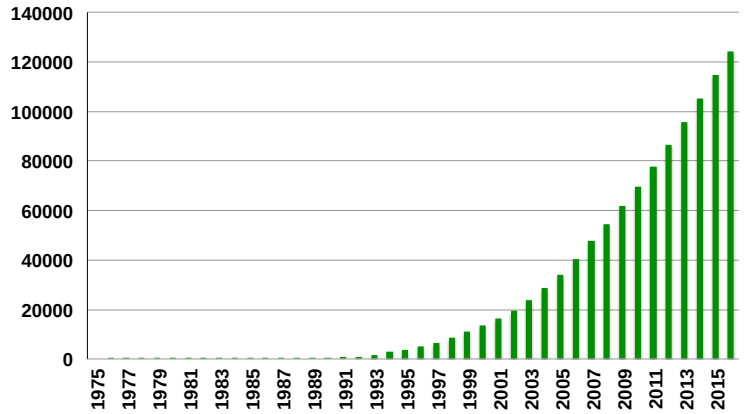
## Number of structures available in the PDB per year

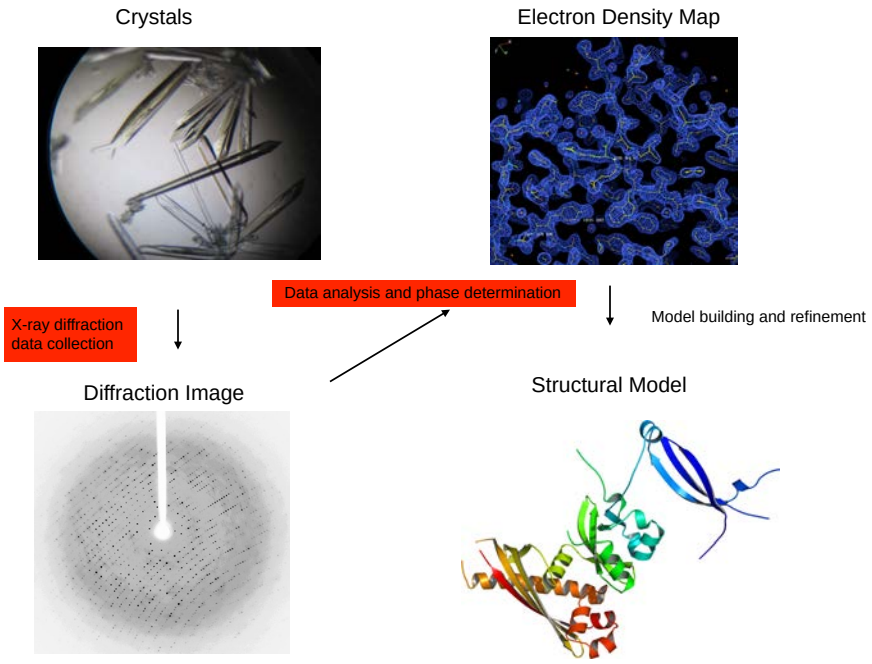



By Nov. 14th, 2016, total structures deposited in PDB: 124,286

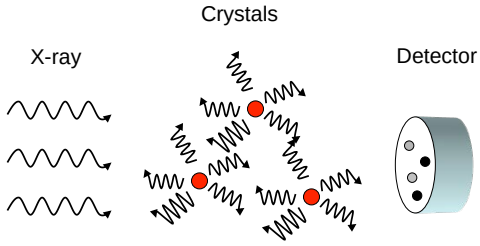
X-ray Crystallography: 111,122 (89%)

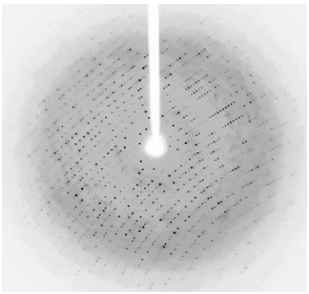
Solution NMR: 11,545 (9%)

Electron Microscopy: 1,232 (1%)



# X-ray Diffraction



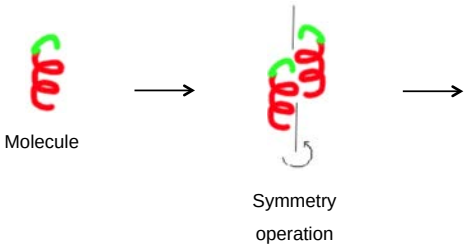



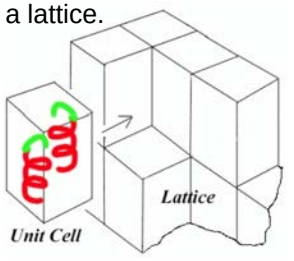
Each diffraction spot represents a *wave*.

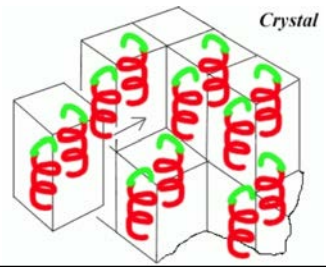
The diffraction pattern (the *position* and *intensity* of each diffraction spot) gives information on the arrangement of the atoms in the crystal.

# Crystal

Three-dimensional *periodic* arrangement of a molecule in a repeating unit (so called the unit cell) into a lattice.

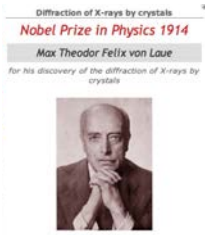


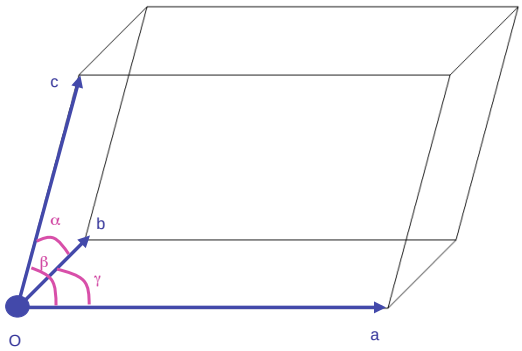



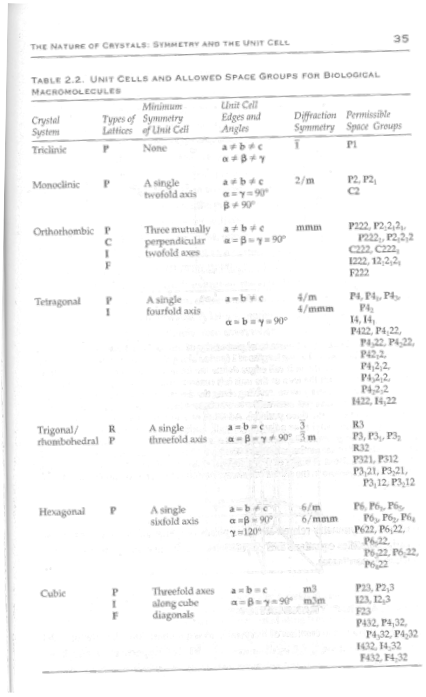




Only the size and contents of *one unit cell* are necessary to describe the entire crystal.

Unit Cell



THE NATURE OF CRYSTALS: SYMMETRY AND THE UNIT CELL 35

TABLE 2.2. UNIT CELLS AND ALLOWED SPACE GROUPS FOR BIOLOGICAL MACROMOLECULES

| Crystal System | Types of Lattices | Minimum Symmetry of Unit Cell | Unit Cell Edges and Angles | Diffraction Symmetry | Permissible Space Groups |
|---|---|---|---|---|---|
| Triclinic | P | None | $a \neq b \neq c$<br>$\alpha \neq \beta \neq \gamma$ | $\bar{1}$ | P1 |
| Monoclinic | P | A single twofold axis | $a \neq b \neq c$<br>$\alpha = \gamma = 90°$<br>$\beta \neq 90°$ | 2/m | P2, $P2_1$<br>C2 |
| Orthorhombic | P<br>C<br>I<br>F | Three mutually perpendicular twofold axes | $a \neq b \neq c$<br>$\alpha = \beta = \gamma = 90°$ | mmm | P222, $P2_12_12_1$<br>$P222_1$, $P2_12_12$<br>C222, $C222_1$<br>I222, $I2_12_12_1$<br>F222 |
| Tetragonal | P<br>I | A single fourfold axis | $a = b \neq c$<br>$\alpha = b = \gamma = 90°$ | 4/m<br>4/mmm | P4, $P4_1$, $P4_2$, $P4_3$<br>I4, $I4_1$<br>P422, $P4_12_12$, $P4_22_12$, $P4_32_12$, $P42_12$, $P4_122$, $P4_222$, $P4_322$, $P4_22_12$<br>I422, $I4_122$ |
| Trigonal/ rhombohedral | R<br>P | A single threefold axis | $a = b = c$<br>$\alpha = \beta = \gamma \neq 90°$ | $\bar{3}$<br>$\bar{3}$m | R3<br>P3, $P3_1$, $P3_2$<br>R32<br>P321, P312<br>$P3_121$, $P3_221$, $P3_112$, $P3_212$ |
| Hexagonal | P | A single sixfold axis | $a = b \neq c$<br>$\alpha = \beta = 90°$<br>$\gamma = 120°$ | 6/m<br>6/mmm | P6, $P6_1$, $P6_5$, $P6_2$, $P6_4$, $P6_3$<br>P622, $P6_122$, $P6_522$, $P6_222$, $P6_422$, $P6_322$<br>$P6_122$, $P6_522$, $P6_222$, $P6_422$, $P6_322$ |
| Cubic | P<br>I<br>F | Threefold axes along cube diagonals | $a = b = c$<br>$\alpha = \beta = \gamma = 90°$ | m3<br>m3m | P23, $P2_13$<br>I23, $I2_13$<br>F23<br>P432, $P4_232$, $P4_332$, $P4_132$<br>I432, $I4_132$<br>F432, $F4_132$ |

7 Crystal Systems

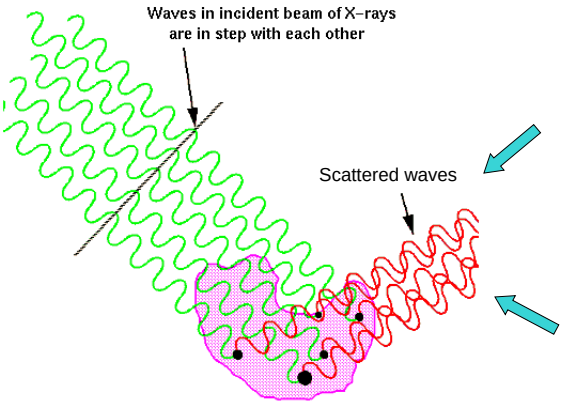
14 Bravais Lattices

65 Space Groups

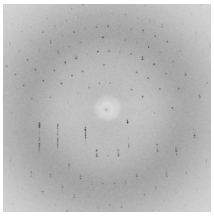
From Crystal to Diffraction

## Wave interference

Scattering from the atoms in a molecule




Waves add in some directions (Constructive)

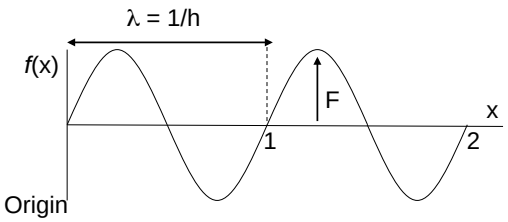


Waves subtract in other directions (Destructive)

***What are the conditions that produce diffraction? – Bragg's Law***

## A simple wave can be described by a *periodic* function

$f(x) = F \sin 2\pi (hx + \alpha)$




$\alpha = 0$

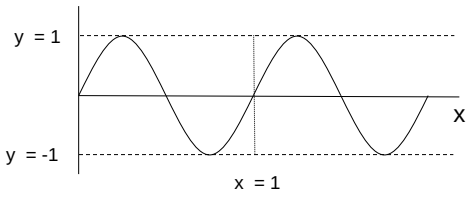
f(x) → vertical height of the wave at any horizontal position x along the wave, measured in wavelengths, where x=1 implies one full wavelength.

*F* → *amplitude* (half its height from peak to valley)

h → *frequency* (number of wavelengths per radian)

$\alpha$ → *phase* (position of the wave, in radian, with respect to the origin)

## $f(x) = F \sin 2\pi(hx+\alpha)$



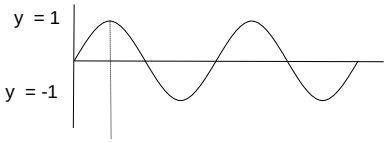

*F* = 1
h = 1
$\alpha = 0$

$f(x) = \sin 2\pi x$

$\lambda = 1/h = 1$

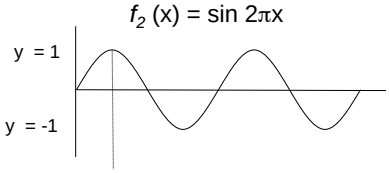
## Constructive interference

$f_1(x) = \sin 2\pi x$



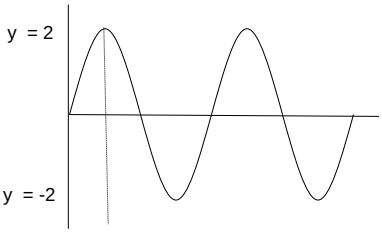

x = 1/4

$f_2(x) = \sin 2\pi x$




x = 1/4

$f_1 + f_2 = 2\sin 2\pi x$

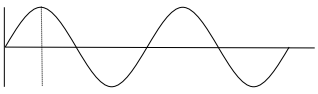



x = 1/4

F = 2

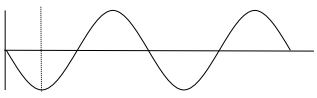
## Destructive interference

$f_1(x) = \sin 2\pi x$



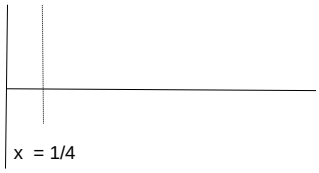
x = 1/4

$f_2(x) = \sin (2\pi x + \pi)$



x = 1/4

$f_1 + f_2 = 0$

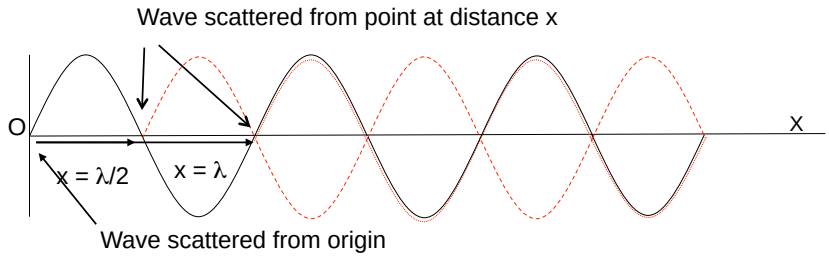


x = 1/4

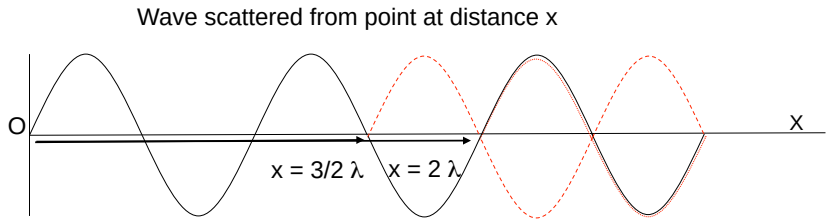
F = 0

*Phase differences* determine the interference.

What is the relationship between *wave phase* differences and *distance* between scattering atoms?




What is the relationship between *wave phase* differences and *distance* between scattering atoms?

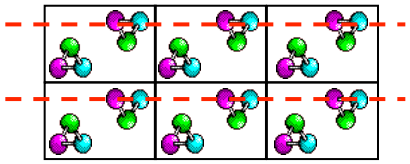



For constructive interference:

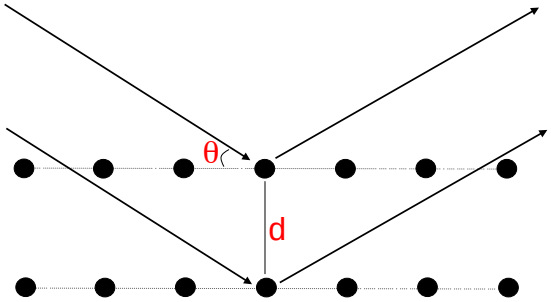
Path distance must be an integral number of wavelengths

$\Delta x = n\lambda$

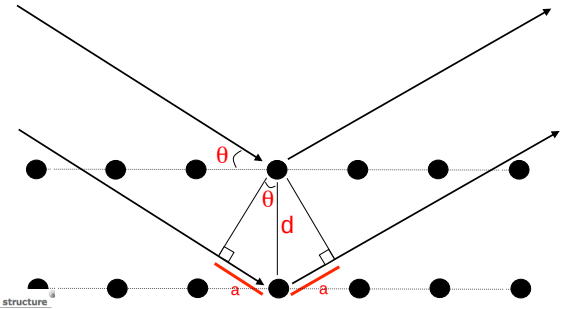
This is the central idea embodied by Bragg's Law



Think of a crystal as a set(s) of equivalent parallel planes of atoms.




What is the relationship between d and θ such that these waves add constructively?

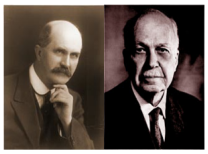



Use of X-rays to determine crystal structure

*Nobel Prize in Physics 1915*

*Sir William Henry Bragg and Sir William Lawrence Bragg*

*for their services in the analysis of crystal structure by means of X-rays*



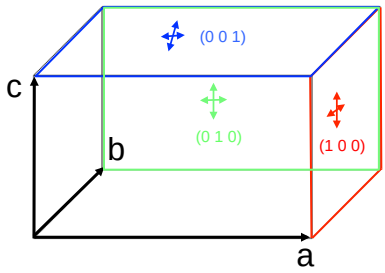
$\Delta x = n\lambda$

$2a = n\lambda$

$2d\sin\theta = n\lambda$ – Bragg's Law

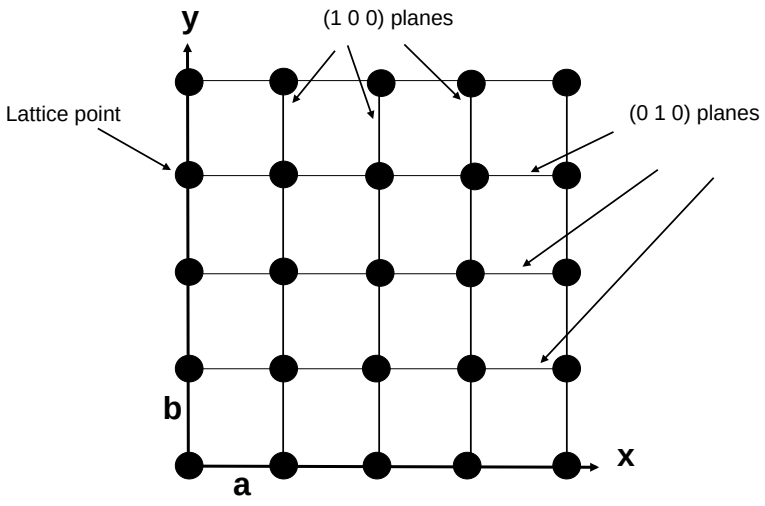
*How to describe the planes in a crystal?*

## *Lattice indices (h k l)* of the atomic planes in a crystal

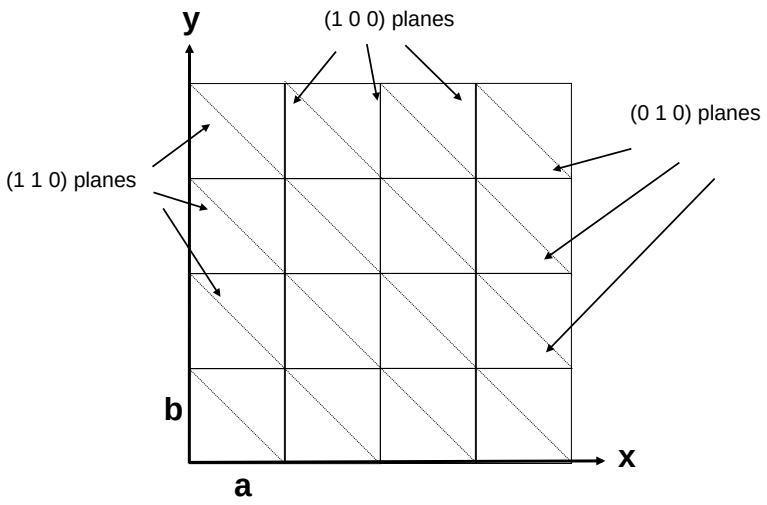


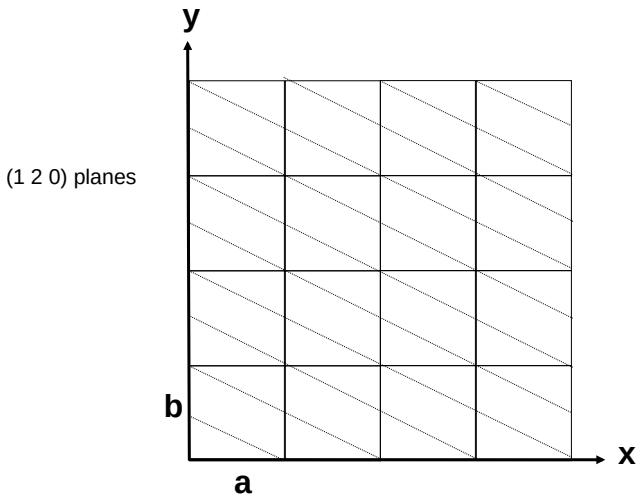

- Three indices **hkl** identify a particular set of equivalent, parallel planes.
- The index **h** gives the number of planes in the set per unit cell in the **a** direction or, equivalently, the number of parts into which the set of planes cut the **a** edge of each cell.
- The indices **k** and **l** specify how many such planes exist per unit cell in the **b** and **c** directions.

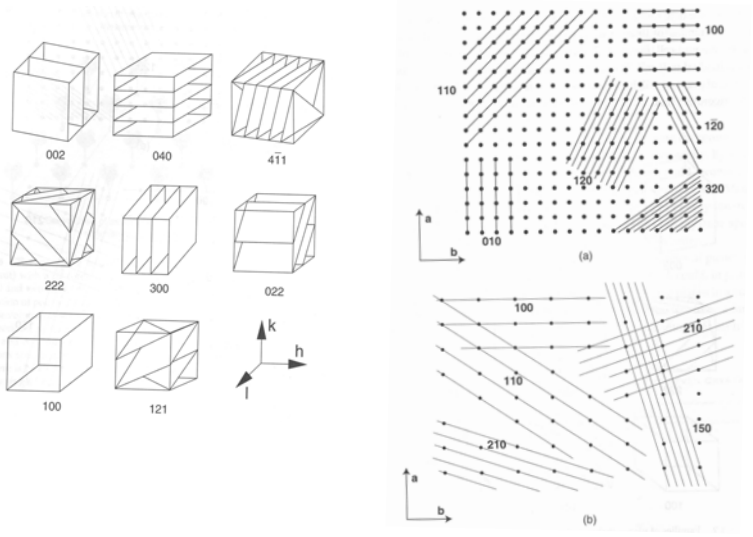
## Planes apply to the whole lattice



Planes apply to the whole lattice






*How to describe the diffraction pattern?*

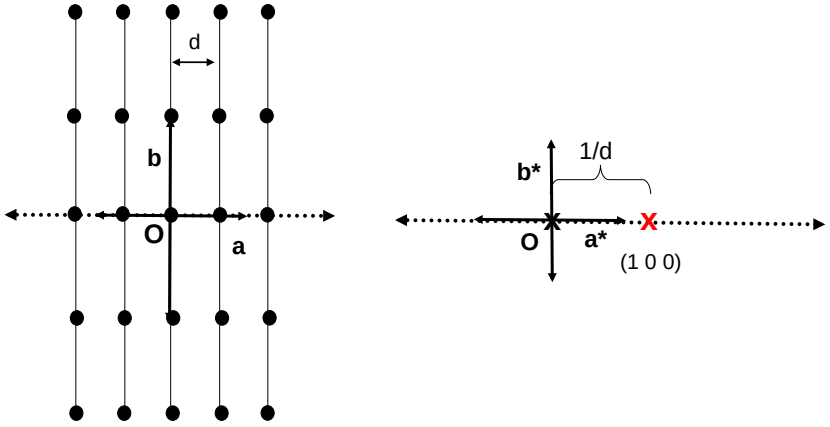
# Construction of *a reciprocal lattice*

A reciprocal lattice:

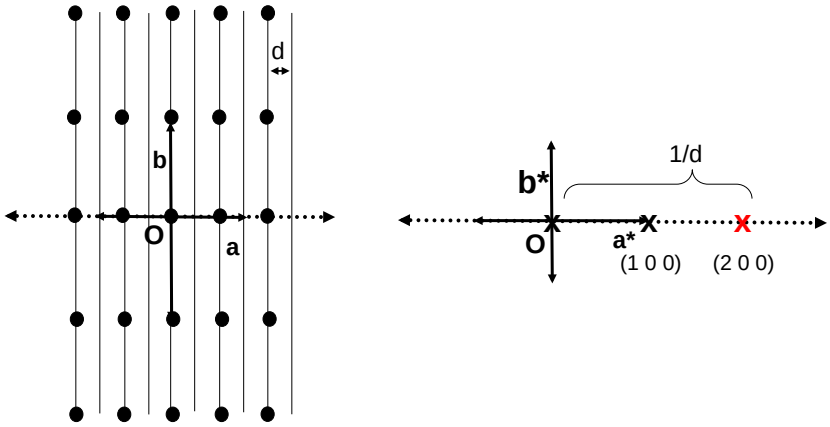
- Its *lattice dimensions* are *reciprocal* to the original cell (and correspond to *the reflection positions*);
- Its 'size' (the *intensity* of the reflection) corresponds to the *contents* of the unit cell.

1) Choose coordinate axes for reciprocal space that are identical to real space, and origin can be arbitrary.
2) For each set of lattice planes (h k l), draw a line from the origin that is orthogonal to the planes.
3) Place reciprocal lattice point (h k l) at a distance 1/d from origin, where d is the interplane spacing.

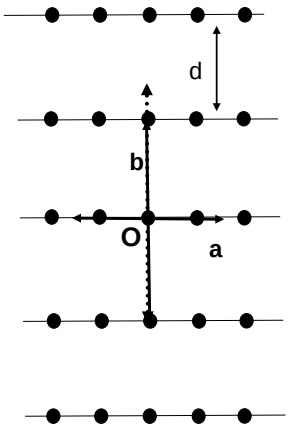
(1 0 0) planes

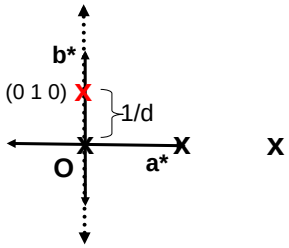



(2 0 0) planes

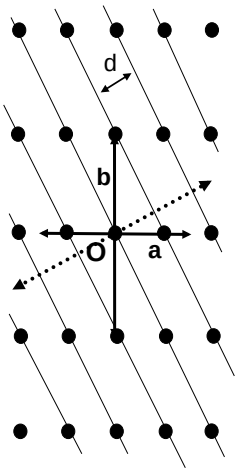


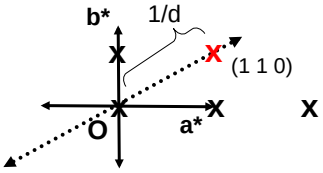
(0 1 0) planes



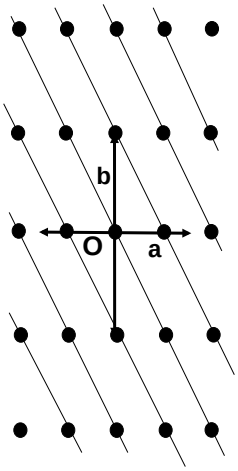



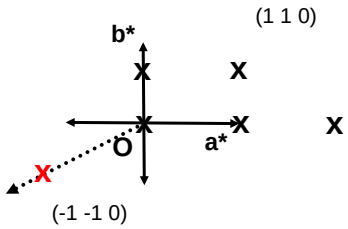

(1 1 0) planes

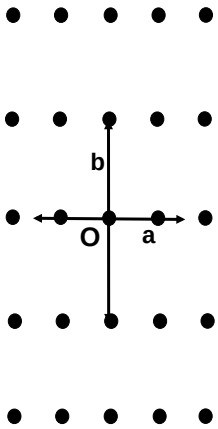


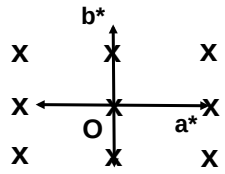



(1 1 0) planes = (-1 -1 0) planes

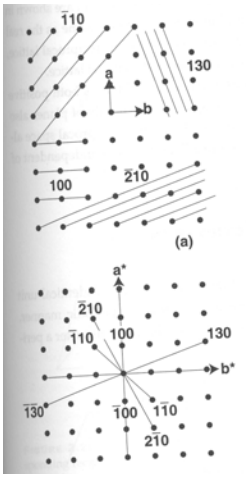



Reciprocal lattice is symmetric about origin




Crystal lattice

(in real space)

Reciprocal lattice

(in diffraction space)

- *An entire set of parallel planes* (h k l), not just one plane, acts as *a single diffractor* and produces *one reciprocal lattice point*.

## The reciprocal lattice

The mathematical definition of the reciprocal lattice constants is:

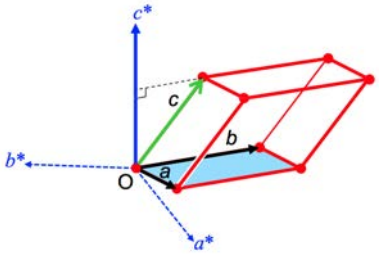
$$a^* = \frac{b \times c}{a \cdot (b \times c)}$$

$$b^* = \frac{c \times a}{b \cdot (c \times a)}$$

$$c^* = \frac{a \times b}{c \cdot (a \times b)}$$

V = unit cell volume

$= a \cdot (b \times c) = b \cdot (c \times a) = c \cdot (a \times b)$

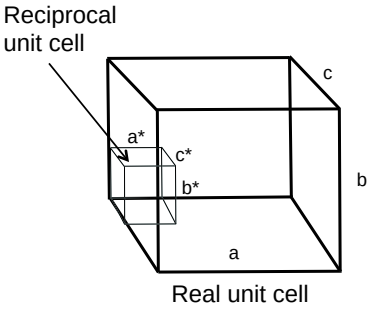



The reciprocal lattice has *the same symmetry* as the crystal lattice.

*The larger* the crystal unit cell, *the smaller* the reciprocal lattice.

In orthogonal crystal lattices ($\alpha = \beta = \gamma = 90°$)

$$a^* = \frac{1}{a} = \frac{1}{d_{100}}$$
$$b^* = \frac{1}{b} = \frac{1}{d_{010}}$$
$$c^* = \frac{1}{c} = \frac{1}{d_{001}}$$

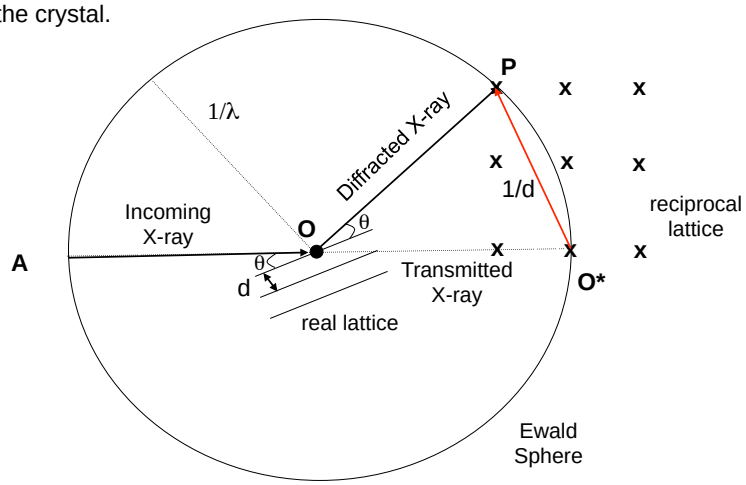



The reciprocal lattice is spatially linked to the crystal because of the way the lattice points are defined, so *if we rotate the crystal, the reciprocal lattice rotates with it*.

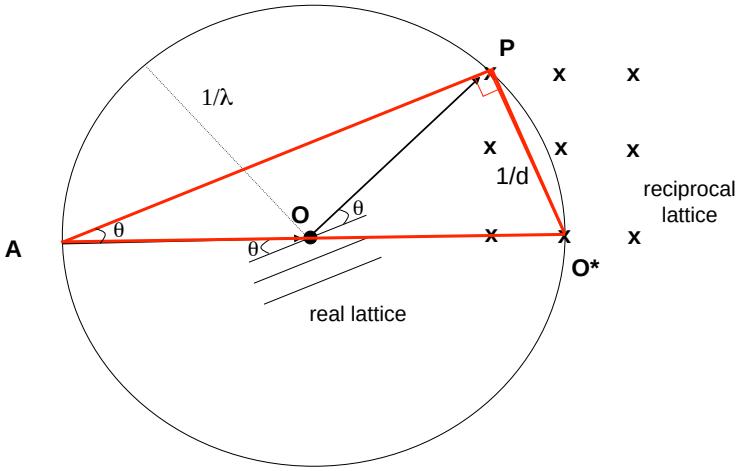
*How to link the reciprocal lattice with the diffraction pattern?*

## Bragg's law in reciprocal space – *Ewald Sphere*

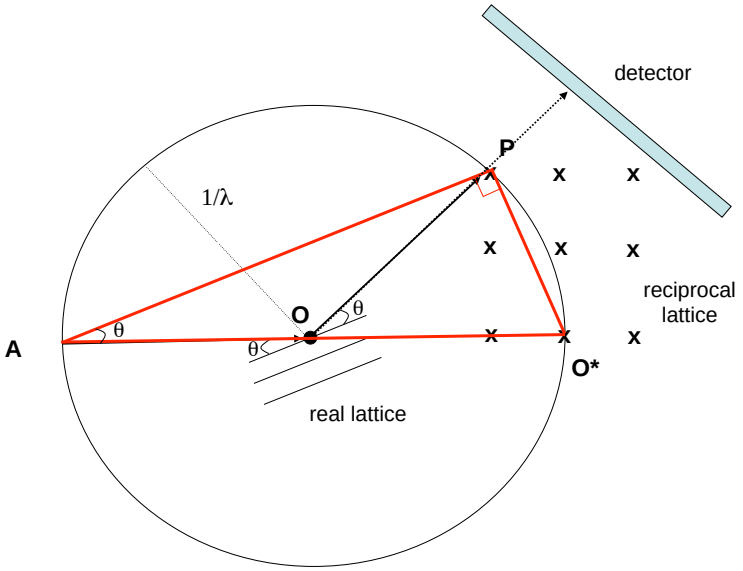
– To show how each reciprocal-lattice point must be arranged with respect to the X-ray beam in order to satisfy Bragg's law and produce a reflection from the crystal.



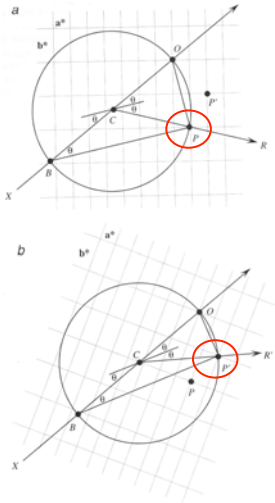
When is Bragg's law ($n\lambda = 2d\sin\theta$) satisfied?




$$\sin\theta = \frac{PO^*}{AO^*} = \frac{1/d}{2/\lambda} = \frac{\lambda}{2d} \rightarrow 2d\sin\theta = \lambda$$

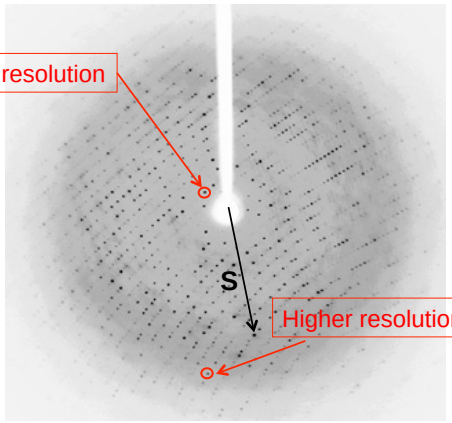



Any reciprocal lattice point on the *Ewald sphere* follows the Bragg's law and produces diffraction!



- *An entire set of parallel planes* (h k l), not just one plane, acts as *a single diffractor* and produces *one reflection hkl*.
- Whenever crystal is rotated so that a reciprocal lattice point comes in contact with the *Ewald sphere*, Bragg's law is satisfied and a reflection occurs.
- At a certain orientation of the crystal, only a small number of reflections are in "diffraction condition". Therefore, the crystal has *to be rotated* to collect all the reflections.
- The *directions* of the reflections, as well as *number* of the reflections, depend only on the *unit-cell dimensions* and not upon the contents of the unit cell.
- The *intensity* of that reflection depends upon the *electron distribution and density* along the planes that produce the reflection.

## Diffraction pattern is an image of the reciprocal lattice



$|S| = 1/d$

$2d \sin\theta = \lambda$

$d = \lambda/2\sin\theta$

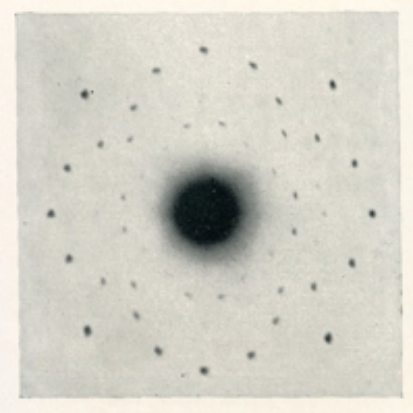
when $\theta = 90°$,

$\sin\theta = 1$

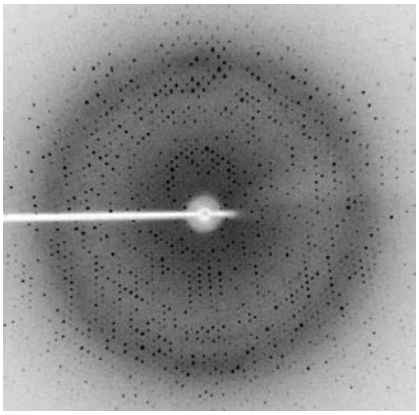
$d_{min} = \lambda/2$

1) d is spacing between real crystal lattice planes ~ Small d means high resolution.
2) Diffraction position reflects only geometry of crystal and experiment, not contents of unit cell.
3) The relative intensities of the diffractions contain information about the structure of the molecule.

Sea salt



Protein



In reciprocal space, the lattice spacing is inversely proportional to the interplanar spacing within the crystal. *The larger* the crystal unit cell, *the smaller* the reciprocal lattice.

## From diffraction data to electron density

The Fourier transform describes precisely the mathematical relationship between an *object* and its *diffraction pattern*.

The Fourier transform allows us to convert a Fourier-series description of the *reflections* to a Fourier-series description of the *electron density map*.

A reflection can be described by a *structure-factor* equation, containing one term for each atom (or each volume element) in the unit cell. In turn, the *electron density of each atom* is described by a Fourier series in which each term is a *structure factor*.

We use the Fourier transform to convert the structure factors $F(h, k, l)$ to $\rho(x, y, z)$, the desired electron density equation.

*How to construct a Fourier series?*

## One-dimensional waves

Fourier term (a simple wave)

$f(x) = F \cos 2\pi(hx + \alpha)$

Or

$f(x) = F \sin 2\pi(hx + \alpha)$

$f(x)$ the vertical height of the wave at any horizontal position x;
$F$ the amplitude of the wave;
$h$ frequency;
$\alpha$ phase.

Fourier series (Sum of simple waves)

$$f(x) = F_0 \cos 2\pi(0x + \alpha_0) + F_1 \cos 2\pi(1x + \alpha_1) + \ldots + F_n \cos 2\pi(nx + \alpha_n)$$
$$= \sum_{h=0}^{n} F_h \cos 2\pi(hx + \alpha_h)$$

As $f(x) = F[\cos 2\pi(hx) + i \sin 2\pi(hx)]$, $\cos\theta + i \sin\theta = e^{i\theta}$, $\theta = 2\pi(hx)$,

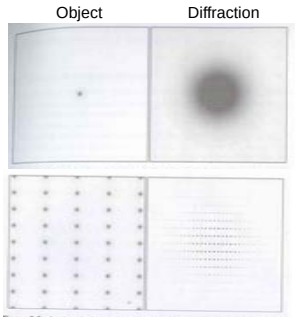
$$f(x) = \sum_h F_h[\cos 2\pi(hx) + i \sin 2\pi(hx)] = \sum_h F_h e^{2\pi i(hx)}$$

## Three-dimensional waves

$$f(x, y, z) = \sum_h \sum_k \sum_l F_{hkl} e^{2\pi i(hx + ky + lz)}$$

## The Fourier transform:

- Is an operation that transforms a function containing variables of one type (say time or a length in Å) into a function whose variables are *reciprocals* of the original type (in this case, 1/time or frequency, or reciprocal length in $Å^{-1}$).
- Is a precise mathematical description of diffraction. The diffraction patterns (in terms of structure factor) are Fourier transforms of the corresponding objects and arrays (in terms of electron density map).




$$F(h, k, l) = \int_x \int_y \int_z f(x, y, z) e^{2\pi i(hx + ky + lz)} dx dy dz$$

Inverse Fourier transform

$$f(x, y, z) = \int_h \int_k \int_l F(h, k, l) e^{-2\pi i(hx + ky + lz)} dh dk dl$$

**Diffraction pattern (Structure factor)** in reciprocal space

Inverse Fourier transform

**Unit cell content (Electron density) in real space**

## Structure factor as a Fourier series

A structure factor describes *one diffracted x-ray*, which produces *one reflection* received at the detector. It can be written as a Fourier series in which each term gives the contribution of one atom to the reflection hkl.

Fourier term
(One atom)

$$f_{hkl} = f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

Fourier series
(Sum of all atoms)

$$F_{hkl} = \sum_{j=0}^{n} f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

$f_{hkl}$ atomic structure factor, the contribution of the single atom *j* to reflection hkl;

$f_j$ scattering factor of atom *j* – a function that amounts to treating the atom as a simple sphere of electron density.

$x_j$, $y_j$, $z_j$ the coordinates of atom *j* in the unit cell (real space).

*h*, *k*, *l* the indices of a specific reflection in the reciprocal lattice satisfying Bragg's law.

phase $2\pi(hx_j + ky_j + lz_j)$ for atom *j*.

- Each diffraction ray is the sum of diffractive contributions from *all atoms* in unit cell.
- The contribution of each atom *j* to $F_{hkl}$ depends on (1) what *element* it is, which determines $f_j$, the amplitude of the contribution, and (2) its *position* in the unit cell ($x_j$, $y_j$, $z_j$), which establishes the phase of its contribution.

$F_{hkl}$ is a complex number!

## Structure factor as a Fourier series

Alternatively, $F_{hkl}$ can be written as the sum of contributions from each volume element of electron density in the unit cell.

$$F_{hkl} = \int_x \int_y \int_z \rho(x,y,z) e^{2\pi i(hx + ky + lz)} dx dy dz$$

$$= \int_{cell\ V} \rho(x,y,z) e^{2\pi i(hx + ky + lz)} dV$$

$\rho(x,y,z)$: the three-dimensional electron density of the molecules that give the diffraction.

V: the unit-cell volume

- Each volume element contributes to $F_{hkl}$ with a phase determined by its coordinates (*x, y, z*).
- $F_{hkl}$ is the Fourier transform of $\rho(x,y,z)$ on the set of real-lattice planes (*hkl*). All of the $F_{hkl}$ s together compose the transform of $\rho(x,y,z)$ on all sets of equivalent, parallel planes throughout the unit cell.

## Electron density as a Fourier series

The three-dimensional electron density of the molecules is a wave equation or periodic function because it repeats itself in every unit cell.

*Inverse Fourier transform* →

$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l F_{hkl}\, e^{-2\pi i(hx + ky + lz)}$$

⇨ If we measure *F(h,k,l)*, we can calculate ρ(x,y,z).

*F(h,k,l)* is a wave that has amplitude, frequency and *phase*.

*What we measure in the experiments: indices of each reflection and its intensity.*

$$I(h,k,l) = |F(h,k,l)|^2$$

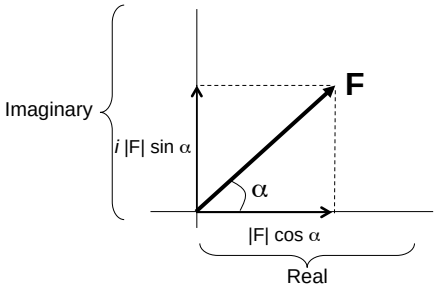
I (h,k,l): the measured reflection intensity of the reflection hkl.

Frequency: determined by the X-ray source.

⇨ We can effectively measure |F(h,k,l)|, not F(h,k,l).

## Phase problem

$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l F_{hkl}\, e^{-2\pi i(hx + ky + lz)}$$




$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l |F_{hkl}|\, e^{i2\pi\alpha'_{hkl}} e^{-2\pi i(hx + ky + lz)}$$

$$= \frac{1}{V}\sum_h\sum_k\sum_l |F_{hkl}|\, e^{-2\pi i(hx + ky + lz - \alpha'_{hkl})}$$

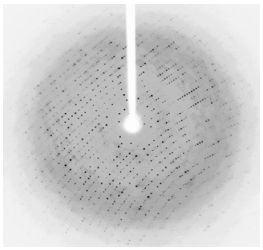
$\mathbf{F} = |\mathbf{F}|\cos\alpha + i\,|\mathbf{F}|\sin\alpha = |\mathbf{F}|\, e^{i\alpha} = |\mathbf{F}|\, e^{i2\pi\alpha'}$

**F**: the structure factor as a vector

|**F**|: the length/amplitude of **F,** which is proportional to $I^{1/2}$

α: the phase angle, $\alpha = 2\pi\alpha'$

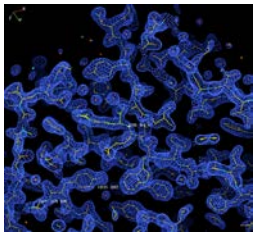
Diffraction Image



I(h k l)

$\alpha$(h k l)

Electron Density Map

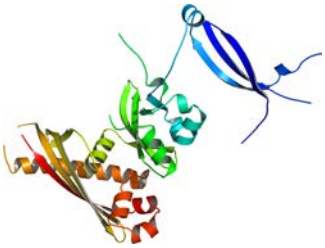


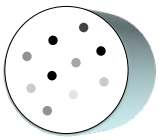
1) Isomorphous Replacement

2) Anomalous Scattering

3) Molecular Replacement

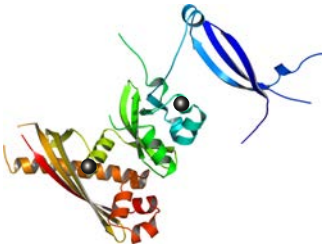
## Isomorphous Replacement

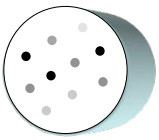




$F_P$

Native Data





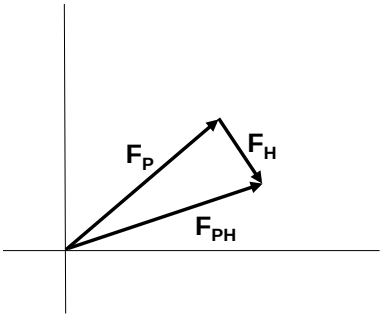
$F_{PH}$

Derivative Data

## Heavy-atom derivatives

- Crystals of the protein are *soaked* in solutions of *heavy ions* (strong diffractors), such as ionic complexes of Hg, Pt, or Au, so that such ions bind to one or a few specific sites on the protein.
- "*Isomorphic*": the heavy atom must not disturb crystal packing or the conformation of the protein, likely with the same unit cell dimensions and diffraction patterns.
- There must be *measurable changes* in at least a modest number of reflection intensities. These changes are the handle by which phase estimates are pulled from the data, so they must be clearly detectable, and large enough to measure accurately.

$$\mathbf{F_{PH} = F_P + F_H}$$




Quantities we want:

$\mathbf{F_P}$

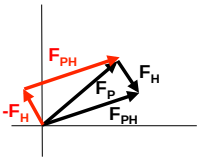
Quantities we know:

$|\mathbf{F_P}|$: intensities of native diffraction

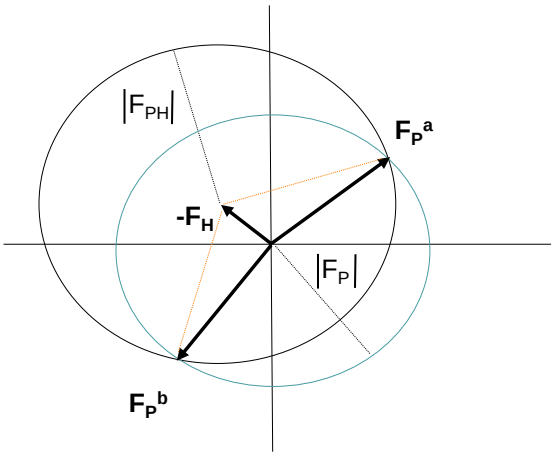
$|\mathbf{F_{PH}}|$: intensities of derivative diffraction

$\mathbf{F_H}$: structure factors with phases of heavy atoms

*How to* get the phase of Fp?



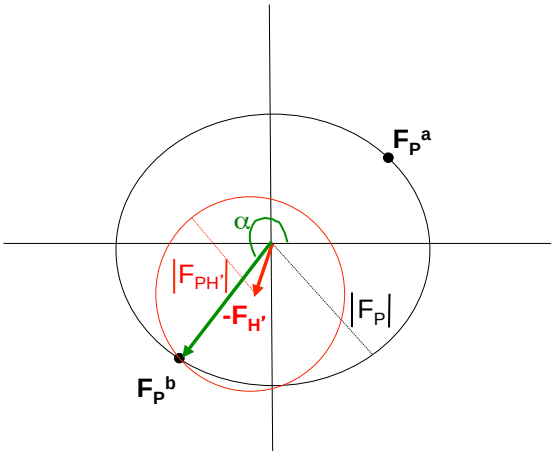

$$F_P = F_{PH} - F_H$$




How do we choose between $F_P^a$ and $F_P^b$?

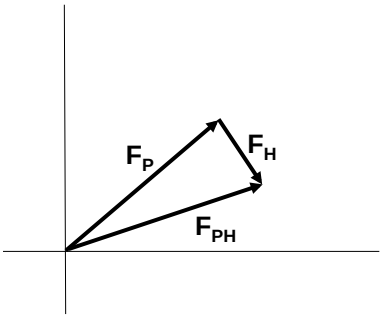
Make a second derivative!

$$F_P = F_{PH'} - F_{H'}$$




$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_P^b \, e^{-2\pi i(hx + ky + lz)}$$

$$\mathbf{F_{PH} = F_P + F_H}$$




Quantities we want:

$\mathbf{F_P}$

Quantities we know:

$|\mathbf{F_P}|$: intensities of native diffraction

$|\mathbf{F_{PH}}|$: intensities of derivative diffraction

$\mathbf{F_H}$: structure factors with phases of heavy atoms

*How to* get $\mathbf{F_H}$?

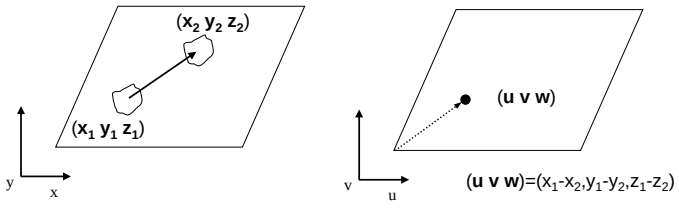
## Locating heavy atoms in the unit cell

$$F_H(h,k,l) = \int_{cell\ V} \rho_H(x,y,z)\, e^{2\pi i(hx + ky + lz)} dV$$

$\rho_H(x,y,z)$ requires knowing the position of heavy atoms in the unit cell.

## The Patterson Function

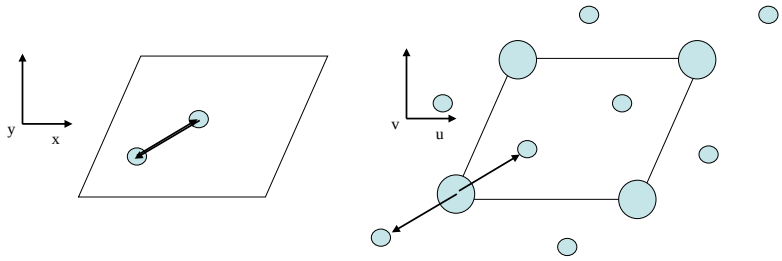
By definition: $$P(u,v,w) = \int_{cell\ V} \rho(x,y,z) \bullet \rho(x+u,y+v,z+w) dV$$



## Patterson Function

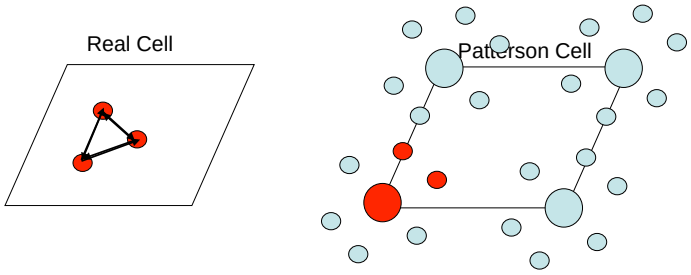
$$P(u,v,w) = \int_{cell\ V} \rho(x,y,z) \bullet \rho(x+u,y+v,z+w) dV$$

Patterson peaks will contain points corresponding to vectors between atoms in the real cell, i.e. inter-atomic distances (not atomic positions though).




Real Cell Patterson Cell

## Patterson Function




1) Patterson is symmetric about origin (centrosymmetry).

2) Can see pattern of real cell in patterson cell repeated.

3) Contains N(N-1) non-origin peaks (not counting origin) → gets complicated!

## Patterson Function

$$P(u,v,w) = \int_{cell\ V} \rho(x,y,z)\ \rho(x+u,y+v,z+w)dV$$

Key point: can calculate P(u,v,w) from experimental data

$$\rho(x) = \frac{1}{V}\sum_h F_h e^{-2\pi i h x} \qquad \rho(x+u) = \frac{1}{V}\sum_{h'} F_{h'} e^{-2\pi i h'(x+u)}$$

$$P(u) = \frac{1}{V^2}\sum_h\sum_{h'} F_h F_{h'} e^{-2\pi i h u} \int_{cell\ V} e^{-2\pi i(h+h')x}dV$$

The integration is equal to zero, unless h=h' when it is equal to V,

By Friedel's Law $F_h=F_{-h}$,

$$P(u) = \frac{1}{V}\sum_h F_h^2 e^{-2\pi i h u}$$

## Patterson Function

$$P(u,v,w) = \frac{1}{V}\sum_h\sum_k\sum_l |F(h,k,l)|^2 \cos 2\pi(hu + kv + lw)$$

It is a Fourier summation with intensities as coefficients and phase angles equal to zero.

Patterson analysis is simplified for heavy atoms:

1) Use $(|F_{PH}(h\ k\ l)|- |F_P(h\ k\ l)|)^2$ as coefficients

→ "difference map" reflects heavy atom contribution

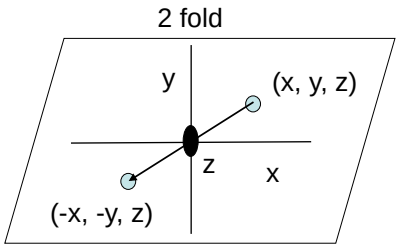
2) If atom *i* contains $Z_i$ electrons and atom *j* contains $Z_j$ electrons, the corresponding vector $r_{ij}$ will have a weight proportional to $Z_iZ_j$.

→heavy atoms, high Z, strong peaks

3) Calculate (x,y,z) of heavy atoms directly from Harker section peaks

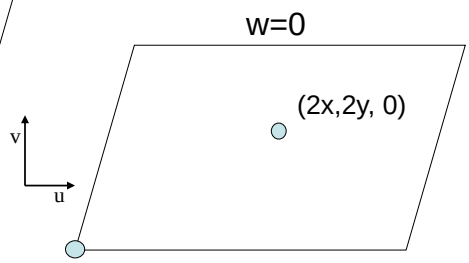
## Harker Peaks

Symmetry related atoms give rise to peaks in Patterson map in specific locations. Each space group has its own Harker planes.

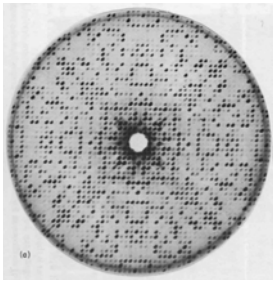



(u,v,w) = (x-[-x], y-[-y], z-z)

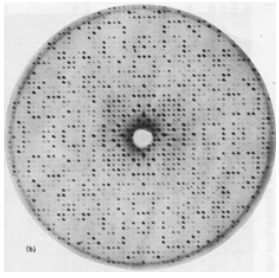
(u,v,w) = (2x,2y, 0)

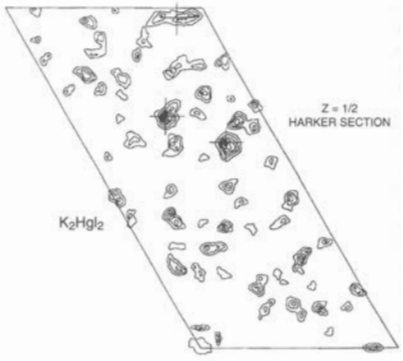
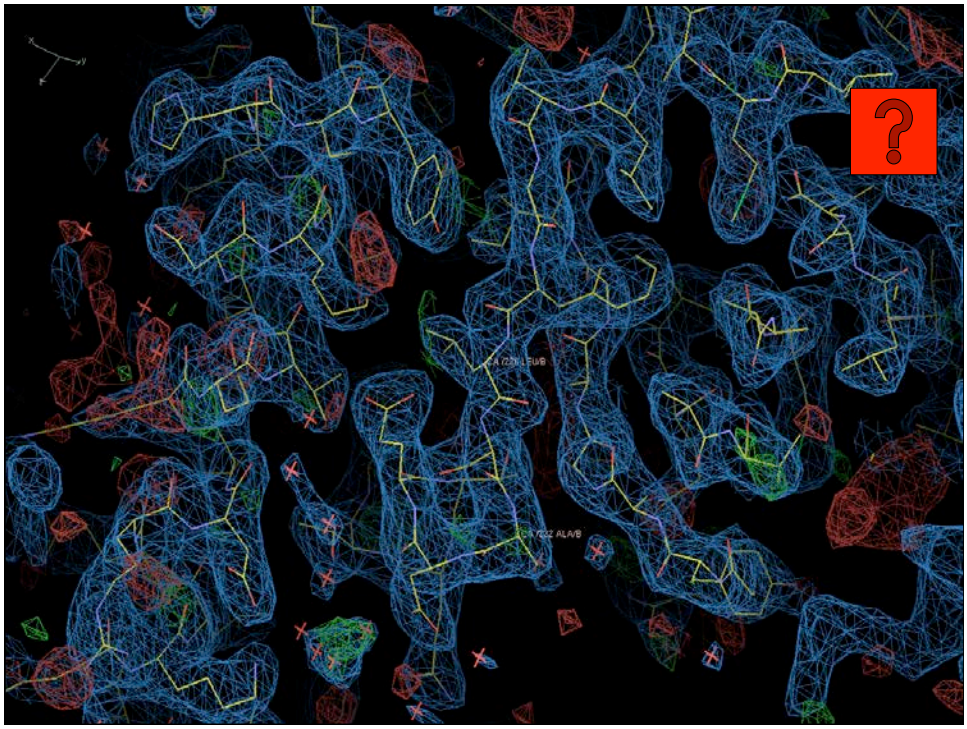



w=0 called Harker plane (section)



Native



A Hg derivative

Difference map