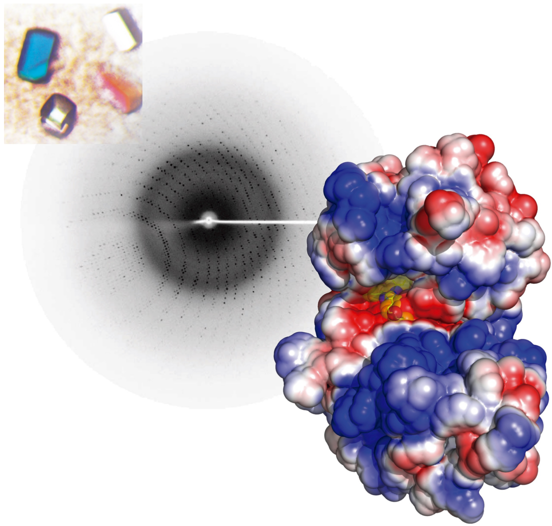
Unit of Protein Crystallography

Institut Pasteur de Montevideo

Some basic math for crystallography



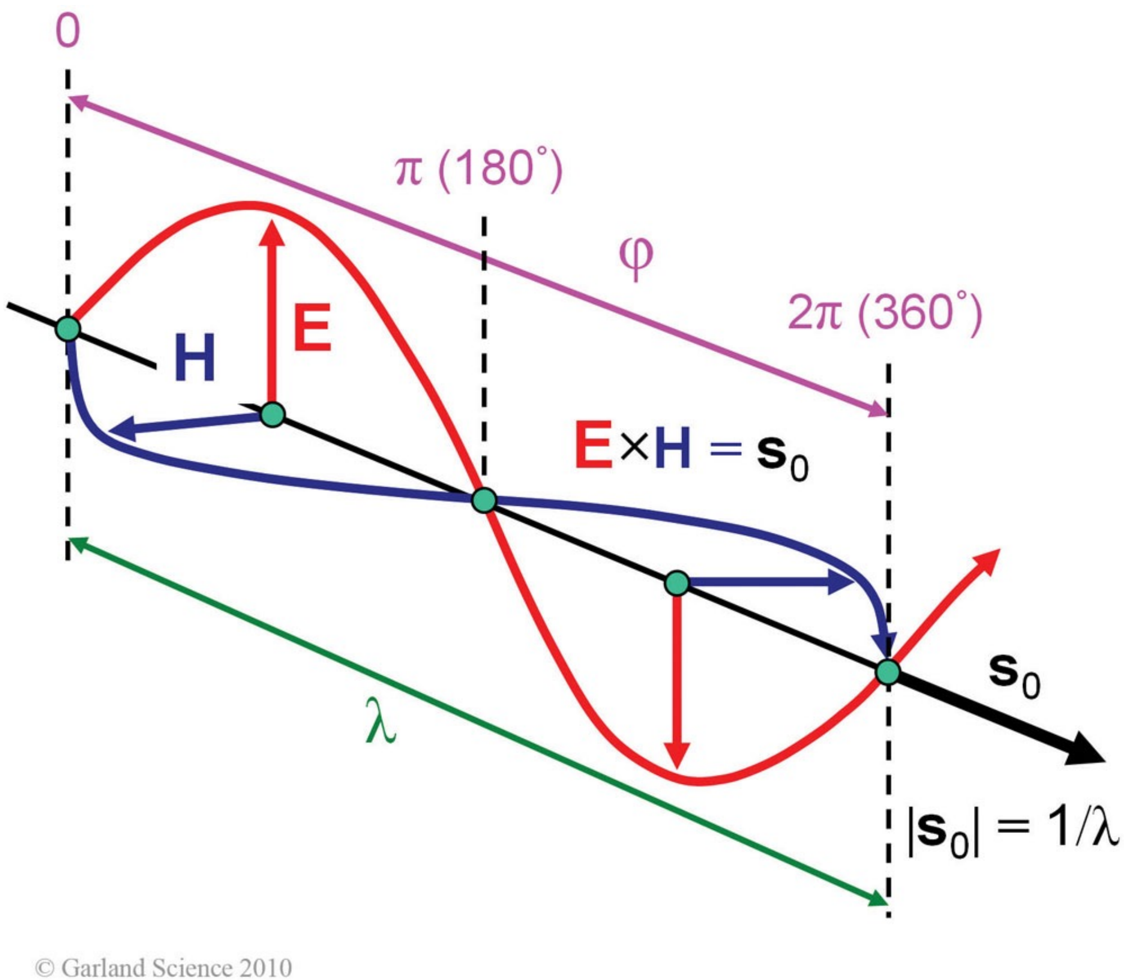
Macromolecular Crystallography School 2018
November 2018 - São Carlos, Brazil

Why sines and cosines, and complex numbers, can be any good for crystallography??

X-ray diffraction of a protein crystal, is light interacting with matter: scattering and interference...

light can be represented as electromagnetic waves (sinusoidal functions describe the electric and the magnetic fields!)

Electromagnetic waves



from “Biomolecular Crystallography” Bernhard Rupp, Garland Science 2010

…so this introduction, will try and refresh very few math elements, about

- complex numbers and some trigonometry: both useful when it comes to dealing with waves!

- a little bit of vectors (structures have lots to do with positions & distances…and these are well described with vectors!)

…so this introduction, will try and refresh very few math elements, about

- complex numbers and some trigonometry: both useful when it comes to dealing with wave

what could be the risk?…

- a little bit of vectors (structures have lots to do with positions & distances…and these are well described with vectors!)

一种语言永远不够

“One language is never enough”

马 [illegible]听杂

…only garbage
I made up!!! 😬

$$\mathbf{F}_{hkl} = V \int_{xyz} \rho_{xyz}\, e^{[2\pi i(hx+ky+lz)]}\, \delta xyz$$

?????

Fourier theory

The diffraction pattern is related to the object that made waves diffract, by a direct mathematical operation called *Fourier transform*

$$\boxed{\mathbf{F}_{hkl} = V \int_{xyz} \rho_{xyz}\, e^{[2\pi i(hx+ky+lz)]}\, \delta xyz}$$

F is a complex number

...and it is a function of h,k,l (reciprocal space)

The complex numbers

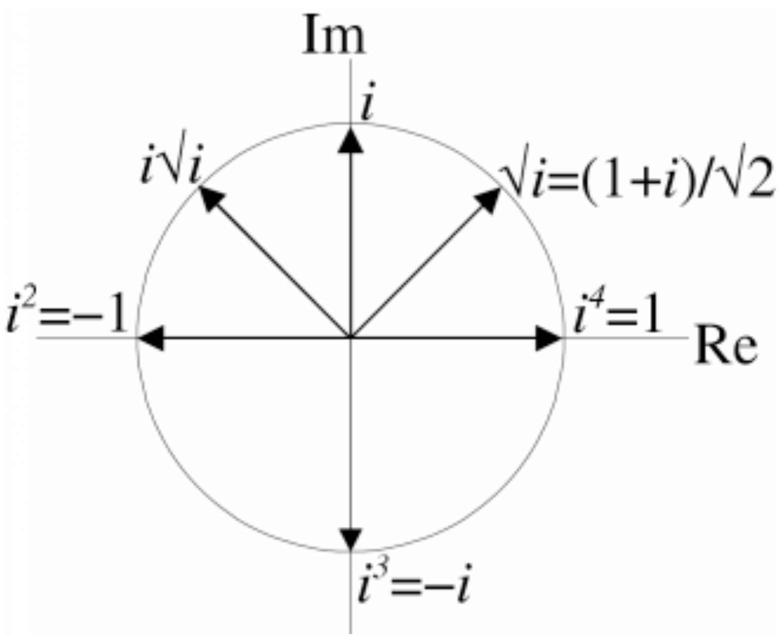
a + *i*b

They arise from certain solutions of quadratic equations :

…when $b^2 < 4ac$…

$$\boxed{ax^2 + bx + c = 0} \Rightarrow \boxed{x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}}$$

The complex numbers



Argand diagram
(complex plane)

Complex numbers are of the type

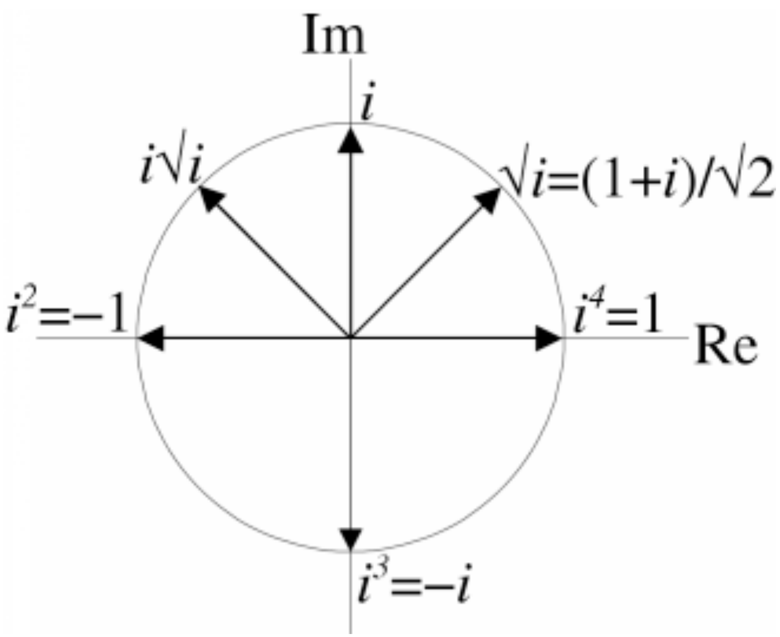
$$\mathbf{F} = a + ib$$

where $i = \sqrt{-1}$

Rotation is posible by multiplying vectors

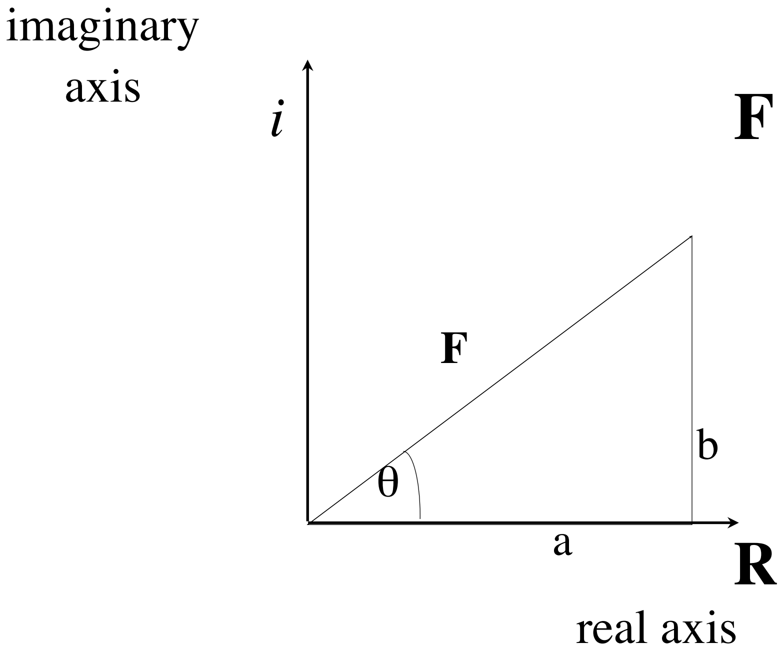
$\sqrt{i} = 45°$ ($i = 90°$)

The complex numbers



Argand diagram
(complex plane)

- They are not vectors, but they share with 2-dimensional vectors a similar notation (see e.g. the Argand diagram)
- Adding complex numbers and vectors is similar
- ..but multiplication is different! no scalar or cross products for complex numbers…



$\mathbf{F} = a + ib$

$f = |F|$

$$a + i\,b = f\cos\theta + i\,f\sin\theta =$$

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

...implying that $\cos\theta + i\sin\theta = e^{i\theta}$

Euler's theorem…

The sum of cosine α plus i times sine α is equal to the exponent of i times α.

Euler's theorem…

$$e^{i\alpha} = \cos\alpha + i\sin\alpha$$

Given,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$
$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$
$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots.$$

…and knowing,

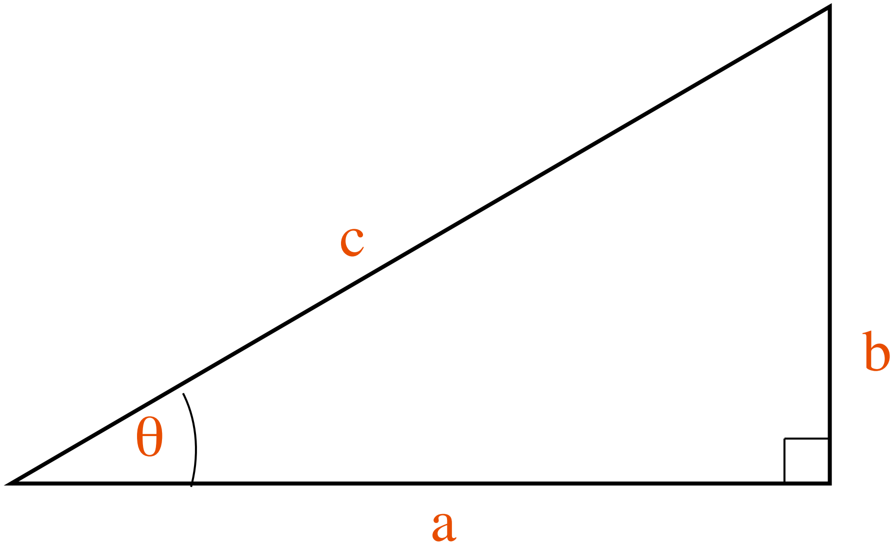
$$i^0 = 1, \quad i^1 = i, \quad i^2 = -1, \quad i^3 = -i,$$
$$i^4 = 1, \quad i^5 = i, \quad i^6 = -1, \quad i^7 = -i,$$

$$\begin{aligned}
e^{iz} &= 1 + iz + \frac{(iz)^2}{2!} + \frac{(iz)^3}{3!} + \frac{(iz)^4}{4!} + \frac{(iz)^5}{5!} + \frac{(iz)^6}{6!} + \frac{(iz)^7}{7!} + \frac{(iz)^8}{8!} + \cdots \\
&= 1 + iz - \frac{z^2}{2!} - \frac{iz^3}{3!} + \frac{z^4}{4!} + \frac{iz^5}{5!} - \frac{z^6}{6!} - \frac{iz^7}{7!} + \frac{z^8}{8!} + \cdots \\
&= \left(1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \frac{z^8}{8!} - \cdots\right) + i\left(z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots\right) \\
&= \cos z + i\sin z
\end{aligned}$$

Trigonometry

"the measurement of triangles"

$\cos\theta = a/c$ $\quad\quad$ $\sin\theta = b/c$ $\quad\quad$ $\tan\theta = b/a$

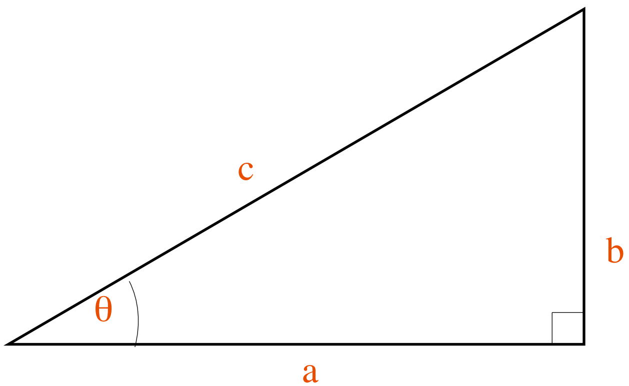


$\tan\theta = \sin\theta/\cos\theta$

Trigonometry

"the measurement of triangles"

$\cos\theta = a/c$ $\sin\theta = b/c$



$a^2 + b^2 = c^2$ ⟹ $\cos^2\theta + \sin^2\theta = 1$

$$c^2 = a^2 + b^2 - 2ab\cos\gamma$$

Sinusoidal functions are symmetrical :

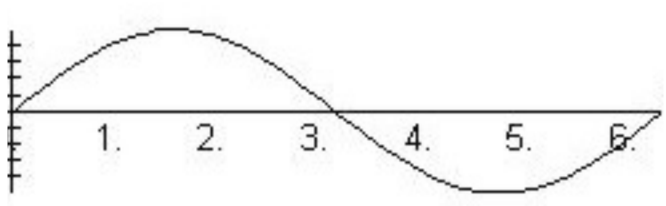
$\cos\theta = \cos -\theta$ (*“symmetric”*)

$\sin\theta = -\sin -\theta$ (*“antisymmetric”*)

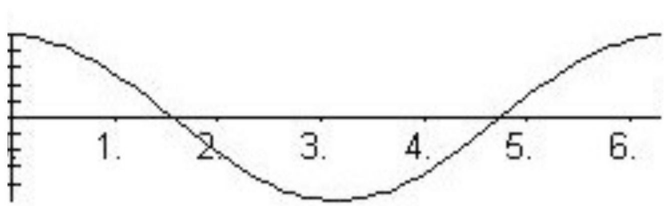
Differentiation of sinusoidal functions

$$\frac{d(\sin(\theta))}{d\theta} = \cos(\theta) \qquad \frac{d(\cos(\theta))}{d\theta} = -\sin(\theta)$$

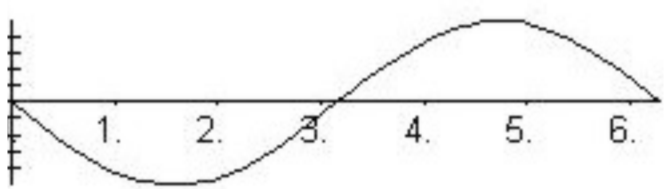
sin θ



cos θ



-sin θ



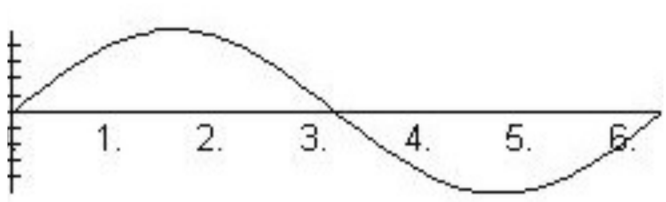
perimeter/radius = 2π

one full cycle is 2π radians

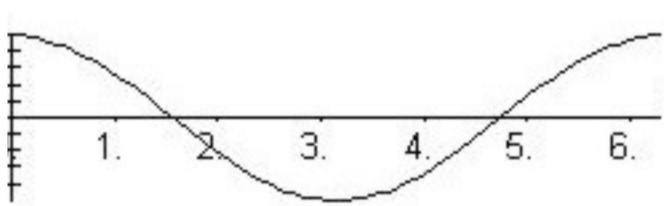
Differentiation of sinusoidal functions

$$\frac{d(\sin(\theta))}{d\theta} = \cos(\theta) \qquad \frac{d(\cos(\theta))}{d\theta} = -\sin(\theta)$$

sin θ



cos θ



perimeter/radius = 2π

one full cycle is 2π radians

Some properties of complex numbers are important for the manipulation of structure factors.

A simple operation is to take the complex conjugate, which means changing the sign of the imaginary part.

Some properties of complex numbers are important for the manipulation of structure factors.

A simple operation is to take the complex conjugate, which means changing the sign of the imaginary part.

Thus, the complex conjugate of the complex number $a + ib$ is written as $(a + ib)^*$, and is equal to $a - ib$

You should be able to confirm that multiplying a complex number by its complex conjugate gives a real number which is the square of the modulus:

$$(\mathbf{a} + i\,\mathbf{b})\,(\mathbf{a} + i\,\mathbf{b})^* = (\mathbf{a} + i\,\mathbf{b})\,(\mathbf{a} - i\,\mathbf{b}) = \mathbf{a}^2 - i\,\mathbf{a}\,\mathbf{b} + i\,\mathbf{a}\,\mathbf{b} - i^2\,\mathbf{b}^2 = \mathbf{a}^2 + \mathbf{b}^2 = \mathbf{r}^2$$

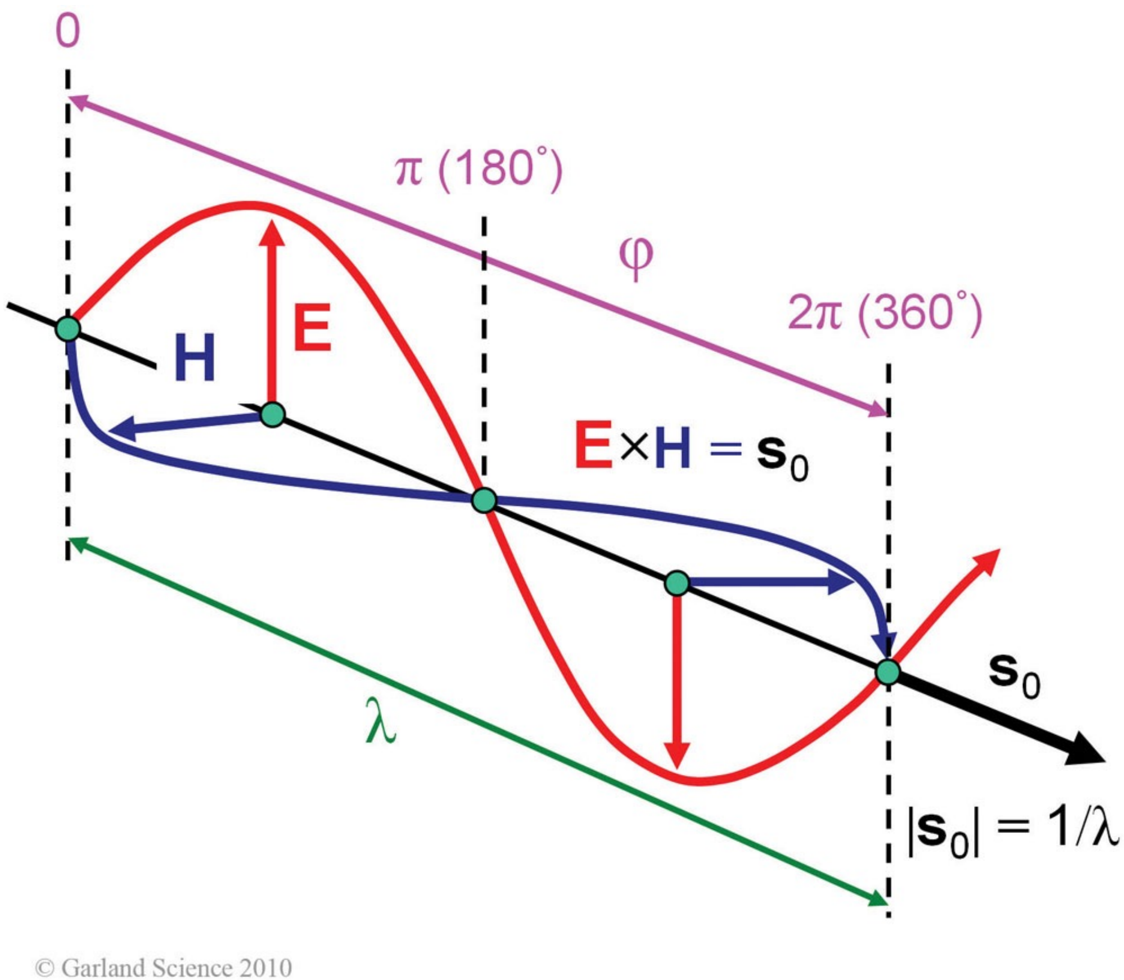
...remember this!

Why sines and cosines, and complex numbers, can be any good for crystallography??

X-ray diffraction of a protein crystal, is light interacting with matter: scattering and interference...

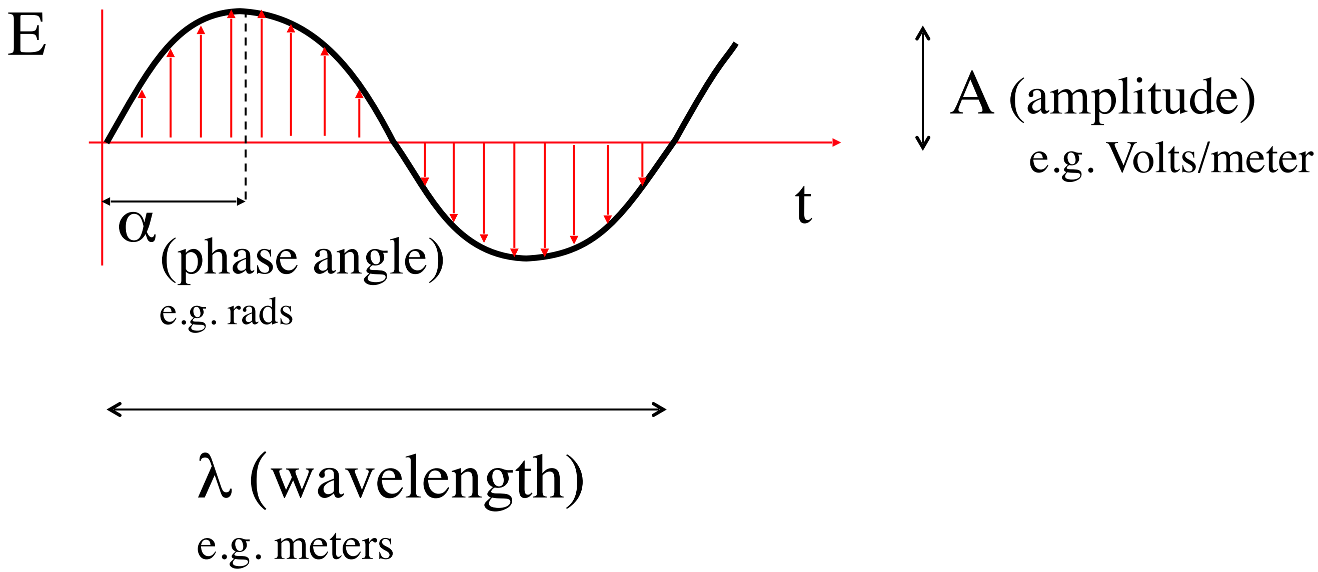
light can be represented as electromagnetic waves (sinusoidal functions describe the electric and the magnetic fields!)

Electromagnetic waves



from “Biomolecular Crystallography” Bernhard Rupp, Garland Science 2010

Waves

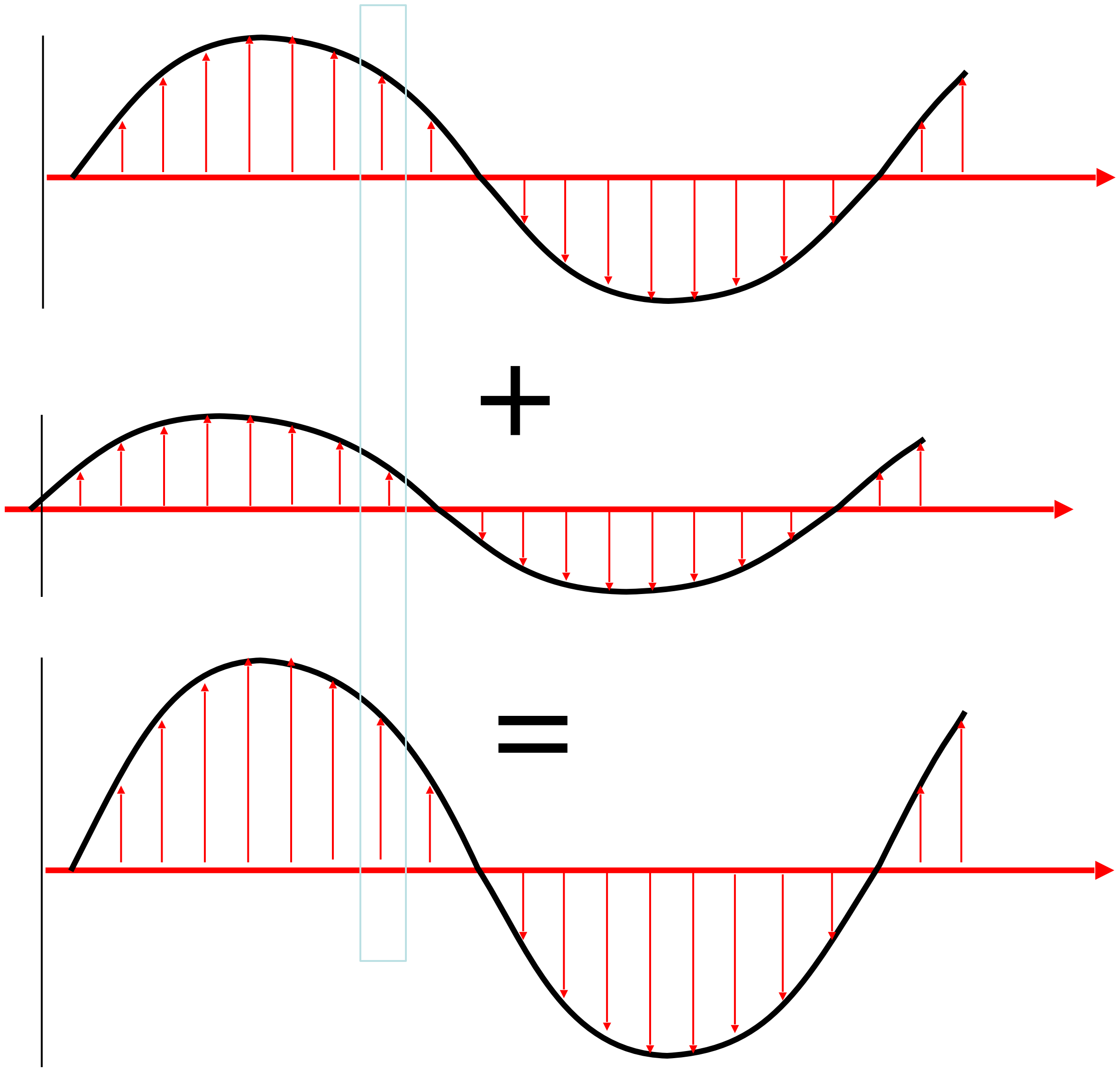


$$E(t) = A\cos(\omega t + \alpha)$$

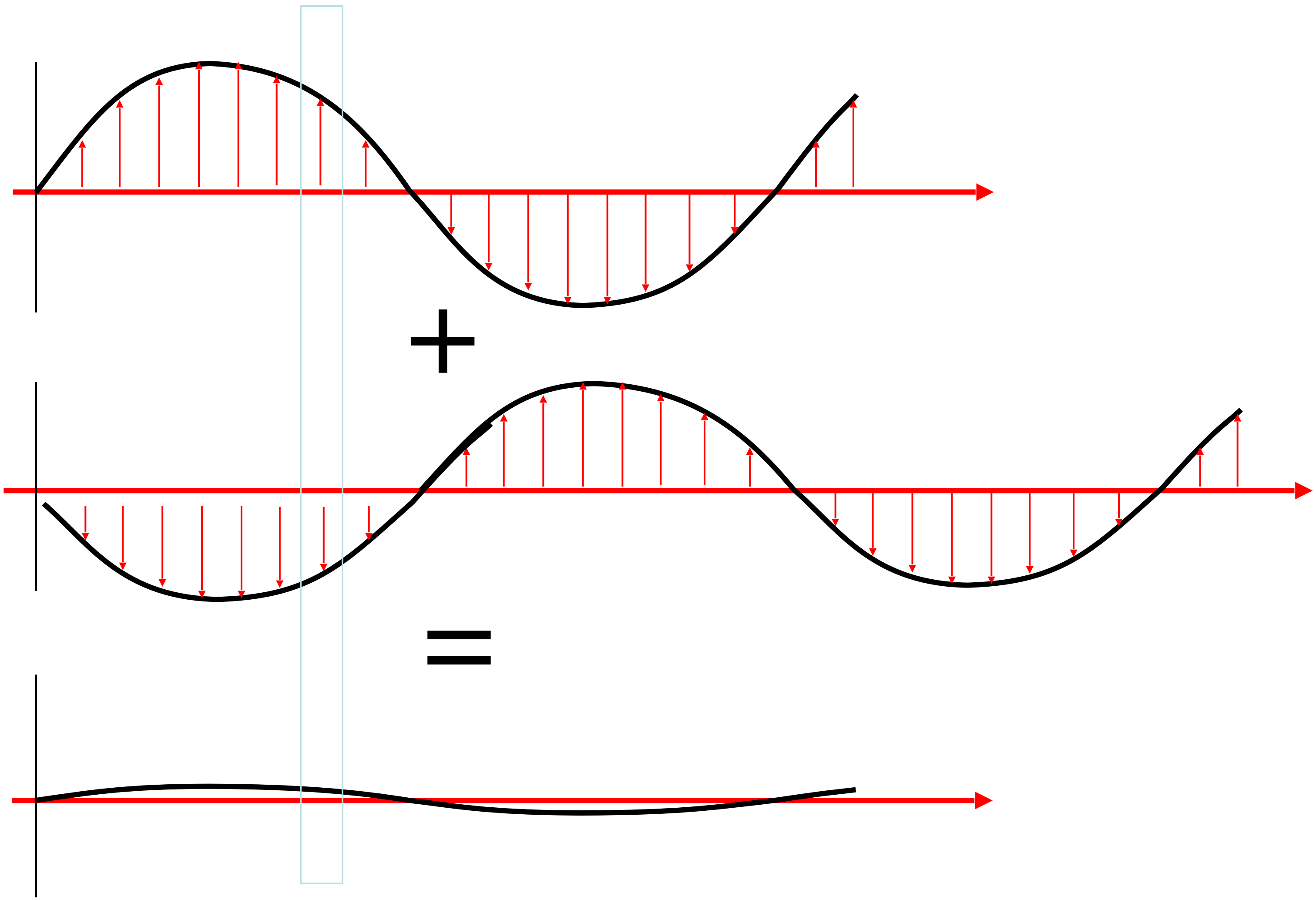
$$\omega = 2\pi\nu$$

adapted from http://www-structmed.cimr.cam.ac.uk/Course/Basic_diffraction/Diffraction.html

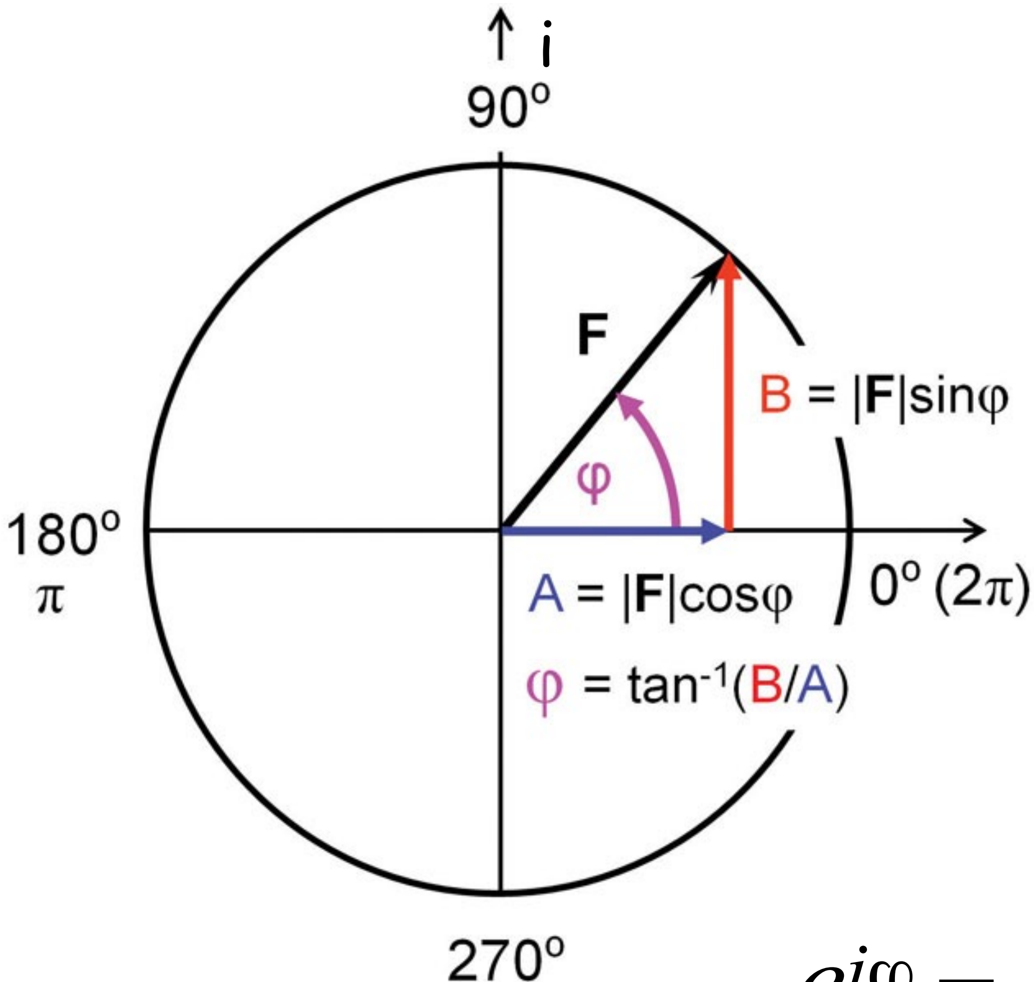
Addition of plane waves



Addition of plane waves (2)



The Argand diagram



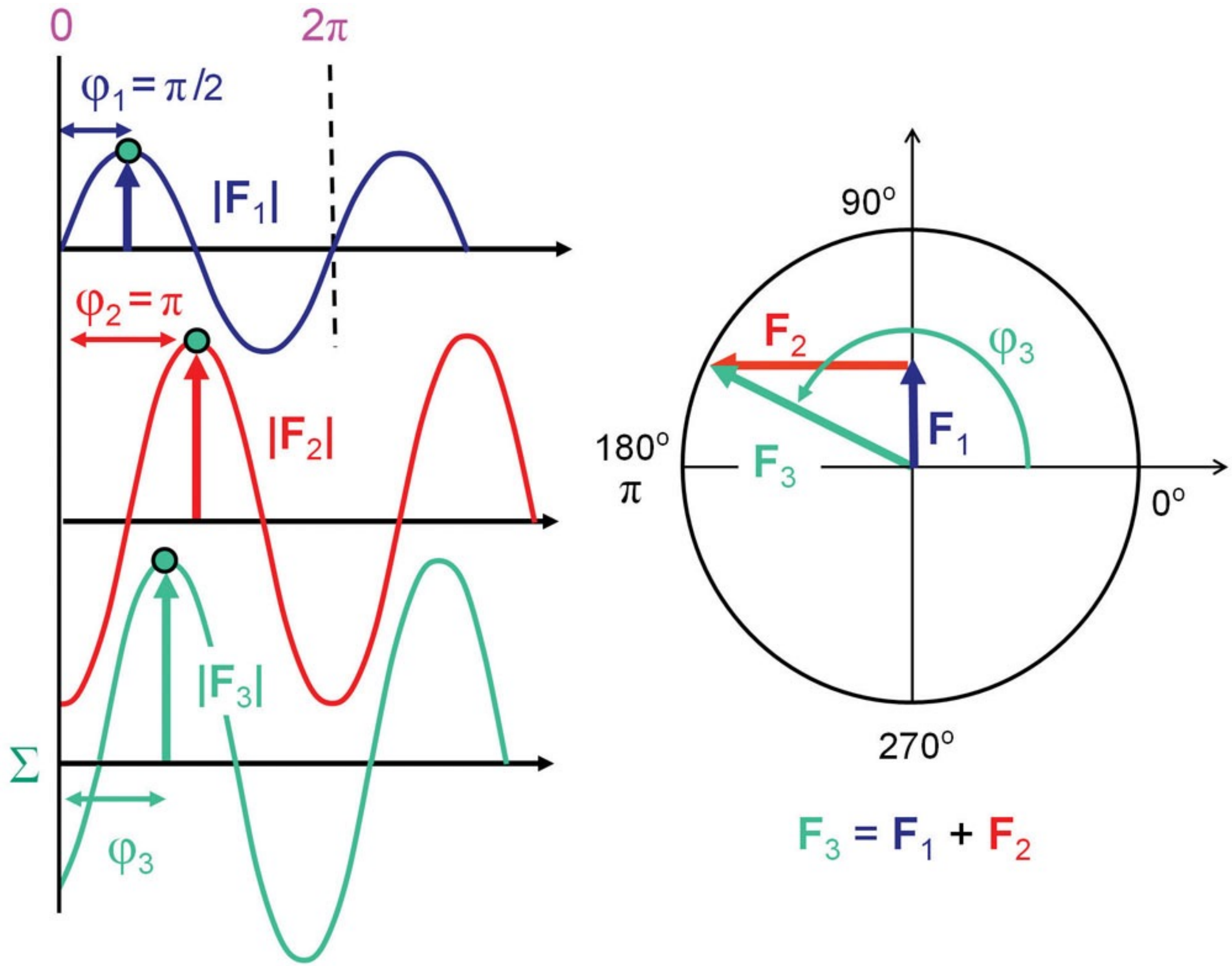
$$\mathbf{F} = F\, e^{i\varphi}$$

where

$$e^{i\varphi} = \cos\varphi + i \sin\varphi$$

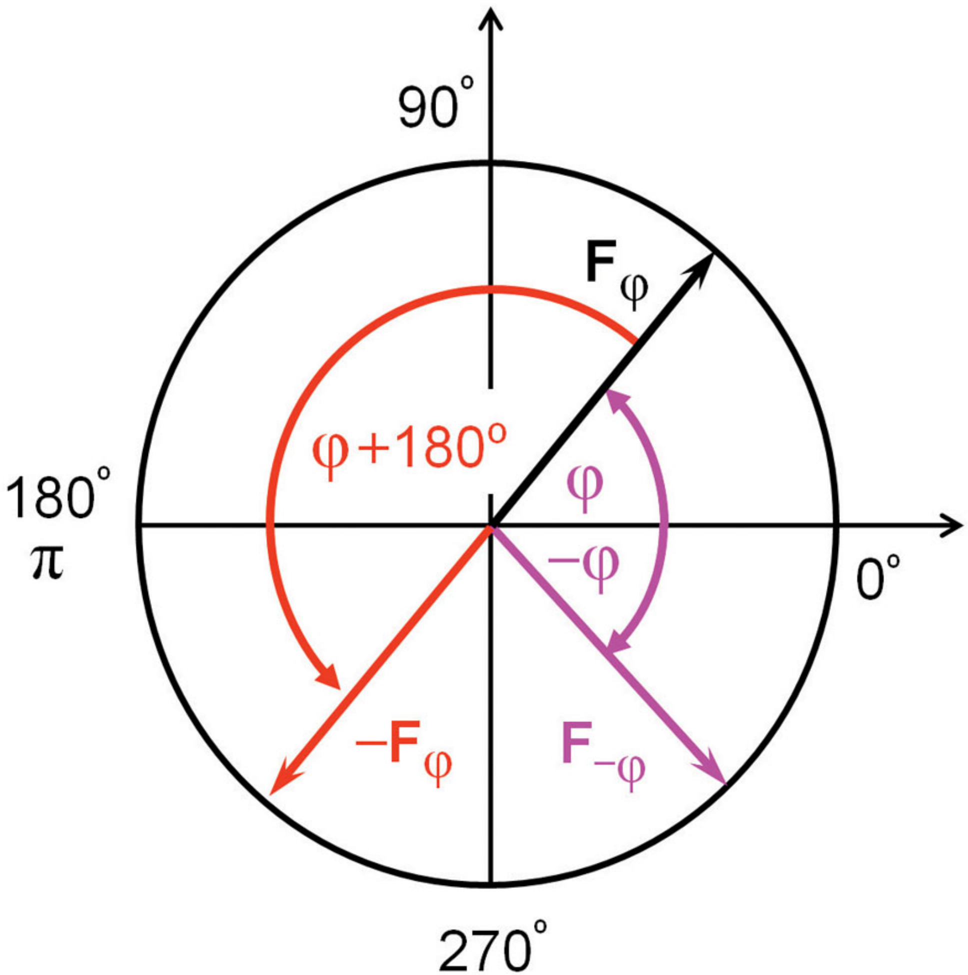
from "Biomolecular Crystallography" Bernhard Rupp, Garland Science 2010

Addition of plane waves



from “Biomolecular Crystallography” Bernhard Rupp, Garland Science 2010

Waves and complex numbers



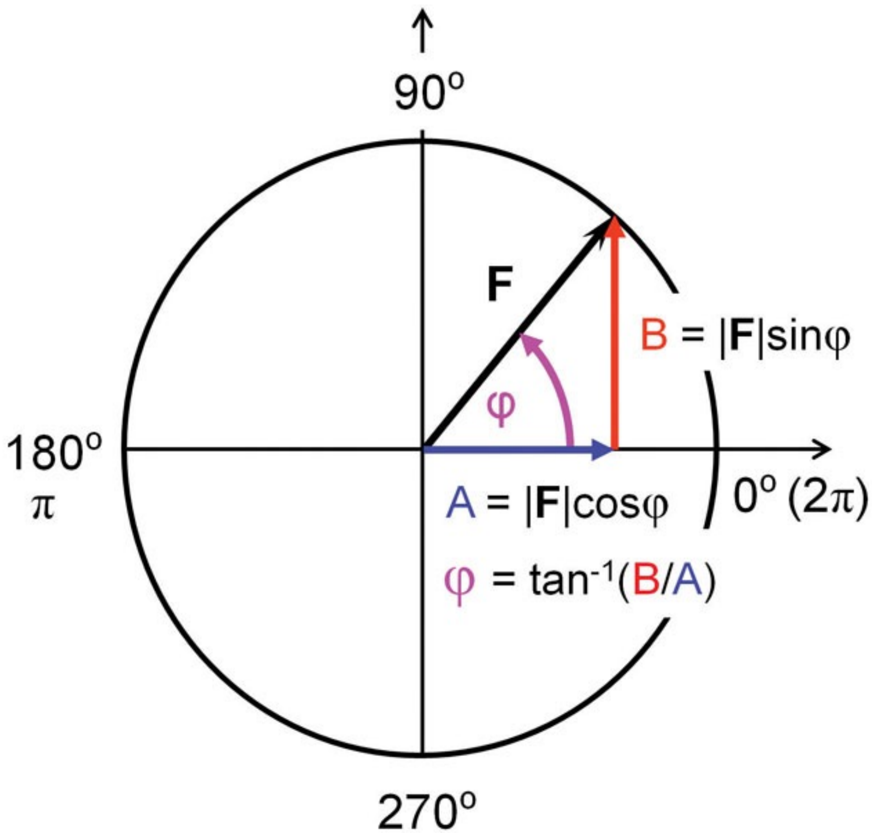
$$\mathbf{F}(\phi) = F\, e^{i\varphi}$$

$$\mathbf{F}(-\phi) = \mathbf{F}^{*}(\phi)$$

$$-\mathbf{F}(\phi) = \mathbf{F}(\phi+\pi)$$

from “Biomolecular Crystallography” Bernhard Rupp, Garland Science 2010

Waves and complex numbers



$$\mathbf{F}(\phi) = F\,e^{i\varphi} = \mathrm{A} + i\mathrm{B}$$

$$F = |\mathbf{F}|$$

$$|\mathbf{F}| = (\mathrm{A}^2 + \mathrm{B}^2)^{1/2}$$

$$F = ((\mathrm{A} + i\mathrm{B})(\mathrm{A} - i\mathrm{B}))^{1/2}$$

$$\therefore$$

$$F = (\mathbf{FF^*})^{1/2}$$

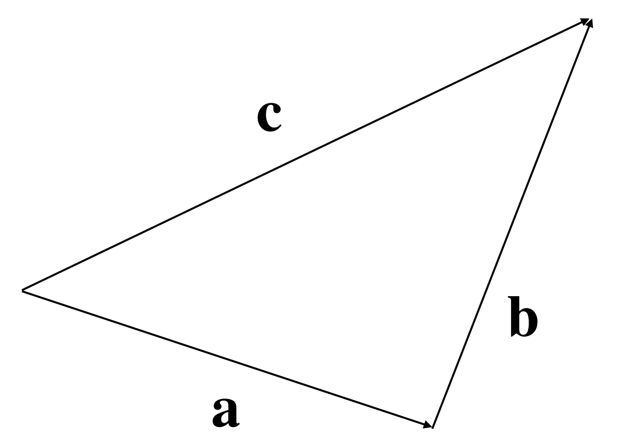
from "Biomolecular Crystallography" Bernhard Rupp, Garland Science 2010

…coming back to vectors, they are of course also present and useful in physics (in particular crystallography)

they have magnitude and direction, as opposed to a scalar which only has magnitude

Vector addition :
algebraically expressed as

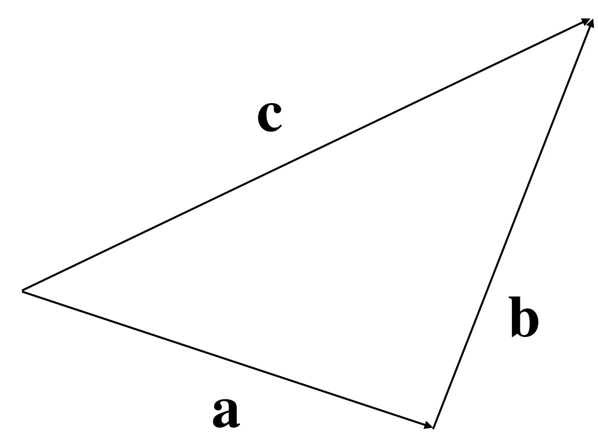
a + **b** = **c**



b = **c** - **a** think of **b** = **c** + (-**a**)

…coming back to vectors, they are of course also present and useful in physics (in particular crystallography)

If vectors **a** and **c** give the positions of two atoms in the cell, they are known as position vectors; **b** is known as a displacement vector, as it gives the displacement of one atom relative to the other.



Vectors

In the unit cell, the position vector **x** has components (x, y, z) such that

$$\mathbf{x} = \mathbf{a}\,x + \mathbf{b}\,y + \mathbf{c}\,z$$

where **a**, **b** and **c** are the lattice translation vectors (the edges of the unit cell) and x, y and z are the fractional coordinates of the point.

Similarly, the position of a point in reciprocal space is given by the vector **h**, which has components (h, k, l) such that:

$$\mathbf{h} = \mathbf{a}^*h + \mathbf{b}^*k + \mathbf{c}^*l$$

Vectors

$$\mathbf{F}_{hkl} = V \int_{xyz} \rho_{xyz}\, e^{[2\pi i(hx+ky+lz)]}\, \delta xyz$$

The scalar (dot) product of the two vectors **x** and **h** is a scalar:

$$\mathbf{h}.\mathbf{x} = h\,x + k\,y + l\,z$$

…an expression found in both the structure factor and electron density equations.

Vectors

The scalar (dot) product of the two vectors **x** and **h** is a scalar:

$$\mathbf{h}.\mathbf{x} = \mathrm{h\,x + k\,y + l\,z}$$

…an expression found in both the structure factor and electron density equations.

The vector (cross) product gives a third vector (normal $\hat{\mathbf{n}}$ to the multiplied ones), used e.g. in the relationships between the direct and reciprocal lattices:

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{V} \qquad \mathbf{b}^* = \frac{\mathbf{a} \times \mathbf{c}}{V} \qquad \mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{V}$$

$$V = \mathbf{a}.\mathbf{b} \times \mathbf{c}$$

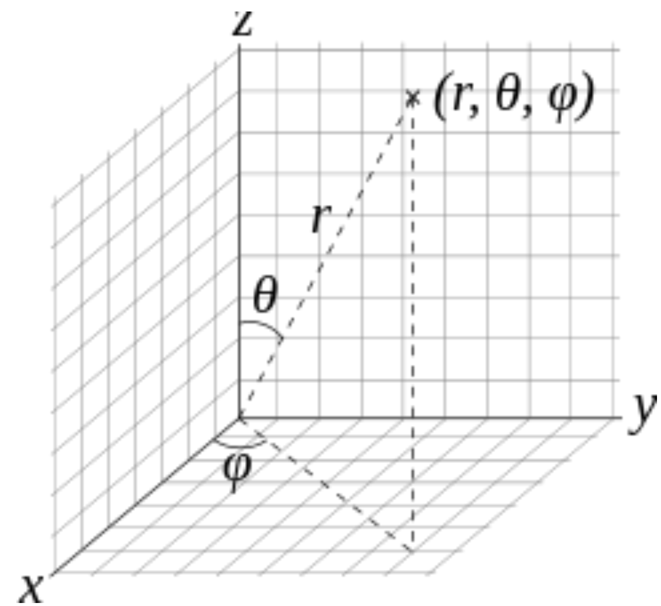
remember…

$$\mathbf{a}.\mathbf{b} = ab\cos\gamma$$

$$\mathbf{a} \times \mathbf{b} = ab\sin\gamma\hat{\mathbf{n}}$$

Vectors

3D vectors and dot product



then for any given vector
$\mathbf{a}=(r,\varphi,\theta)=(a_1,a_2,a_3)$ it follows

$a_1 = r\cos\varphi\sin\theta$ $\quad$ $b_1 = r'\cos\varphi'\sin\theta'$

$a_2 = r\sin\varphi\sin\theta$ $\quad$ $b_2 = r'\sin\varphi'\sin\theta'$

$a_3 = r\cos\theta$ $\quad$ $b_3 = r'\cos\theta'$

such that

$$\mathbf{a}\cdot\mathbf{b} = r\,r'\,[\cos(\varphi-\varphi')\sin\theta\sin\theta' + \cos\theta\cos\theta']$$

$\therefore$ it holds that

$$\mathbf{a}\cdot\mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3$$

remember…

$$\mathbf{a}.\mathbf{b} = ab\cos\gamma$$

$$\mathbf{a}\times\mathbf{b} = ab\sin\gamma\,\hat{\mathbf{n}}$$

This…

$$V = \mathbf{a}.\mathbf{b} \times \mathbf{c}$$

the volume of the unit cell

…is easier than this…

$$V = abc\sqrt{1-\cos^2\alpha-\cos^2\beta-\cos^2\gamma+2\cos\alpha\cos\beta\cos\gamma}$$

coming back to our Fourier transform equation, connecting the electron density function to a reciprocally related structure factor function…

$$\mathbf{F}_{hkl} = V \int_{xyz} \rho_{xyz}\, e^{[2\pi i(hx+ky+lz)]}\, \delta xyz$$

it's getting a little bit clearer…

we'll see later why the physical diffraction actually comprises this Fourier transform in real action, transforming the electron density of the crystal into an image function (structure factors $\mathbf{F}_{hkl}$) that we measure in the experiment

$$\boxed{\mathbf{F}_{hkl} = V \int_{xyz} \rho_{xyz}\, e^{[2\pi i(hx+ky+lz)]}\, \delta xyz}$$

we'll not go now into calculus, but think of the integral as a summation over all real space xyz so you can see, that for each h,k,l adopting a particular **F** value (it is a complex number value, with amplitude and phase), you need to sum ALL the values of ρ for all the x,y,z space!!! and it is non linear, because we have wave (sinusoidal) components

$$\mathbf{F}_{hkl} = V \int_{xyz} \rho_{xyz}\, e^{[2\pi i(hx+ky+lz)]}\, \delta xyz$$

also cool is that the Fourier transform can be inverted :

$$\rho_{xyz} = \frac{1}{V} \sum_{hkl} \mathbf{F}_{hkl}\, e^{[-2\pi i(hx+ky+lz)]}$$

also cool is that the Fourier transform can be inverted :

watch out for the change of sign here

$$\rho_{xyz} = \frac{1}{V} \sum_{hkl} \mathbf{F}_{hkl}\, e^{[-2\pi i(hx+ky+lz)]}$$

also cool is that the Fourier transform can be inverted :

$$\boxed{\rho_{xyz} = \frac{1}{V} \sum_{hkl} \mathbf{F}_{hkl}\, e^{[-2\pi i(hx+ky+lz)]}}$$

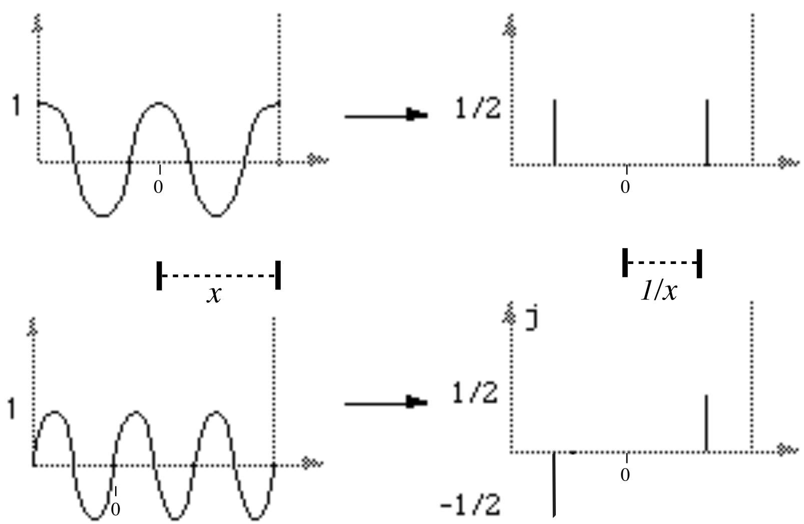
this is interesting! since we are actually faced to this problem, once we've measured the diffracted structure factors, and want to reconstruct the 3D electron density of the protein xtal!!

Waves, structure factors and how Fourier helps us

X rays have wave properties, hence the utility of mathematical descriptors of waves : the sinusoidal functions describe "simple" waves (those that have a single frecuency)

space or time domain
(…periodic for these functions)

frequency domain
(…discrete)



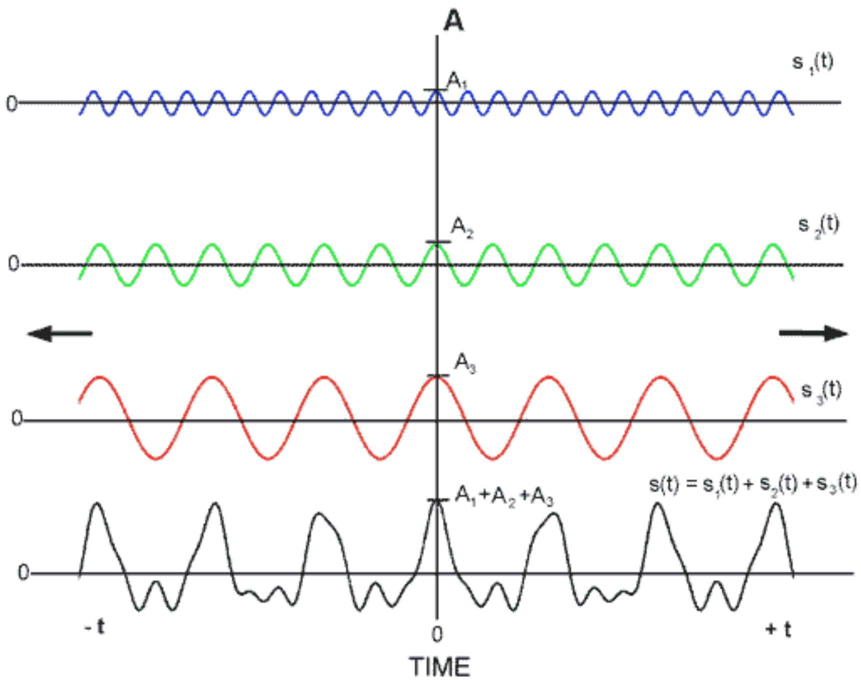
cosine function and its Fourier transform

sine function and its Fourier transform

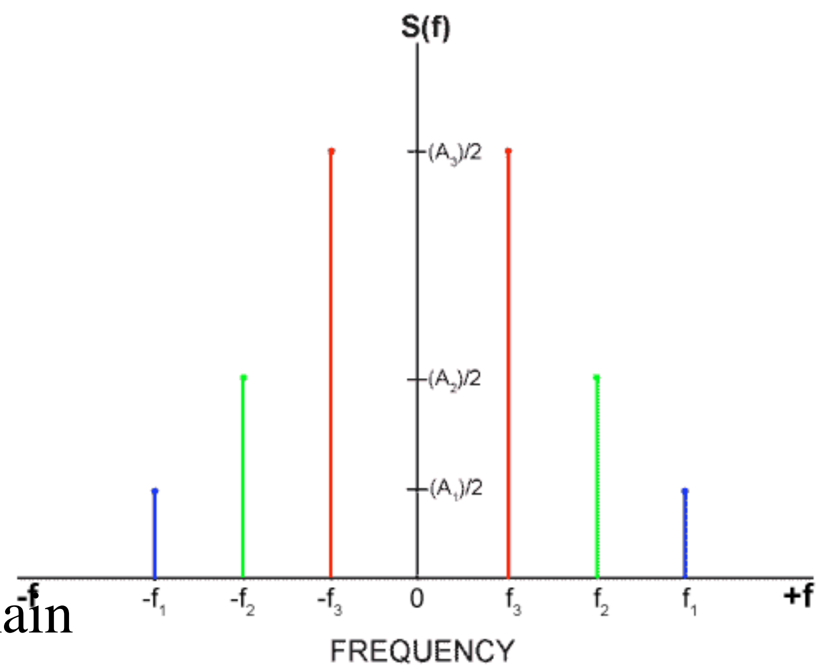
Waves and structure factors

(periodic) space or

time domain



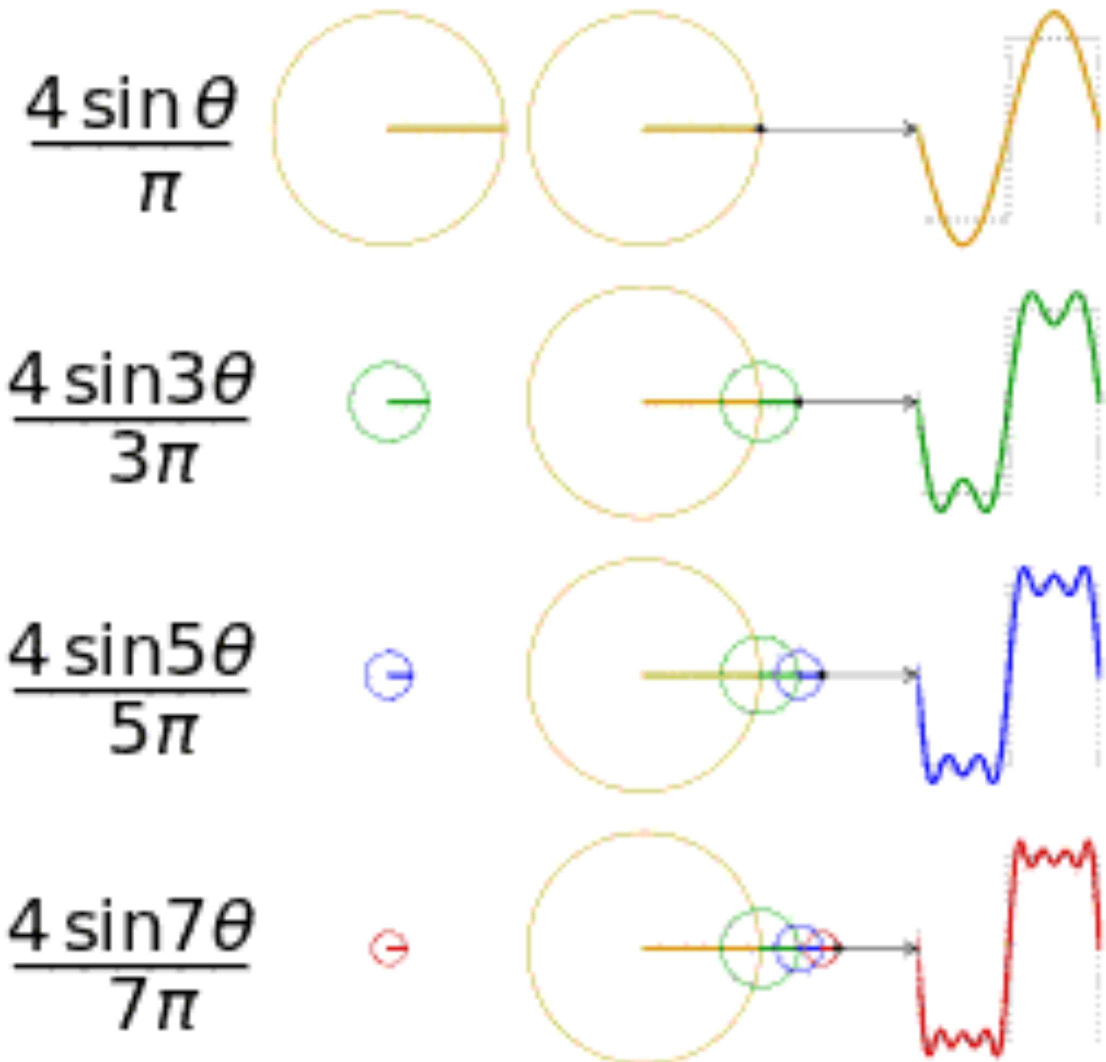
3 cosine functions combined in one "composite" wave



frequency domain

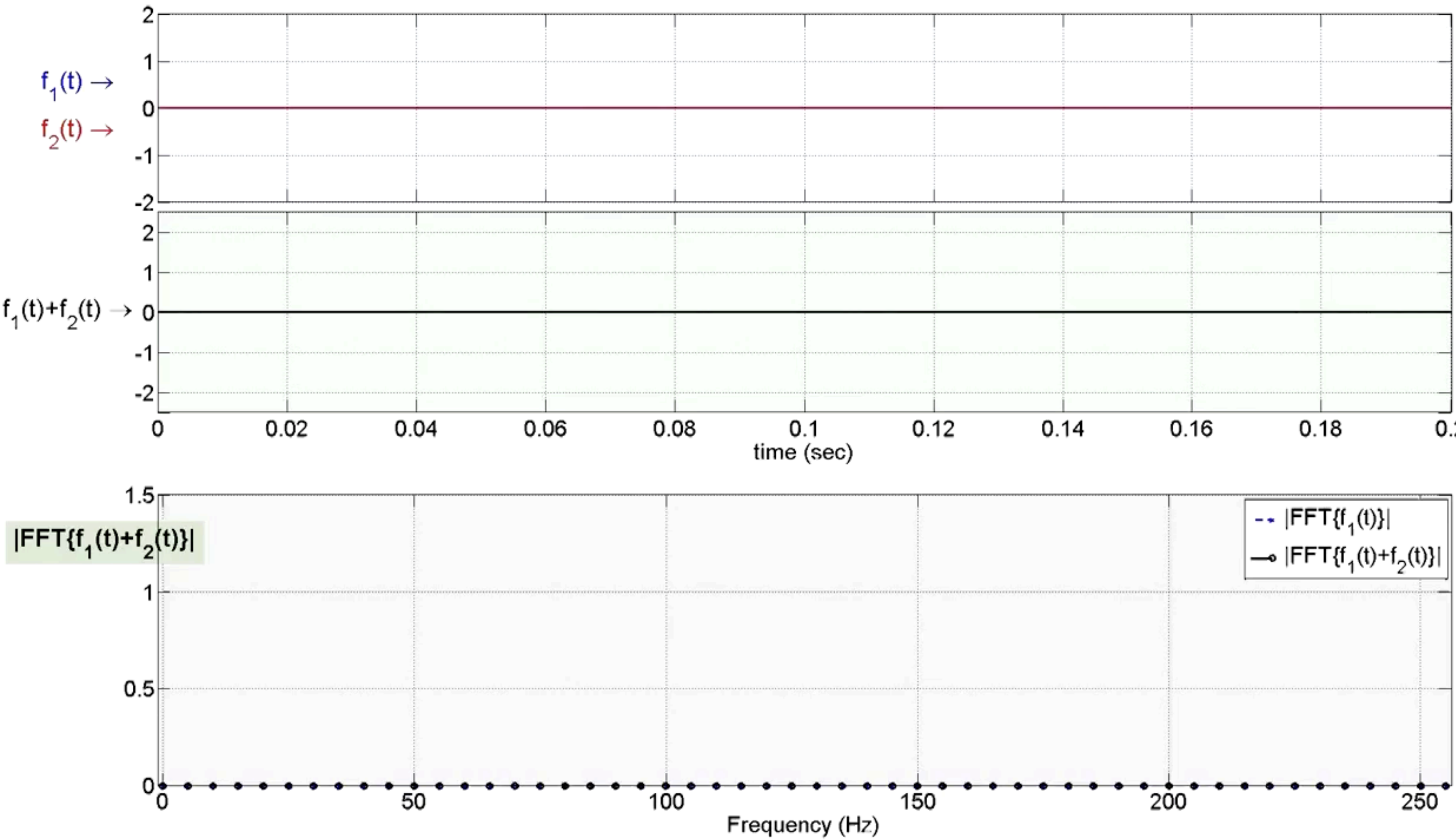
Waves and structure factors

taken from https://en.wikipedia.org/wiki/User:LucasVB/Gallery#/media/File:Fourier_transform_time_and_frequency_domains.gif
Further material at http://1ucasvb.tumblr.com/post/44489240563/the-continuous-fourier-transform-takes-an-input



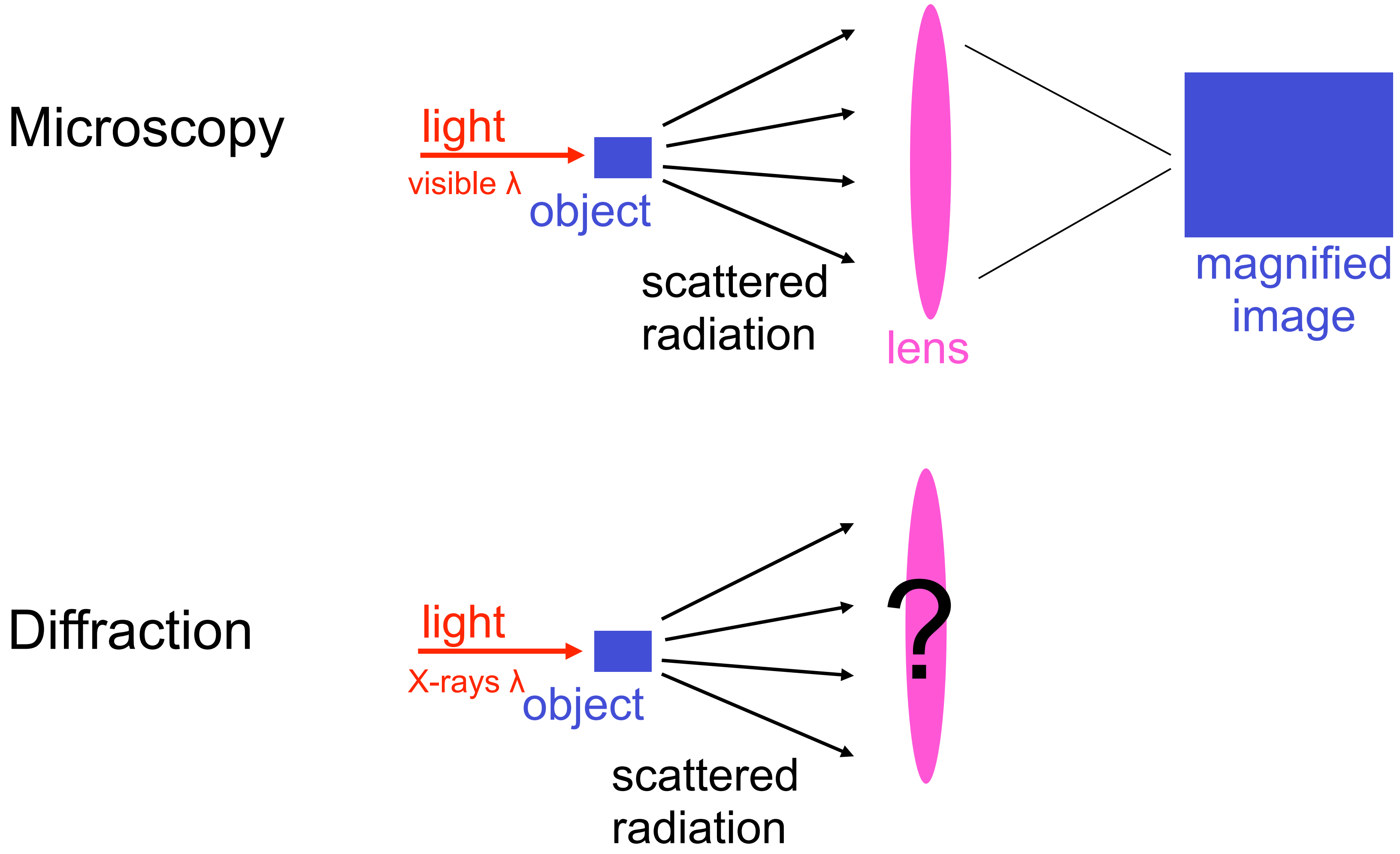
adapted from https://www.youtube.com/watch?v=LznjC4Lo7lE
Matlab code available there

yet one more example of how Fourier works for us…

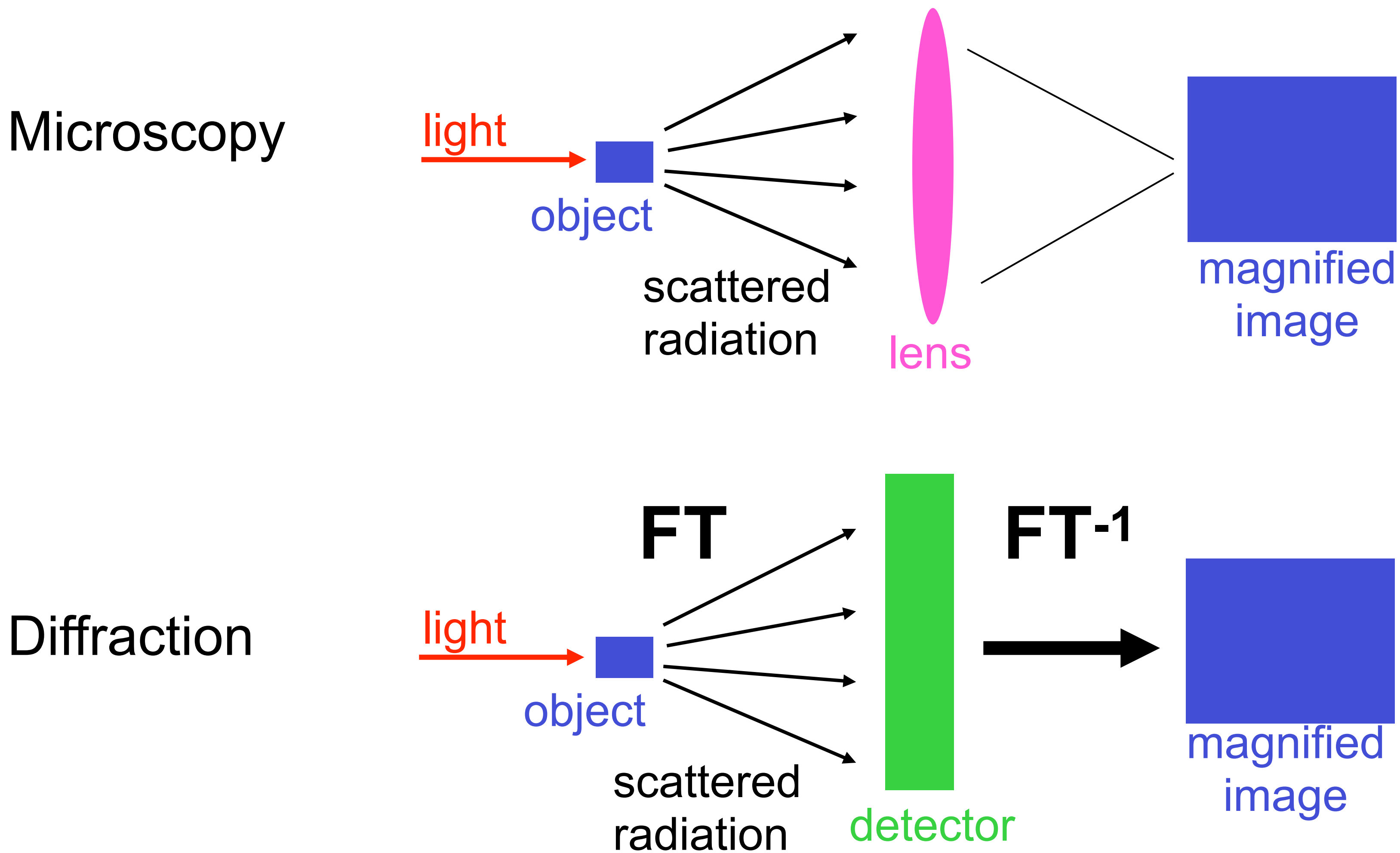


taken from https://www.youtube.com/watch?v=-GYB7khbIA0
Matlab code available there

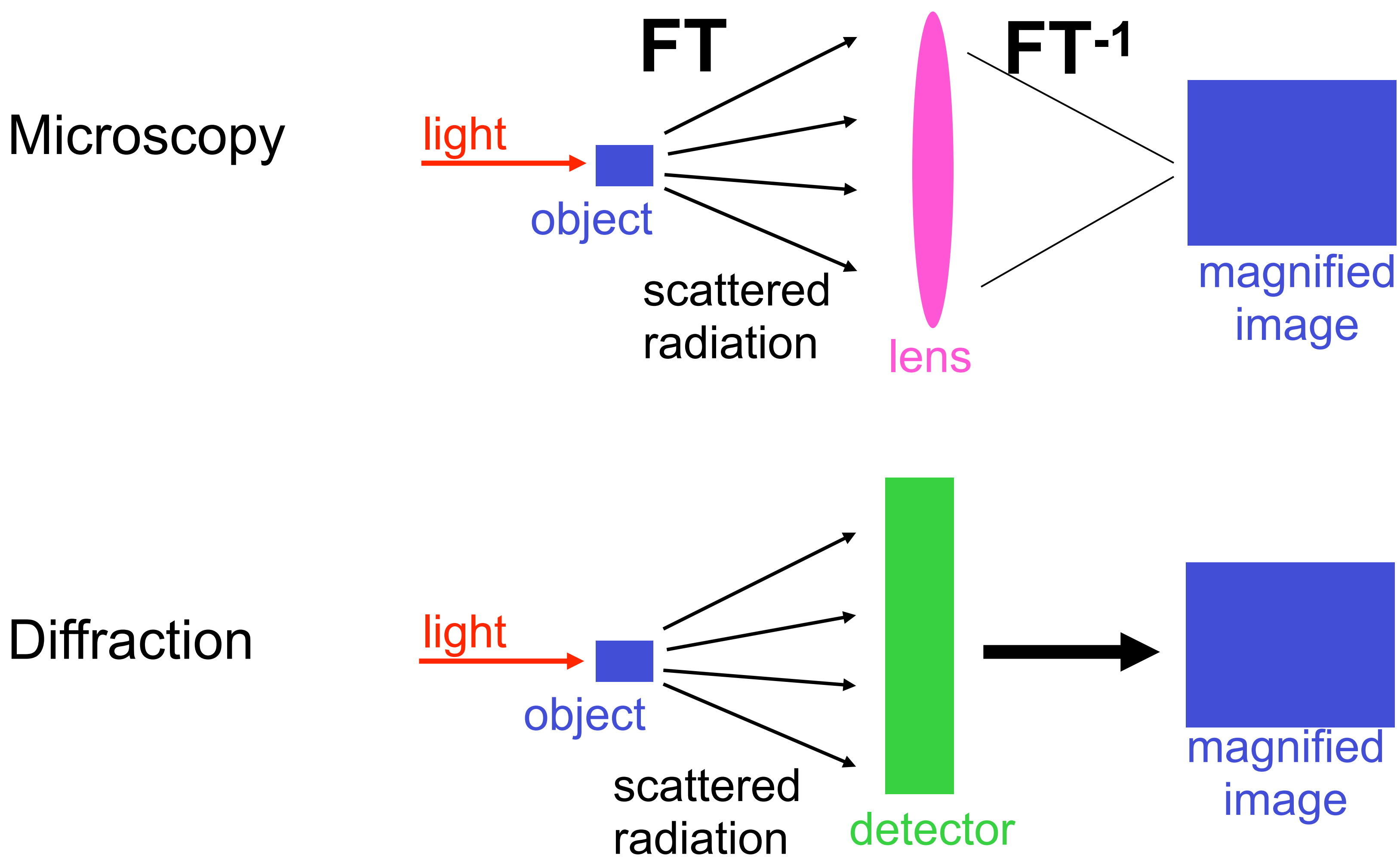
Microscopy and diffraction: cousins



Microscopy and diffraction: cousins



Microscopy and diffraction: cousins



Apart from these concepts, we don't have time now to go into a few other tools in Math that can help in several different aspects and stages of the process:

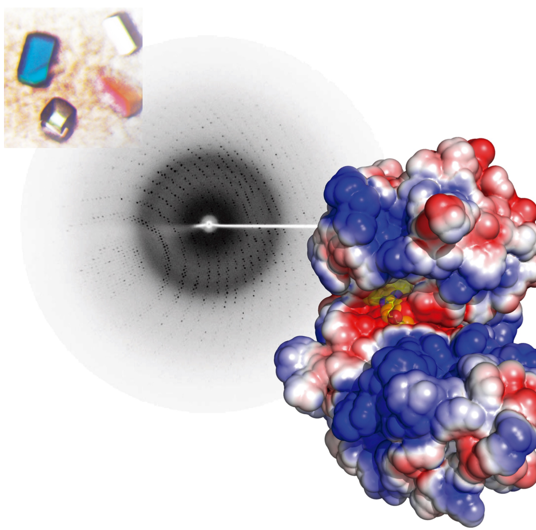
- Matrix algebra (to deal with vector manipulations, as in calculating atomic distances, as in rotating coordinate references and objects, etc)

- Statistics: maximum likelihood, probabilities, distributions, random and systematic errors (extremely valuable in all stages, since we are typically collecting many observations in each experiment, data processing, phasing, direct methods, refinement, etc)

Unit of Protein Crystallography

Institut Pasteur de Montevideo

thank you!



Macromolecular Crystallography School 2018
November 2018 - São Carlos, Brazil