With slides from Silvia Onesti

X-ray biopolymer crystallography: an overview

Metodi sperimentali per determinare la struttura delle macromolecole

Alta risoluzione	cristallografia a raggiX | NMR | cristallografia elettronica
Media risoluzione	microscopia elettronica | diffrazione di fibre | Spettro-metria di massa | SAXS | Microscopia a Forza Atomica
Metodi Spettroscopici	NMR | dicroismo circolare | Assorbanza | Fluorescenza | Fluorescenza anisotropa | diffusione della luce
Metodi Chimici	scambio H-D | mutagenesi sito specifica | modificazioni chimiche | proteomica
Metodi termodinamici	Equilibrio di (un)folding
Metodi computazionali	predizione della struttura delle Proteine | Docking molecolare

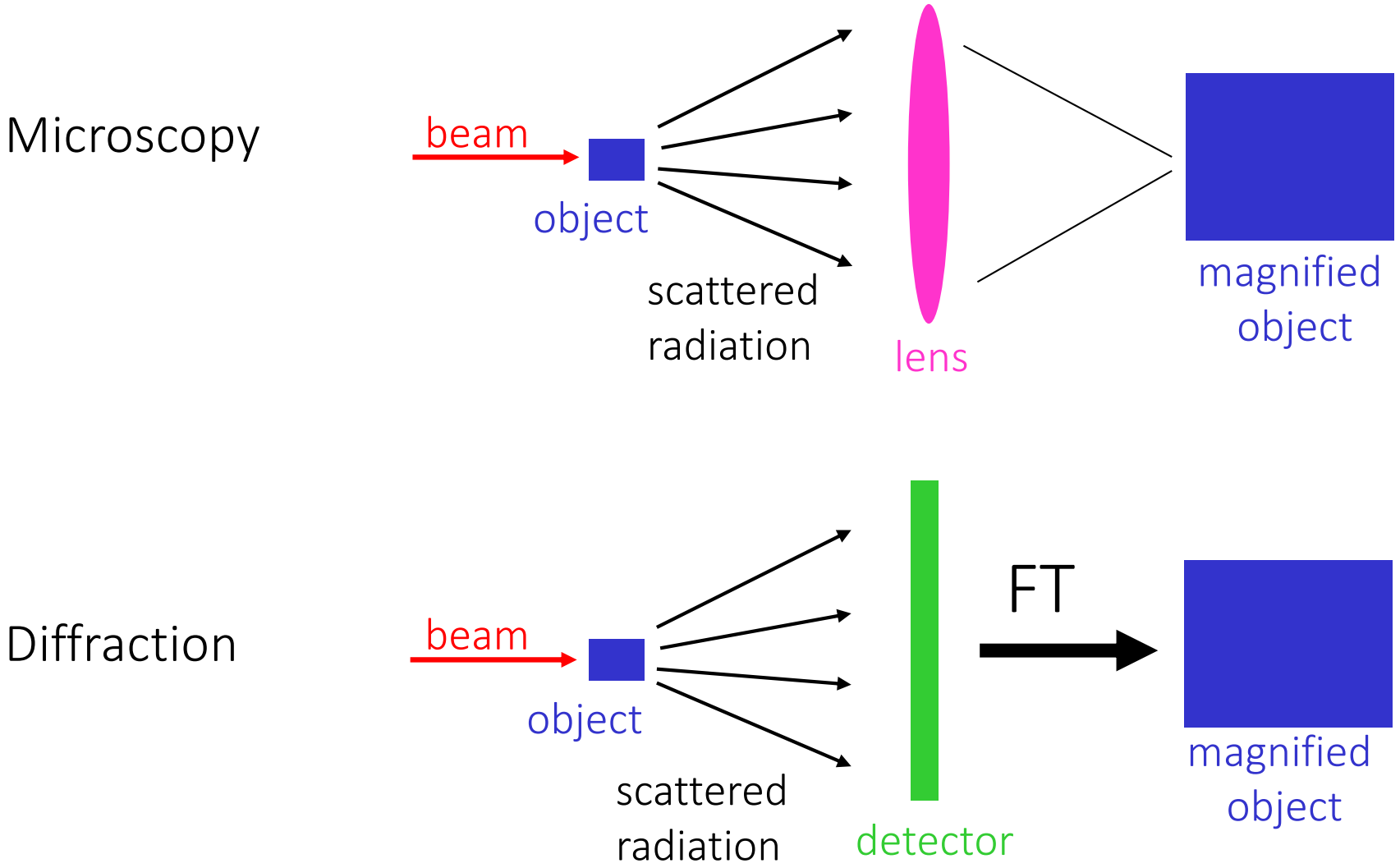
Structure determination methods

- X-ray crystallography (protein crystallography)
- X-ray fiber diffraction
- Small angle X-ray scattering (SAXS)

- Scanning electron microscopy (SEM)
- Transmission electron microscopy (TEM):
  - 2D crystals (electron diffraction)
  - Fibers/helices/tubular crystals
  - Single particle electron microscopy
- Atomic Force Microscopy (AFM)

- Nuclear magnetic resonance (NMR)

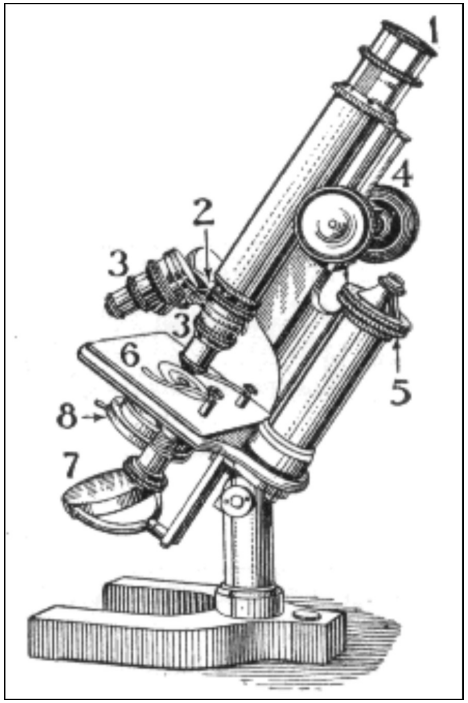
Microscopy vs diffraction

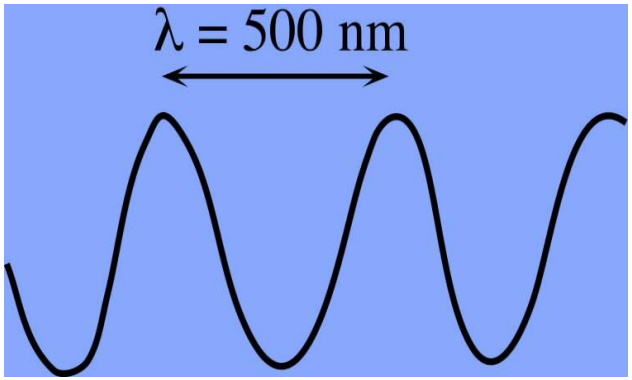


Why can't we use a microscope?

Normally, to look at small objects we use microscopes...

...but they can only provide images of things larger than the wavelength of light





Protein (10 nm)

Atom (0.1 nm=1Å)

In theory, we could use X-rays (light of λ=0.1 nm, the right size for looking at atoms)...

X-rays

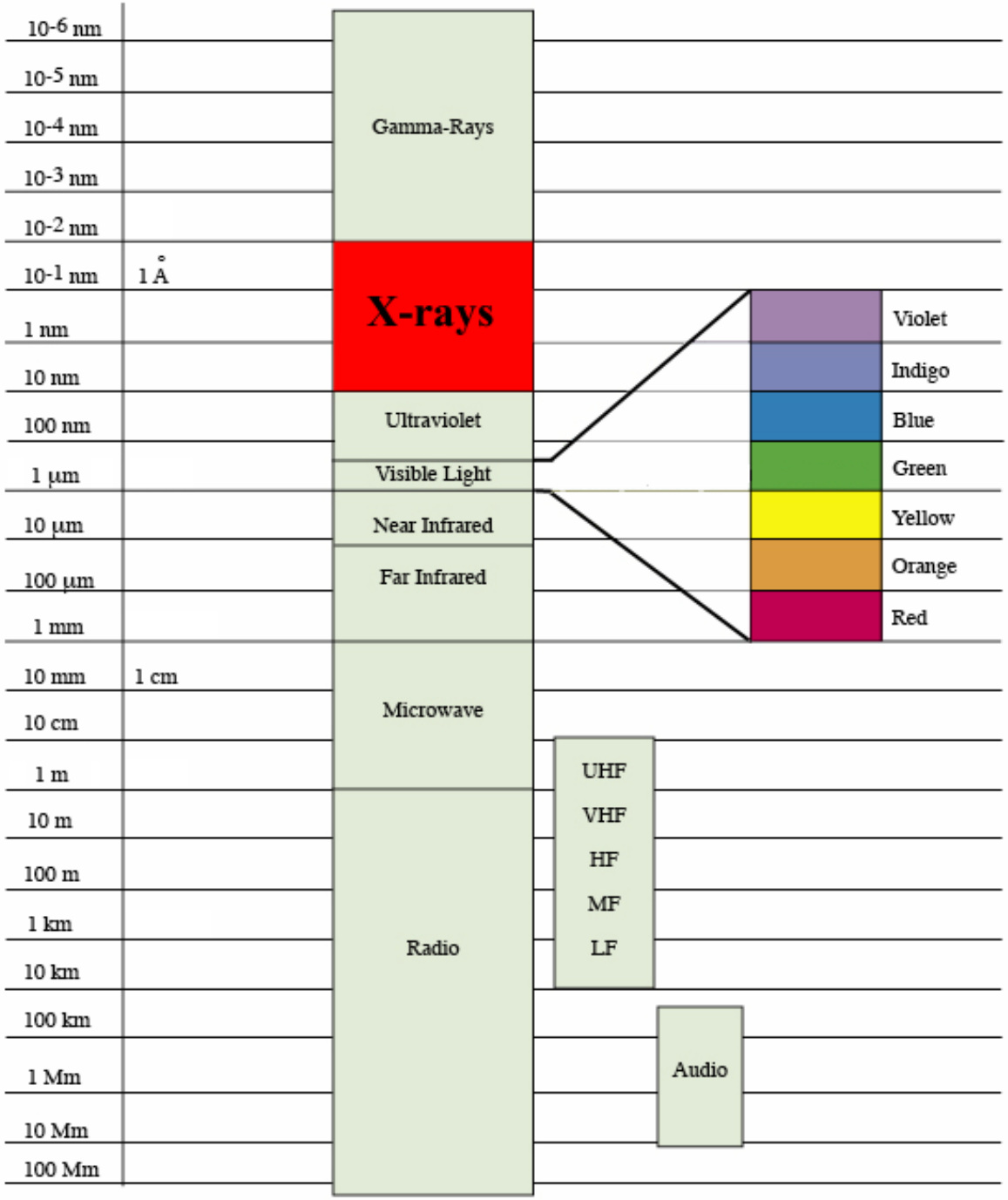
X-rays:
E = 1-100 keV
λ = 10 nM-0.01 nm

$E=h\nu$

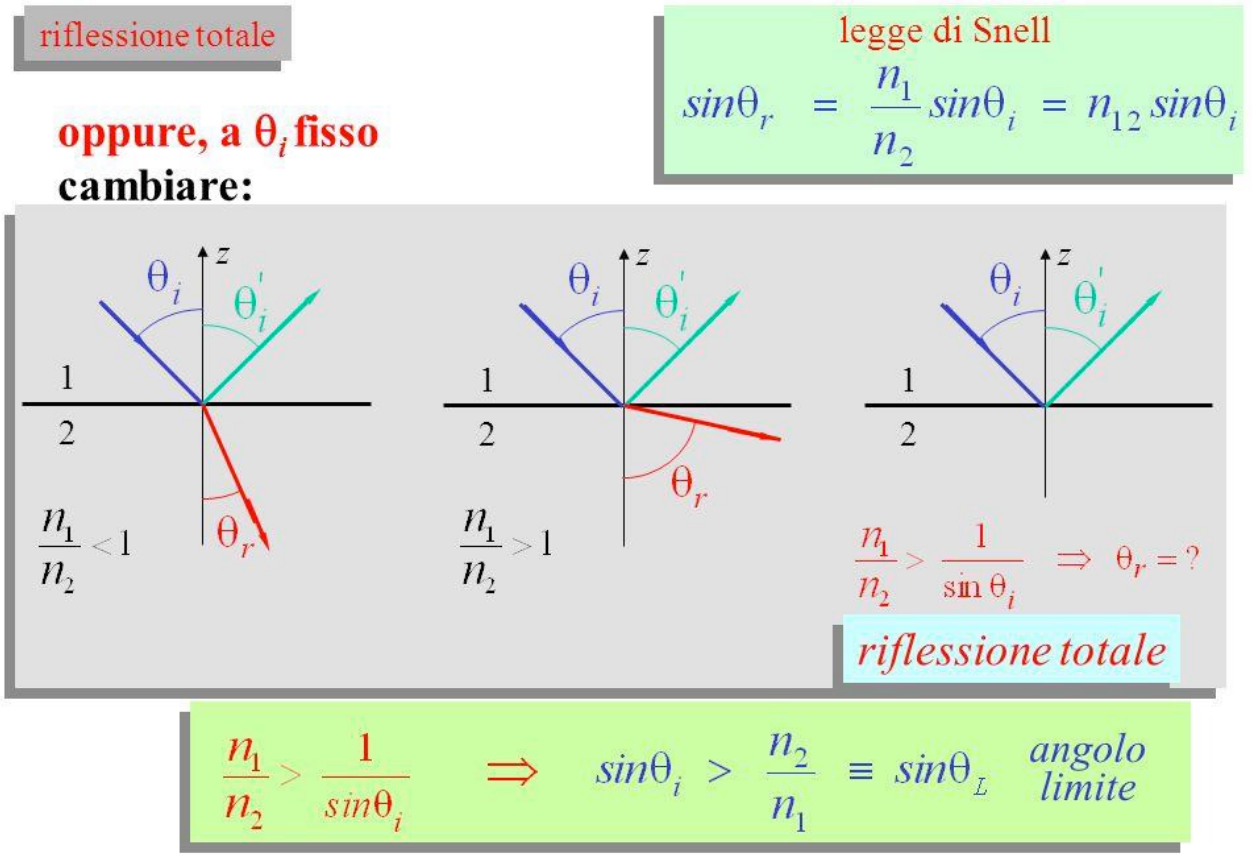
$\lambda=c/\nu$

Soft X-rays (lower E) are used in medical imaging.
Hard X-rays (higher E) are used in crystallography.

...but we can't make X-ray microscopes as hard X-rays do not reflect or refract easily.



X-rays

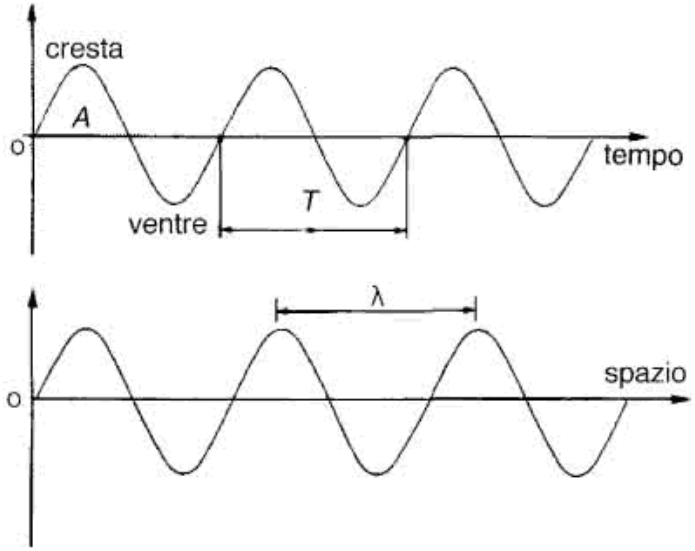


X-rays change their speed only little when moving from one medium to another

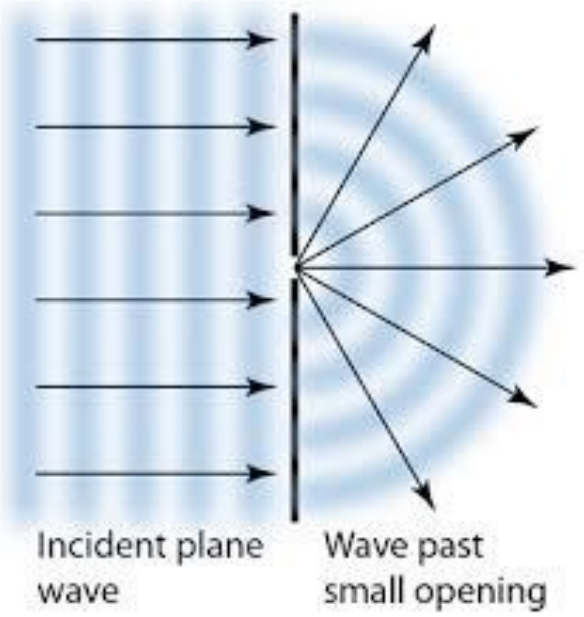
Interaction of radiation with matter

$A(t)=A_0\cos(\omega t)$
$B(t)=B_0\cos(\omega t+\varphi_0)$

According to $\varphi = \omega t+\varphi_0$ the two waves can be in phase or not.
Coherence: phase difference maintained, either vs. t or space



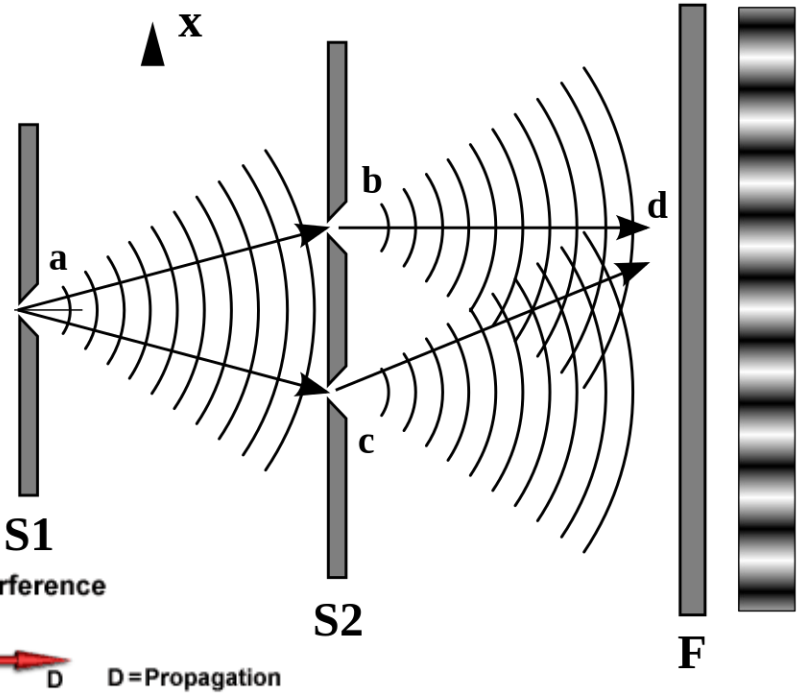
Huygens' principle:
When a plane wave hits a slit, each point of the slit is a source of a spherical wave



Interference between coherent waves

Interference between waves generated by two slits:

It can occur if the distance between the slits is comparable with λ: two spherical waves coherent generated from b and c



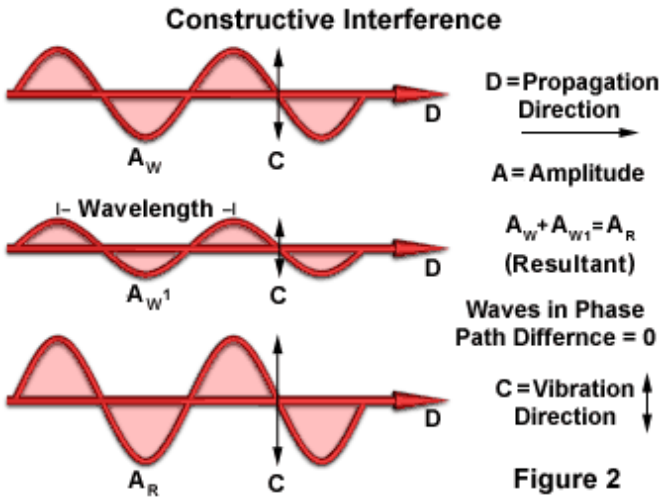


Figure 2

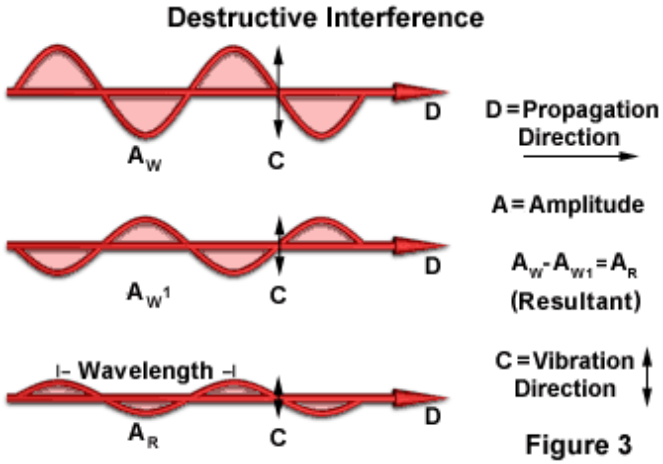
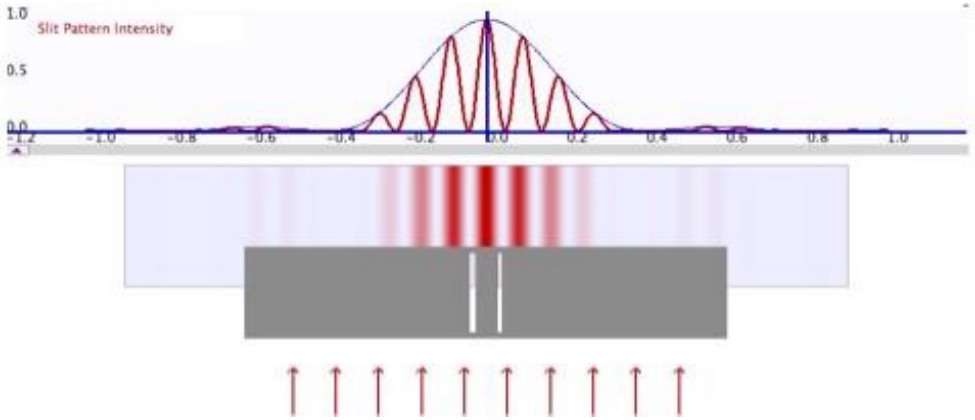
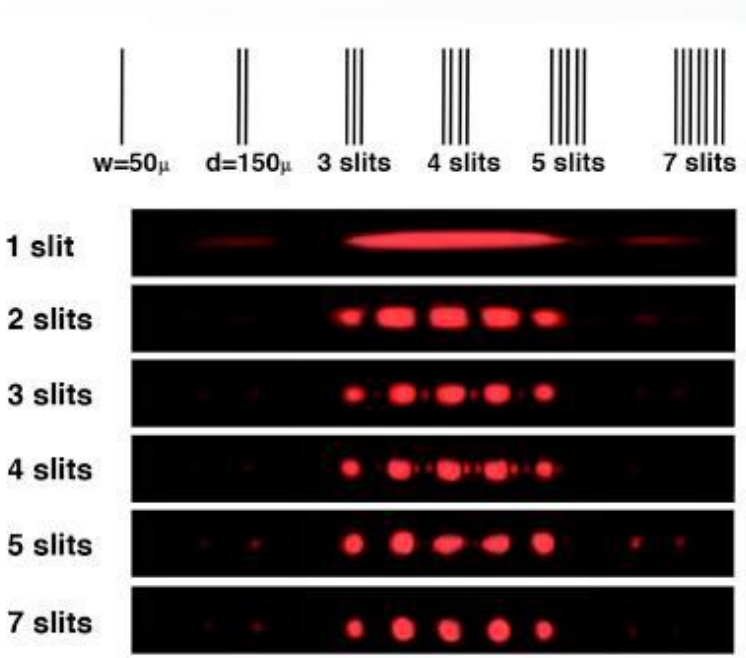


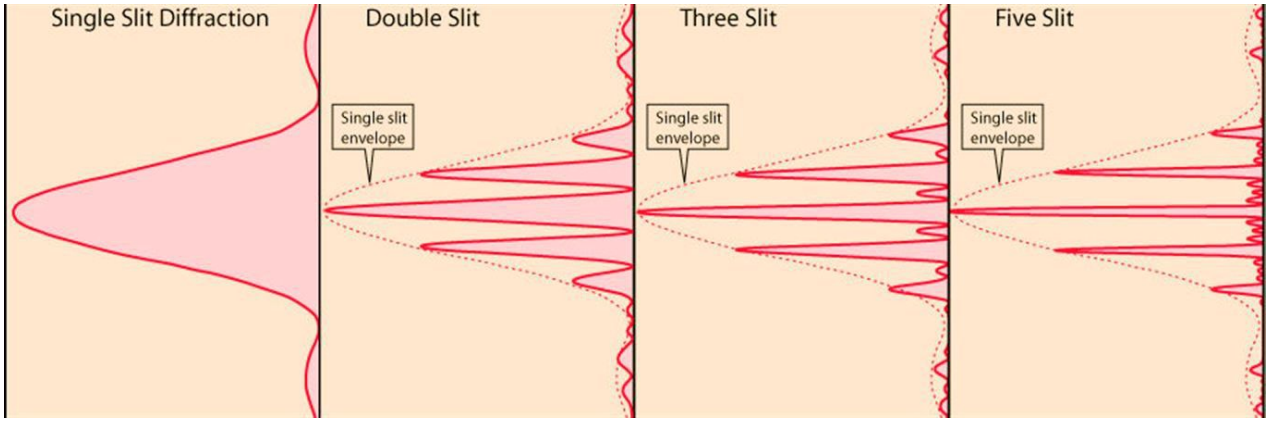
Figure 3

Dark and light zones depending on the direction of observation

Interference between coherent waves



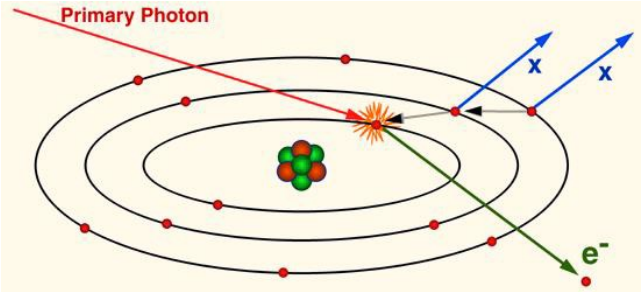




More intensity, diffraction spots

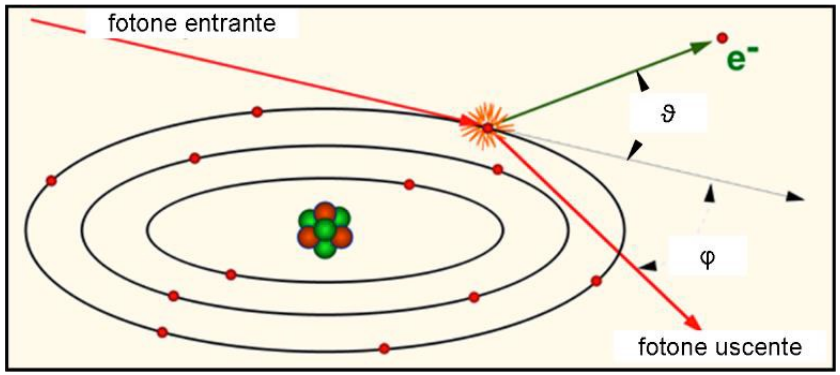
Interaction of x-rays with matter

Photoelectric effect



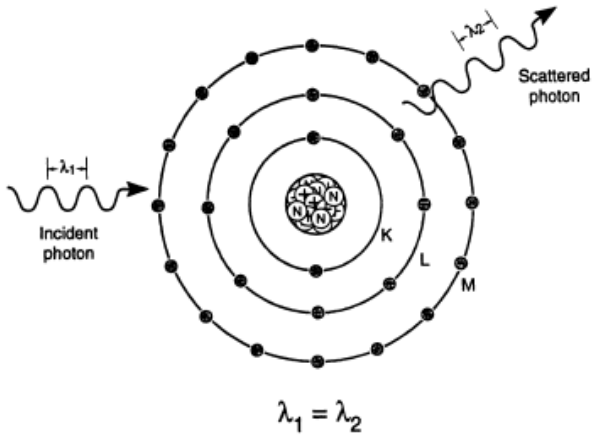
Soft X-rays, E comparable with the energies of deep electrons in the atoms (tens-hundreds eVs)

Compton Scattering



Inelastic: radiation partially absorbed by a valence electron and re-emitted at lower E

Elastic Scattering (diffusion)



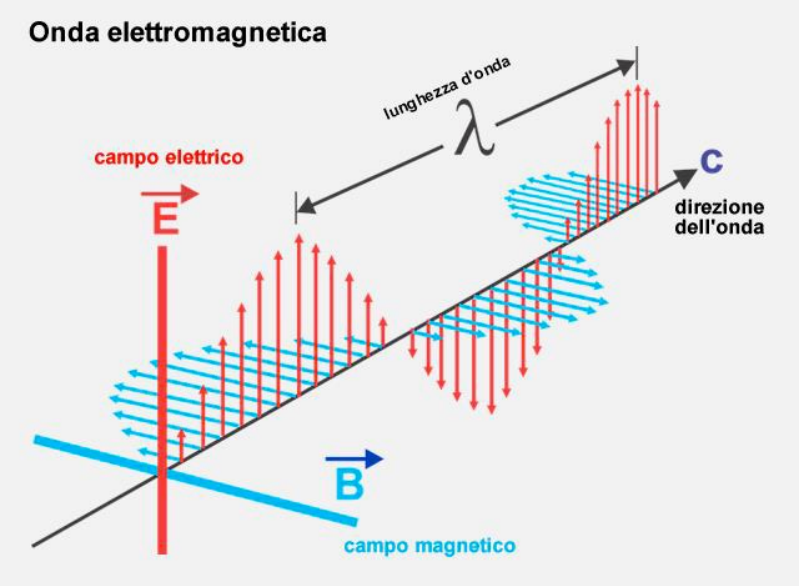
$\lambda_1 = \lambda_2$

Elastic (Rayleigh): radiation re-emitted with the same E ----DIFFRACTION IN CRYSTALS

Interaction of x-rays with matter

The solution of Maxwell equation (propagation of EM field) are the electromagnetic waves. For the E field:

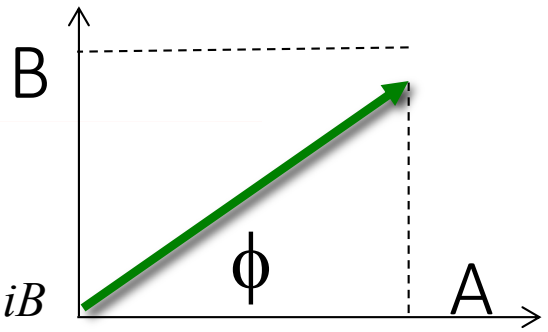
$E(t,x)=E_0 cos[2\pi\nu(t-x/c)]$



Introducing complex numbers (and Eulero formula):

$$E(t,x) = E_0\, exp[2\pi\nu i\left(t-\frac{x}{c}\right)]$$

$$\mathbf{Z} = |Z|\exp(i\phi) = |Z|(\cos\phi + i\sin\phi) = A + iB$$



X-ray scattering by a single electron (Thompson)

$\boldsymbol{F=eE}_i$

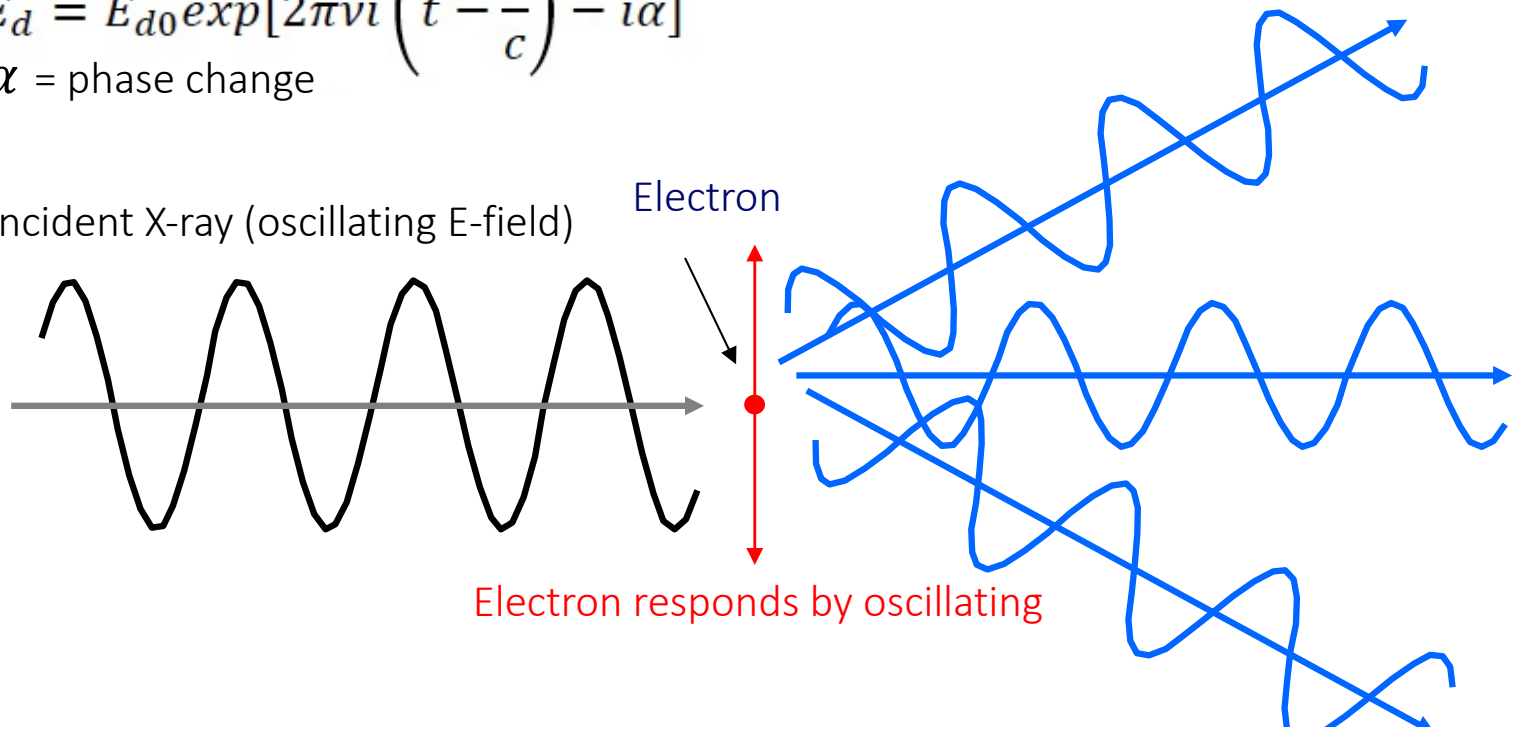
force exerted by the electric field E on the charge e, which cause an oscillatory motion with frequency equal to the incident wave, and acceleration:

$$\vec{a} = \frac{\vec{F}}{m} = e\frac{\overrightarrow{E_i}}{m}$$

Oscillating electron emits X-rays over a wide angle

$$\vec{E}_d = E_{d0} exp[2\pi\nu i\left(t - \frac{\vec{r}}{c}\right) - i\alpha]$$

α = phase change



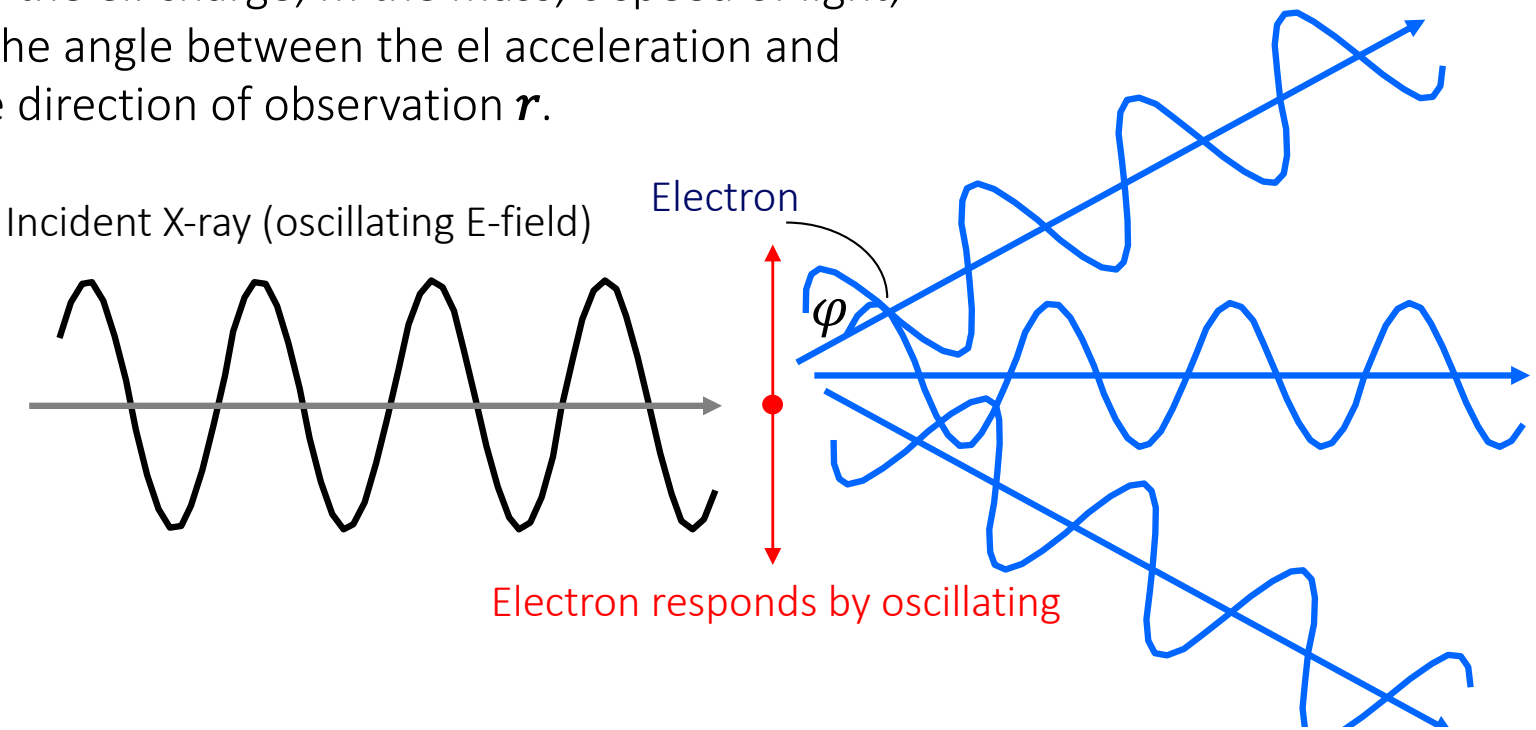
X-ray scattering by a single electron (Thompson)

$$\vec{E}_d = E_{d0} exp[2\pi\nu i\left(t - \frac{\vec{r}}{c}\right) - i\alpha]$$

According to Thompson:

$$\boldsymbol{E_{d0} = \frac{1}{r} E_{i0} (\frac{e^2}{mc^2}) \sin \varphi}$$

e is the el. charge; m the mass; c speed of light;
φ the angle between the el acceleration and
the direction of observation $\boldsymbol{r}$.



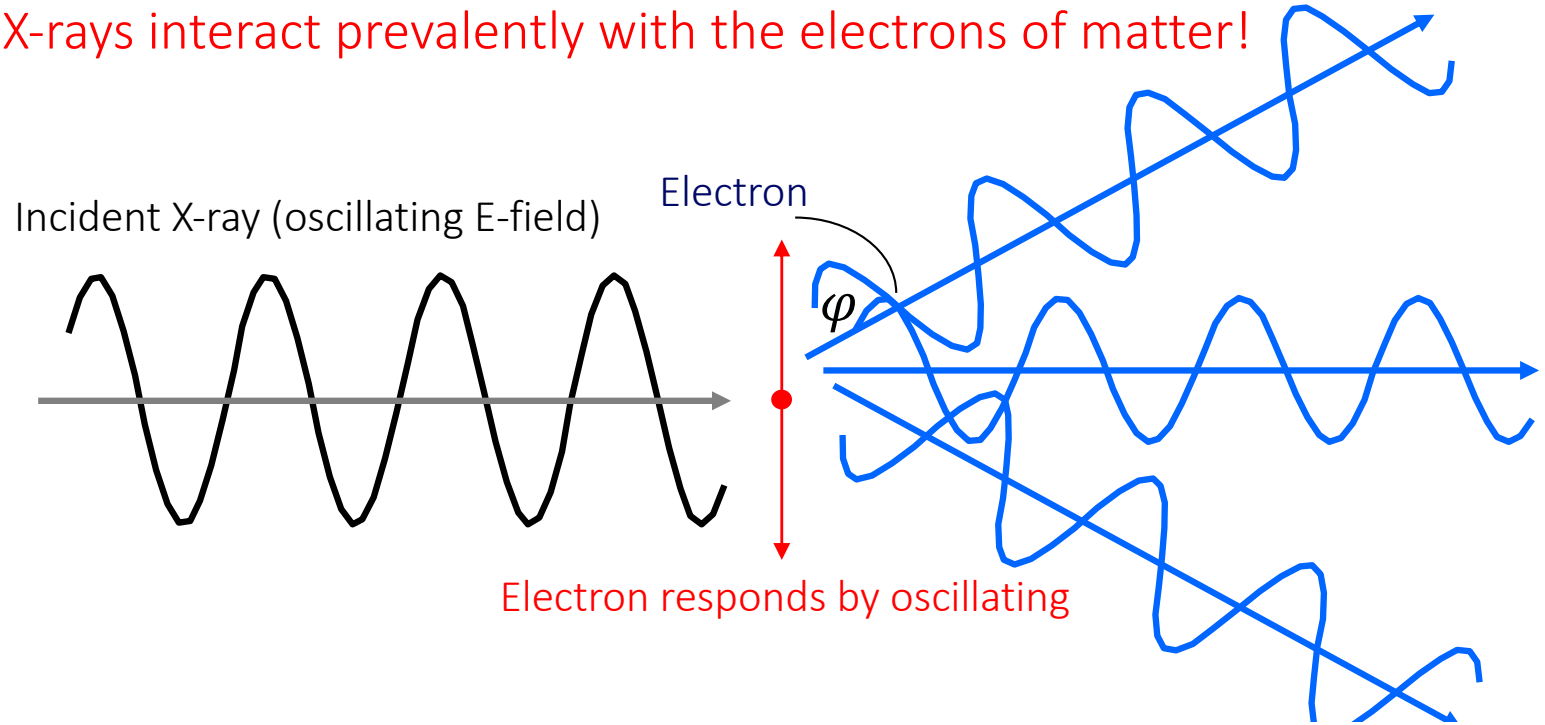
X-ray scattering by a single electron (Thompson)

$I = |E|^2$

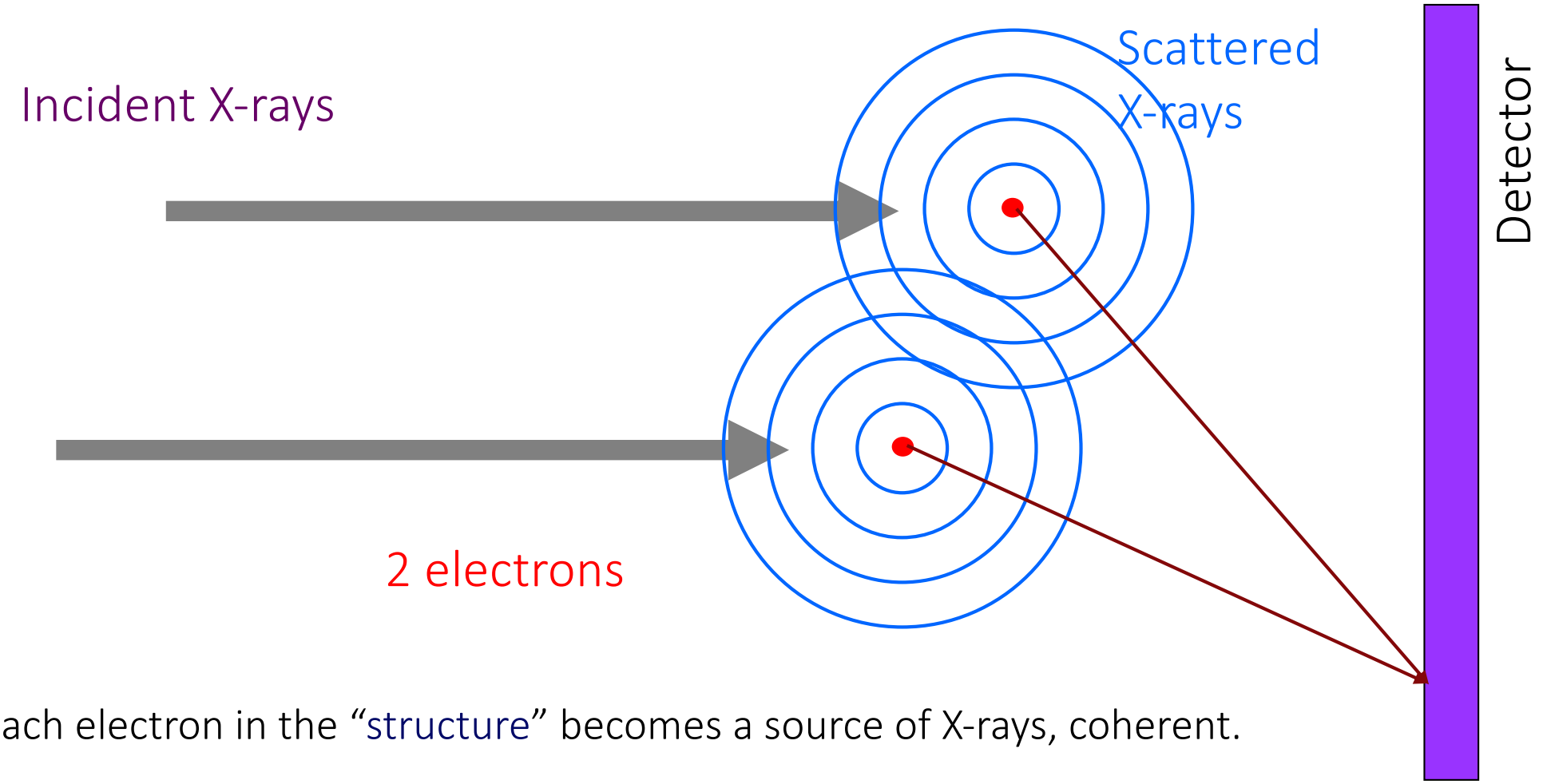
$$I_{eTh} = I_{d0} = |E_{d0}|^2 = \left|\frac{1}{r} E_{i0} \left(\frac{e^2}{mc^2}\right) \sin\varphi\right|^2 = I_{i0} \frac{e^4}{m^2 r^2 c^4} \sin^2\varphi$$

I_{eTh} depends on e/m. , 1837 times fainter for a proton than an electron!

X-rays interact prevalently with the electrons of matter!



X-ray scattering by two electrons



Each electron in the “structure” becomes a source of X-rays, coherent.

Diffused, coherent waves from different electrons interfere.

The diffraction pattern (which would be observed on a detector) depends on how these sacattering centres are arrayed, I.e. on the structure.

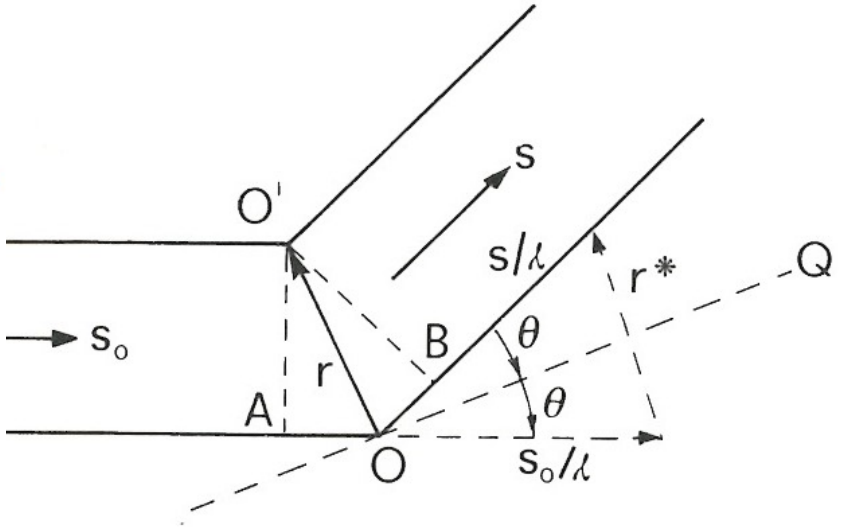
X-ray scattering by two electrons

Conditions for interference depend on the difference between the optical path of the two waves:

$$\overline{OA} + \overline{OB} = differenza\ di\ cammino\ ottico$$

$\overline{OA} = -\vec{r} \cdot \overrightarrow{s_o}$
$(proiezione\ di\ \vec{r}\ lungo\ \overrightarrow{s_o};\ segno$
$-\ perchè\ opposto\ alla\ direzione\ di\ \vec{s})$

$\overline{OB} = \vec{r} \cdot \vec{s}\ (proiezione\ di\ \vec{r}\ lungo\ \vec{s})$



Esprimendo la differenza di cammino ottico come **differenza di fase**:

$$\delta = 2\pi \frac{\overline{OA} + \overline{OB}}{\lambda} = 2\pi \frac{\vec{r} \cdot (\vec{s} - \overrightarrow{s_0})}{\lambda}$$

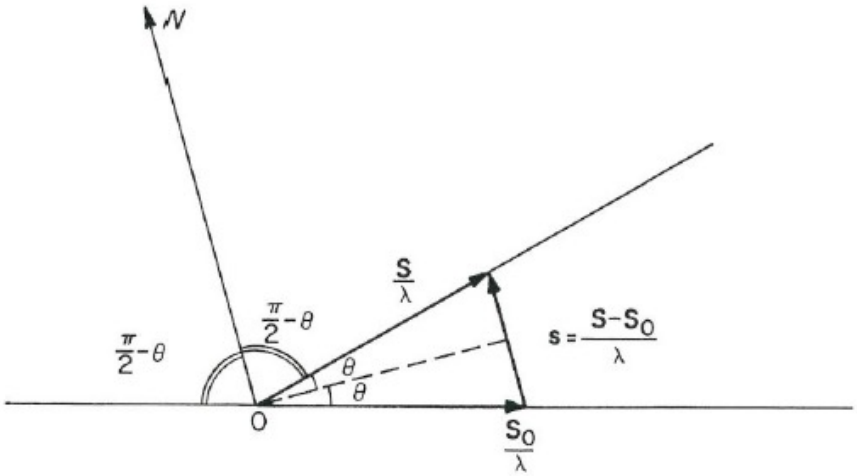
è pari al numero di 'periodi'

X-ray scattering by two electrons

$$2\pi \frac{\vec{r} \cdot (\vec{s} - \vec{s_0})}{\lambda}$$

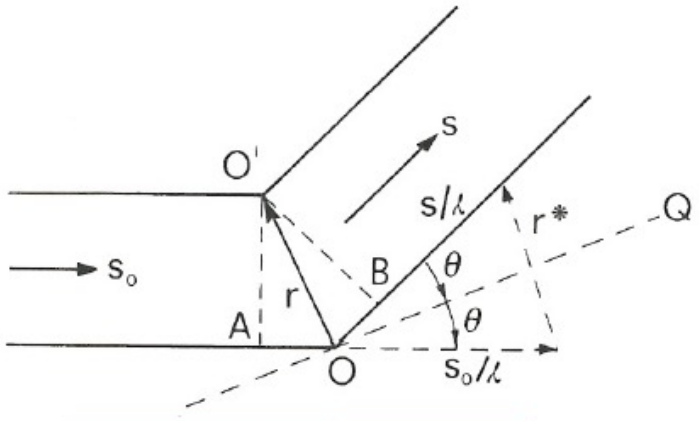
$$\vec{r}^* = \frac{(\vec{s} - \vec{s_0})}{\lambda}$$

$$\delta = 2\pi \vec{r} \cdot \vec{r^*}$$



r* is perpendicular to the **s** and s_0 bisector, forming an angle θ. Its module:

$$|\vec{r^*}| = \frac{2sin\theta}{\lambda}$$



r and λ should have the same order of magnitude

X-ray scattering by two electrons

$$\vec{E}_d^1 = E_{d0} exp[2\pi\nu i\left(t - \frac{\vec{r}}{c}\right) - i\alpha]$$

$$\vec{E}_d^2 = E_{d0} exp[2\pi\nu i\left(t - \frac{r}{c}\right) - i\alpha + i\delta]$$

$$\vec{E}_d^2 = \boxed{E_{d0} exp\left[2\pi\nu i\left(t - \frac{\vec{r}}{c}\right) - i\alpha\right]} \exp(i\delta) \leftarrow \vec{E}_d^1$$

$$\boldsymbol{\vec{E}_d^2 = \vec{E}_d^1 exp(i\delta)}$$

$$\vec{E}_d = \vec{E}_d^1 + \vec{E}_d^2 = \vec{E}_d^1[1 + \exp(i\delta)]$$

$$\boldsymbol{\vec{E}_d(r^*) = \vec{E}_d^1[1 + exp(2\pi i\vec{r} \cdot \vec{r^*})]}$$

An additive term which determines if the waves are summed up or sottracted

X-ray scattering by N electrons

Se abbiamo N elettroni, tutti uguali nelle posizioni $r_1, r_2, \ldots, r_N$ dal primo elettrone (r_1 sarà 0, origine del sistema di riferimento), e applicando lo stesso procedimento utilizzato per 2 elettroni, otteniamo:

$$\vec{E}_d(r^*) = \vec{E}_d^1 \sum_{j=1}^{N} \exp\left(2\pi i \vec{r_j} \cdot \vec{r^*}\right)$$

Dove r* denota la 'direzione di osservazione', mentre r_j denota la posizione del diffusore j-simo

Più in generale, ammettendo che i diffusori abbiamo una carica generica (non necessariamente quella di 1 elettrone), allora ogni diffusore genererà onde di ampiezza E_0 non necessariamente uguali tra loro (maggiore è la carica maggiore maggiore è l'ampiezza dell'onda diffusa,vedi formula di Thomson) e che indicheremo cone $\boldsymbol{A_j}$

Potremo così scrivere:

$$\vec{E}_d(r^*) = \sum_{j=1}^{N} A_j \exp\left(2\pi i \vec{r_j} \cdot \vec{r^*}\right)$$

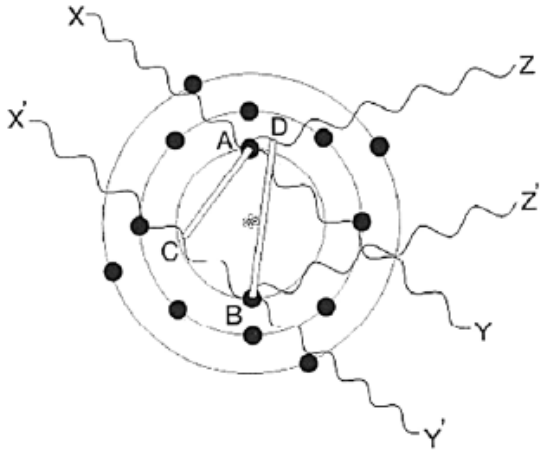


Figure 3.12. The scattering of X-rays from a real atom extended in space.

X-ray scattering by N electrons

$$\vec{E}_d(r^*) = \sum_{j=1}^{N} A_j \exp\left(2\pi i \vec{r_j} \cdot \vec{r^*}\right)$$

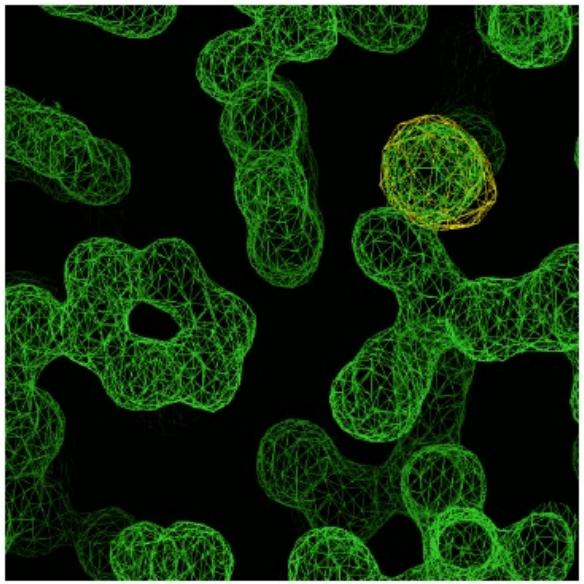
If instead of amplitude A we define the ratio between the intensity of the diffused wave and the one diffused by a single electron (Thomson):

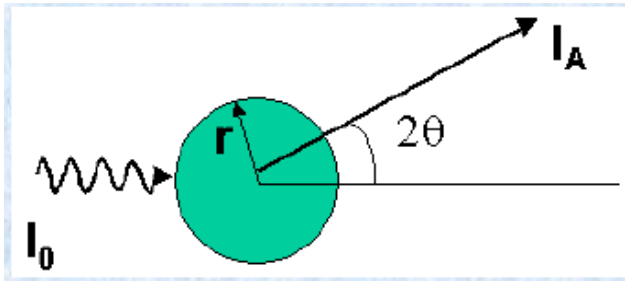
$$\vec{F}(r^*) = \sum_{j=1}^{N} f_j \exp\left(2\pi i \vec{r_j} \cdot \vec{r^*}\right)$$

$$f^2{}_j = \frac{I_j}{I_{Th}}$$

f_j is the **Structure Factor** of the diffusor j

Diffusion from a charge density distribution





$\boldsymbol{\rho(r)}$ charge density distribution

$\boldsymbol{\rho(r)dr}$ infinitesimal charge element with:

$\boldsymbol{2\pi i r_j \square r^*}$ phase difference:

$$\boldsymbol{\rho(\vec{r})\, exp\left(2\pi i \vec{r_j} \cdot \vec{r^*}\right)}$$

Integrating over the space of the charge density with volume V:

$$\vec{F}(\vec{r^*}) = \int_V \rho(\vec{r}) \exp\left(2\pi i \vec{r_j} \cdot \vec{r^*}\right) d\vec{r}$$

Structure Factor

Which is the Fourier transform of $\boldsymbol{\rho(r)}$ integrated over V : $\boldsymbol{F(r*)}$=FT[$\boldsymbol{\rho(r)}$]

X-ray scattering by N electrons

FT put two spaces in relation:

the **direct space** (r) and the **reciprocal space** ($r*$).

La $\boldsymbol{\rho}(\boldsymbol{r})$ is defined in the direct space, **F(r*)** in the reciprocal space

The scattered wave result of the interference of x-rays with a charge distribution, along a direction defined by r*, is the FT of the charge distribution itself.

FT [FT[$\boldsymbol{\rho}(\boldsymbol{r})$]] gives $\boldsymbol{\rho}(\boldsymbol{r})$: **from the phase and module of diffused waves we can determine the charge distribution in the real space**

Summary

- Electrons diffuse x-Rays
- diffuse waves interfere
- the sum of diffuse waves depends on the difference of optical path which is function of charge distribution in the real space
- for a charge density distribution, we introduce infinitesimal volumes, each acting as a unit charge
- the total diffuse wave is the Fourier Transform of the charge distribution

Atomic form factor

$\boldsymbol{\rho_{el}(r)}$ function of $|\psi_r|^2$

$$\overrightarrow{f_{el}}(\overrightarrow{r^*}) = \int_V \rho_{el}(\vec{r}) \exp(2\pi i \vec{r_j} \cdot \overrightarrow{r^*})\, d\vec{r}$$

is the **electronic scattering function**. For one atom with N electrons we define the **atomic scattering function**:

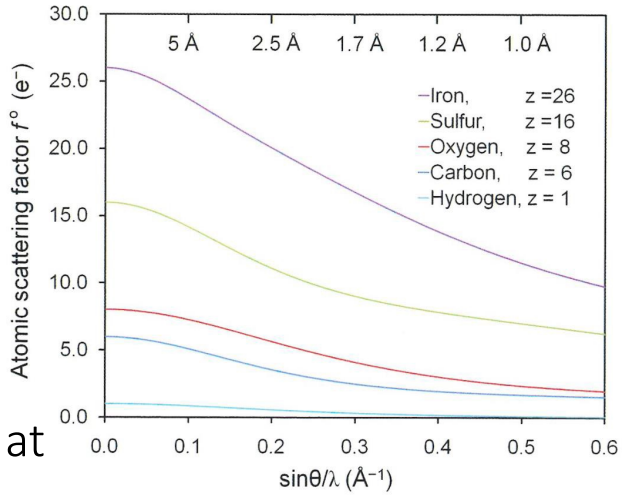
$$f_{at}(r^*) = \sum_{i=1}^{N} f_i^{el}$$

and the intensity diffused by an atom is:

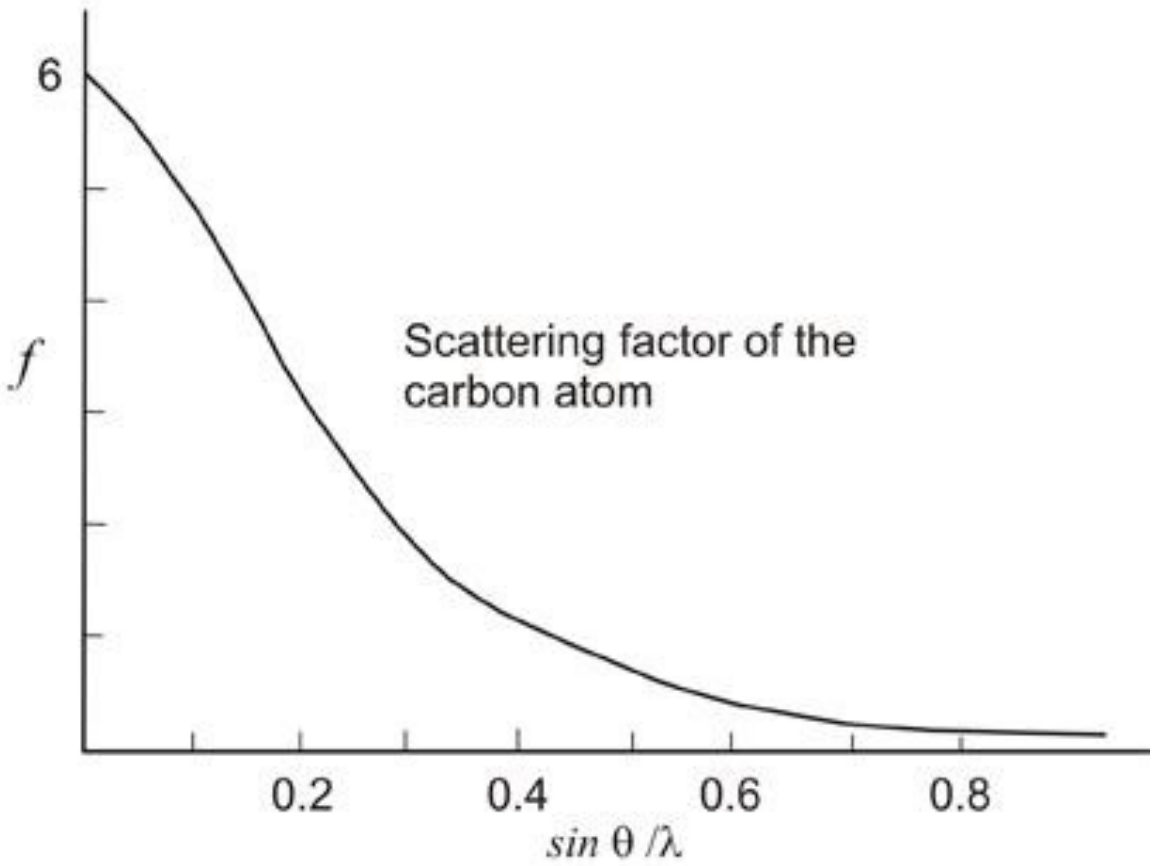
$$\boldsymbol{I_{at} = f_{at}^{\,2}}$$

Which depends on $|r^*|=2sin\theta/\lambda$

The more efficiently one electron scatters, the higher the θ at which diffraction is measured!



Atomic form factor

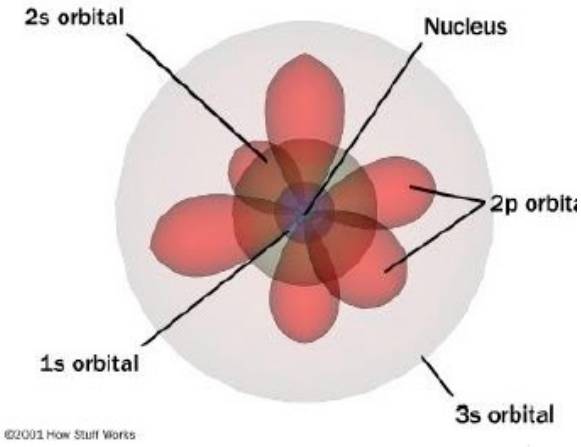


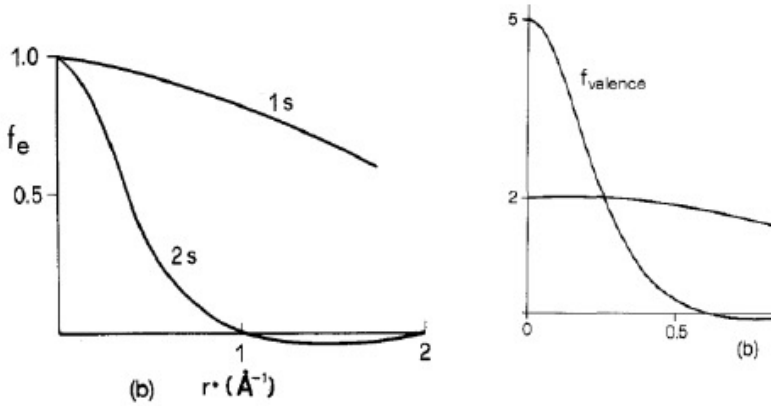
$$f_{at}(r^*) = \sum_{i=1}^{N} f_i^{el}$$

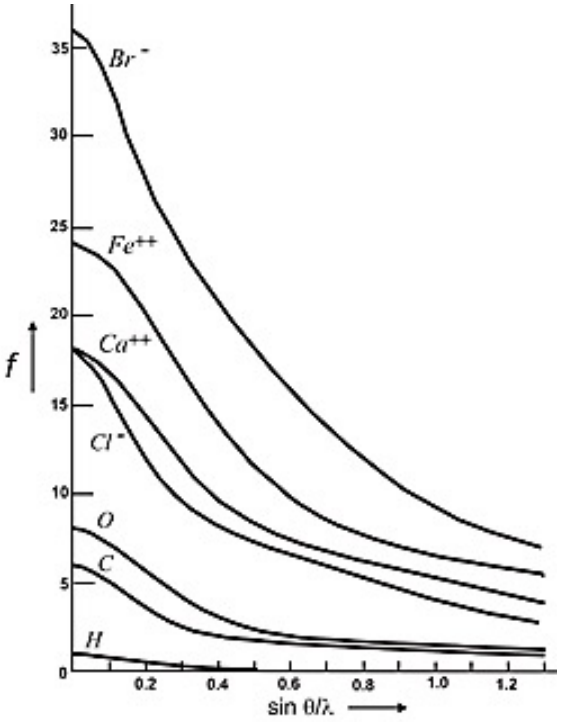
$$f_{at}(0) = \sum_{i=1}^{N} 1 = Z$$

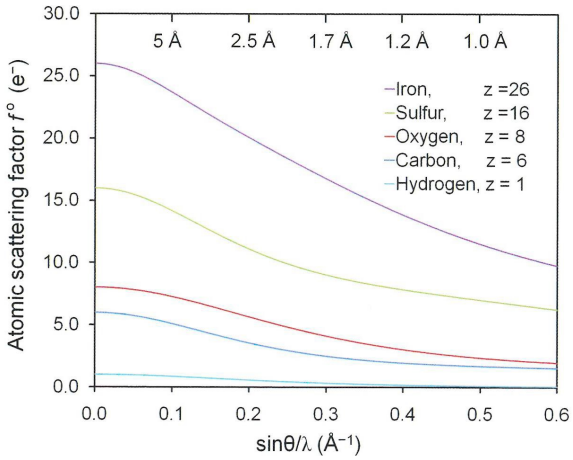
At r* = 0, or ϑ=0, the atomic form (or scattering) factor is Z, the atomic number!

Atomic form factor









The higher Z, the higher the diffusion of X-rays from that atom!

Molecular scattering factor

$$\rho_{mol}(r) = \sum_{j=1}^{N} \varrho(r - r_j)$$

$\rho(r\text{-}r_j)$ r_j is the distance from the origin

$$\rho_{mol}(r) = \sum_{j=1}^{N} \varrho(r - r_j) \longrightarrow \overrightarrow{F_M}(\overrightarrow{r^*}) = \sum_{j=1}^{N} \int_V \rho_{el}(\vec{r} - \vec{r_j}) \exp(2\pi i \vec{r} \cdot \overrightarrow{r^*})\, d\vec{r}$$

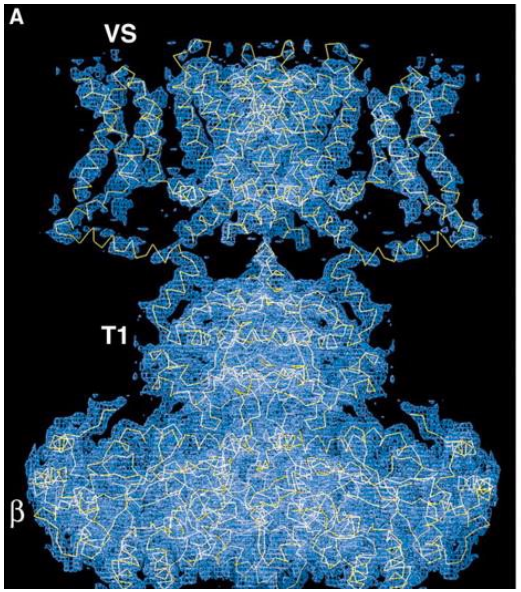
r-rj=R_j

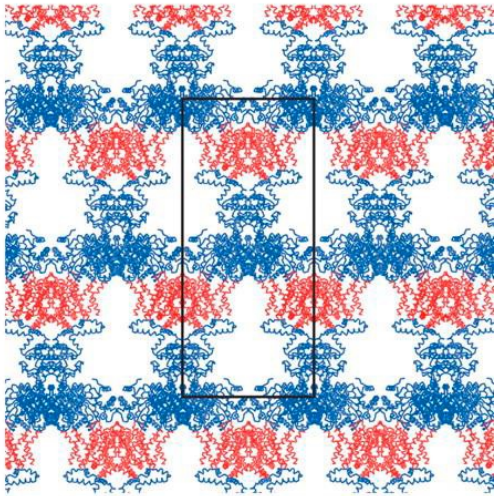
$$\overrightarrow{F_M}(\overrightarrow{r^*}) = \sum_{j=1}^{N} \int_V \rho_{el}(\overrightarrow{R_j}) \exp[2\pi i(\vec{r_j} + \vec{R_j}) \cdot \overrightarrow{r^*}\ d\overrightarrow{R_j}$$

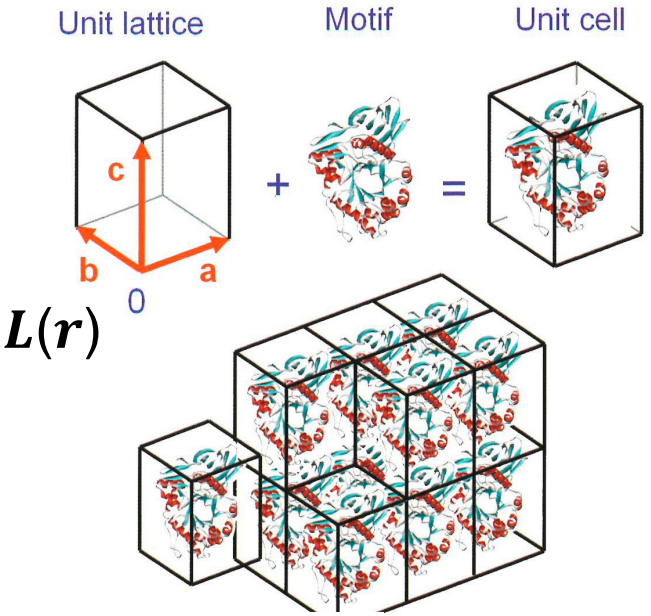
$$\sum_{j=1}^{N} \underbrace{\left\{ \int_V \rho_{el}(\overrightarrow{R_j}) \exp(2\pi i \vec{R_j} \cdot \overrightarrow{r^*})\, d\overrightarrow{R_j} \right\}}_{\boldsymbol{f_j^{at}}} \exp(2\pi i \vec{r} \cdot \overrightarrow{r^*})$$

$$\boldsymbol{\overrightarrow{F_M}(\overrightarrow{r^*}) = \sum_{j=1}^{N} f_j^{at}(r^*)\, exp(2\pi i \vec{r} \cdot \overrightarrow{r^*})}$$

Diffraction from a crystal







$$\rho_{crys}(r) = \rho_{mol}(r) * L(r)$$

$\boldsymbol{L(r)}$ is a mathematical function which describes the crystal lattice

Diffraction from a crystal

$$\rho_{crys}(r) = \rho_{mol}(r) * L(r)$$

$$\overrightarrow{F_{crys}}(\overrightarrow{r^*}) = TF[\rho_{crys}(r)]$$

$$TF[\rho_{crys}(r)] = TF[\rho_{mol}(r) * L(r)]$$

$$TF[\rho_{mol}(r) * L(r)] = TF[\rho_{mol}(r)] \cdot TF[L(r)]$$

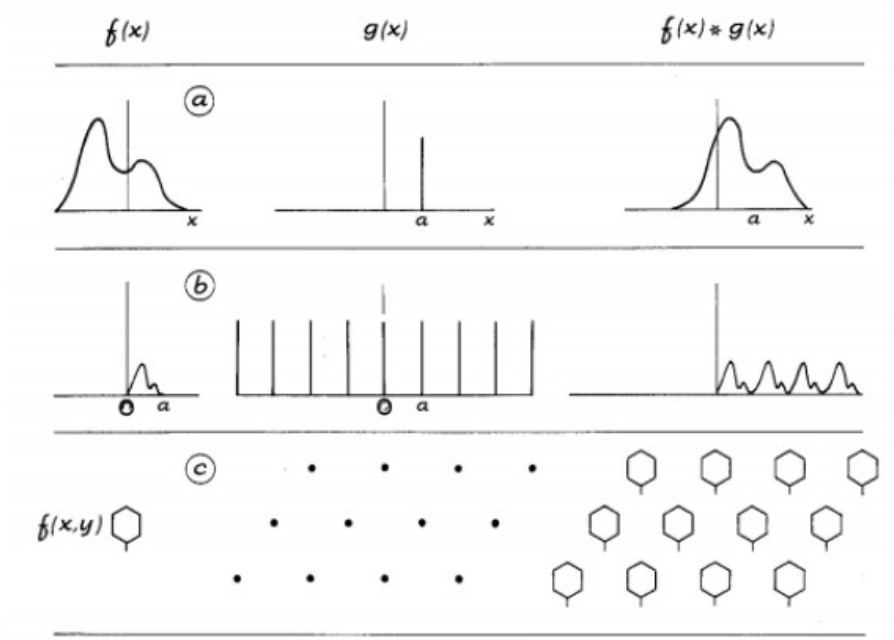
$\boldsymbol{TF[\rho_{mol}(r)]}$ is the structure factor of the unit cell

Diffraction from a crystal

La convoluzione di una funzione f(x) con una funzione che è diversa da zero per valori predefiniti, ha come effetto la ripetizione della funzione f(x) in tutti i punti diversi da zero

La TF[L(r)] è diversa da zero solo in corrispondenza dei nodi del reticolo reciproco, identificati da 3 indici h k l (numeri interi),

Di conseguenza avremo che la funzione F(r*) assume valori diversi da zero solo quando r* identifica un nodo del reticolo reciproco, definito come r*(hkl)

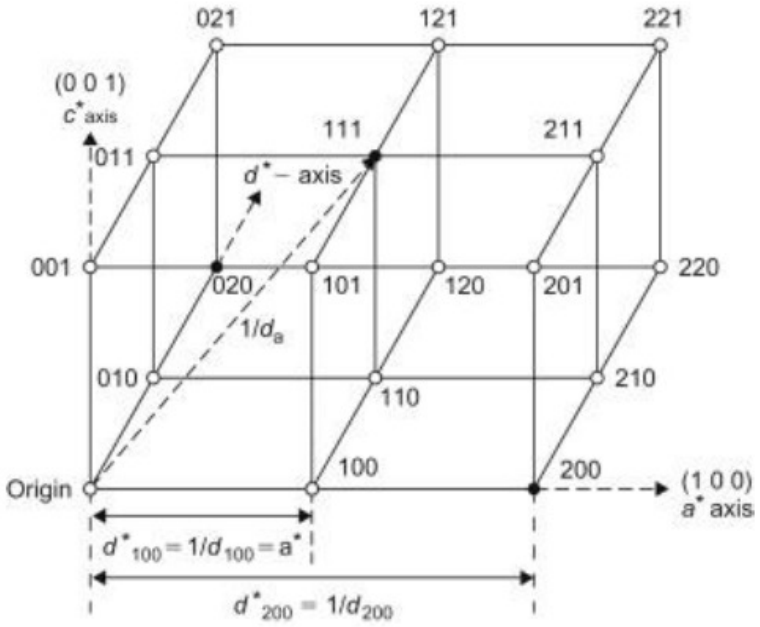


Diffraction from a crystal

L'esistenza del reticolo e la ripetizione ordinata della densità elettronica secondo la geometria del reticolo ha una conseguenza importante, **le onde diffuse non sono più diffuse in tutte le direzioni, ma solo lungo le direzioni definite dai nodi del reticolo reciproco identificati con la terna di numeri interi (h k l).**

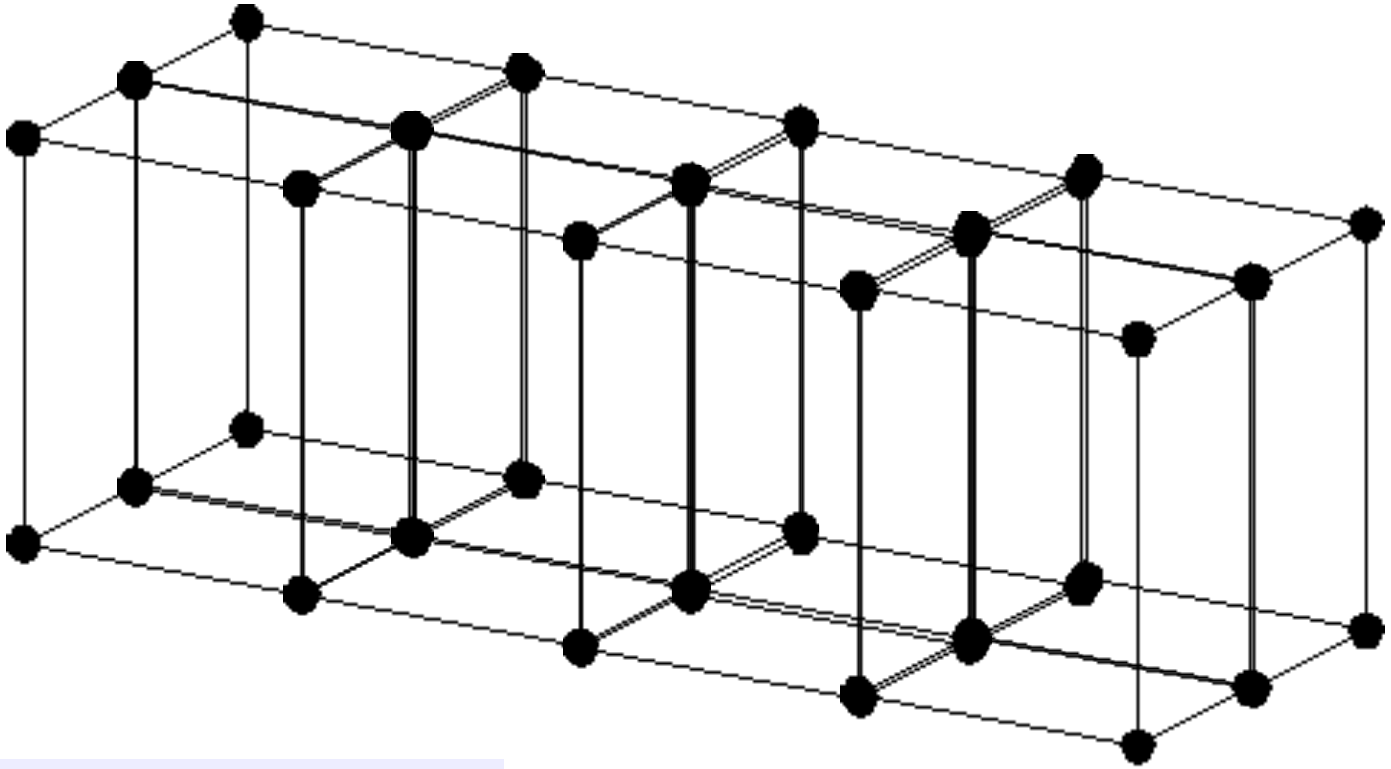
Quindi osserverò ampiezze diffratte diverse da zero, esclusivamente lungo direzioni che obbediscono alla seguente relazione:

$$\boldsymbol{r}^* = \boldsymbol{r}^*_{h,k,l}$$



ovvero **quando il vettore nello spazio reciproco coincide con un vettore che identifica un nodo del reticolo reciproco**, l'onda diffusa dal cristallo sarà dato dalla somma dei termini per i quali vale la relazione:

Crystals: unit cells

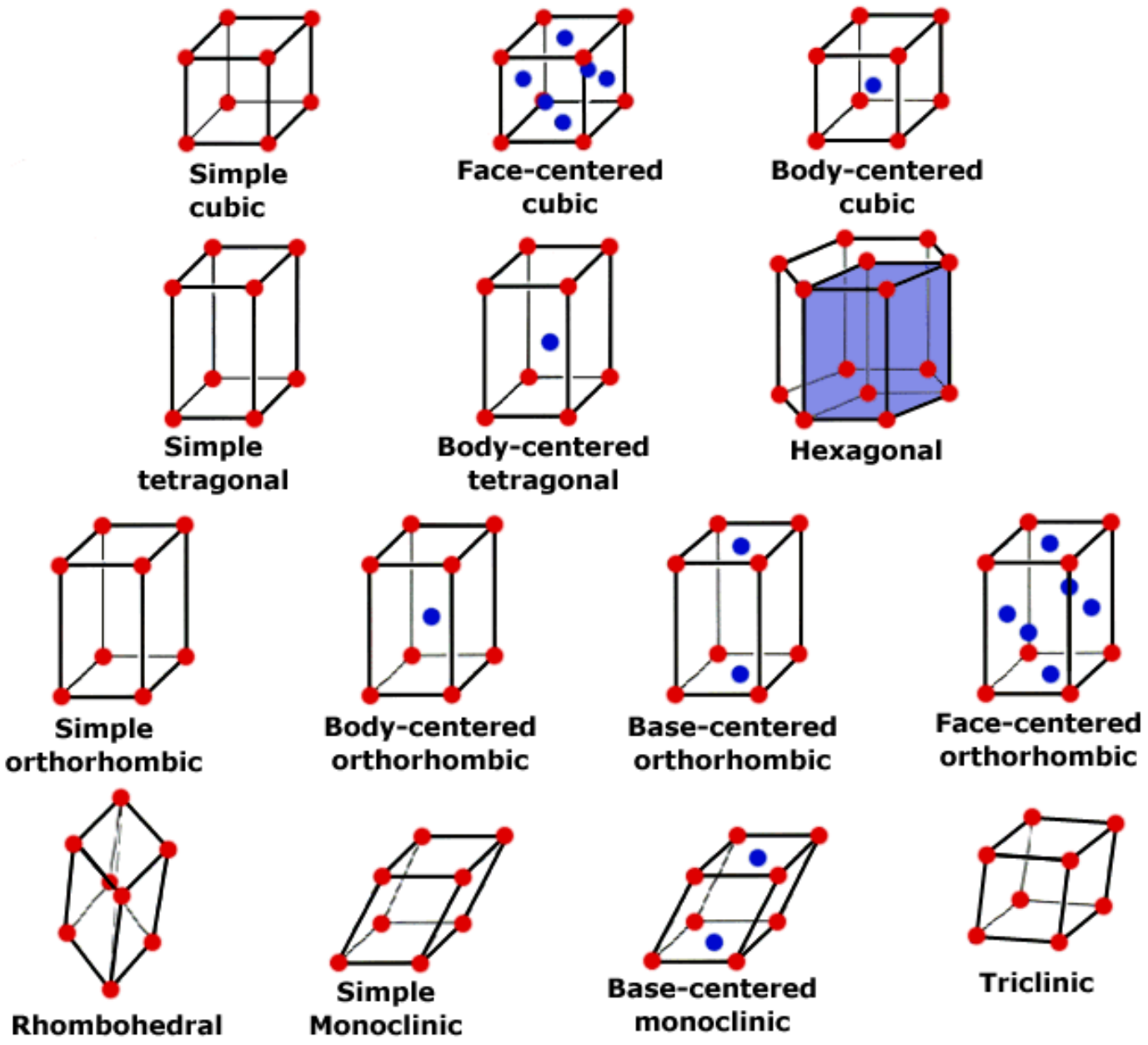


$$\vec{T} = u\vec{a} + v\vec{b} + w\vec{c}$$

u, v, w interi

$\vec{T}$ is a vector of the Bravais Lattice

The 14 Bravais Lattices

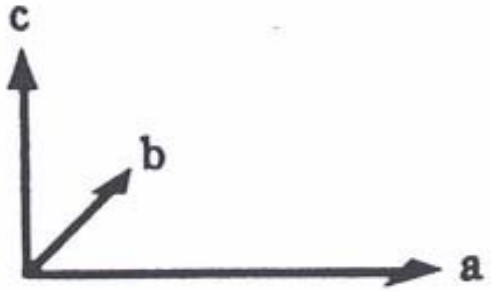


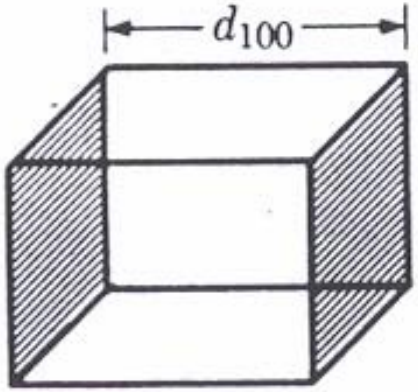
The space groups

Table 1.2 *Symmetry elements*

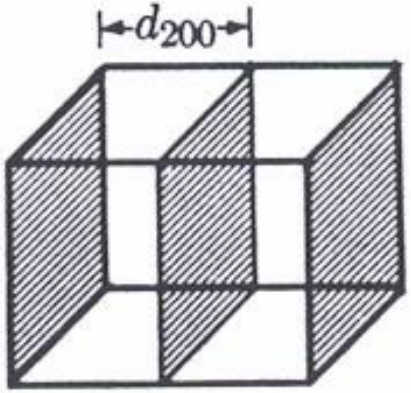
	Symmetry element	Hermann–Mauguin symbols (crystallography)	Schönflies symbols (spectroscopy)
Point symmetry	Mirror plane	m	σ_v, σ_h
	Rotation axis	$n = 2, 3, 4, 6$	$C_n (C_2, C_3, \text{etc.})$
	Inversion axis	$\bar{n} (= 1, 2, \text{etc.})$	—
	Alternating axis*	—	$S_n (S_1, S_2, \text{etc.})$,
	Centre of symmetry	$\bar{1}$	i
Space symmetry	Glide plane	a, b, c, d, n	—
	Screw axis	$2_1, 3_1$, etc.	—

* The *alternating axis* is a combination of rotation (n-fold) and reflection perpendicular to the rotation axis. It is little used in crystallography.

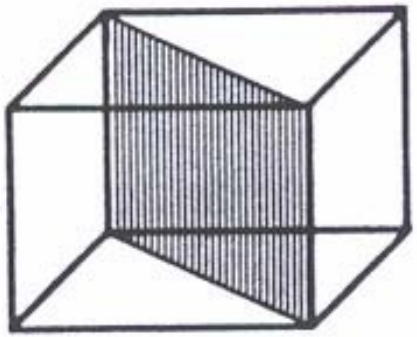




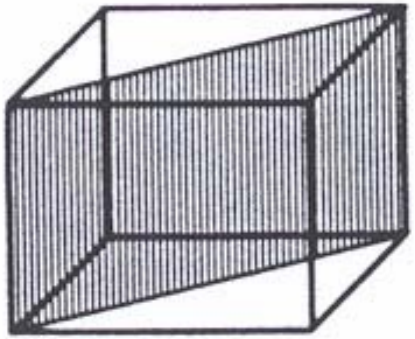
(100)



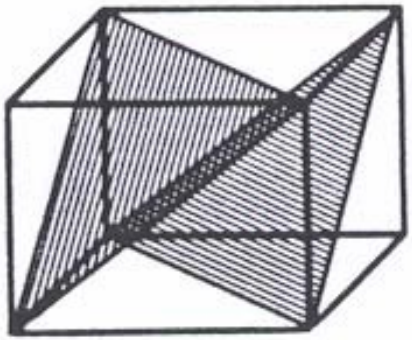
(200)



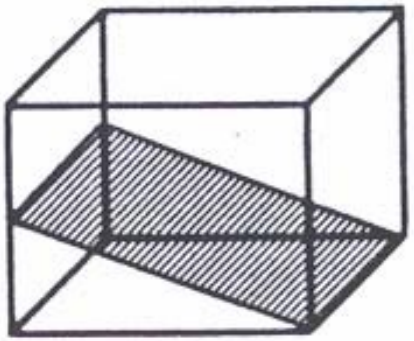
(110)



($\bar{1}10$)

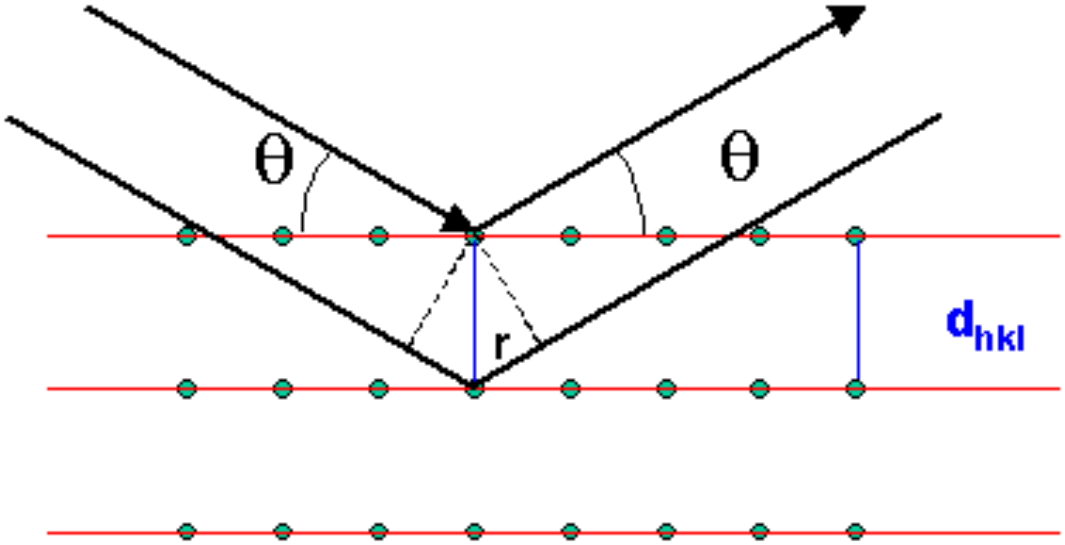


($11\bar{1}$)



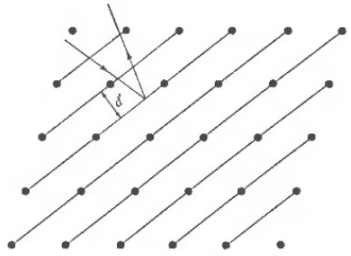
(102)

A practical approach: Bragg diffraction



Conditions for a sharp peak:

1) X-rays specularly reflected by atoms/ions/molecules in one crystal plane
2) reflected waves from successive planes interfere constructively

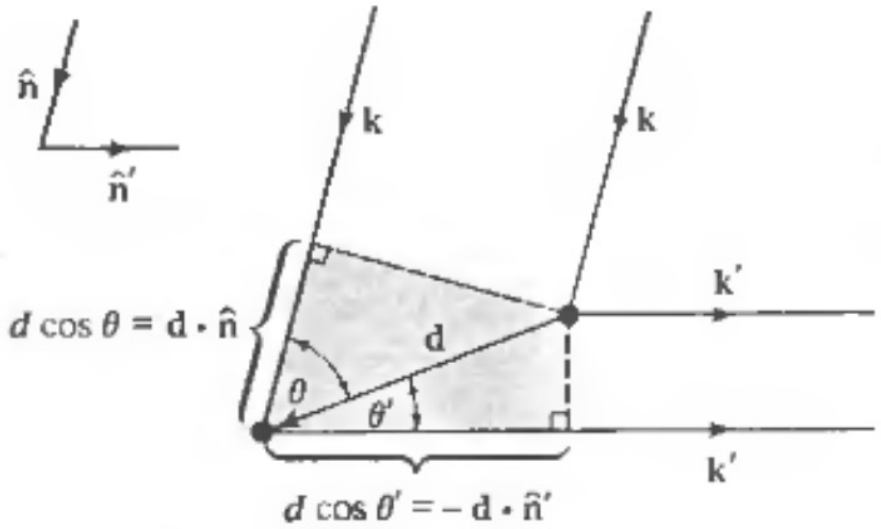


$2d_{hkl} \sin(\theta) = n \lambda$ Bragg's law

$1/d = (2/\lambda) \sin\theta$
$1/d \propto \sin\theta$

When d is large, diffraction pattern is compressed

von Laue condition for diffraction



No *ad hoc* consideration on specularity
No crystal plane
Scattering from points of a Bravais Lattice (each reradiate in all directions).

$$d\cos\theta + d\cos\theta' = \mathbf{d}\cdot(\hat{\mathbf{n}} - \hat{\mathbf{n}}'). \tag{6.3}$$

The condition for constructive interference is thus

$$\mathbf{d}\cdot(\hat{\mathbf{n}} - \hat{\mathbf{n}}') = m\lambda, \tag{6.4}$$

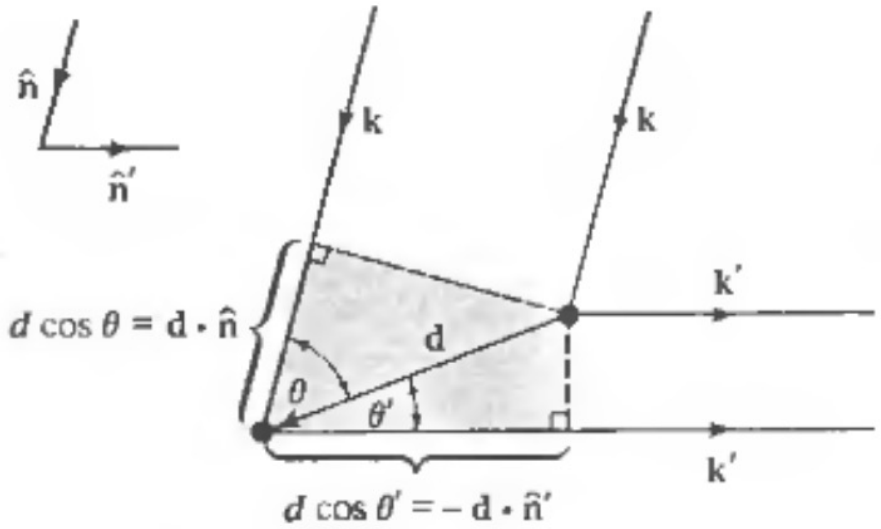
for integral m. Multiplying both sides of (6.4) by $2\pi/\lambda$ yields a condition on the incident and scattered wave vectors:

$$\mathbf{d}\cdot(\mathbf{k} - \mathbf{k}') = 2\pi m, \tag{6.5}$$

for integral m.

Ashcroft and Mermin

von Laue condition for diffraction



No *ad hoc* consideration on specularity
No crystal plane
Scattering from points of a Bravais Lattice (each reradiate in all directions; elastic scattering).

Next, we consider not just two scatterers, but an array of scatterers, at the sites of a Bravais lattice. Since the lattice sites are displaced from one another by the Bravais lattice vectors **R**, the condition that all scattered rays interfere constructively is that condition (6.5) hold simultaneously for all values of **d** that are Bravais lattice vectors:

$$\mathbf{R} \cdot (\mathbf{k} - \mathbf{k}') = 2\pi m, \quad \text{for integral } m \text{ and all Bravais lattice vectors } \mathbf{R}. \tag{6.6}$$

This can be written in the equivalent form

$$e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{R}} = 1, \quad \text{for all Bravais lattice vectors } \mathbf{R}. \tag{6.7}$$

Ashcroft and Mermin

The reciprocal space

R = Points of a Bravais Lattice

$e^{i\mathbf{k}\cdot\mathbf{r}}$ Plane wave. For a set of **K**, it has the periodicity of the BL. These **K** vectors define the Reciprocal Lattice

$$e^{i\mathbf{K}\cdot(\mathbf{r}+\mathbf{R})} = e^{i\mathbf{K}\cdot\mathbf{r}}$$ for each **r**

$$e^{i\mathbf{K}\cdot\mathbf{R}} = 1$$ The RL is a BL in the space of wave vectors. The RL of the RL is the original BL

$$\mathbf{K} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3$$

$$\mathbf{b}_1 = 2\pi\frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}$$

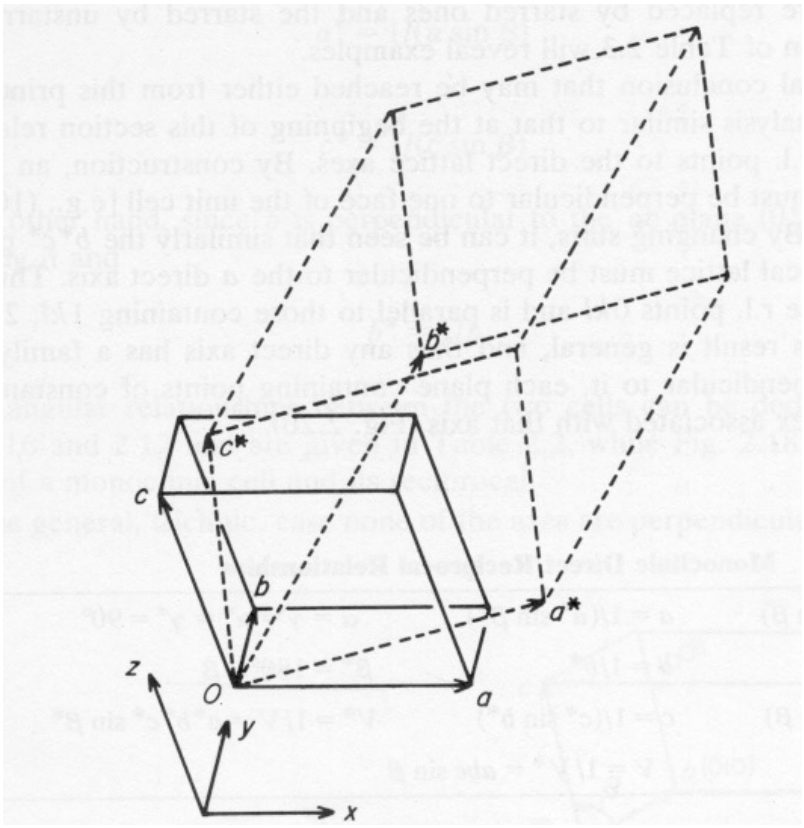
$$\mathbf{b}_2 = 2\pi\frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}$$

$$\mathbf{b}_3 = 2\pi\frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}.$$

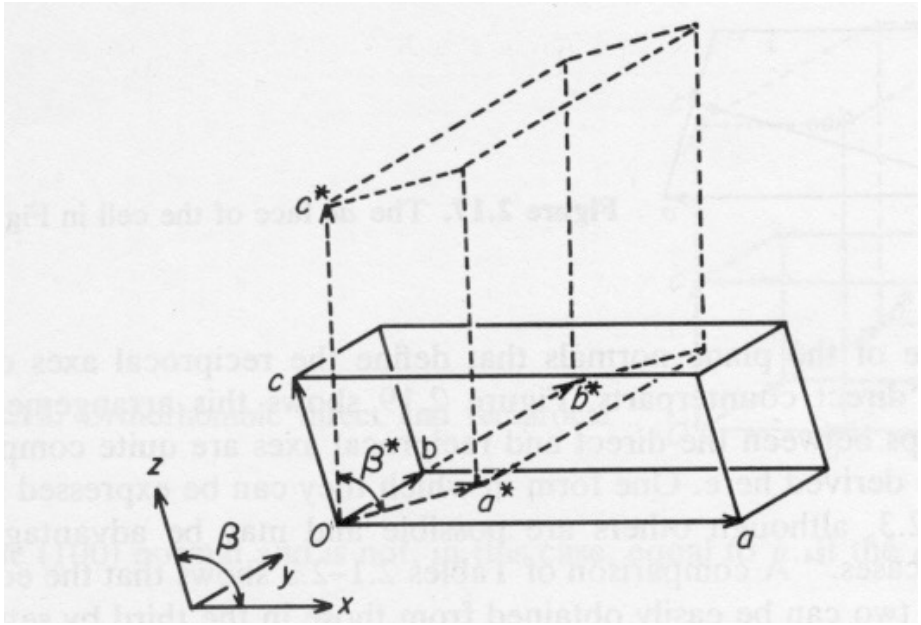
RL vectors have dimension of l^{-1}

Examples of direct and reciprocal lattices

(Figures from Jensen and Stout "*X-Ray structure determination. A pratical guide*"

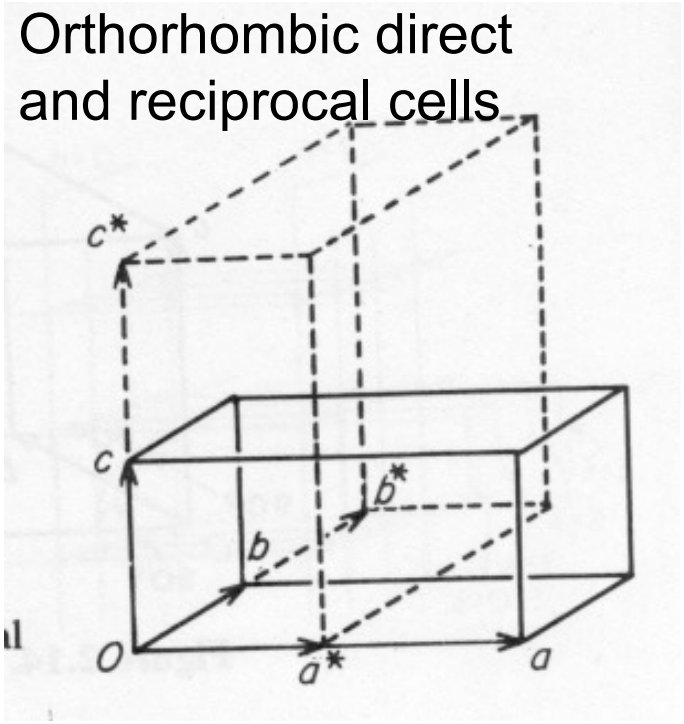


Triclinic direct and reciprocal cells

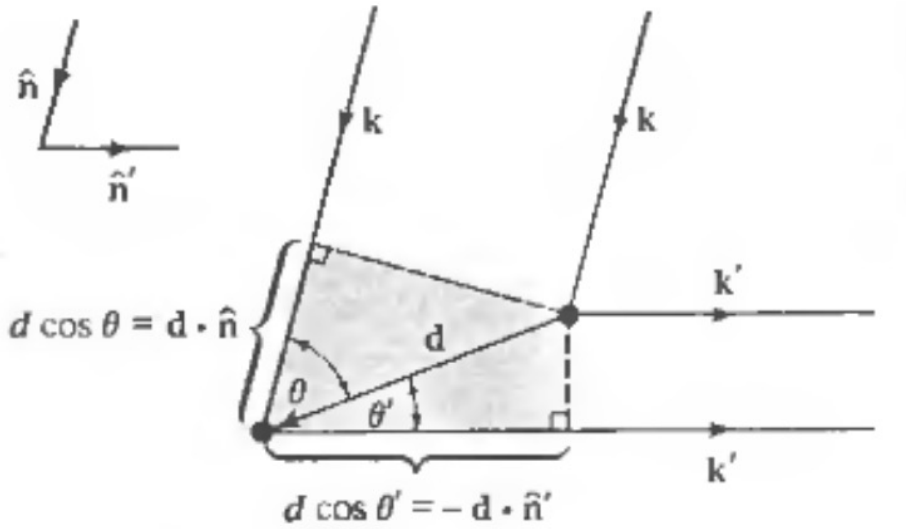


Monoclinic direct and reciprocal ce

Orthorhombic direct and reciprocal cells

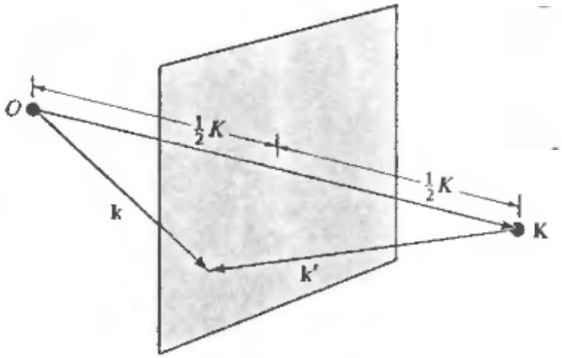


von Laue condition for diffraction



The Laue condition then states that to have constructive interference, the change in the wave vector **K = k'-k** is a vector of the RL. Therefore **k'-k=k-k'** and |**k**'|=|**k**| or |**k**|=|**K-k**|

Figure 6.5
The Laue condition. If the sum of k and $-\mathbf{k}'$ is a vector K, and if k and k′ have the same length, then the tip of the vector k is equidistant from the origin O and the tip of the vector K, and therefore it lies in the plane bisecting the line joining the origin to the tip of K.



$$\mathbf{k} \cdot \hat{\mathbf{K}} = \tfrac{1}{2}K;$$

Ashcroft and Mermin

Bragg-von Laue equivalence

For **K = k'-k** vector of the RL, given **|k|=|k'|**, they form the same angle with the plane perpendicular to **K**.
This can be seen as a **Bragg diffraction** from the family of planes perp. to **K**

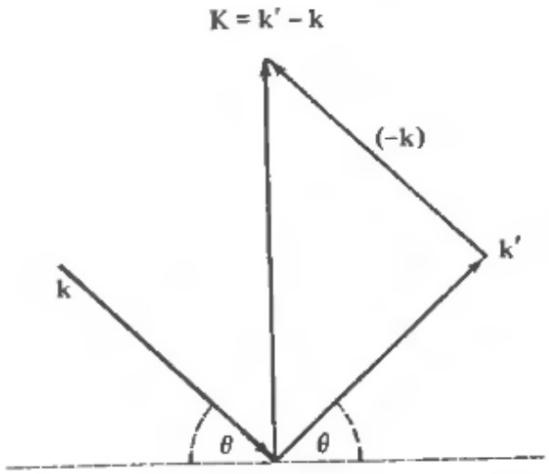


Figure 6.6
The plane of the paper contains the incident wave vector k, the reflected wave vector k', and their difference K satisfying the Laue condition. Since the scattering is elastic ($k' = k$), the direction of K bisects the angle between k and k'. The dashed line is the intersection of the plane perpendicular to K with the plane of the paper.

*A Laue diffraction peak corresponding to a wave vector change **K** (RL vector) corresponds to a Bragg diffraction from the family of BL planes perpendicular to **K.***
The order n of the Bragg reflection is **|K|** divided by the shortest **|k|** vector parallel to **K**

Bragg-von Laue equivalence

For **K = k'-k** vector of the RL, given **|k|=|k'|**, they form the same angle with the plane perpendicular to **K**.
This can be seen as a **Bragg diffraction** from the family of planes perp. to **K**

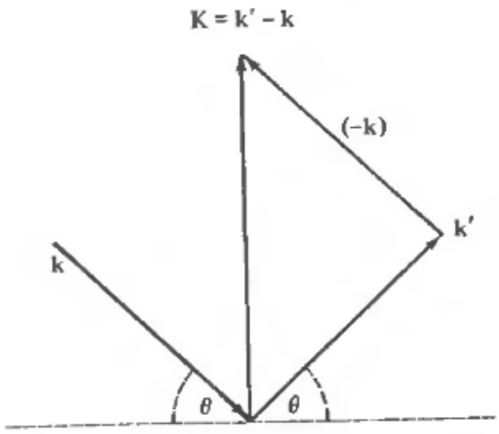
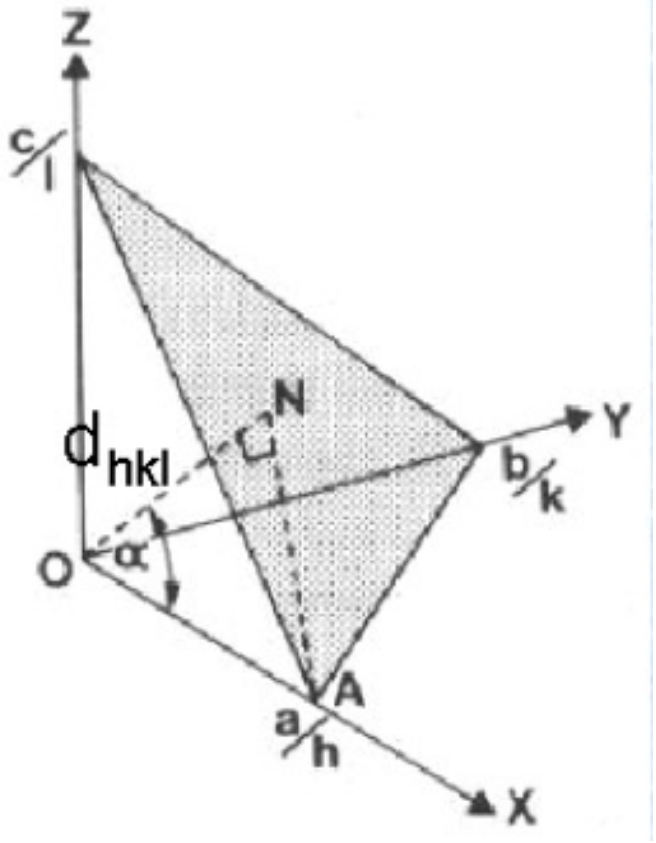


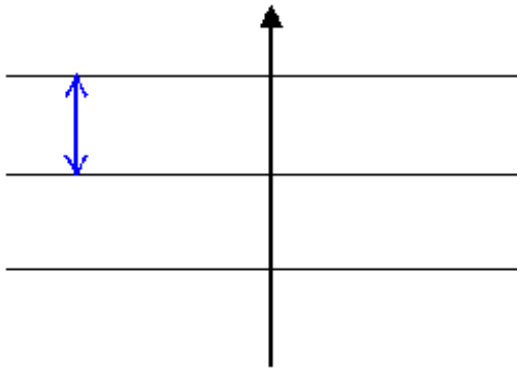
Figure 6.6
The plane of the paper contains the incident wave vector **k**, the reflected wave vector **k'**, and their difference **K** satisfying the Laue condition. Since the scattering is elastic ($k' = k$), the direction of **K** bisects the angle between **k** and **k'**. The dashed line is the intersection of the plane perpendicular to **K** with the plane of the paper.

To design all the possible plane families is in fact too complex.
Instead, a family of planes is represented by the $\mathbf{d^*_{hkl}}$ vector,
i.e. **a unique point in the reciprocal space!**

$$\mathbf{d_{hkl}^* = K / d_{hkl}}$$

The reciprocal space





$(a/h)\cos\alpha = d_{hkl}$

$\cos\alpha = (h/a)\, d_{hkl}$
$\cos ß = (k/b)\, d_{hkl}$
$\cos\gamma = (l/c)\, d_{hkl}$

For orthorombic lattice:

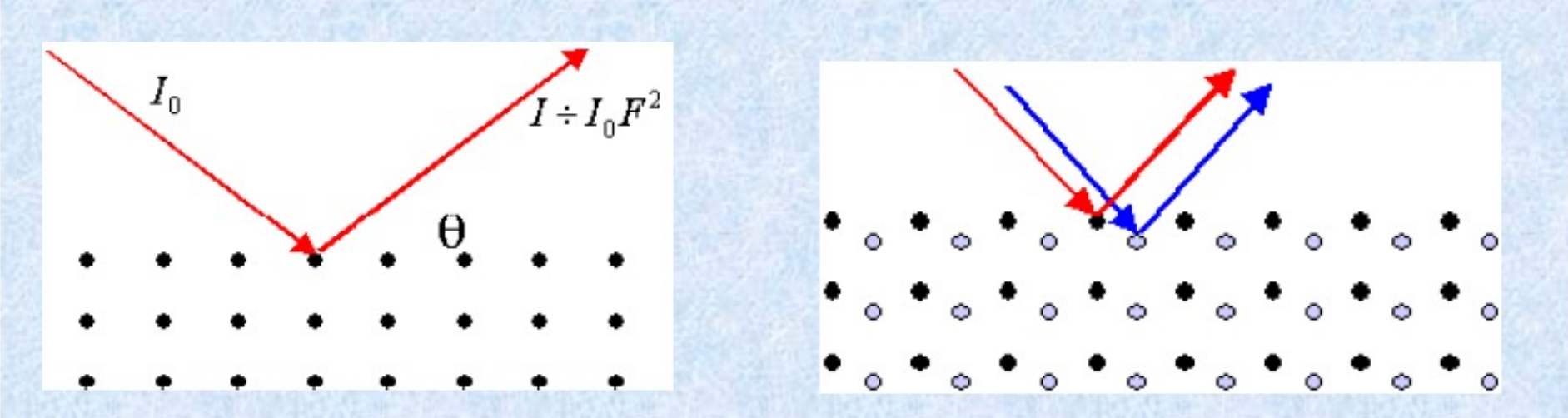
$(\cos\alpha)^2+(\cos ß)^2+(\cos\gamma)^2 = 1$

$(h/a)^2\, d^2_{hkl} + (k/b)^2\, d^2_{hkl} + (l/c)^2\, d^2_{hkl} = 1$

For a cubic crystal:

$1/d^2_{hkl} = 1/a^2 * (h^2+k^2+l^2)$

Is Bragg's law still valid for two or more atoms in a unit cell?



Two atoms in a unit cell (reflect) waves from their respective planes.

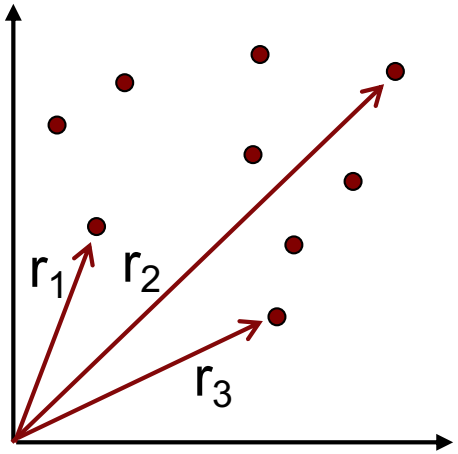
The waves combine and form a resultant wave, that looks like it has been reflected from the original unit cell lattice plane.

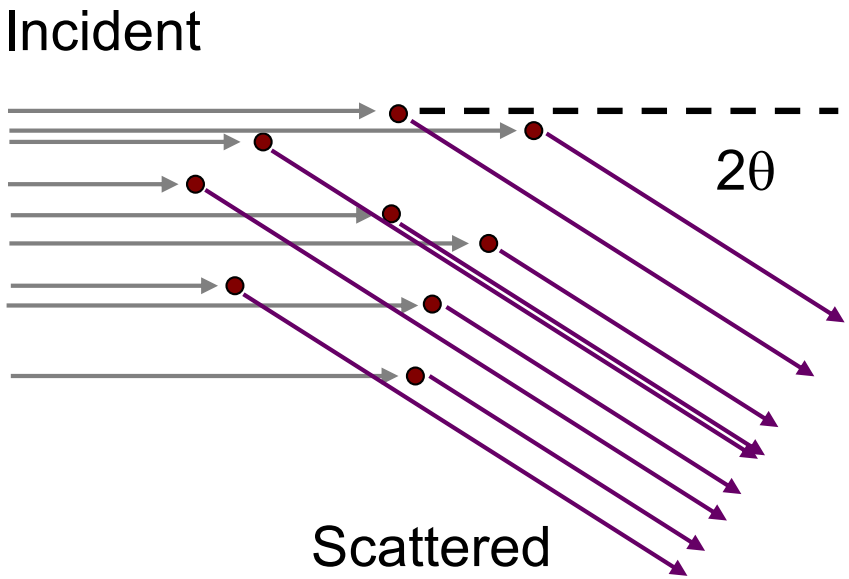
The 2 lattice are the same, but shifted one with respect to the other

Diffraction is spot is in the same place, but has different intensity (intensity of resultant wave).

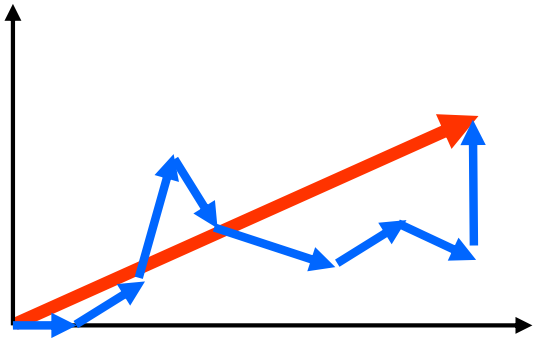
We assumed the electron density is in planes. In reality it is spread throughout the unit cell. Nevertheless, the derivation is still valid, since it can be shown that waves scattered from electron density not lying in a plane P, can be added to give a **resultant as if reflected from the plane**.

Scattering from many electrons (i.e. a molecule):





$$f(S) = \sum_{j=1}^{N} A_j e^{i 2\pi r_j \cdot s}$$



It is like adding many waves/vectors with magnitude A_j and phase (**rj**.**s**) - I can calculate this from the wavelength, angle and the positions of the scatterers.
In practise I can consider all the electrons in one atom as a "block" – I sum together the contributions of the atoms.

Scattering from a 3D crystal

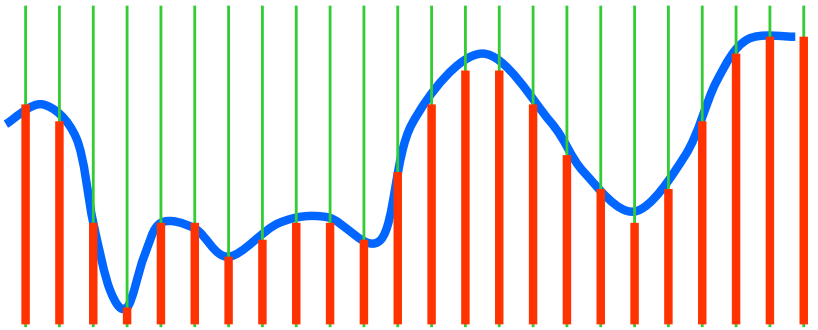
$$F_{3D}(s) = F(s) \left[\frac{\sin(2N+1)\pi(a\cdot s)}{\sin \pi(a\cdot s)} \right] \left[\frac{\sin(2N+1)\pi(b\cdot s)}{\sin \pi(b\cdot s)} \right] \left[\frac{\sin(2N+1)\pi(c\cdot s)}{\sin \pi(c\cdot s)} \right]$$

scattering from one molecule

3D fringe function that depends on the lattice spacing **a, b, c**

The fringe function has two effects:
- makes the pattern “discrete”
- amplifies the signal

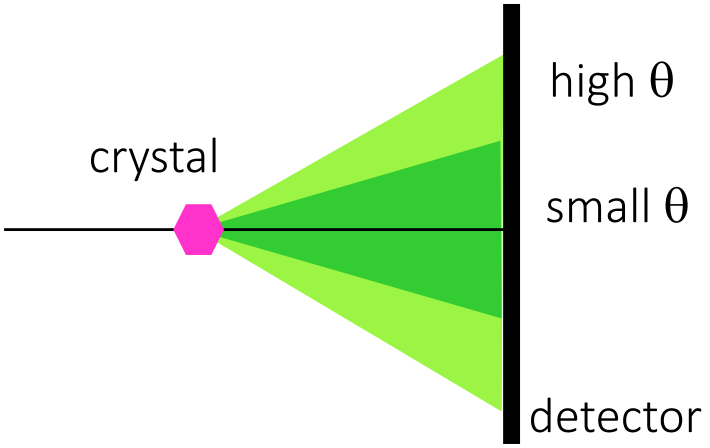
The diffraction pattern from a crystal is the diffraction from the molecule **sampled and amplified** according to the fringe function



The electron density equation: problem # 1:resolution limits

Theoretical limit

There are theoretical limits that depends on the wavelength of the X-ray radiation. Typically one uses a wavelength of roughly 1 A, which would allow for resolution of 0.5 A.



Practical limit

In practice for macromolecules the resolution limit is usually set by the intrinsic degree of order of the crystal typically one sees diffraction to 2.0-3.5 A – this is even more true of membrane protein crystals which tend to be more disordered due to the less directional nature of hydrophobic interactions.

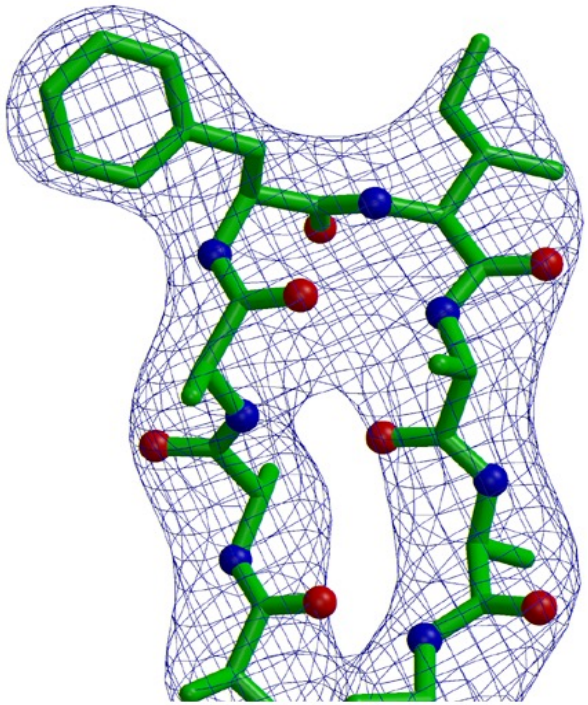
Why is resolution important?

"Resolution" is related to the level of details that can be visualised: a high resolution map gives a far more detailed picture of the atomic structure:

low resolution → high resolution

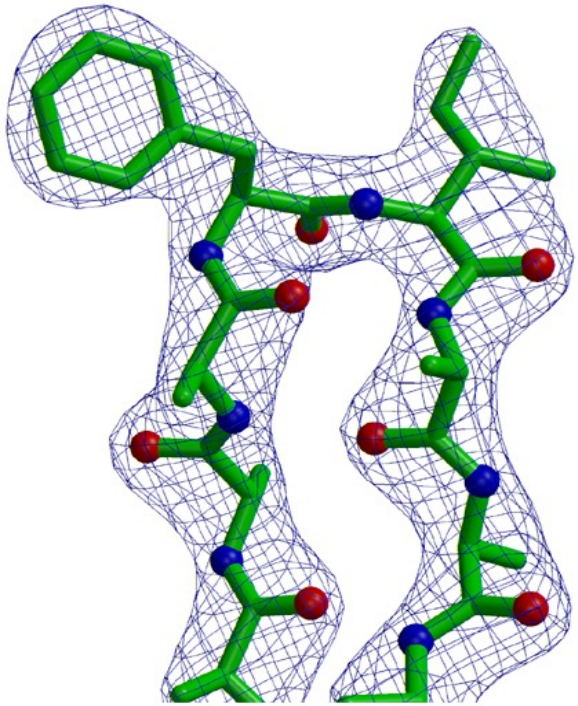
low s, small θ, large d

high s, large θ, small d

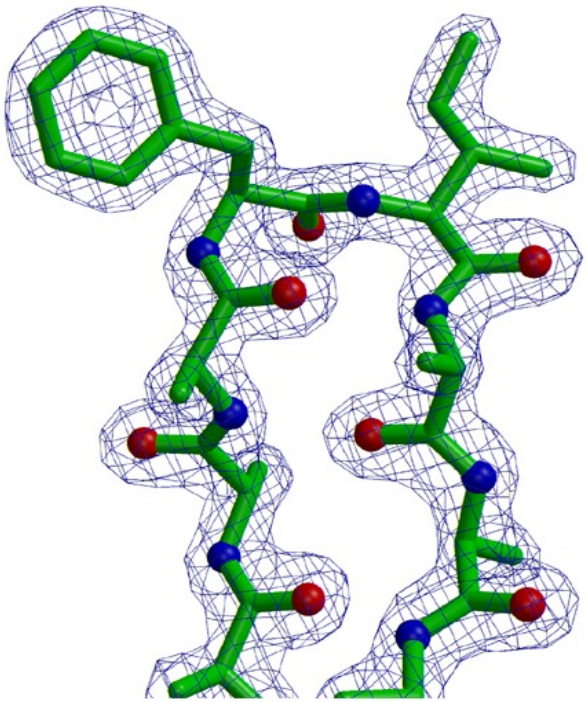


Here is a 4 Å map

things which are less than 4Å apart tend to be blurred



Here is a 2.9 Å map

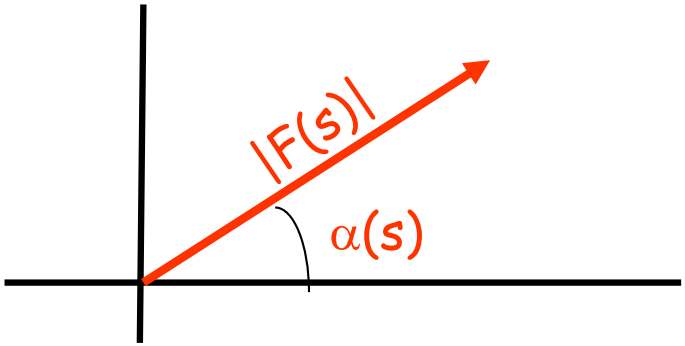


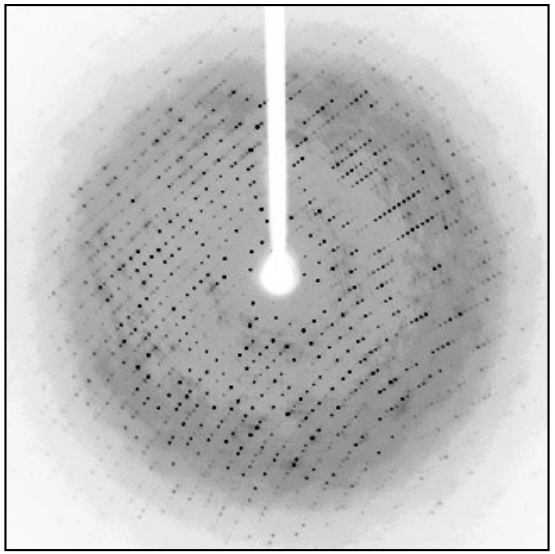
Here is a 2 Å map

almost atomic detail

The electron density equation: problem # 2:the phase problem

◆ F(s) is a complex number with modulus |F(s)| and phase α





In a diffraction experiment we measure the intensity of each spot I(s)

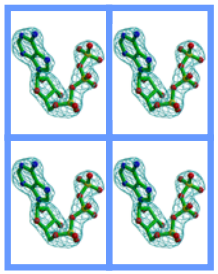
$I(s) = |F(s)|^2$

We can derive the amplitude F(s) but we have lost the information about the relative phase

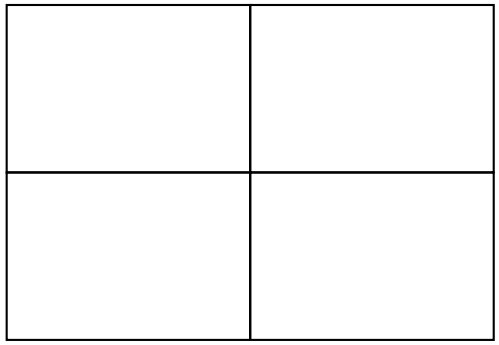
➡ ***the phase problem!!***

Universal problem in crystallography – also for small molecules.

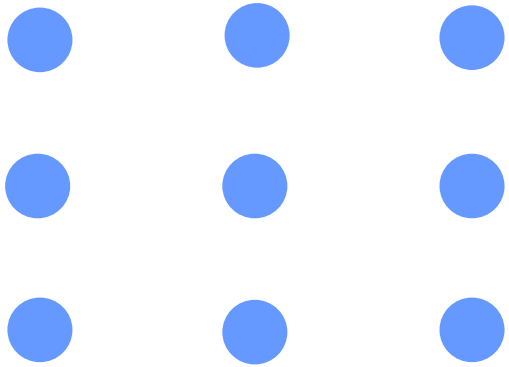
Small molecule crystal



small unit cell

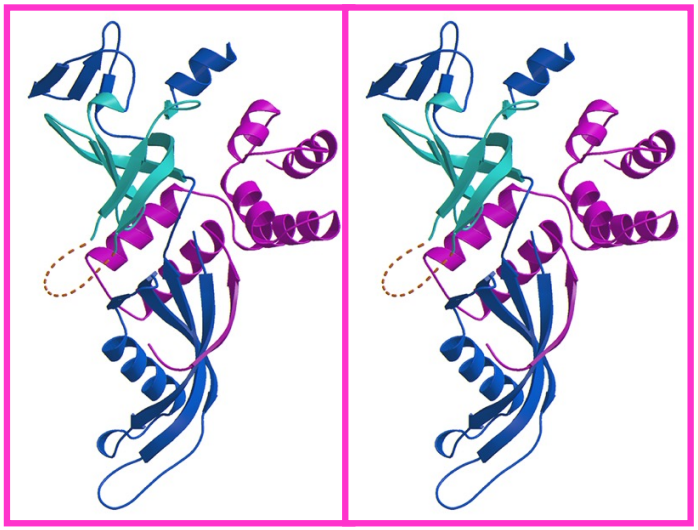


large “reciprocal lattice”
(coarse fringe function)

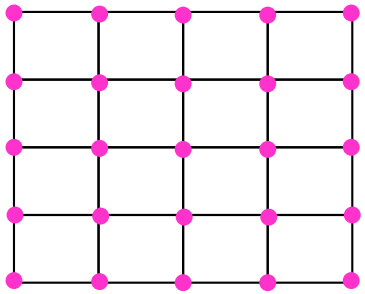


a few, very strong reflections;
diffract to high resolution

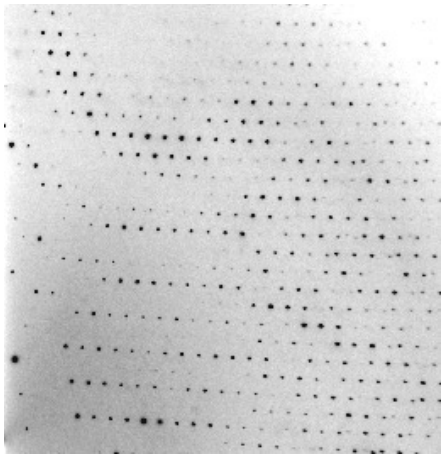
Protein crystal



large unit cell

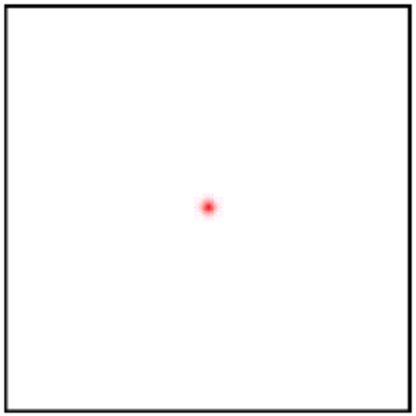


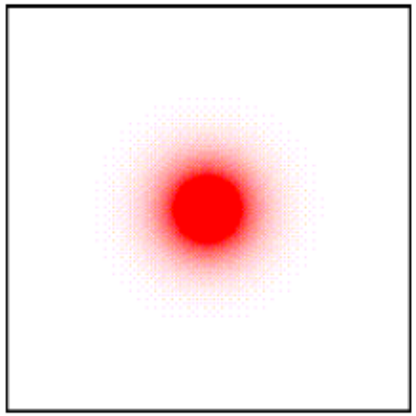
small “reciprocal lattice”
(fine fringe function)



many, very weak reflections;
diffract to low resolution

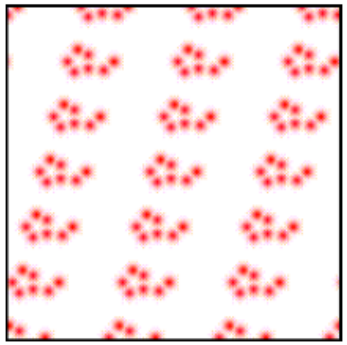
- An atom, and its Fourier Transform:

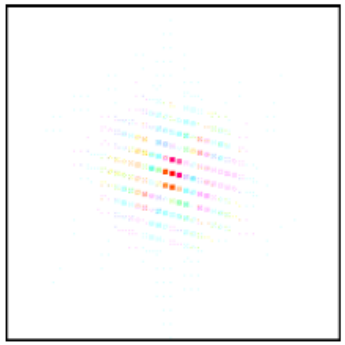




- Note the both functions have circular symmetry. The atom is a sharp feature, whereas its transform is a broad smooth function. This illustrates the reciprocal relationship between a function and its Fourier transform.

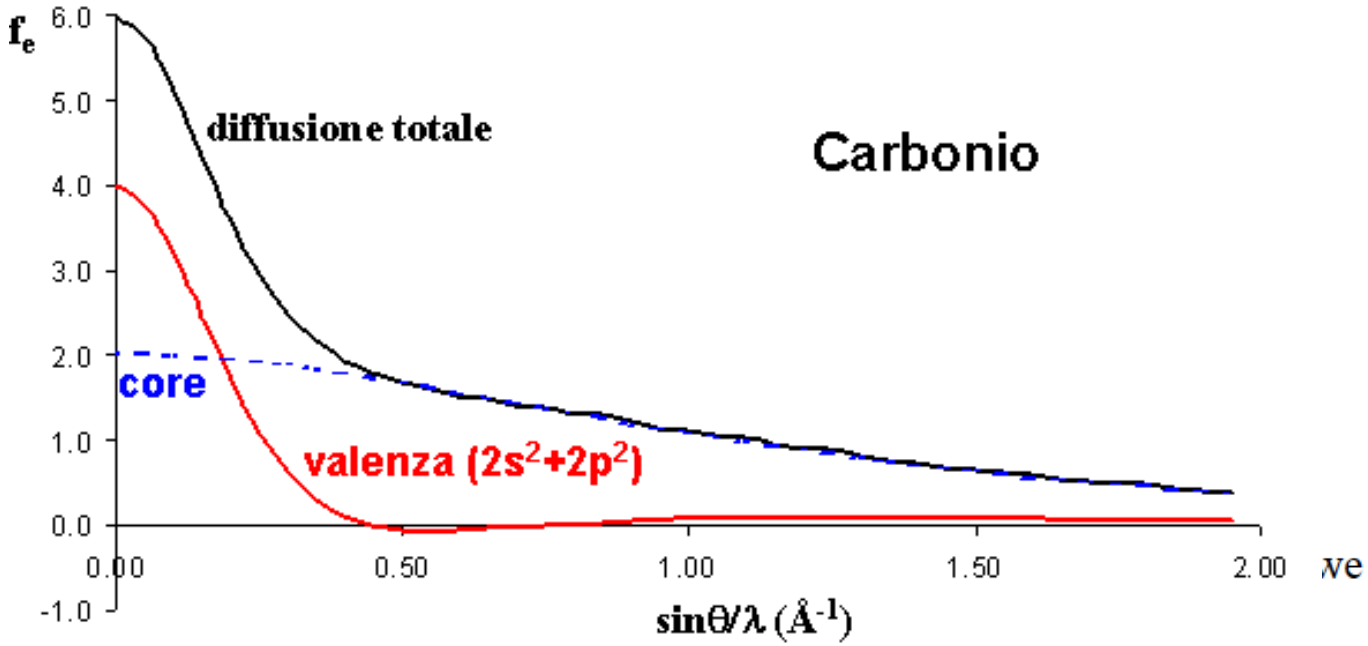
- A crystal, and its Fourier Transform:

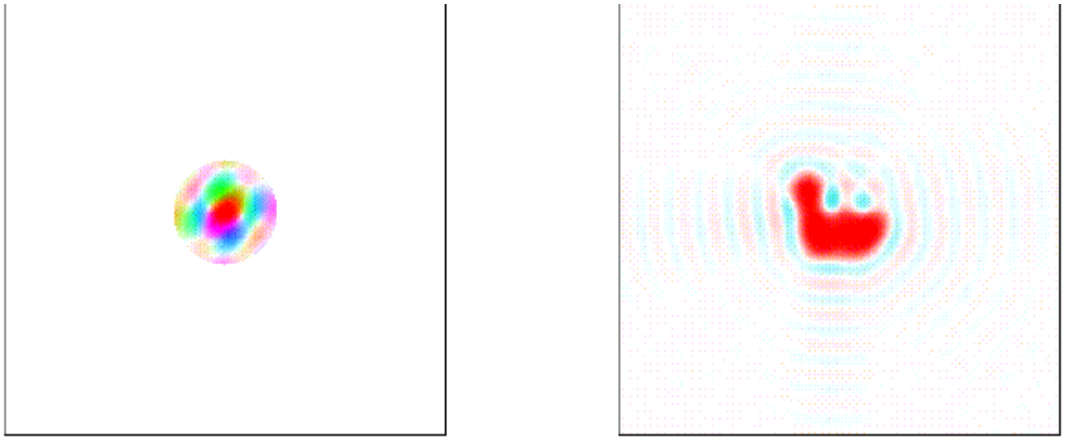




- Finally, we build up a crystal by convoluting the *molecule* with the *grid*. The result is a crystal structure. The Fourier transform of the crystal is thus the product of the *molecular transform* and the *reciprocal lattice*. This is the *diffraction pattern*.

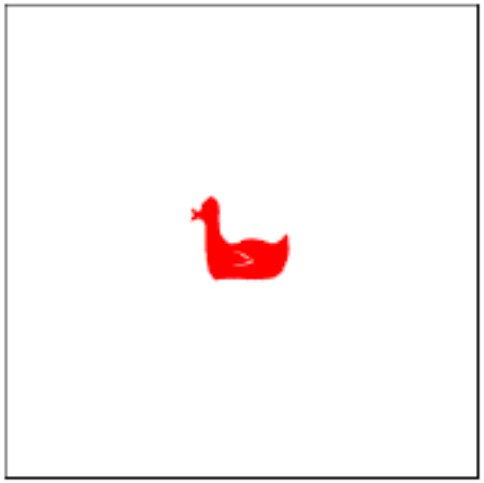
- A duck and its Fourier Transform

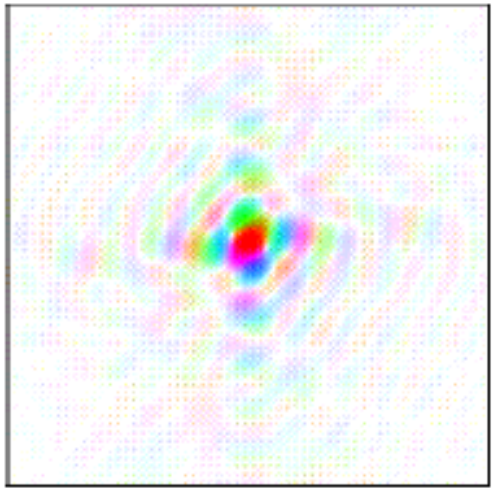




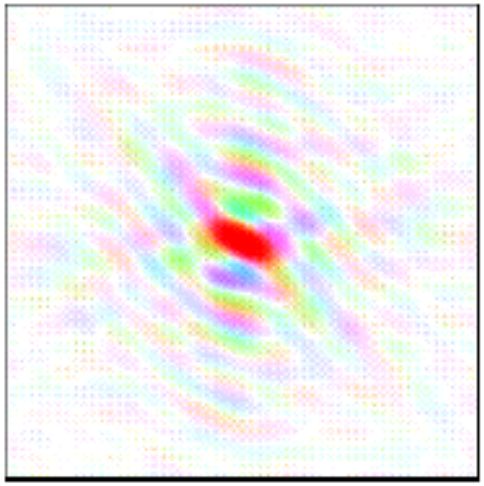
- **Crystallographic Interpretation:**
- There is considerable loss of detail. At low resolution, your atomic model may reflect more what you expect to see than what is actually there.

- Here is our old friend; the Fourier Duck, and his Fourier transform:

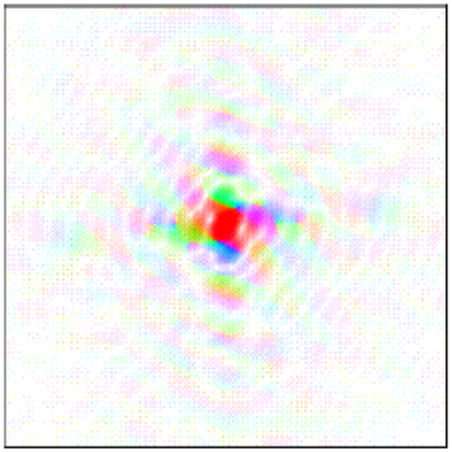


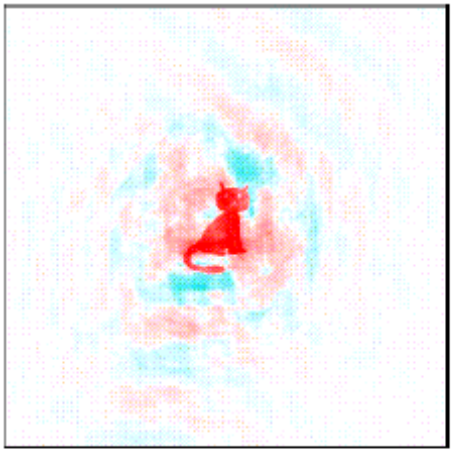


- And here is a new friend; the Fourier Cat and *his* Fourier transform:

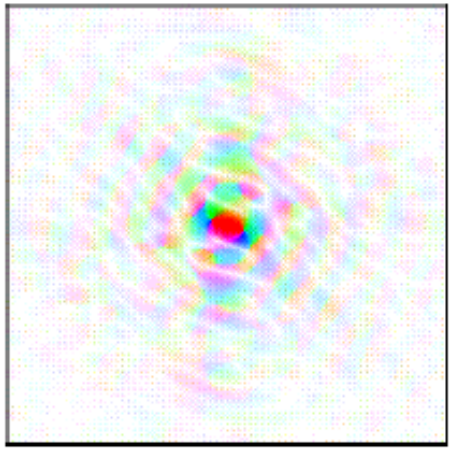


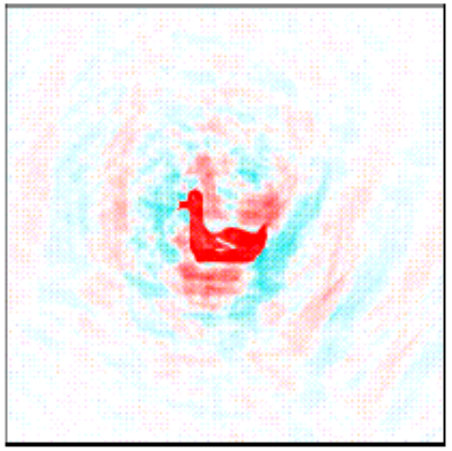
- Now we will mix them up. Let us combine the the magnitudes from the Duck transform with the phases from the Cat transform. (You can see the brightness from the duck and the colours from the cat). If we then transform the mixture, we get the following:





- We can do the same thing the other way round. Using the magnitudes from the Cat transform and the phases from the Duck transform, we get:





How to solve macromolecular structures:

MIR (multiple isomorphous replacement)

Older method (Cambridge, 60') – relies on binding "heavy" atoms to the crystal and compare he diffraction pattern to the native. Trial and error search for good heavy atoms, it may take longer to get it right

MR (molecular replacement)

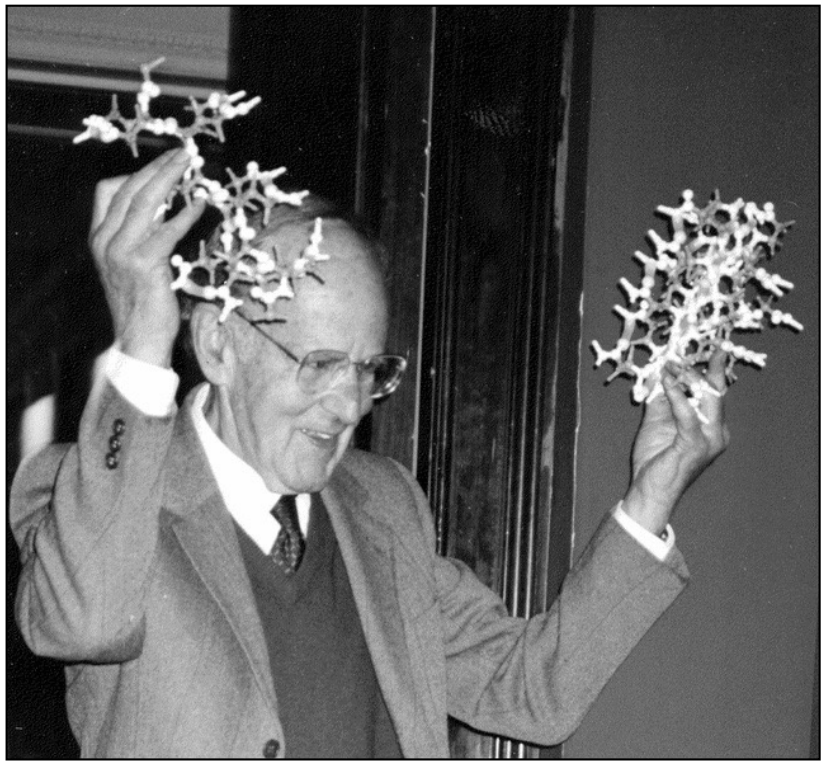
Older method (Cambridge, 70'-80') – relies on the expected similarity between the protein and another whose structure is known. Cannot solve de novo structures. Requires high homology (30% sequence identity?)

MAD (multiwavelength anomalous dispersion)

Relies on the absorption of specific wavelengths due to electronic transitions within the atom core. Similar to MIR but generally far quicker and more accurate. Requires high specification synchrotron radiation.

Some of the players...

Max Perutz (1914-2002), the inventor of MIR



In 1953, Perutz showed that the diffracted X-rays from protein crystals could be phased by comparing the patterns from crystals of the protein with and without heavy atoms attached.
In 1959, he employed this method to determine the molecular structure of hemoglobin. This work resulted in his sharing with John Kendrew the 1962 Nobel Prize for Chemistry.

Error treatment in MIR

The real breakthrough in using MIR for the determination of protein structures came when people learned how to deal with errors.

Blow D.M. & Crick F.H.C. (1959) “The treatment of errors in the isomorphous replacement method”. Acta Crystallogr. 12, 794-802

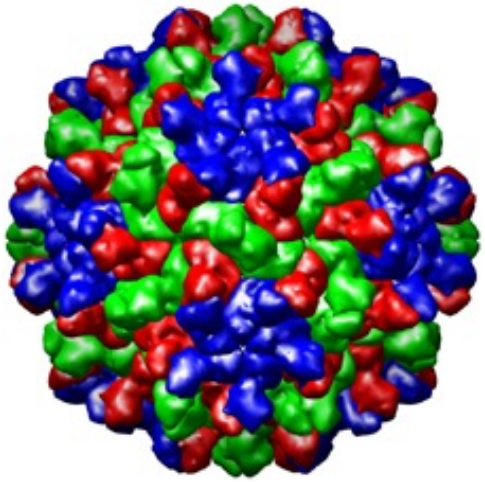
David Blow
(1931-2004)

Francis Crick
(1916-2004)

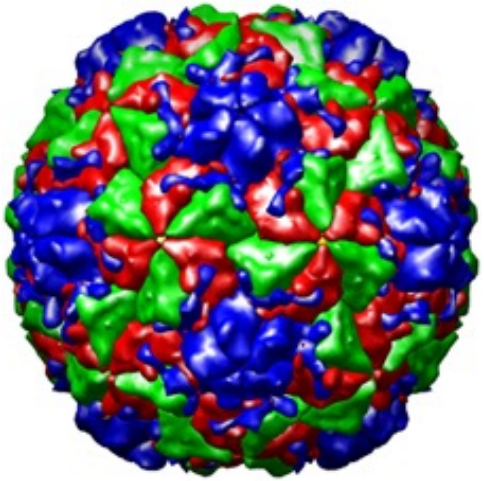
Error treatment is very complex – but it was an essential step in solving protein structures.

Molecular Replacement

The Molecular replacement method was mostly developed by Michael Rossmann. He used the structure of Tomato Bushy Stunt Virus, a plant virus, to determine the crystal structure of the Human Rhinovirus 14 (the common cold virus) in the early 80's.



Tomato Bushy Stunt Virus



Human Rhinovirus 14

However the theoretical basis were developed much earlier:

Rossman, M. G. and Blow, D. M. (1962). ActaCryst. 15:24-31.

Phase problem solved?
Calculating an electron density map

Once we have decent experimental estimates of the phases (from MIR, MR or MAD) we can obtain an electron density map by calculating a weighted Fourier transform with coefficients F_{obs} (i.e. the amplitudes of the diffracted rays and phases $\alpha_{MIR}/\alpha_{MR}/\alpha_{MAD}$.

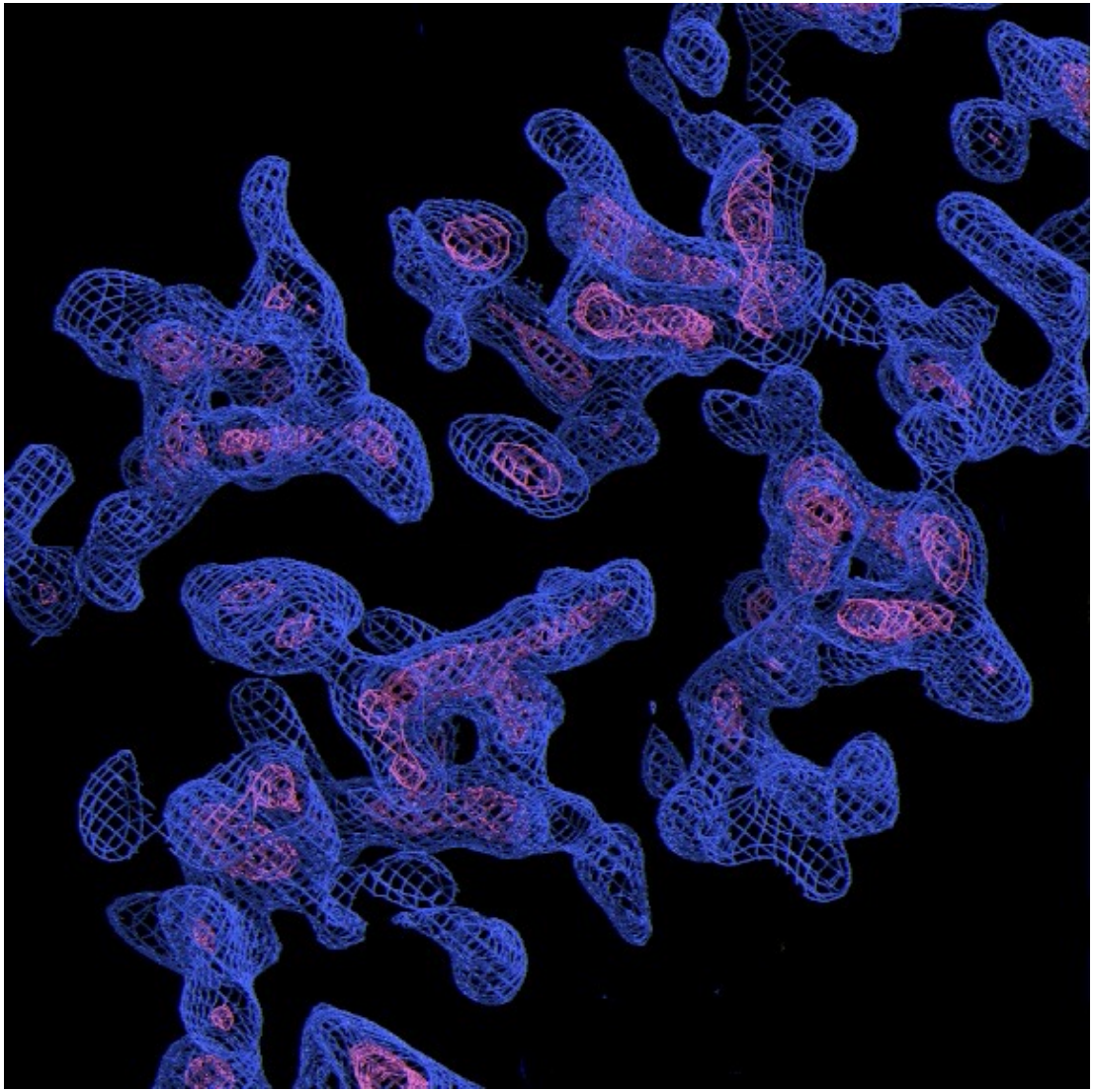
$$\rho(xyz) = \sum_{hkl} W(hkl) * |F_P(hkl)| \exp[i\,\alpha(hkl)] \exp[-2\pi\, i(hx + ky + lz)]$$

At the end of the day, the only criterion for determining how good is your MIR/MR/MAD solution is whether the map is interpretable.

Can you build a polypeptide chain?

Electron density maps

After all this effort, we have a 3D map showing the shape of the protein:

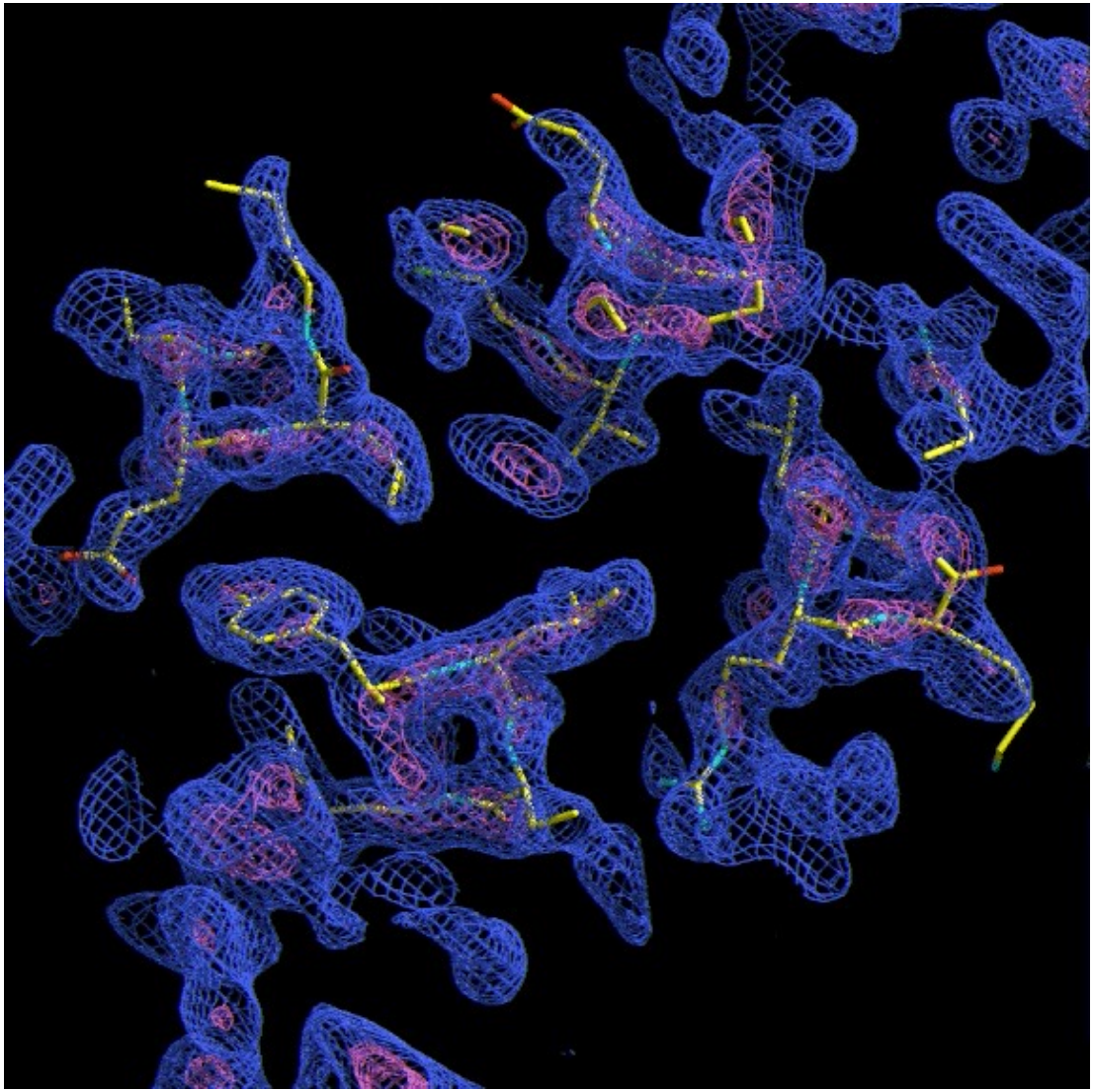


Electron density map displayed at two contour levels:

blue = 1 r.m.s (1σ);
magenta = 3 r.m.s (3σ)

Electron density maps

The task now is to try to fit an atomic model of the protein to the map…



Electron density map displayed at two contour levels:

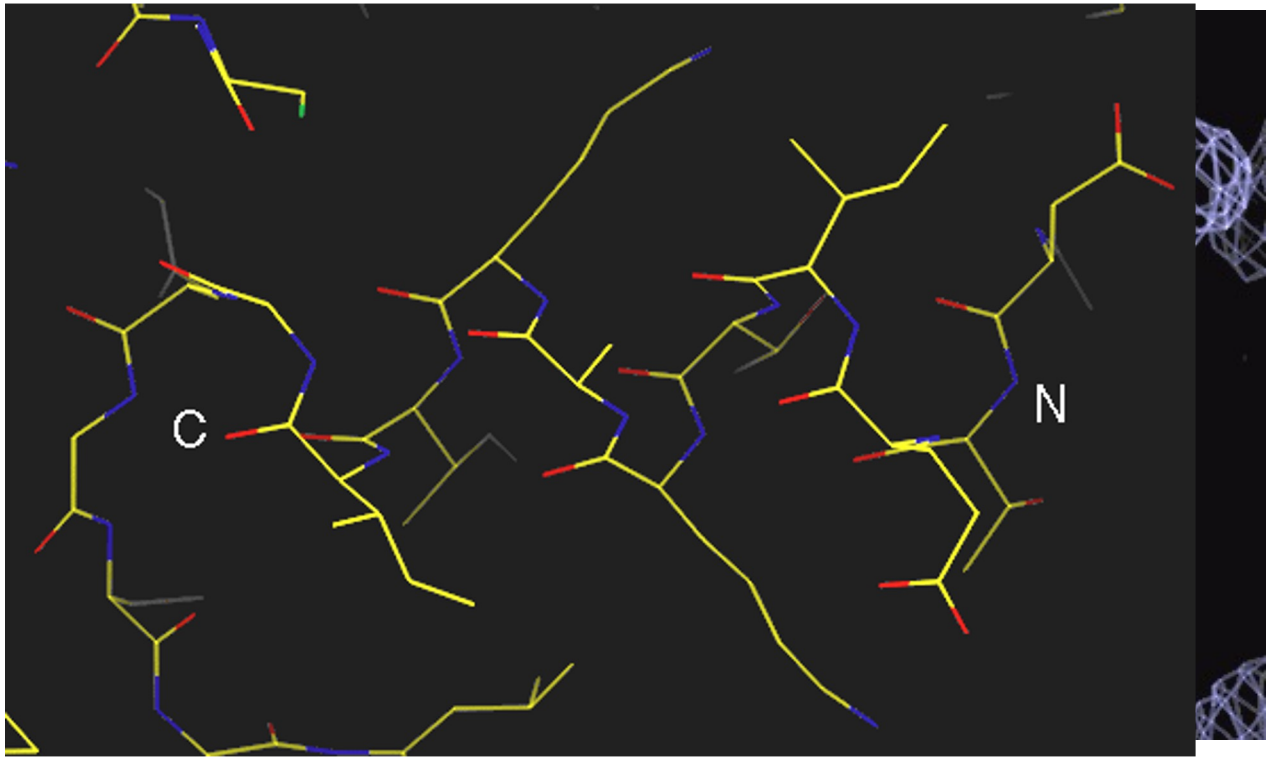
blue = 1 r.m.s (1σ);
magenta = 3 r.m.s (3σ)

Atomic model fitted to the map (in yellow)

Displaying the electron density map

Visualising a 3D electron density map over the entire unit cell can be daunting.

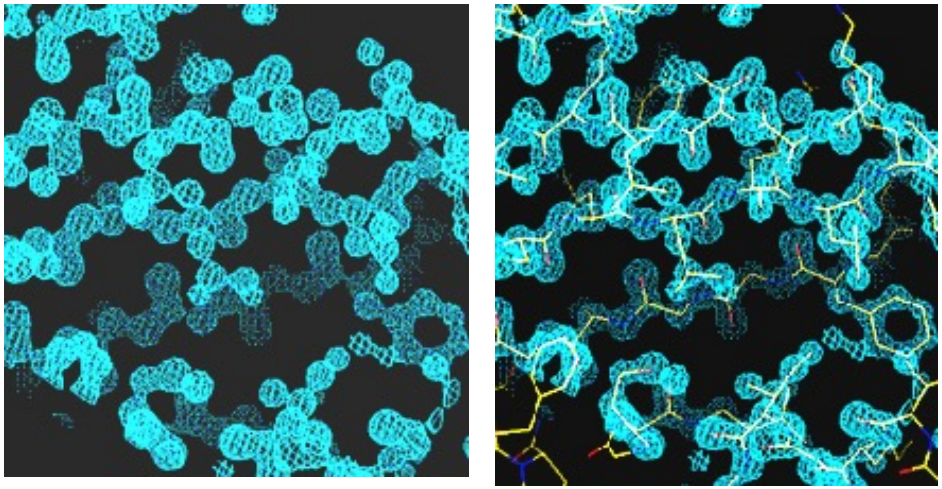
To visualise electron density maps we use high resolution interactive graphics, which display the map as a chicken wire. A map is a 4-dimensional object (each point x,y,z has an associated value) and to display it in 3D we have to choose a contour level that allows us to see about 70 % of the van der Waals radius of the atoms (typically one contours at the 1σ level)



Here is a rather good bit of density (much better than the density you are likely to obtain from the initial phases) and still it is not easy for the “untrained eye” to see what is going on...

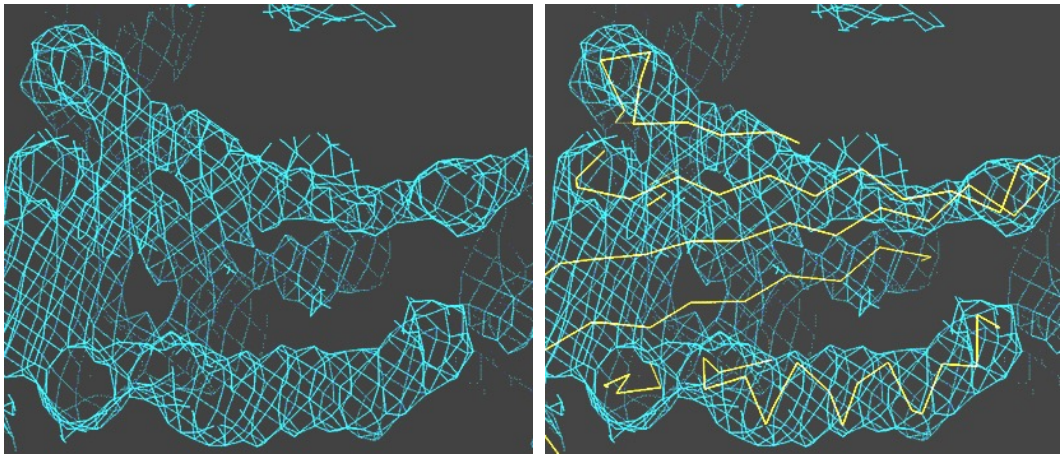
Maps and resolution

The task of model building is to interpret the electron density maps in light of chemical knowledge, basic stereochemistry, chemical sequence, etc...
The level of interpretation depends on the resolution of the map:



Here is a 1 Å map

At very high resolution, individual atoms can be seen and fitted in the electron density blobs: the problem therefore is reduced to 'join-the-dots'



Here is a 6 Å map

At very low resolution only large features can be seen - for example helices look like rods and β-sheets can barely be detected.

Refinement

Manual model building is not sufficient to build a completely accurate model but is required to get to the starting point for refinement.
Refinement is a process of optimisation of the atomic model to match the observed data and to conform to ideal stereochemistry

Successful structure determination usually requires several alternated rounds of model-building and refinement.

During refinement one calculate the expected diffraction pattern from the current model, compare it to the experimental diffraction data, and minimise the square of the differences.

Two problems:

- non linearity and presence of multiple minima
- low ratio observation parameters (especially at low res.); compensate by using stereochemical constraints.

The crystallographic R-factor

During the cycles of refinement, we calculate "R factors" to assess the progress. These are similar to the Q values used in minimisation.

$$Q = \sum \left[F_{obs} - F_{calc} \right]^2$$

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

R = "residual" = fractional difference between observed and calculated diffraction - a sort of "fractional error"

To monitor the refinement, we calculate R after each cycle (the summation is over all reflections). If things are going well, R reduces.

However, with complicated refinements like these, it is possible to "over-fit" the data – for cross-validation we take away 5% of the data (which we do not use in the refinement, to monitor the agreement. This is known as R_{free}

Practical considerations

Protein production for crystallography
- Natural sources
- Recombinant technology

Crystallisation
- The theory
- Techniques
- Crystallisation robots

X-rays generation and data collection

How to read a PX paper

How to use a PDB

Natural vs recombinant sources

Recombinant DNA technology makes it easier to produce large quantities of purified protein.

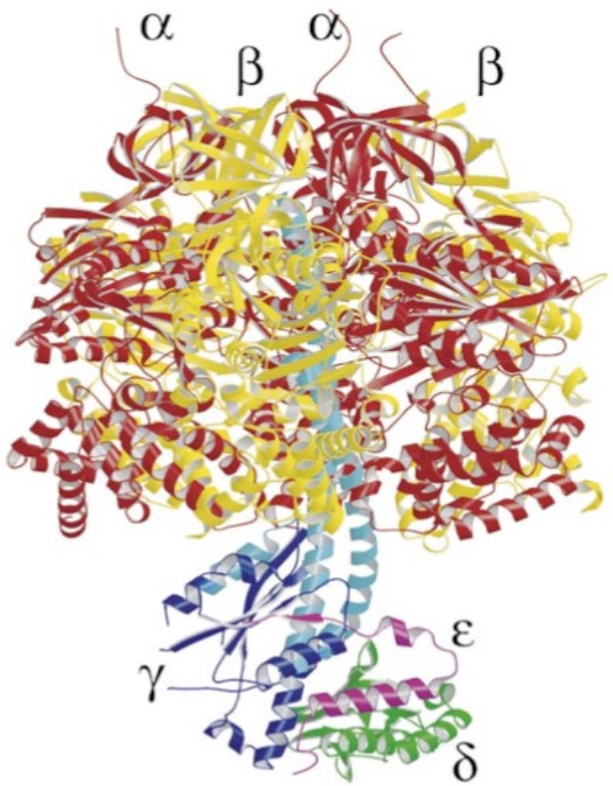
However “natural” sources still useful especially for big complexes, since these cannot be easily re-constituted by recombinant expression.

Examples:

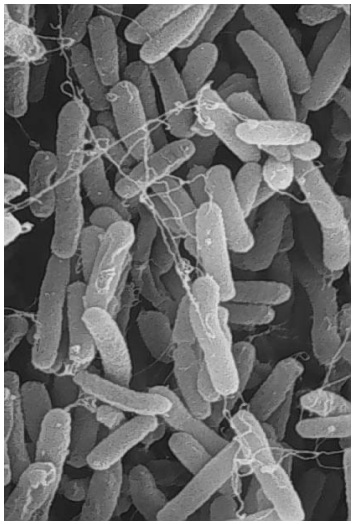
F_oF_1 ATP synthase - purified from
bovine heart muscle/yeast cells

ribosomes - purified from archaea

RNAPII - purified from yeast cells



Expression in Escherichia coli



First choice organism:

- cloning and handling is easy
- grows rapidly in liquid cultures (doubling time < 30 min)
- 1-6 L of culture can yield 5-50 mg of purified protein
- well characterised organism

Not foolproof…

- Difficult to make proteins with disulphide bridges
- Protein may over-express well but be unfolded
- Control folding with temperature and/or chaperones
- May be able to “refold” proteins by controlled denaturation/renaturation during purification - but tricky!
- Proteins may be toxic for bacteria - try to get around this by tight control of expression

Easy/difficult to crystallise?

A lot of interesting proteins are difficult to crystallise.

Best cases:

- single proteins
- rigid domain structure
- one dominant conformation

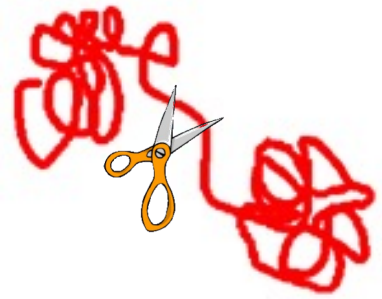
Worst cases:

- multi-domain proteins with flexible unstructured linkers
- proteins with flexible N- and/or C-termini
- proteins that are part of large macromolecular complexes
- presence of posttranslational modifications

“Optimise” proteins for crystallisation

Use bioinformatics (database searches, sequence alignments) to identify “core domains” that can be expressed in a soluble form.

Use limited proteolysis to identify compact domains.



beware:
you can
cut flexible
loops).

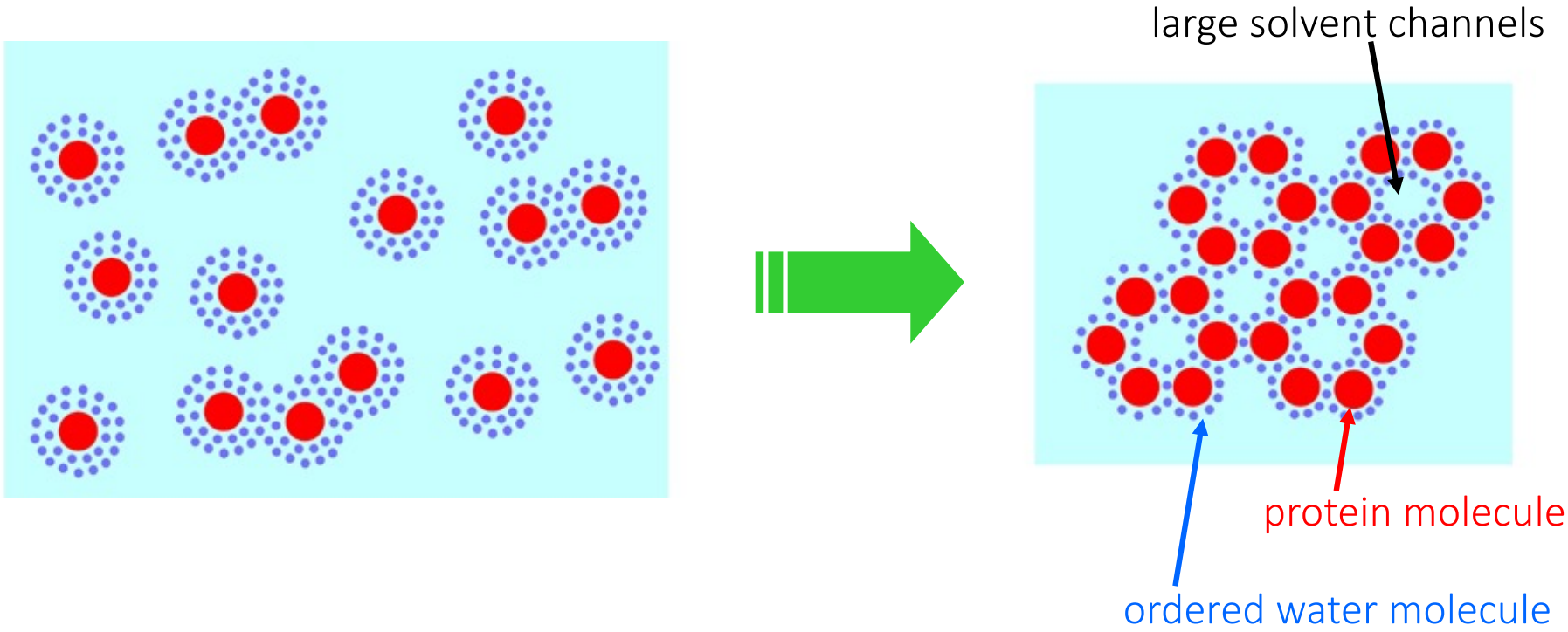
Co-express proteins that are part of the same complex.

Add ligands/inhibitors/cofactors/metals
to stabilise one conformation.

Avoid/encourage post-translational modifications (mutagenesis of target residues, mimic modifications by mutagenesis, change expression system/cell lines, so that the process does/does not occur).

Crystallisation

Protein crystallisation is the transfer of protein molecules from an aqueous solution to an ordered solid phase.



Protein crystallisation is controlled by:

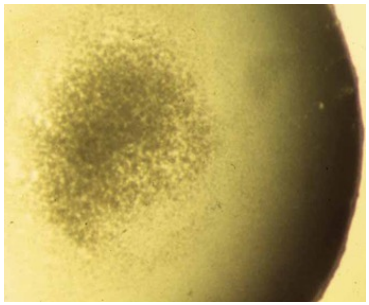
- thermodynamic factors governing the solubility
- kinetic factors governing nucleation and growth

Solubility and crystals

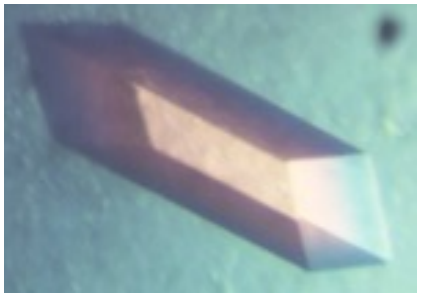
To coax a protein into forming crystals, one has to decrease its solubility to the point where the solution becomes saturated. This is done by changing pH, ionic strength or temperature, or by adding organic solvents.

In a saturated solution the protein in solution is in thermodynamic equilibrium with one or more solid states:

- amorphous precipitate →

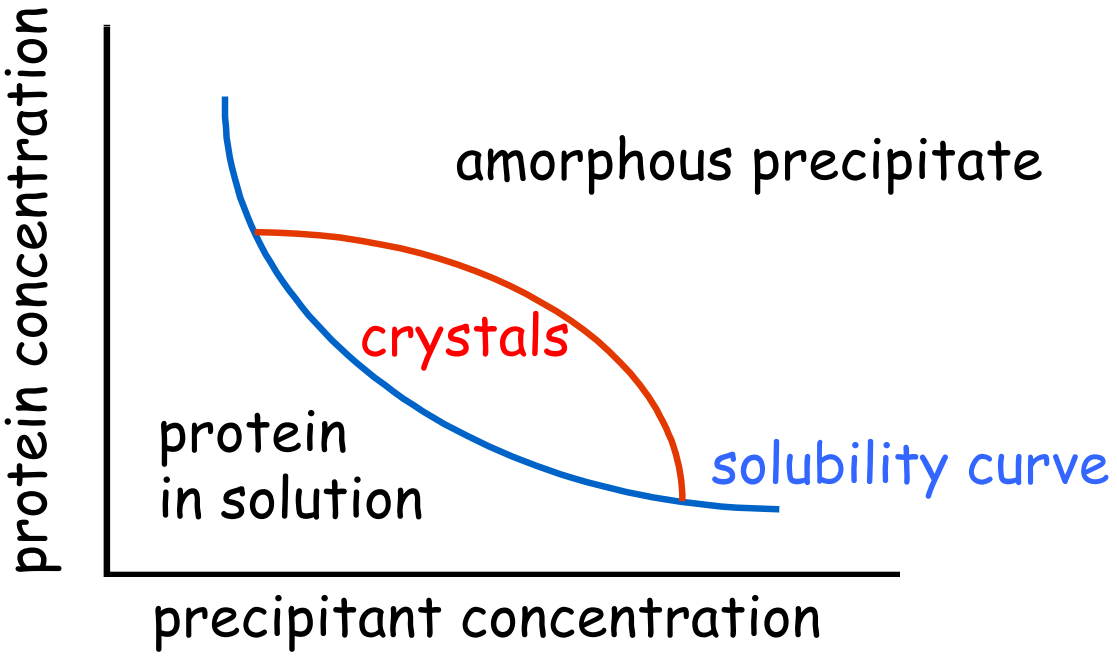


- crystalline state
  - micro-crystals
  - needles
  - plates
  - large single crystals →



Phase diagram

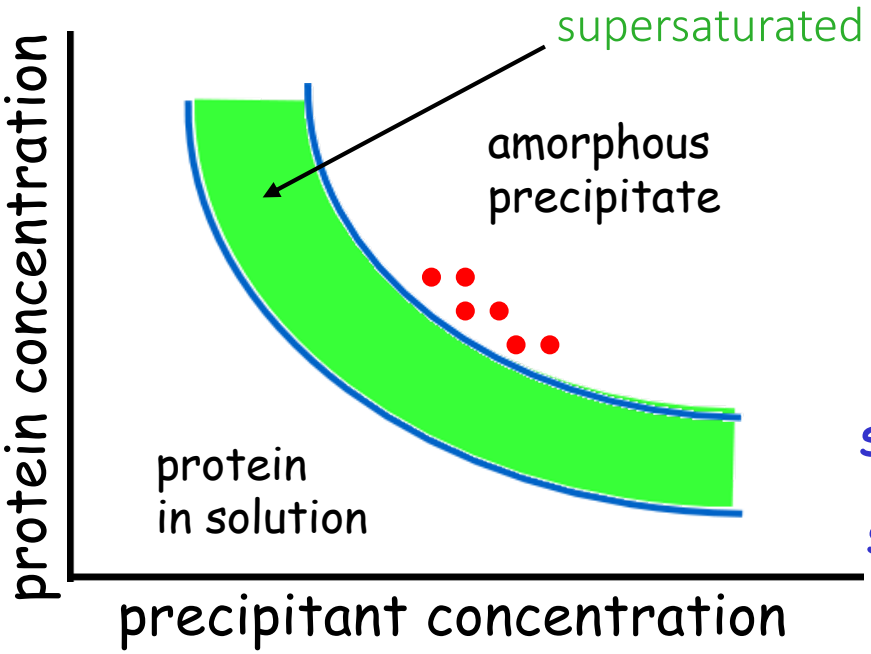
The process of crystal formation can be understood by using a phase diagram. Because several factors influence protein solubility, the phase diagram has many dimensions. A very important factor is the precipitant concentration, and a two-dimensional diagram is shown here:



Supersaturation

Above the solubility curve the solid phase is thermodynamically favoured; yet, because of kinetic barriers to the formation of aggregates, the protein may remain in solution, forming a supersaturated solution.

A supersaturated solution is thermodynamically metastable: its conversion to a stable solid state is kinetically controlled and can be triggered by the presence of nucleation centres.



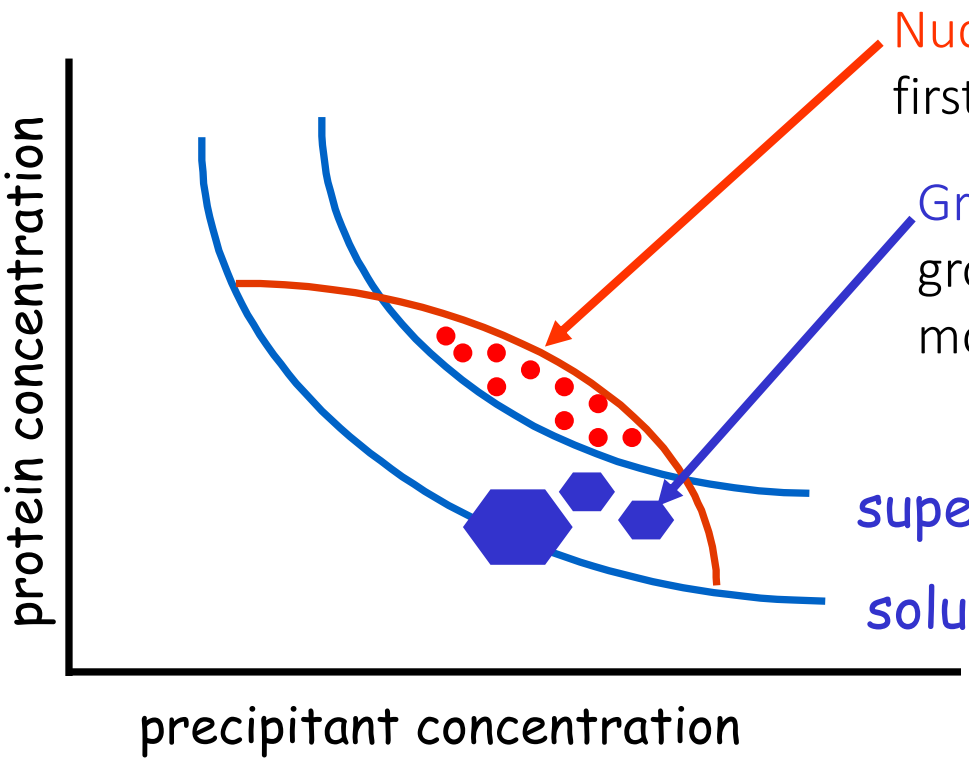
Above a certain level a supersaturated solution will spontaneously give rise to the formation of nuclei, without any need of 'external' nucleation centres.

supersolubility curve

solubility curve

Nucleation and growth

The formation of crystals requires aggregates of a critical size (nuclei). Aggregates smaller than the critical size are unstable and will dissolve, while aggregates larger than the critical size will grow.
The process of crystallisation can be divided into two distinct stages:



Nucleation: the formation of the first few ordered aggregates

Growth: the "most fit" aggregates grow in size by addition of protein molecules from the solution

supersolubility curve

solubility curve

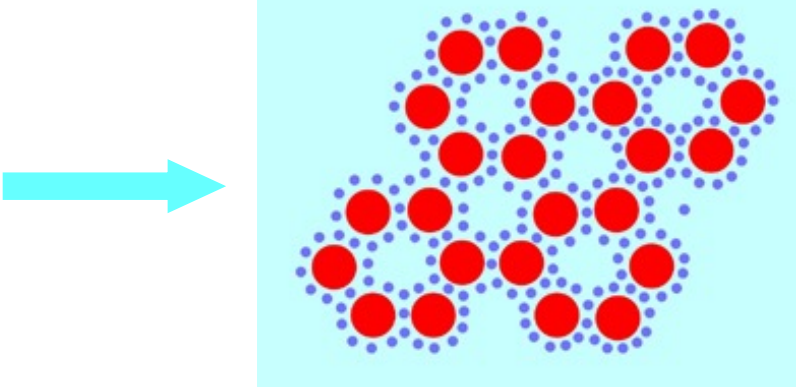
Crystallising proteins

Why proteins are difficult to crystallise:

- most proteins are labile and easily denatured
- large proteins often exist in multiple conformations
- complex behaviour -> polymorphism
- many proteins are difficult to obtain in large amounts
- proteins needs to be highly purified for crystallisation

Why protein crystals are difficult to handle:

- high solvent content (30-80%)
- mechanically fragile
- not well ordered -> resolution limits
- sensitive to radiation damage



Crystallisation of membrane proteins may presents additional problems such as homogeneity of the purified samples, choice of detergents, presence of micelles, tendency to form hydrophobic interactions which are less directional and ordered, etc..

Crystallising proteins

Purity is not an absolute requirement for crystallisation, but often crystals obtained from impure solutions are small or of poor quality.

The protein solution has to be homogeneous not only in terms of composition, but also in terms of structure and conformation.

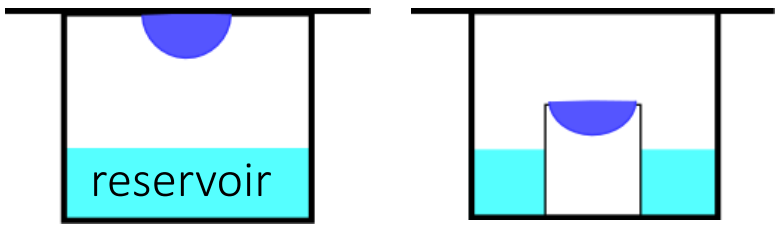
There are many factors to consider. Some of the more important ones are:
- presence of ligands (substrates, cofactors, metal ions etc)
- flexible domains and loops
- post-translational modifications (glycosylation, phosphorylation etc)
- aggregation

Highly homologous proteins will typically crystallise in very different conditions.

One needs large amounts of very pure protein that can be concentrated to several mg/ml without aggregation.
Robotic screening allows testing of ~1000 conditions/mg of protein.

Vapour diffusion

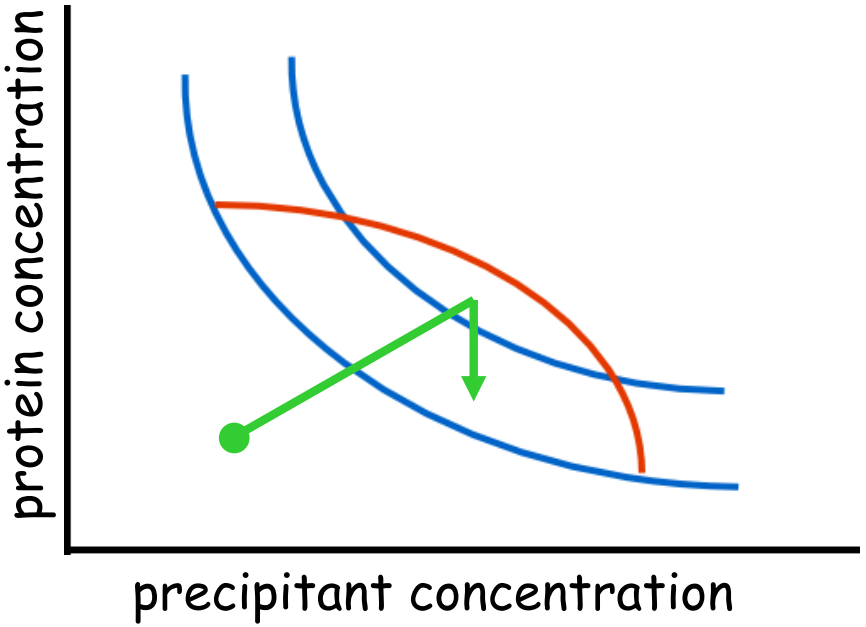
A droplet containing the protein and the precipitating agent is equilibrated against a reservoir containing the precipitant at higher concentration. Equilibrium is reached through the vapour phase, usually by loss of water from the droplet.



hanging drop

sitting drop

Typically the drop is made by mixing equal volumes of protein and reservoir solution.

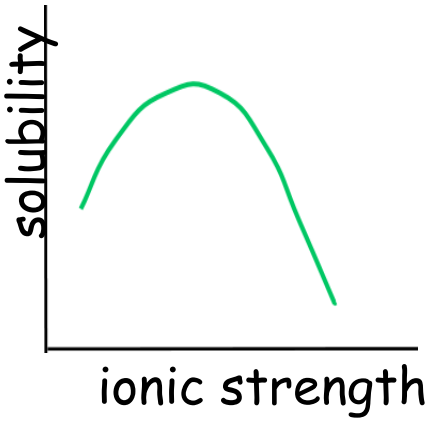


Vapour diffusion is best suited for robotic nanolitre crystallisation.

Precipitants

SALTS

Solubility changes with the ionic strength of the solution, causing salting in/salting out effects
Common ionic precipitants include $(NH4)_2SO_4$, various phosphates, NaCl, LiCl, and many others.



ORGANIC SOLVENTS

Reduce the dielectric constant of the medium, thereby enhancing the electrostatic interactions between the protein molecules and lowering the solubility. May denature proteins.
Commonly used organic solvents include ethanol and isopropanol.

POLYETHYLENE GLYCOLS (PEGs)

Believed to act by a combination of effects: volume exclusion, modification of the dielectric properties of the solution, electrostatic interaction with proteins, depletion of the hydration sphere of the protein. Very effective

Salts and PEGs are the most common precipitants

High throughput crystallisation

Automated methods for crystallisation (and crystal visualization) are now routinely used by most labs. These development are driven by the needs of pharmaceutical companies and structural genomics projects.

Robotic crystallization systems relies on the same principles described so far, but carry out each step of the procedure quickly, accurately, in smaller volumes.

Manual crystallisation:

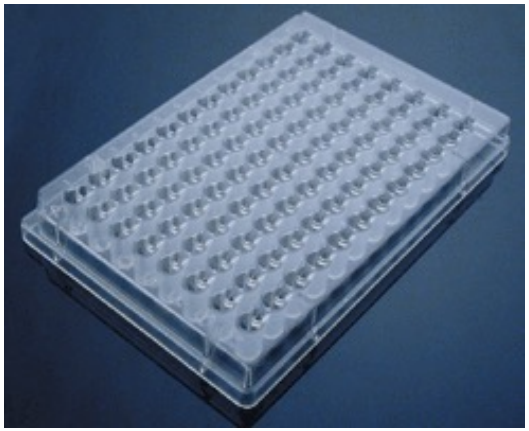
- slow and time consuming
- error prone and not always reproducible
- expensive in terms of amount of purified protein (drops: 1-2 μl)

Robotic crystallisation:

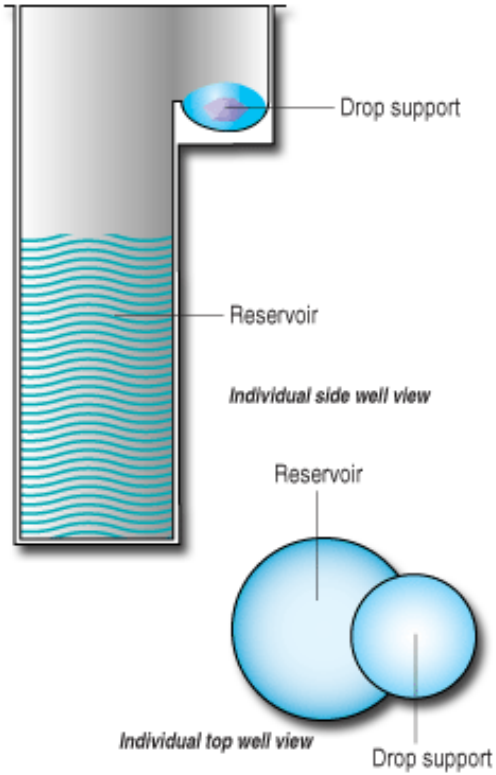
- faster and more efficient
- more accurate and reproducible
- smaller sample sizes (down to 100 nl drops) cut down on expenditure of purified protein

96-wells plates

The standard format for automated crystallisation are plates with 96 reservoir wells. These plates are designed for the sitting drop vapor diffusion method, and are sealed with clear tape.



8 x 12 matrix

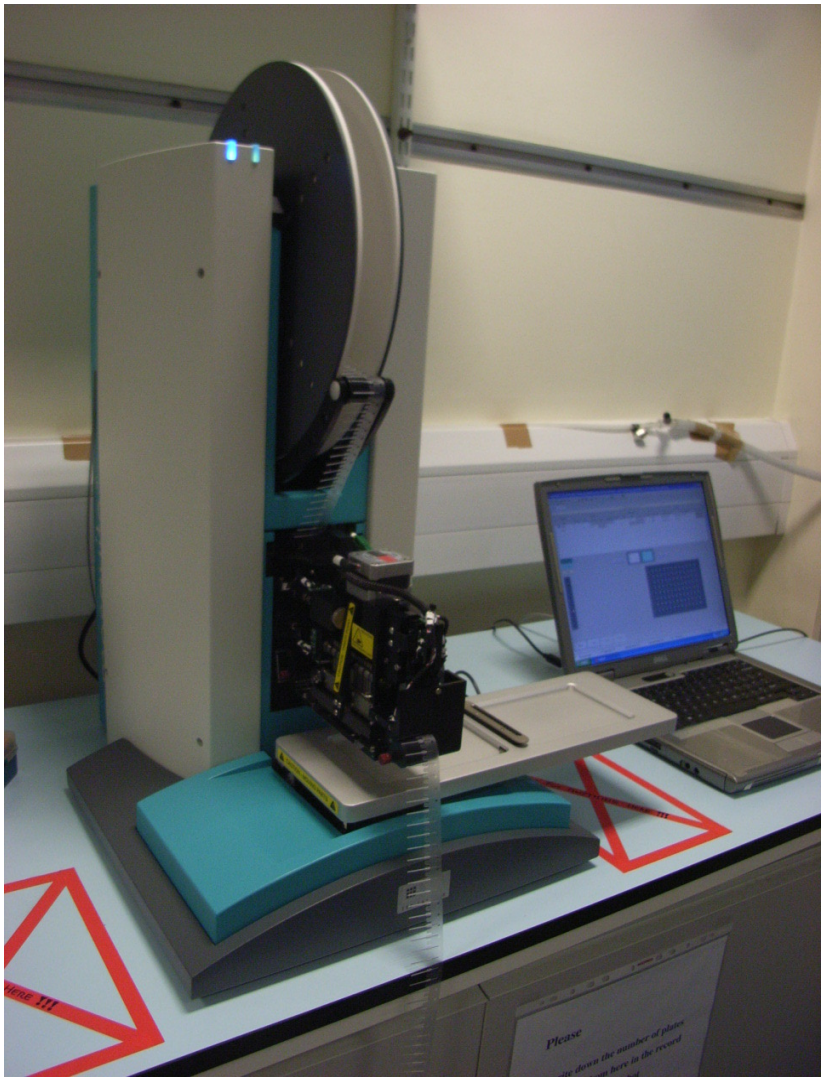


Mosquito crystallisation robot

To set up the crystallisation drops containing nanolitre volumes of protein and well solution.

Employs disposable tips and pipettes 100-200 nl drops.

Takes only 2 min to set up 96 drops.



Use of specialised 96-well plates