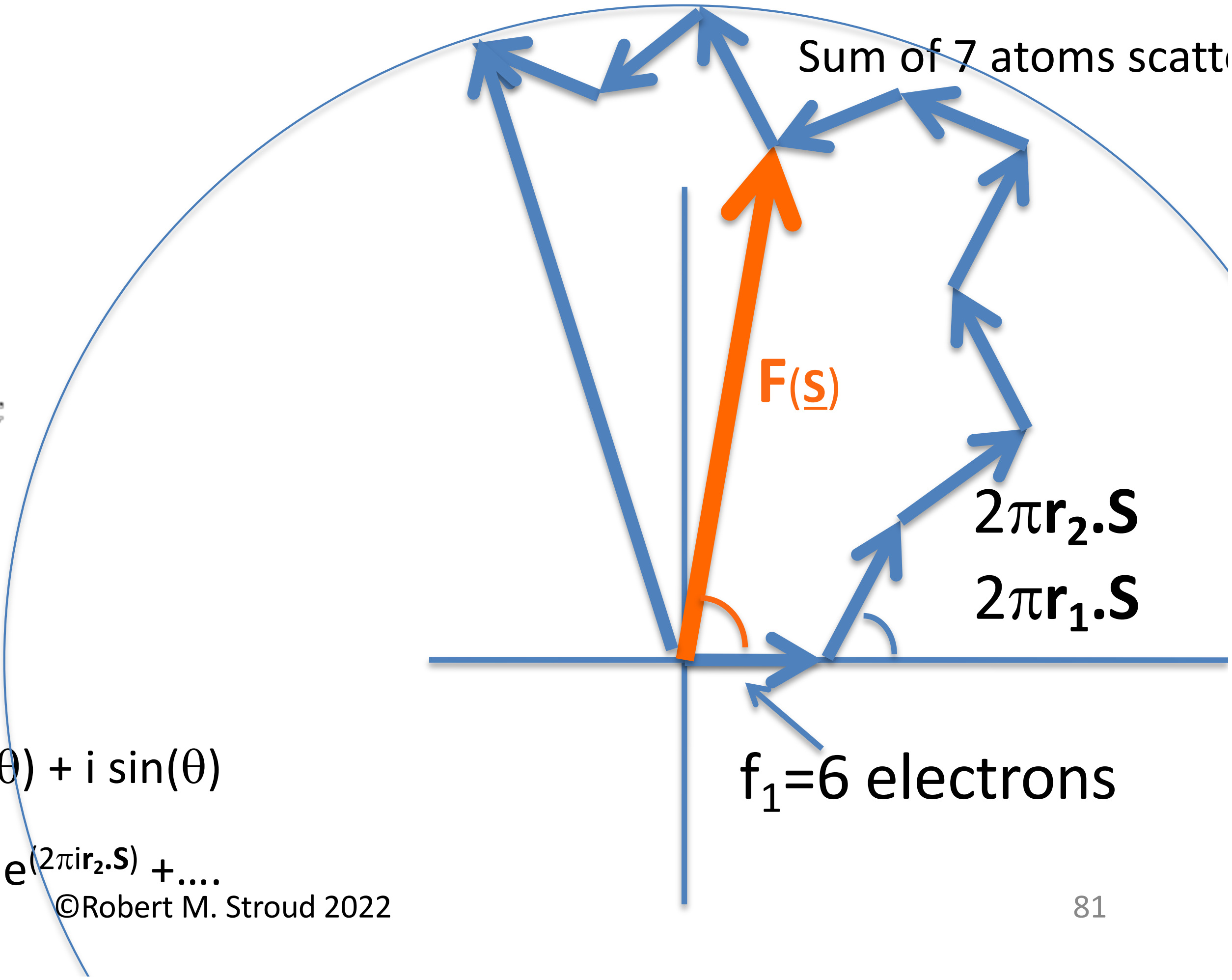
In reality, maybe 3 atoms are missing.
How to see what is missing?

$i = \sqrt{(-1)}$



$$e^{(i\theta)} = \cos(\theta) + i \sin(\theta)$$

$$\mathbf{F}(\underline{\mathbf{S}}) = f_1 e^{(2\pi i \mathbf{r_1}.\mathbf{S})} + f_2 e^{(2\pi i \mathbf{r_2}.\mathbf{S})} + \ldots$$



USES: 1. Determining missing regions

$|\mathbf{F}(\underline{\mathbf{S}})|_{obs}$ Compare with $|\mathbf{F}(\underline{\mathbf{S}})|_{calc}$

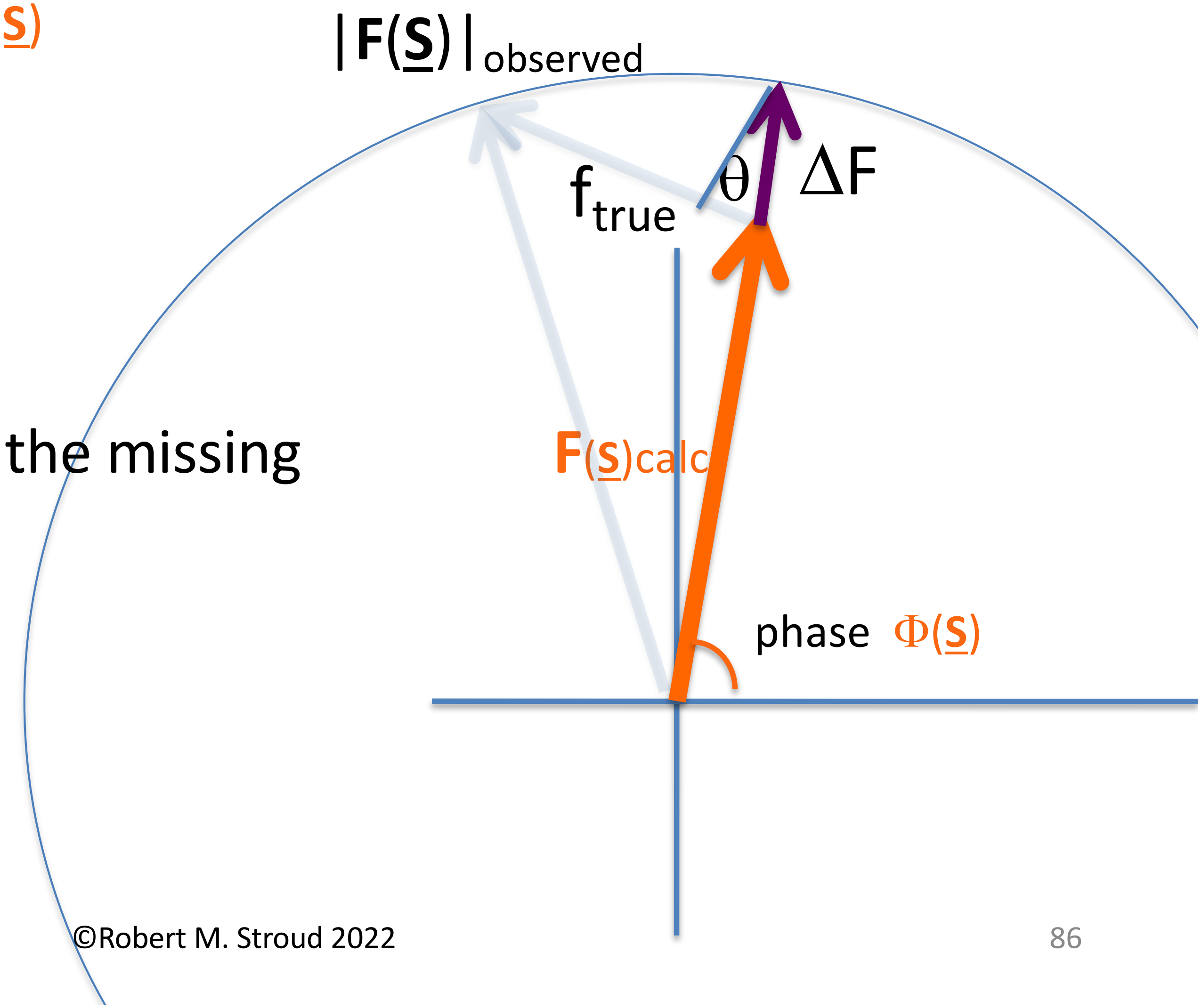
Transform $\Delta F = \Big|\, |\mathbf{F}(\underline{\mathbf{S}})|_{obs} - |\mathbf{F}(\underline{\mathbf{S}})|_{calc} \,\Big|\ \Phi(\underline{\mathbf{S}})$

or

$[2|\mathbf{F}(\underline{\mathbf{S}})|_{obs+substrate} - |\mathbf{F}(\underline{\mathbf{S}})|_{obs}]\ \Phi(\underline{\mathbf{S}})$

= a '$2F_0$-F_o map'

It is **<u>unbiased</u>** as to where the missing Atoms are.



USES: 2. Add a substrate, Grow a new crystal

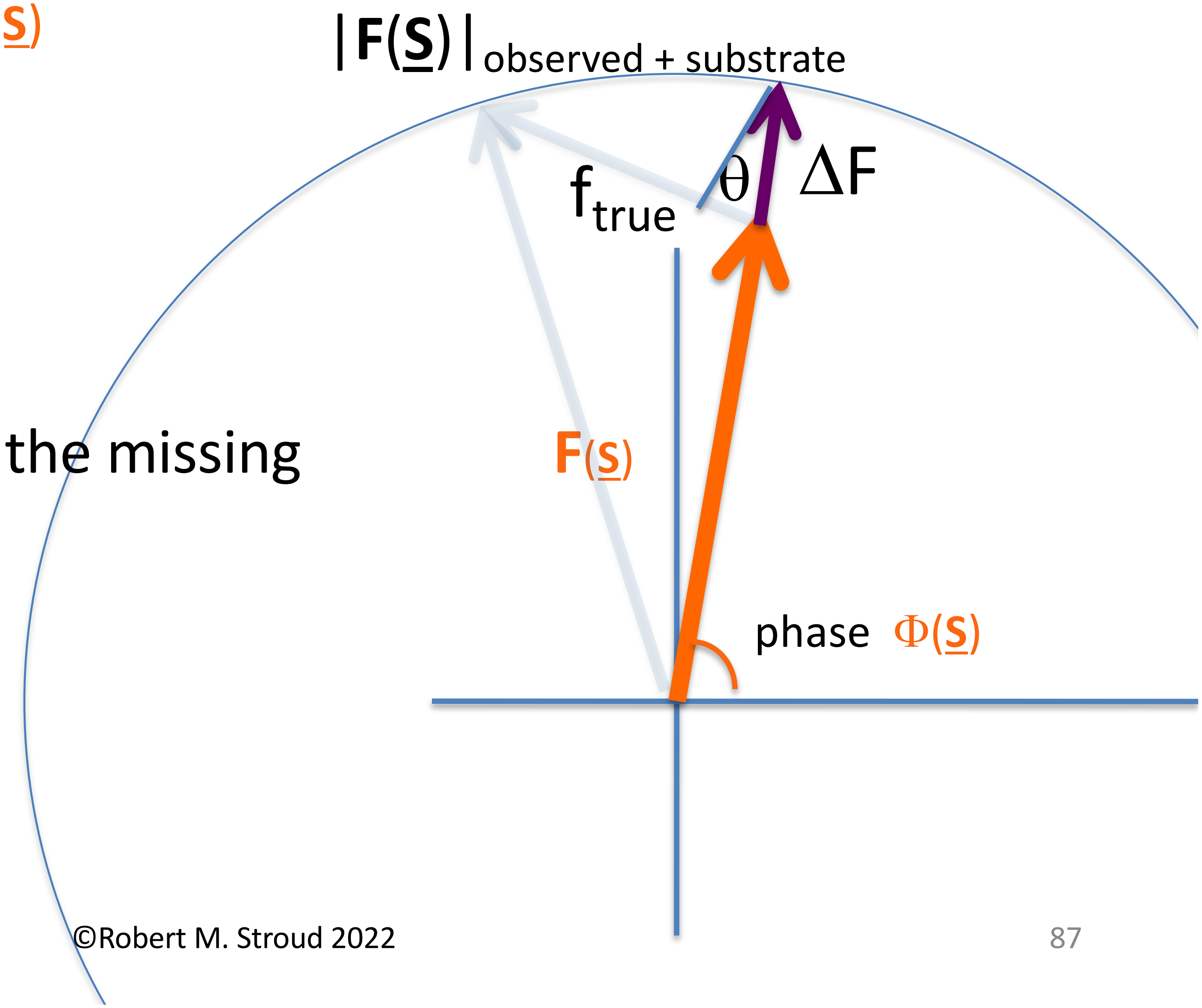
Measure New $|\mathbf{F}(\underline{\mathbf{S}})|_{obs+substrate}$ Compare with the apo-protein.

Transform $\Delta F = \Big| \, |\mathbf{F}(\underline{\mathbf{S}})|_{obs+substrate} - |\mathbf{F}(\underline{\mathbf{S}})|_{obs} \Big| \; \Phi(\underline{S})$

or

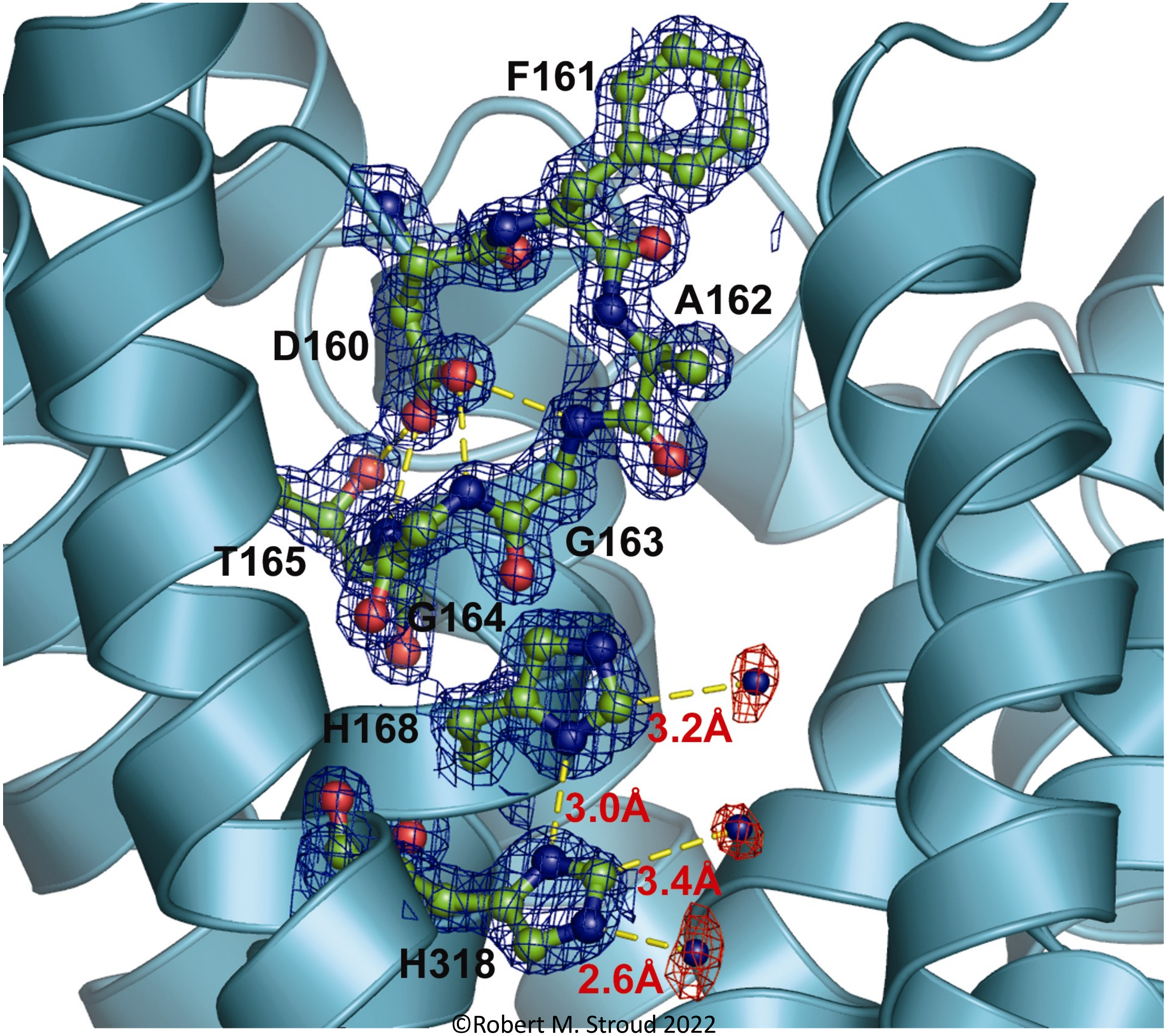
$[2|\mathbf{F}(\underline{\mathbf{S}})|_{obs+substrate} - |\mathbf{F}(\underline{\mathbf{S}})|_{obs}] \; \Phi(\underline{S})$

= a ‘$2F_0$-F_o map’



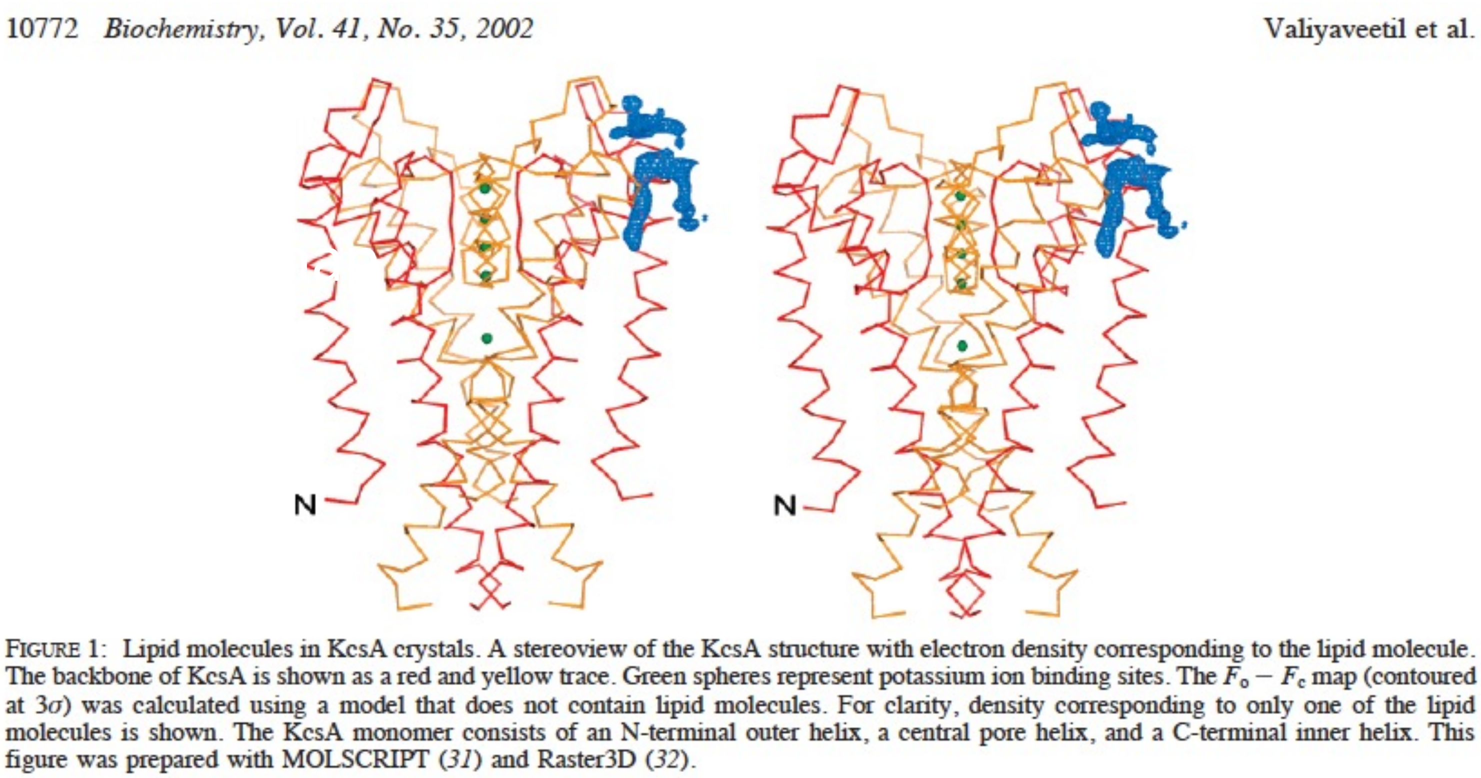
It is **unbiased** as to where the missing substrate is.

A Dfference map shows 1/3 occupied NH3 sites and the role of D160 at 1.35Å Resolution. Here are 0.3 NH_3 peaks!



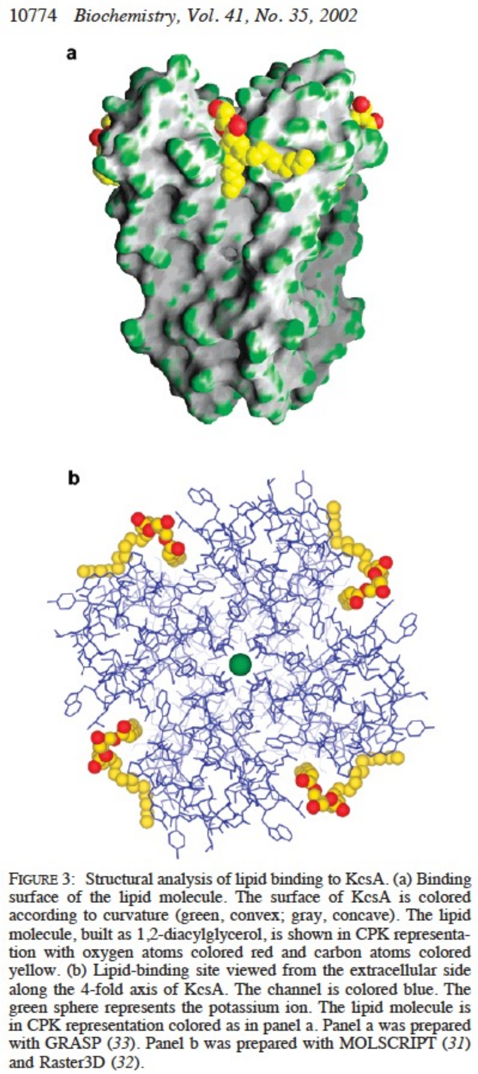
Khademi..Stoud 2003

Fo-Fc maps identify everything ordered that is 'missing'



10772 *Biochemistry, Vol. 41, No. 35, 2002* Valiyaveetil et al.

FIGURE 1: Lipid molecules in KcsA crystals. A stereoview of the KcsA structure with electron density corresponding to the lipid molecule. The backbone of KcsA is shown as a red and yellow trace. Green spheres represent potassium ion binding sites. The $F_o - F_c$ map (contoured at 3σ) was calculated using a model that does not contain lipid molecules. For clarity, density corresponding to only one of the lipid molecules is shown. The KcsA monomer consists of an N-terminal outer helix, a central pore helix, and a C-terminal inner helix. This figure was prepared with MOLSCRIPT (*31*) and Raster3D (*32*).



10774 *Biochemistry, Vol. 41, No. 35, 2002*

FIGURE 3: Structural analysis of lipid binding to KcsA. (a) Binding surface of the lipid molecule. The surface of KcsA is colored according to curvature (green, convex; gray, concave). The lipid molecule, built as 1,2-diacylglycerol, is shown in CPK representation with oxygen atoms colored red and carbon atoms colored yellow. (b) Lipid-binding site viewed from the extracellular side along the 4-fold axis of KcsA. The channel is colored blue. The green sphere represents the potassium ion. The lipid molecule is in CPK representation colored as in panel a. Panel a was prepared with GRASP (*33*). Panel b was prepared with MOLSCRIPT (*31*) and Raster3D (*32*).

-Eliminate Bias

-Half electron content

-See electrons

The closer you get –the lower the noise.
Can see single electrons.

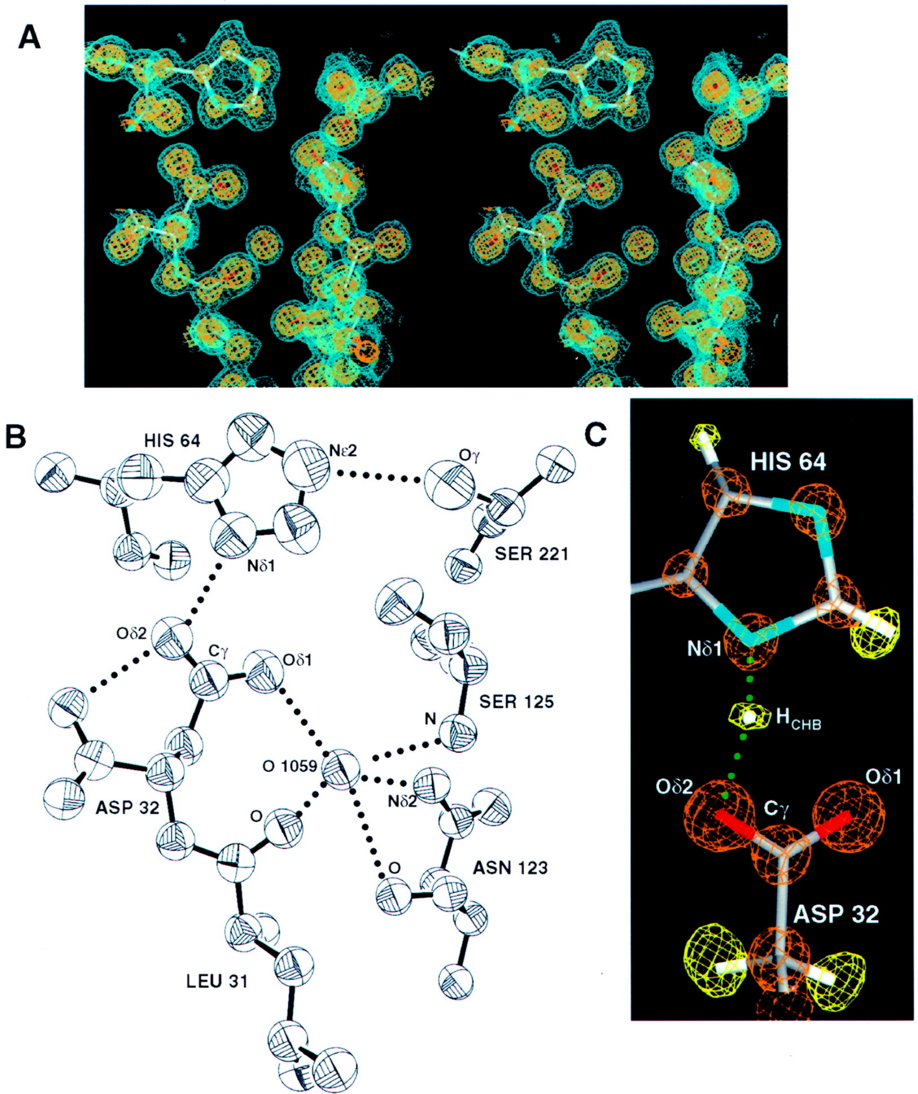
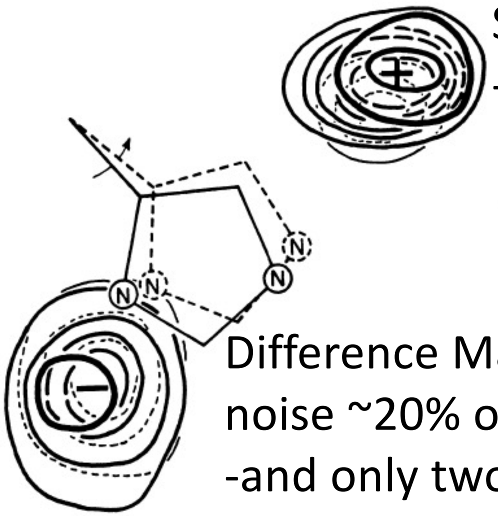


Figure 3 The catalytic triad. **(A) Stereoview displaying Model H superimposed on the 2*Fo Fc* (**model H phases) at 1 (aqua) and 4 (gold). The densities for C and N in His 64 are weaker than in Asp 32. The Asp 32 CO2 bond at 4 is continuous, while the density for the C and O1 are resolved. (B) Schematic of the catalytic residues and hydrogen bonded neighbors with thermal ellipsoid representation countered at 50% probability (*29*). Catalytic triad residues Ser 221 and His 64 show larger thermal motion than the Asp 32. Solvent O1059 appears to be a relatively rigid and integral part of the enzyme structure. (C) Catalytic hydrogen bond (CHB)**. A *Fo Fc* (model H phases) difference map** contoured at +2.5 (yellow) and 2.5 (red) and a 2*Fo Fc* (model H phases) electron density map contoured at 4 (gold). The position of the short hydrogen atom (labeled HCHB) in the CHB is positioned in the positive electron density present between His 64 N1 and Asp 32 O2.

Published in: Peter Kuhn; Mark Knapp; S. Michael Soltis; Grant Ganshaw; Michael Thoene; Richard Bott; *Biochemistry* **1998,** 37, 13446-13452.
DOI: 10.1021/bi9813983



Seeing small shifts
-down to ±0.1Å

Difference Map
noise ~20% of noise in the parent protein
-and only two peaks!

FIG. 7. The peaks associated with His57 on the difference map. The lower peak is negative density (−) while the other one is positive (+). The latter peak is a composite with a solvent molecule density (see text).

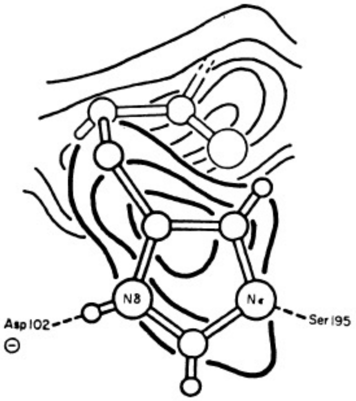


FIG. 8. Electron density for His57 in the DIP-trypsin Fourier map, computed for the plane parallel to the imidazole ring.

TABLE 1

Analysis of Fourier maps

Map	$\langle\lvert F_{\text{obs}}\rvert\rangle$ (e)	σ (e)	Calculated† $\langle \Delta\rho^2\rangle^{\frac{1}{2}}$ (e Å^{-3})	Observed‡ r.m.s. error (e Å^{-3})	Observed highest noise (e Å^{-3})	s.d.§	Observed highest peak (e Å^{-3})	s.d.
BA-trypsin—DIP-trypsin	84·7	2·3	0·069	0·059	0·17	2·5	0·75	11
DIP-trypsin	573·0	21·0	0·38	—	—	—	—	—

† ΔF: $\langle \Delta\rho^2\rangle = \frac{1}{2V^2}\sum_{hkl} \Delta F^2\,(2-m^2)$,

F_{DIPT}: $\langle \Delta\rho^2\rangle = \frac{1}{V^2}\sum_{hkl} F^2_{\text{DIPT}}\,(1-m^2)$,

(after Henderson & Moffat, 1971).

‡ The observed root mean-square density error is based on a relatively featureless region of the map.
§ s.d., the electron density given as a multiple of the calculated r.m.s. error.

Difference maps; to the last electron! Why? Supplementary Proof of the ‘Random Walk’ calculation

The 'Random Walk' problem? (p33.1-33.3)

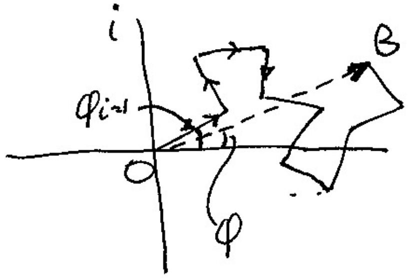
What is the average sum of
n steps in random directions?

(What is the average amplitude
<|F(s)|> from an n atom structure?)

-AND why do we care?!........

How much difference in <|F(s)|> from adding a
4 carbon atom substrate?
mercury atom (f=80)?

"Random Walk", in 2d - -
(What is scattering amplitude from
n atoms, each with f electrons)



n steps, of length f
relatively randomly
different angles.
ie. Probability
$P(\varphi)$

Easiest to consider as the expected value maximum for $|OB|^2$ (rather than OB)

$$|OB|^2 = \underline{OB} . \underline{OB}^*$$

[Review: $\underline{OB} = |OB| (\cos(\varphi) + i \sin\varphi)$
$\underline{OB}^* = |OB| (\cos(\varphi) - i \sin\varphi)$

So $\underline{OB} . OB^* = |OB|^2 (\cos^2\varphi + \sin^2\varphi) = |OB|^2$]

$$\text{So } \underline{OB} = \sum_{i=1}^{n} f e^{2\pi i \varphi_i}$$

$$\underline{OB}^* = \sum_{i=1}^{n} f e^{-2\pi i \varphi_i}$$

The average intensity for an
n atom structure, each of f electrons
is <I>= nf^2

The average amplitude is Square root
of n, times f

so

$$|OB|^2 = \left(\sum_{i=1}^{n} e^{2\pi i \varphi_i}\right)\left(\sum_{j=1}^{n} e^{-2\pi i \varphi_j}\right) \cdot f^2$$

$$= \sum_{i}, \sum_{j} e^{2\pi i(\varphi_i - \varphi_j)} \cdot f^2$$

since $P(\varphi_i) = P(\varphi_j)$ = constant (all equally probable)

the sums ..

- for $i \neq j$ $P(\varphi_i)$ all equally probable

  so $\sum_i \sum_j \Rightarrow \int_0^{2\pi} \underset{0}{\underset{\shortparallel}{\cos 2\pi(\varphi_i - \varphi_j)}} + \underset{0}{\underset{\shortparallel}{i \sin 2\pi(\varphi_i - \varphi_j)}}$

- for $i = j$ $e^{2\pi i(\varphi_i - \varphi_j)} = e^{2\pi i \, 0} = 1$

$$\langle |OB|^2 \rangle = n f^2$$

$$OB = \sqrt{n}\, f \quad \text{———} \ ①$$

Can we "see" an added 4 atoms?

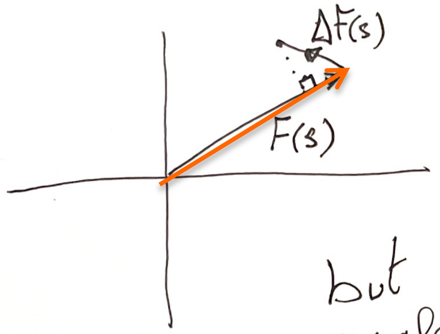
25 kDa Protein ~ 2500 atoms of f ~ 7 electrons each.

We measure $I = |F(s)|^2$

where $\langle |F(s)| \rangle \sim \sqrt{2500} \,.\, 7 = 350$ electrons

Change in $\langle |F(s)| \rangle$ from adding 4 atoms?

$\langle |\Delta \underline{F}(s)| \rangle \quad \sim \sqrt{4} \,.\, 7 = 14$ electrons

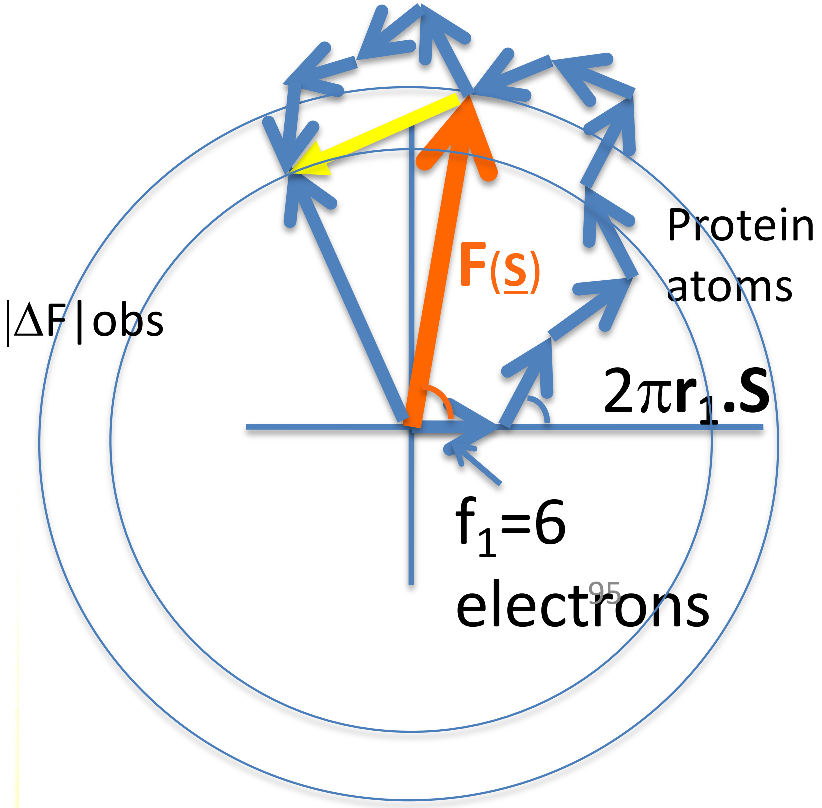


but ΔF(s) is at a 'random' angle to F(s) so the difference in |F(s)| will only be

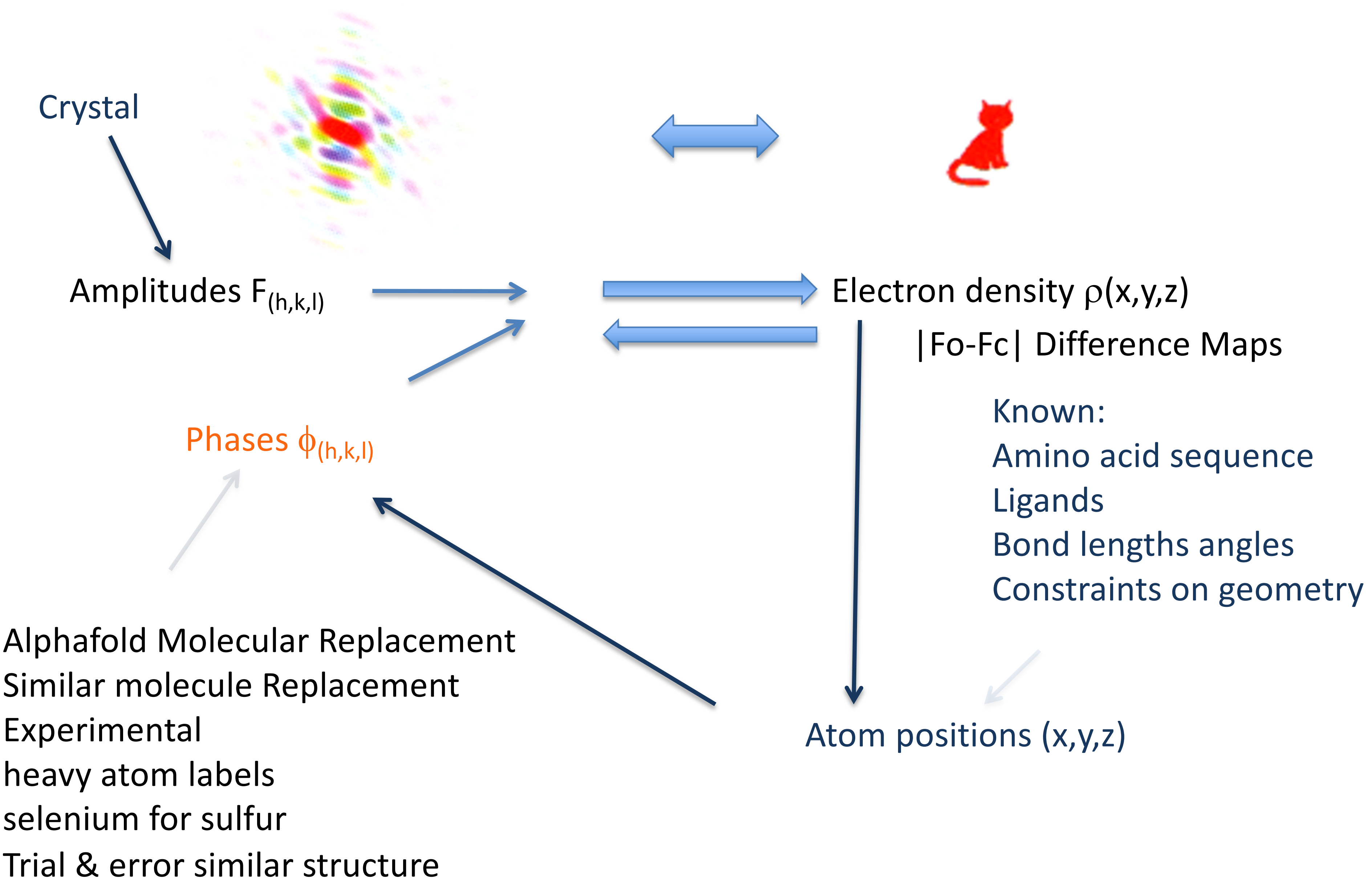
$\frac{2}{\pi} \Delta F(s) \sim 9$ electrons

so $\langle |\Delta F|_{obs} \rangle / \langle |F_{obs}| \rangle = \frac{9}{350} \sim 2.5\%$

$\langle \Delta I \rangle / \langle I \rangle = 5\%$



AXIOM: Forward FT ↔ Back FT-1 are Truly Inverse



Crystal

Amplitudes $F_{(h,k,l)}$

Electron density $\rho(x,y,z)$

|Fo-Fc| Difference Maps

Phases $\phi_{(h,k,l)}$

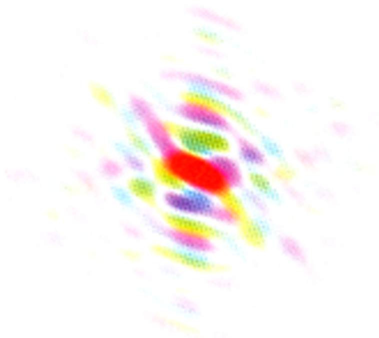
Known:
Amino acid sequence
Ligands
Bond lengths angles
Constraints on geometry

Alphafold Molecular Replacement
Similar molecule Replacement
Experimental
heavy atom labels
selenium for sulfur
Trial & error similar structure

Atom positions (x,y,z)

Change one side ↔ Change the other

Crystal



↔

Amplitudes $F_{(h,k,l)}$

Electron density $\rho(x,y,z)$

|Fo-Fc| Difference Maps

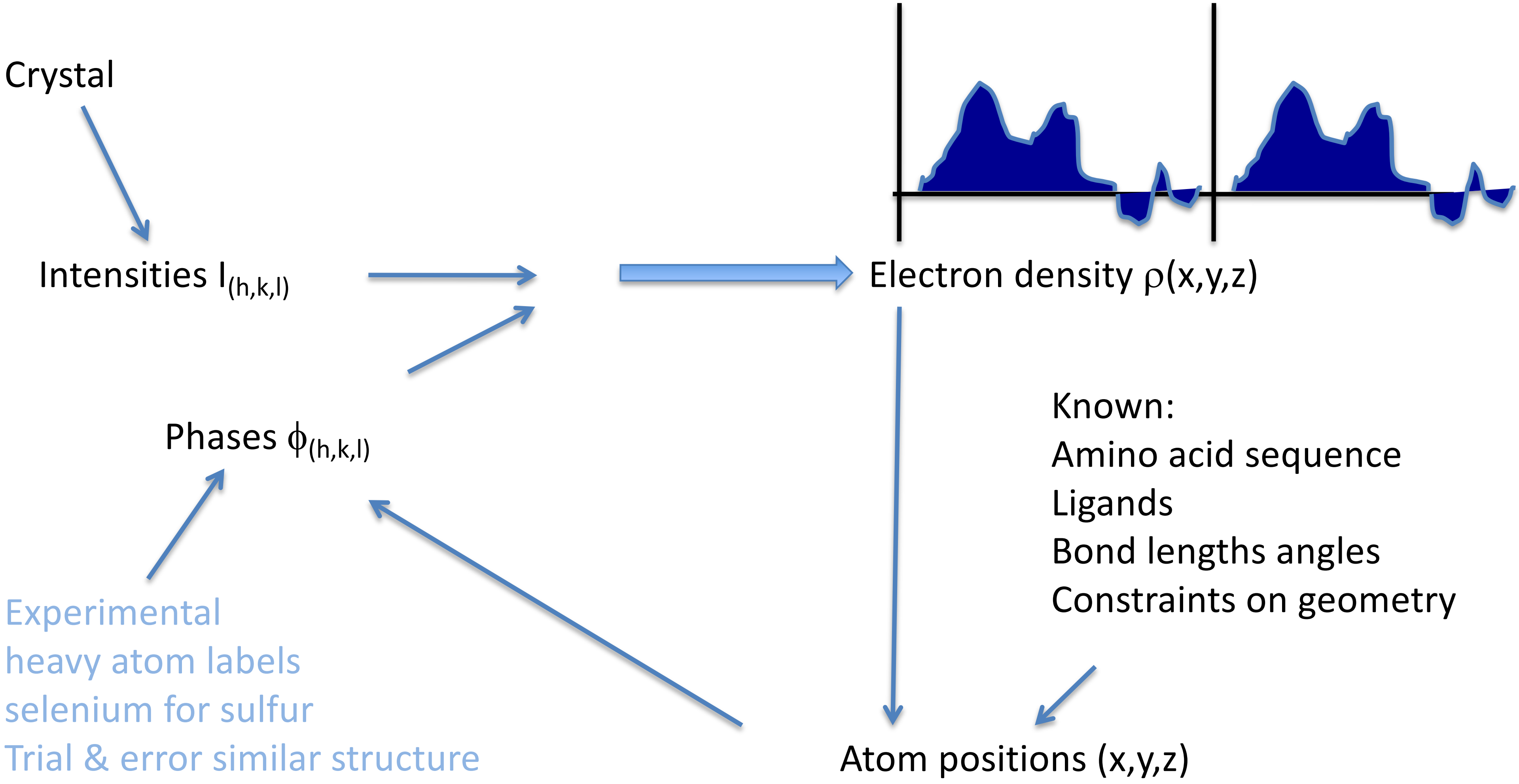
Phases $\phi_{(h,k,l)}$

Known:
Amino acid sequence
Ligands
Bond lengths angles
Constraints on geometry

Alphafold Molecular Replacement
Similar molecule Replacement
Experimental
heavy atom labels
selenium for sulfur
Trial & error similar structure

Atom positions (x,y,z)

Density Modification: Solvent Flattening



Crystal

Intensities $I_{(h,k,l)}$

Electron density $\rho(x,y,z)$

Phases $\phi_{(h,k,l)}$

Known:

Amino acid sequence

Ligands

Bond lengths angles

Constraints on geometry

Experimental

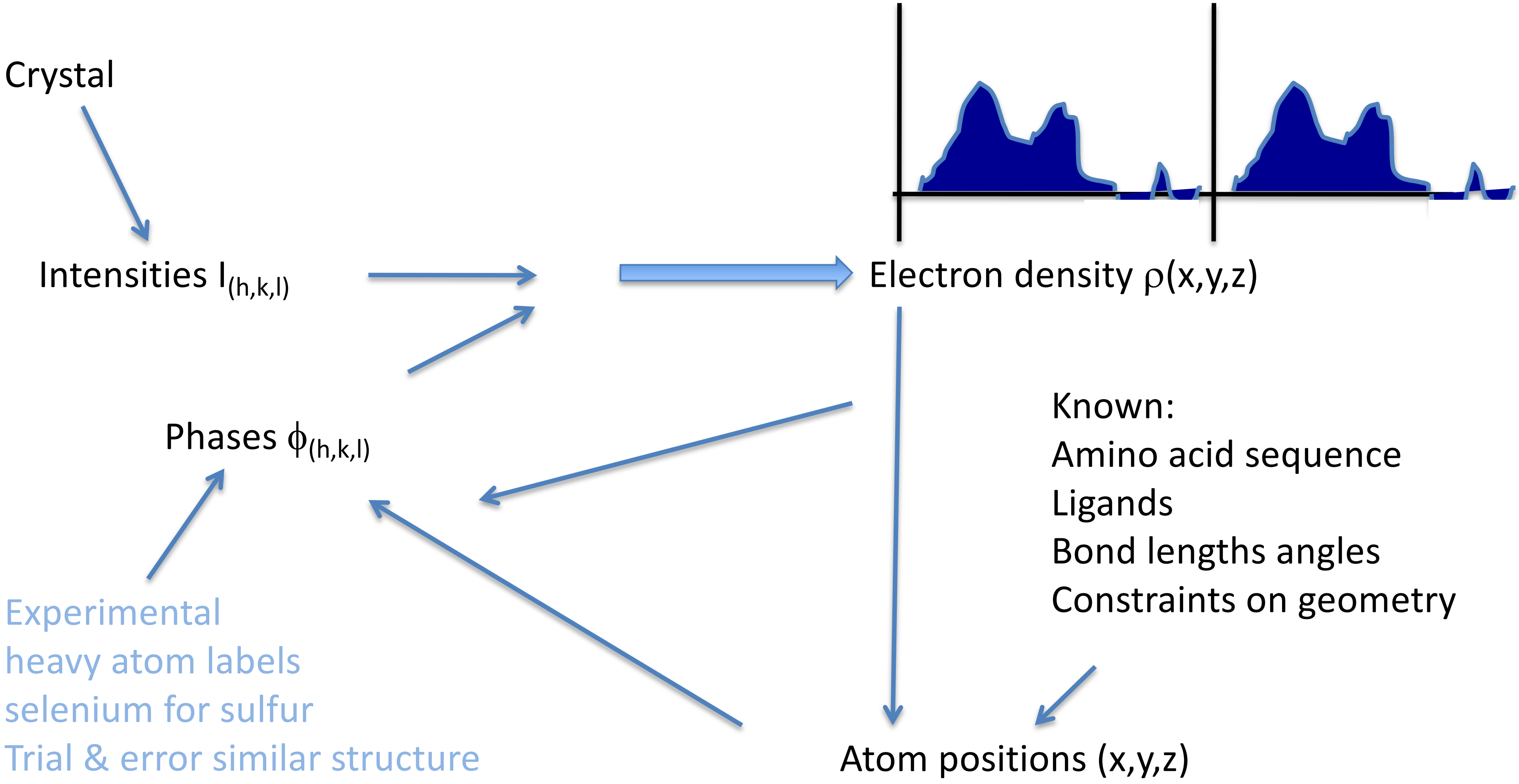
heavy atom labels

selenium for sulfur

Trial & error similar structure

Atom positions (x,y,z)

Density Modification: Solvent Flattening



Crystal

Intensities $I_{(h,k,l)}$

Electron density $\rho(x,y,z)$

Phases $\phi_{(h,k,l)}$

Known:
Amino acid sequence
Ligands
Bond lengths angles
Constraints on geometry

Experimental
heavy atom labels
selenium for sulfur
Trial & error similar structure

Atom positions (x,y,z)

Refinement

- Least Squares Refinement is common when errors in observations are presumed to be random errors that obey Gaussian statistics.
- Refine x_i,y_i,z_i, B_i with respect to the F_o

Minimize $E = \Sigma_{hkl} 1/\sigma^2(k|F_{obs}|-|F_{calc}|)^2$ with respect to $(xyzB)_i$ of all atoms.

To include an energy term, that constrains the structure toward acceptable geometry

Minimize $E = (1-w)\ \mathrm{Energy} + w\ \Sigma_{hkl} 1/\sigma^2(k|F_{obs}|-|F_{calc}|)^2$ where w is the fractional weighting on geometry versus X-ray terms. Energy has vdW, torsional restraints, bond length and dihedral angles.

Maximum Likelihood refinement seeks the most probable solution most consistent with all observations. ie Least squares refinement alone minimizes the difference between |Fo| and |Fc|.

Validation? R factors

- Use Current structure to calculate Amplitudes
- $F_{(h,k,l)}$calc and Phase $_{(h,k,l)}$calc
- Compare differences between Observed and Calculated Amplitudes

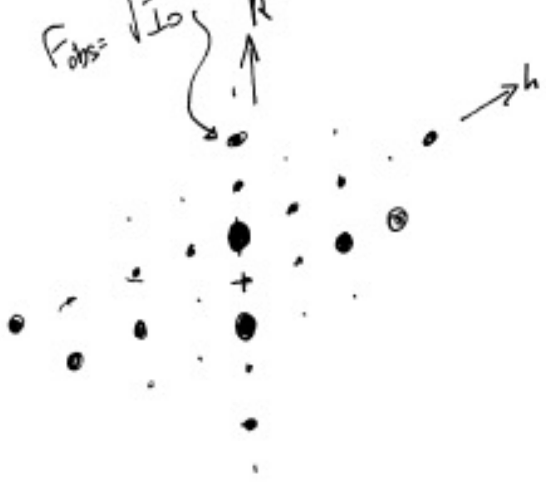
Intensity = Amplitude^2

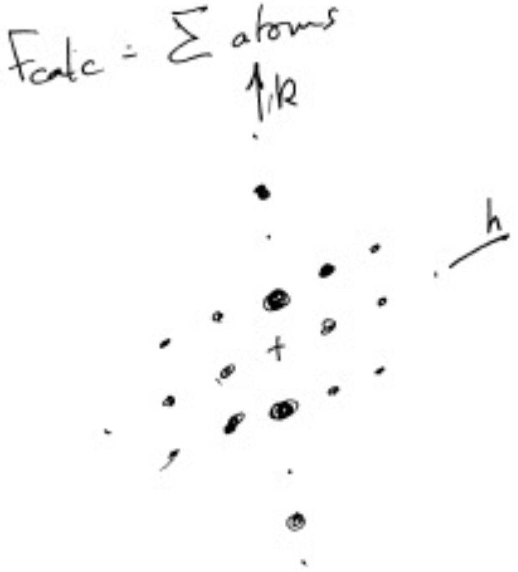
Bottom Lines: $I_{hkl} = |F_{hkl}|^2$.

- "R factor" = Agreement between Amplitudes calculated, $F_{calc_{hkl}} = \sum_j f_j e^{2\pi i(hx+ky+lz)}$ and Amplitudes observed

$$\sqrt{I_{obs_{hkl}}} = F_{obs_{hkl}}$$

$$R = \frac{\sum_{hkl} \big||F_o| - |F_c|\big|}{\sum_{hkl} |F_o|}$$





How do we judge the Quality of structure?

2. Overall quality criteria: agreement of observations with diffraction calculated from the interpreted structure.

3. Since we refine the structure To match the I_{hkl} overfitting ?

Define R_{free} for a 'hold-out ' set of observations.

4. OK? R < 20%, R free< 25%

5. But the experimental errors in measuring Fo are ~ 3%. inadequate models of solvent, atom motion, anharmonicisity

6 Accuracy ~ 0.5*res*R

Residual "R" factors

R_{cryst} (or just "R") How well does structure match the observations?

observed vs calculated data (F(s))

R_{free} Remove bias; leave some observations out of determination

cross-check with "random" subset of data

should be < 0.3 and $< R_{cryst} + 0.1$

R_{sym} = R_{merge} How self consistent are observations that should be identical? = measuring errors.

(self-consistency of data: Is)

“R” factors

$$R = \frac{\sum \left| F_{obs} - F_{calc} \right|}{\sum F_{obs}}$$

completely random: 0.59

starting MR solution: 0.4-0.55

something still wrong?: > 0.3

correct chain trace: < 0.2

small molecule: ~ 0.05

“R” factors

R_{cryst} (or just “R”)

observed vs calculated data (Fs)

R_{free}

cross-check with “random” subset of data

should be < 0.3 and $< R_{cryst} + 0.1$

R_{merge}

$$R_{merge} = \frac{\sum \left| I_{obs} - \langle I \rangle \right|}{\sum I_{obs}} \longleftarrow$$

blows up as $I_{obs} \to 0$

completely random: 0.59

weak data (high angle): 0.7- ∞

wrong symmetry choice?: ~0.2-0.55

small or disordered crystal: ~0.1-0.2

typical: ~ 0.05

An optical Diffraction demonstration

Compute a Transform of a series of 50%b/50%w parallel lines

- Ronchi ruling: