Crystallography in a SP mood - rev 7

Alessandro Falaschi

Abstract

The role of Fourier's analysis in X-ray crystallography is addressed as an application of signal processing, for a SP course given as part of a Bioinformatics degree, where the topic is relevant for protein folding determination. Since multidimensional transforms are better explained in two dimensions, the discussion begins from images, but the reader should have adequate knowledge about 1D transforms. After 3D extension of the concepts, crystallography is introduced informally, trying to keep the theory to the bare minimum. After scattering and diffraction, the Ewald's sphere makes its entrance, and some additional practical considerations are given.

1 Monodimensional recap

A time-dependent signal $s(t)$ can be represented as an (infinite, complex, weighted) integral sum of exponential *orthogonal basis functions* $e^{j2\pi ft}$ with weights $X(f)\,df$

$$x(t) = \int_{-\infty}^{\infty} X(f)\, e^{j2\pi ft} df$$

If $x(t)$ is a *real* signal then $X(f) = X^*(-f)$ and the basis functions become[1] *trigonometric*

$$x(t) = \int_{-\infty}^{\infty} |X(f)| \cos\left(2\pi ft + \arg\{X(f)\}\right) df$$

When $x(t)$ is *real* and *periodic* with period T the sum becomes *discrete*, i.e. the *Fourier series*

$$x(t) = X_0 + 2\sum_{n=1}^{\infty} M_n \cos\left(2\pi \frac{n}{T} t + \varphi_n\right)$$

in which $X_n = M_n e^{j\varphi_n}$ are the complex *Fourier coefficients* with $M_n = M_{-n}$ and $\varphi_n = -\varphi_{-n}$

If $x(t)$ is *bandlimited* the Fourier coefficients are *a finite set*, and the series stops at a *finite* index $n = N$, so that the samples $x_n = x(t = nT_c)$ (with sampling interval T_c) of $x(t)$ are given by an *inverse DFT* as

$$x_n = \frac{1}{N} \sum_{m=0}^{N-1} X_m \, \mathrm{e}^{\, j2\pi \frac{m}{N} n}$$

where the *sequence* x_n *repeats* periodically for $n < 0$ and $n \geq N$. Actually, a signal cannot be *time*- and *frequency*- limited at the same time, but if $T \gg T_c$, i.e. N is large enough, the *aliasing* error will be negligible.

---

[1] By denoting $X(f) = |X(f)|\, e^{j\varphi(f)}$ we have

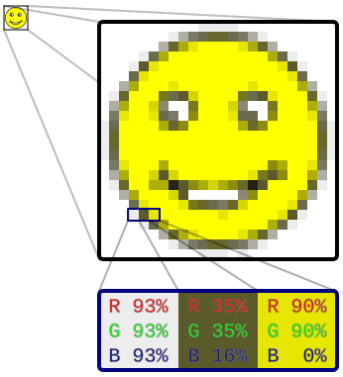
$$\int_{-\infty}^{\infty} X(f)\, e^{j2\pi ft} df = \int_{-\infty}^{\infty} |X(f)|\, e^{j(2\pi ft + \varphi(f))} df =$$
$$= \int_{-\infty}^{\infty} |X(f)| \cos(2\pi ft + \varphi(f))\, df + j \int_{-\infty}^{\infty} |X(f)| \sin(2\pi ft + \varphi(f))\, df =$$

but the integrand in the imaginary part *is odd*, and integrates to zero

2 Raster image signals

Digital images of the *raster* type are made of an $N \times M$ rectangular array $x(x, y)$ of *real* and *positive* values named *pixels* (picture elements), representing the brightness intensity (zero = black, maximum = white) of such a point of the image

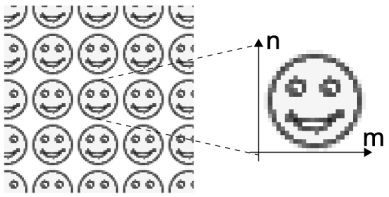
- in this sense images are *sampled* and *quantized* signals, in which the spatial coordinates pair (x, y) take only discrete values, and each pixel can assume only a limited number of values: the brightness dynamic is thus measured in *bits/pixel*, e.g. 8 means a set of 256 different levels
- *color images* are made of *three* arrays, each representing the relative amount of one of three color components by which the image can be decomposed, as for example in the RGB *(Red, Green and Blue)* color model
- *vector graphics* instead build images as a collection of *drawing primitives* (points, lines, curves and poligons) and allow to *zoom in* without loosing resolution - but we will not deal about them



24x24 pixel image of a smiley magnified 10x to demonstrate pixel boundaries. Source Wikipedia

3 Bidimensional DFT

A raster image $s(n, m)$ can be thought of as a *single period* of a bi-dimensional periodic sequence as shown.



This means that we can directly extend the uni-dimensional DFT in two dimension, as

$$S(h, k) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} s(n, m)\, \mathrm{e}^{-j2\pi \frac{hn}{N}} \mathrm{e}^{-j2\pi \frac{km}{M}} = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} s(n, m)\, \mathrm{e}^{-j2\pi \left(\frac{hn}{N} + \frac{km}{M}\right)} \tag{1}$$

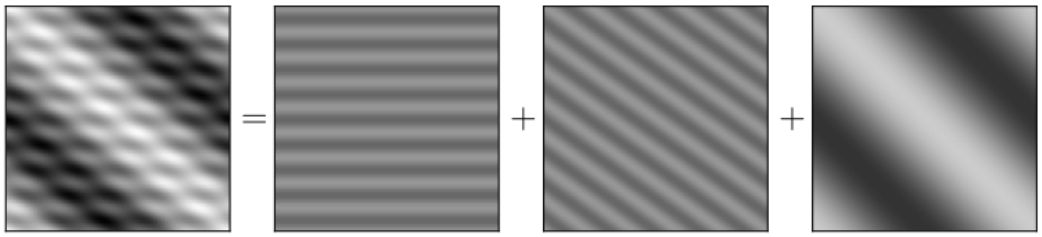
showing that the discrete Fourier transform of a matrix signal is itself a matrix signal of the same dimensions. Well yes, each term $S(h, k)$ of (1) is the result of an inner product between matrices!

Spatial harmonic frequencies Indexes (h, k) have the meaning of *spatial frequencies*, where h refers to the number of vertical cycles (along the n axis) and k expresses the number of (horizontal) cycles along the m axis. As for the uni-dimensional case, bandwidth restrictions apply, so that h (or k) indices greater than $\frac{N}{2} - 1$ (or $\frac{M}{2} - 1$) actually refers to *negative* frequencies, in one or both directions.

2D IDFT The inverse 2D-DFT can be written as

$$s(n, m) = \frac{1}{MN} \sum_{h=0}^{N-1} \sum_{k=0}^{M-1} S(h, k)\, \mathrm{e}^{j2\pi \left(\frac{nh}{N} + \frac{mk}{M}\right)} \tag{2}$$

allowing to get back the original image. This formula has the same meaning as in one dimension, i.e. the original image $s(n, m)$ can be reconstructed as a weighted sum of 2D arrays, as given in the following example (source), in which the original synthetic squared image on the left is built by the sum of the three images on the right.



3.1 Bidimensional basis functions

These are given by the 2D arrays

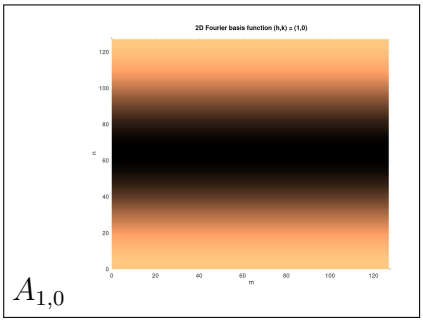
$$A_{h,k}(n,m) = \mathrm{e}^{j2\pi\left(\frac{nh}{N}+\frac{mk}{M}\right)} = \cos\left(2\pi\left(\frac{nh}{N}+\frac{mk}{M}\right)\right) + j\sin\left(2\pi\left(\frac{nh}{N}+\frac{mk}{M}\right)\right) \tag{3}$$

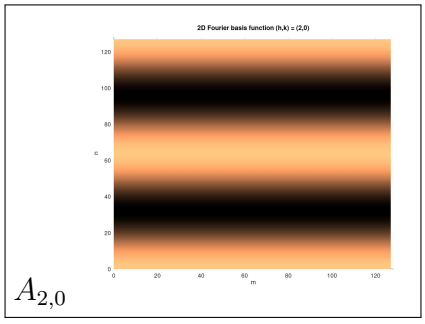
for $h = 0, 1, \cdots, N-1$ and $k = 0, 1, \cdots, M-1$, i.e. there are $N \times M$ basis functions, each one

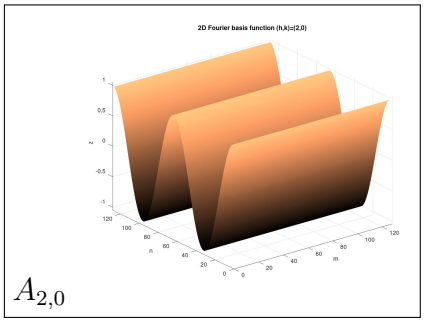
- expressing the combination of spatial frequencies $(h.k)$, and
- made of an $N \times M$ array of complex numbers, obtained by varying (n, m) in the expression of $A_{h,k}(n,m)$

We now take *a tour* examining the actual behavior of some the $A_{h,k}(n,m)$ *planes*, evaluated for different choices of the h, k pair of spatial frequencies, where h refers to vertical variations, and k to horizontal ones.

First of all, let us realize that $A_{0,0} = 1$ for every (n, m), because $\cos(0) = 1$ and $\sin(0) = 0$. Then, by keeping $k = 0$ we will observe only vertical frequencies, as given in the two figures on the left with one $(h = 1)$ or two $(h = 2)$ vertical periods. There and in the following, images refers the real part[2] of $A_{h,k}$ and to a square grid with $M = N = 128$, using a *colormap* with *black* for negative values and *copper* per positive ones, as shown in the 3D image of $A_{2,0}$ on the rigth.

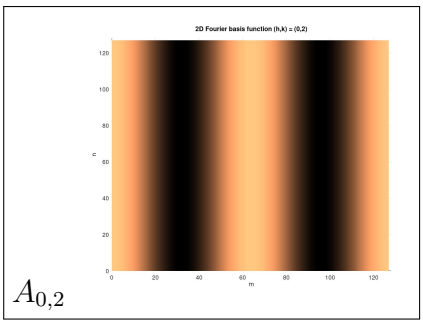


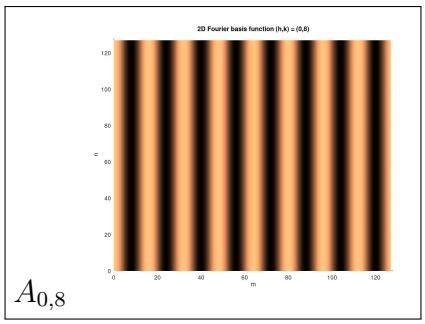


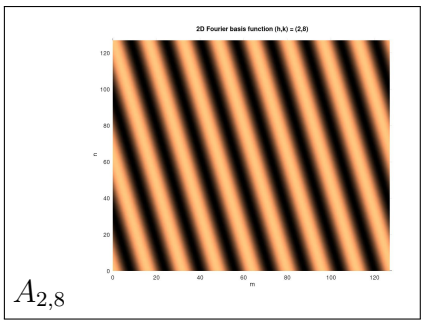


Obviously enough, for $A_{0,2}$ (see the next pictures) we observe two periods of *horizontal* variation, and for $A_{0,8}$ eight, but the interesting thing is the last picture on the right, where two vertical periods are shown together with eight horizontal ones, resulting in the $A_{2,8}$ pattern.

[2]The imaginary part differs only for *an half period* shift of the results.





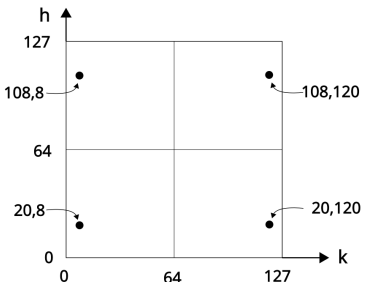


3.2 The frequency (or reciprocal) lattice

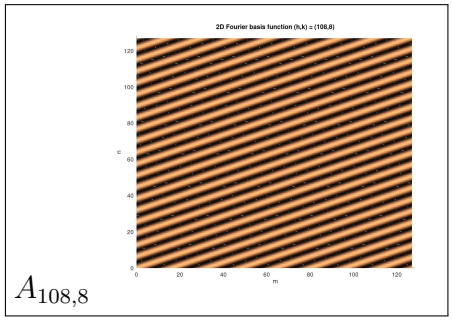
We are now ready for making one step more. Every pair of h, k spatial frequency indexes can be mapped into an array with the same dimensions of the image, called the *reciprocal lattice*, to which the relevant basis function $A_{h,k}$ corresponds. Think at it in this way:

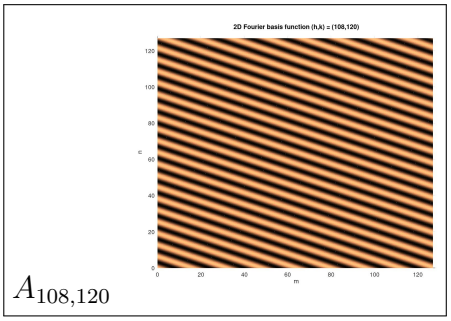
- as every pixel (n, m) in an image array takes a *real* value $s\,(n, m)$ giving its brightness, so
- every element (h, k) of the reciprocal lattice takes a *complex* value $\mathrm{S}(h, k)$ telling *how much* the basis function $A_{h,k}$ does participate to the construction of the image

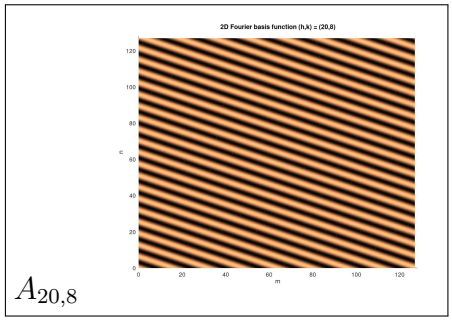
This mapping can be *substantiated* by looking at the *real component* of four basis functions ($A_{108,8}$, $A_{108,120}$, $A_{20,8}$, $A_{20,120}$) shown on the right, drawn in the same ordering of the positions of the respective indexes within the reciprocal lattice, shown on the left.

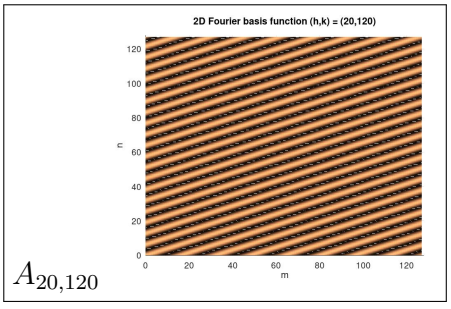


2D frequency lattice









As already noticed, spatial frequencies indexes greater than $\mathrm{N}/2$ are related to *negative frequencies*, and the corresponding basis functions *mirrors* those for positive frequencies. So, the *vertical* frequency of $A_{20,8}$ is the same as for $A_{20,120}$ (index $h = 20 < 64$), while instead the *horizontal* frequency *is mirrored* (index $k = 120 > 64$, and $k = 120 = 128 - 8$), resulting in the change of orientation for the oscillations. The same reasoning applies when comparing $A_{108,8}$ to $A_{20,8}$, while the equality in between $A_{20,8}$ and $A_{108,120}$ can be explained as a double mirroring.

3.3 The DFT values

Let us now put our attention to some properties of the values $S(h,k)$ evaluated by the use of (1).

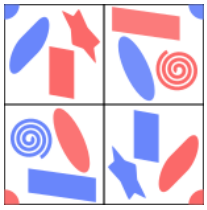
Periodicity and conjugate symmetry First of all, (1) holds true also if h,k indices goes *outside* the reciprocal lattice boundaries, making $S(h,k)$ *periodic* in both the spatial frequency directions, meaning that

$$S(h,k) = S(h+iN, k+jM) \qquad \forall\, i,j \in \mathbb{Z} \tag{4}$$

Moreover, as the pixel values of $s(n,m)$ *are real*, it can be shown that for $S(h,k)$ a *conjugate symmetry* property holds true, which can be conveniently written as

$$S(h,k) = S^*(N-h, M-k) \qquad 0 \le h \le N-1, \quad 0 \le k \le M-1 \tag{5}$$

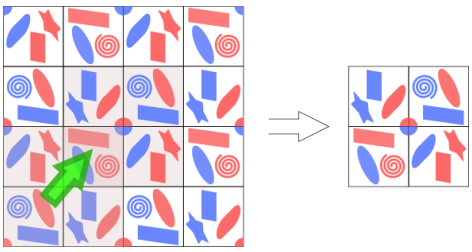
Symmetry in practice The latter properties mean that (for a square image) row $N/2$ and column $M/2$ *break* the DFT spectrum $S(h,k)$ into four quadrants. For the *real part* and the *magnitude*, the upper-right quadrant values are a mirror image of those in the lower-left, while values in the upper-left are a *mirror image* of those in the lower-right. This symmetry also holds for the *imaginary* part and the phase, except that their values in the mirrored quadrants are opposite in sign. In other words, every point in the frequency lattice has a matching point placed symmetrically on the other side of the center of the image (row N/2 and column N/2). One of the points is the positive frequency, while the other is the matching negative frequency.



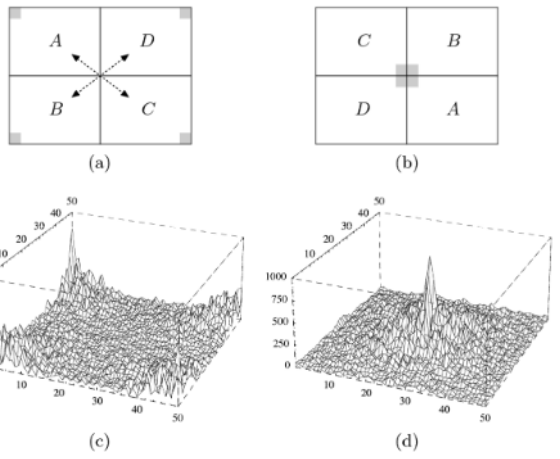
example sketch of a 2D DFT symmetry: shapes are the modulus, colors the phase sign

Furthermore, the more an (h,k) index pair is near to one of the four corner of the reciprocal lattice, the lower will be the frequencies appearing in the associated basis function; on the opposite, the nearer they are to the center, the higher there will be the frequencies, up to the limit (at the reciprocal center) of two contiguous values (in the image space) per period.

Centering the zero frequency By virtue of the periodicity (4) of the DFT, it is customary to reorganize its values by moving them *diagonally* by $N/2$, so that the lattice indices relating to the low frequencies are shown at the center of the DFT array, as is sketched on the side. This is also in accordance with the definition of a *continuous* 2D Fourier transform.



On the side another picture (source) about this concept is shown, in which the gray rectangles at the corners in (a) represents low frequency regions that in (b) are rejoined together, while in (c), (d) the *module* of an actual 2D-DFT is shown, undergoing the same rearrangement. We can observe that in general, most of the energy in a picture is concentrated at low frequencies, giving information about regions in the image with nearly constant brightness values.

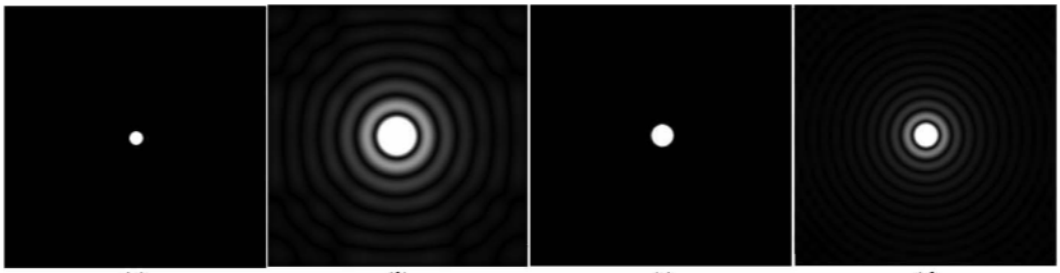


Please keep in mind that such a re-arrangement of the displayed 2D DFT values also modifies the location of the 2D Fourier basis functions on the reciprocal lattice!

3.4 Visual evaluation of the properties of a 2D DFT

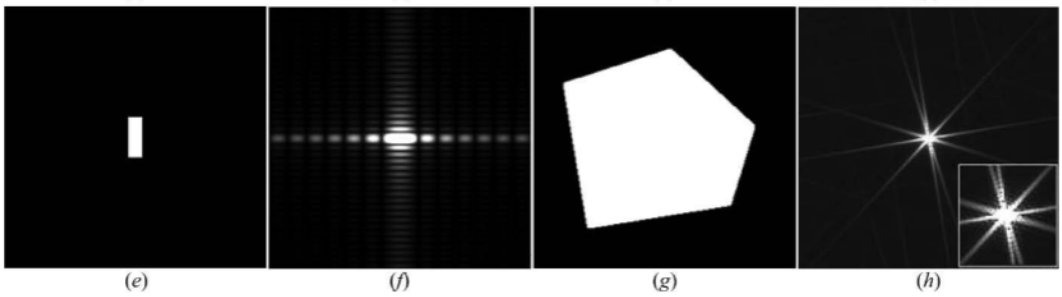
Instead of developing formulas, we try to deepen our knowledge by using a mainly visual approach, on the base of pictures[3] taken from Aubert, E. & Lecomte, C. (2007). Illustrated Fourier Transforms for crystallography. Journal of Applied Crystallography and others, partly reproducible with the aid of the well-done didactic tool available at http://www.jcrystal.com/products/ftlse/.

Inverse relationship of the dimensions in the two domains The small circle image given in fig (a), which can be thought as a sort of (radially symmetric) 2D equivalent of a 1D $rect\,(t)$, has its 2D DFT *module* as given in fig (b): doesn't it look very similar[4] to a $sinc\,(f)$? Please note that also in two dimensions a *widening in space* (fig (c)) is reflected in a *shrinking* in frequency (fig (d): in fact, the energy of major area objects is concentrated in the region relating to lower frequencies.



Shape relationships Further insight can be gained from the following fig (e), showing a true 2D $rect$ (i.e. $rect_{\Delta x}\,(x) \cdot rect_{\Delta y}\,(y)$) with different widths in the two directions: its 2D DFT (fig (f)) is *the product* of two $sinc$ in frequency, with the oscillation period of the tails which is inversely proportional to the $rect$'s width in the same direction. This is because the 2D DFT of factorized images is itself factorized, and given by the product of the two 1D DFT: in fact, in such case (1) factorizes.

The polygon in fig (g) has sharp brightness changes at its five edges, and the corresponding 2D DFT (fig (h)) is characterized by five (double) tails which are perpendicular to the edges, crossing at the origin[5], and magnified at the bottom-right corner.



[3]The following images are composed of 256 x 256 pixels, and coded by 2^8 greyscale values, from 0 (black) to 255 (white).

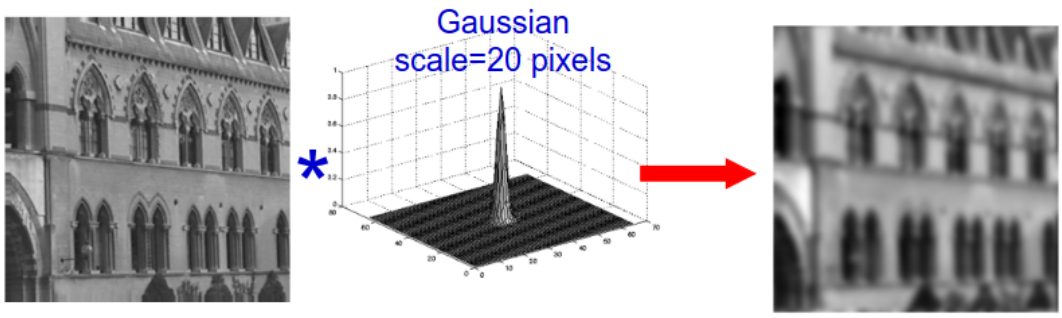
[4]Actually it's a *Bessel* function, but doesn't matter so much.

[5]Remembering images of the sun recorded in movies, often observed as a disk surrounded by six tails, one recalls that usually hexagonal diaphragms are used in front of the camera.

Convolution and filtering An image $f(n,m)$ can be filtered by performing a 2D convolution with a 2D impulse response $h(n,m)$; by denoting the resulting image as $g(n,m)$, the formal expression is

$$g(n,m) = f(n,m) * h(n,m) = \frac{1}{MN} \sum_{h=0}^{M-1} \sum_{k=0}^{N-1} f(h,k)\, h(n-h, m-k) \tag{6}$$

Technically, this means that every pixel of $g(n,m)$ is the result of a *weighted sum* of pixels of $f(n,m)$, where the weights are given by the values of $h(n,m)$, after that it has been shifted in a way[6] that its center is located around the location of the pixel $g(n,m)$ for which we are calculating the convolution. Depending on the values of $h(n,m)$ the resulting $g(n,m)$ will be a low-pass, high-pass or *whatever else* version of $f(n,m)$. As en example, by using a *Gaussian spatial filter* the result will be *a smearing* of the original pixel values, i.e. a low-pass effect, as shown below[7]

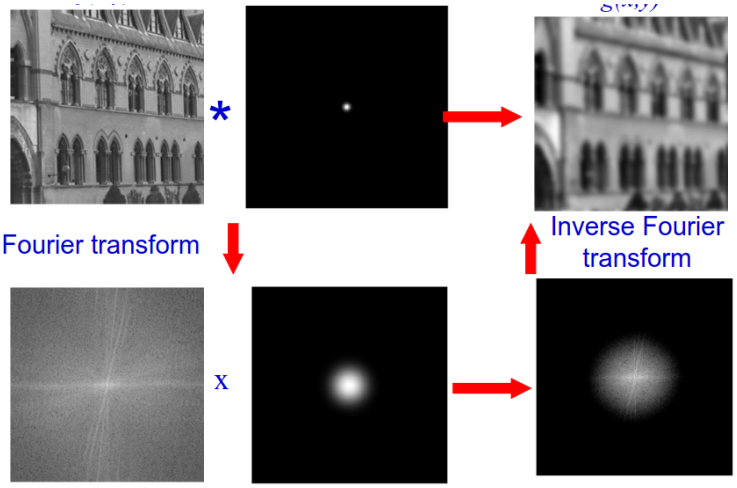


Convolution theorem Also in the 2D case, the transform of a convolution is the product of the transforms of the original signal and that of the impulse response, i.e.

$$g(n,m) = f(n,m) * h(n,m) \quad \overset{\mathcal{F}}{\Longleftrightarrow} \quad G(h,k) = F(h,k) \cdot H(h,k) \tag{7}$$

allowing a great reduction[8] in complexity: whereas (6) exhibits an $O\left((N \cdot M)^2\right)$ complexity, performing the transform by means of an FFT reduces the burden to about $O(N \cdot M \cdot log_2(M \cdot N))$.

The picture on the side illustrates the process: after having calculated the 2D FFT of $f(n,m)$ and $h(n,m)$ (again a *Gaussian spatial filter* whose transform is Gaussian itself), they are element-wise multiplied[9], and the result antitransformed. Please note that multiplication by $H(h,k)$ (center bottom) causes the resultant image to retain only the low-frequency values, thus obtaining a low-pass effect.



[6] More precisely, in analogy with what happens in the 1D convolution, before of the shifting $h(n,m)$ has to be *inverted in direction*, which in this case means *bottom to top, right to left*. Moreover, if either f or h is smaller than the other, it may be thought as zero filled up to the same size. Finally, since images are considered periodic, when the h (or k) index grows, the part of $h(n,m)$ which does not overlap to $f(n,m)$ re-enters at the opposite side.

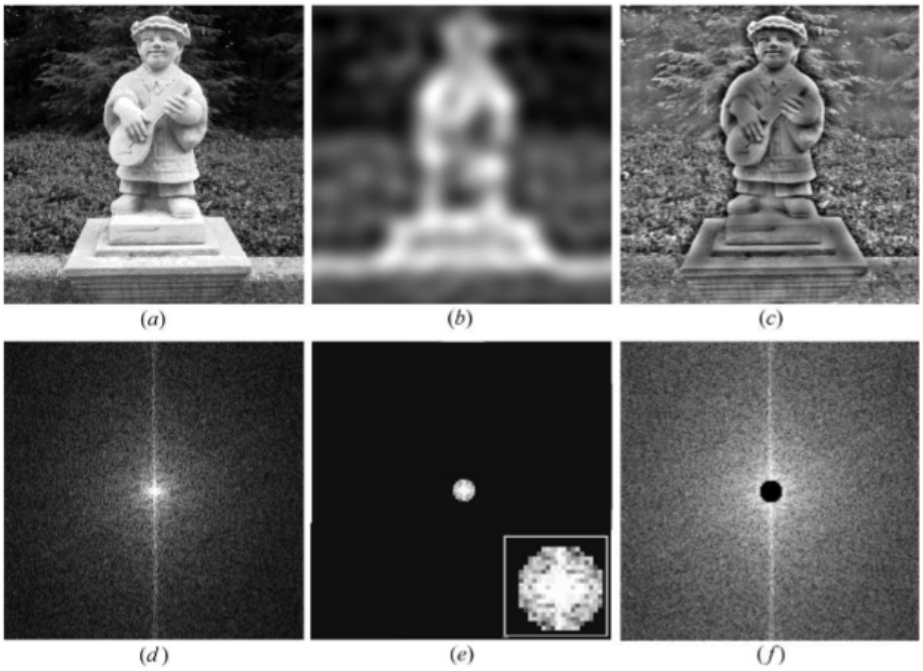
[7] This image is taken from https://www.robots.ox.ac.uk/~az/lectures/ia/lect2.pdf, where many other good insights about the 2D Fourier transform can also be found.

[8] However, if *the kernel* $h(n,m)$ is much smaller than $f(n,m)$, i.e. if its spatial extension is much reduced, complexity decreases, and the operation can be advantageously evaluated in the space domain.

[9] See for instance https://en.wikipedia.org/wiki/Hadamard_product_(matrices)

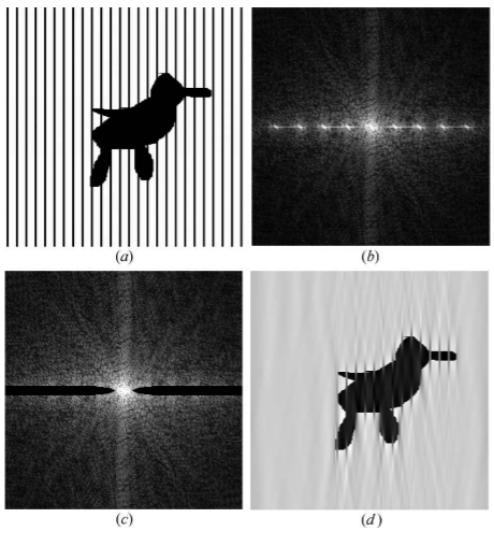
Filters defined in frequency The previous theorem allows us to obtain the desired effect by working directly in the frequency domain: in fact, we can design $H(h,k)$ so that it lets only the desired frequencies to pass, or that suppress the undesired ones, in a similar way to what happens for 1D filters.

The picture below shows an image (a) and its Fourier transform (d); the latter is masked as in (e) in order to retain only a circular central area (magnified in the bottom-right inset) related to low-frequencies. The result is then antitransformed, and gives the blurred image in (b). On the contrary, if low frequencies are masked as in (f)[10], the corresponding antitransform (c) retains only the high frequency components, which encode sharp variation of brightness and edges of the original image.



By remembering that 2D DFT and its inverse assume that images and their transforms both are periodic in the two directions; we can interpret the two vertical tails observable in (d) as reflecting the vertical brightness discontinuity seen in (a), which by converse has no discontinuity in the horizontal direction.

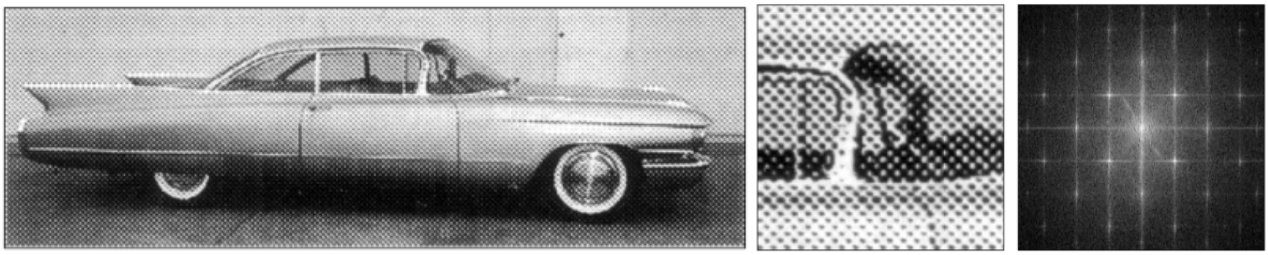
Periodic noise removal Sometimes an image could be affected by some sort of periodic noise (fig a), which is reflected by the presence of bright peaks in the DFT (fig b), located at the reciprocal lattice positions where the basis functions responsible of the periodic noise (and its harmonics) lies. As 2D DFT retains the linearity property, i.e. the transform of a sum is the sum of the transforms, the effect of the noise can be attenuated by trying to remove its contribution to the DFT, as shown in fig c where the contribution of the peaks have been masked. In fact, the 2D IDFT of fig c is fig d, where much of *the cage* has almost disappeared, thus letting to bird *to be free*!



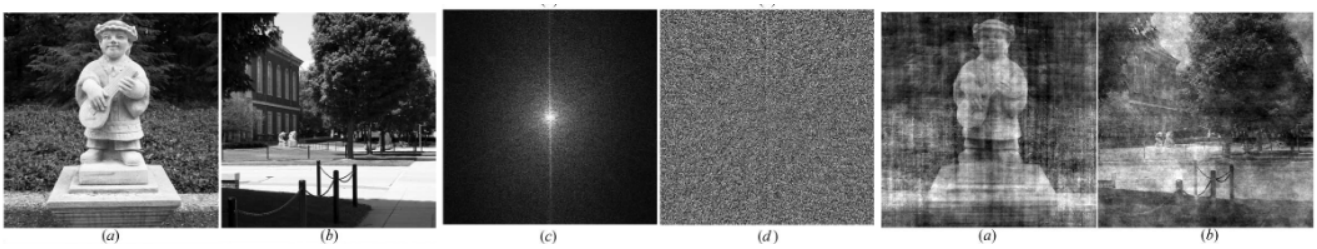
Below it is shown (source) another kind of periodical noise, as arising in photos included in old newspapers: here the noise

[10] Here the magnitudes have been enhanced for clarity reasons.

is a grid of points, as evidenced by the enlargement shown in the centre, which makes the regular pattern of peaks to appear in its 2D DFT on the right.



Relevance of the phase information During the study of 1D Fourier analysis for time-dependent signals, we saw that the actual waveform drastically changes if its phase spectrum is modified. The same applies to 2D transforms, except that in this case the phase modification gives unexpected results about the image appearance. Consider the two images (a) and (b) below on the left: module and phase of (a) are respectively shown as (c) and (d) in the center. As evident, the phase information doesn't seem to give particular information about the original image. But here comes the surprise: if the phase information (d) is substituted to the phase of the DFT of (b-left) and then a IDFT is performed, the resulting image (a-right) much more resembles to (a-left) than to (b-left) from which the module was taken. A similar paradox results, if we take module from (a-left) and phase form (b-left): after the IDFT, the resulting images is (b-right). So, we can affirm that the shapes contained in the image are essentially derived from phase information.



Dirac Pulses A 2D discrete Dirac pulse is defined as $\delta(n,m) = \begin{cases} 1 & \text{if } n = m = 0 \\ 0 & otherwise \end{cases}$ and its 2D DFT (1) equals to 1 for any h, k in the reciprocal space. A *shifted pulse* at position n_0, m_0 is written as $\delta(n - n_0, m - m_0)$, and it is $\neq 0$ only for $n = n_0$ and $m = m_0$. When a shifted pulse is convoluted with an image, the result is the same image, shifted in the location where the pulse is located.

$$f(n,m) * \delta(n - n_0, m - m_0) = f(n - n_0, m - m_0) \tag{8}$$

The picture at the side[11] gives a visual interpretation of such a statement. The left image is a gaussian intensity peak, displaced from the top left corner by some amount. Its convolution with a sum of four Dirac pulses, further displaced in the image plane, is the

[11] In this example, the origin of the image is put at the top left corner.

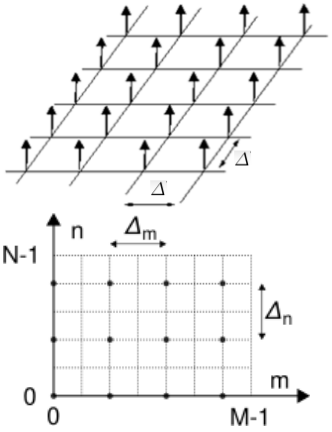
sum of four individual convolutions, each one of which displaces the gaussian peak in a position equal to the sum of the displacements of the gaussian and the pulse.

Obviously enough, convolution with a pulse in the origin leaves the images unaltered, as can be realized by thinking that the 2D DFT of $\delta(n,m)$ is one everywhere, and its product with the DFT of the images leaves it unchanged[12].

Convolution with a bed-of-nails In the development of the theory leading to the sampling theorem which holds for one-dimensional, bandwidth limited signals, we studied the role of the train of pulses $\pi_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$, whose transform is *again* a train of pulses in frequency $\Pi_{1/T}(f) = \frac{1}{T}\sum_{n=-\infty}^{\infty} \delta(f - {}^{n}/_{T})$, with a spacing in frequency that is the reciprocal of the one in time.

A similar results also holds in two dimensions: the discrete 2D, regularly spaced equivalent of an 1D train of pulses is defined as

$$\pi_{\Delta}(n,m) = \sum_{p=0}^{\frac{N}{\Delta_n}-1} \sum_{q=0}^{\frac{M}{\Delta_m}-1} \delta(n - p\Delta_n, m - q\Delta_m) = \mathrm{BON}_{\Delta_n,\Delta_m,N,M}(n,m)$$



and by depicting it like in figure, it takes the name of a *bed-of-nails* or BON. The Δ_n, Δ_m, N, M subscripts indicate its periods (vertical and horizontal) and its extensions, which in the example are $2, 2, 6, 8$ respectively; to ensure that the BON is periodic, as a need for the DFT operations, ${}^{N}/_{\Delta}$ and ${}^{M}/_{\Delta}$ must be integer ratios.

By virtue of the linearity of the convolution, the translation property (8) of an image when it is convoluted with a translated pulse applies to all the pulses that are present in the BON with which a convolution is made. Consequently, the image *is replicated* in the positions of every pulse in the BON, as represented by the following sketch (source), in which the image is supposed to be smaller than the Δ period.

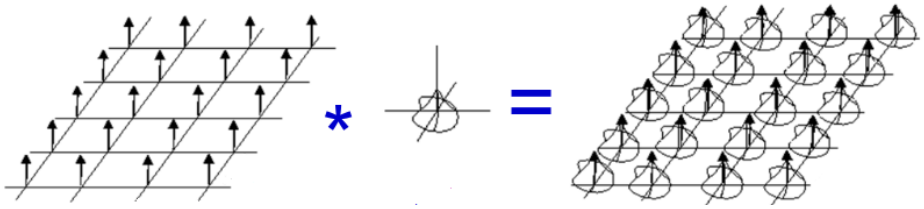


Image repetition makes its spectrum punctured Since the transform of a BON still is a BON in the reciprocal space (see appendix 10.2), and as a consequence of the convolution theorem (7), convolution of an image with a BON in the real space makes the DFT of the result to be *the product* of the 2D DFT of the image, times a BON in the reciprocal space.

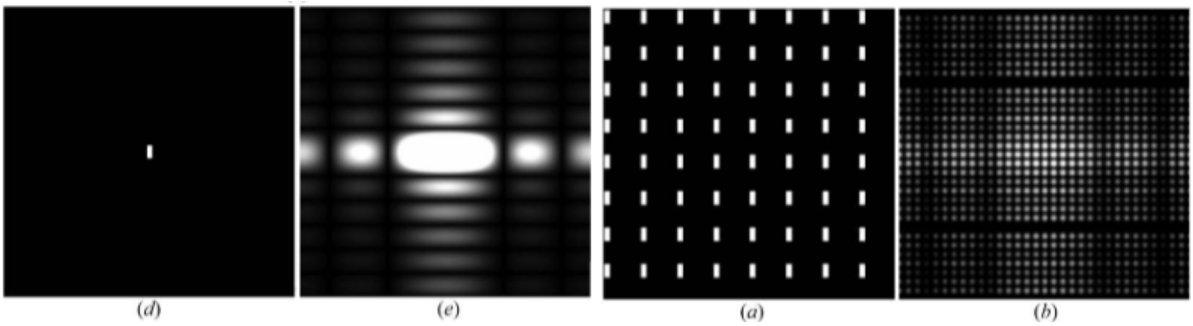
Consider for example the 2D $rect$ shown in the following fig (d), whose 2D DFT module is given in fig (e). When (d) is convoluted with a BON, we obtain the pattern shown as (a), with the 2D $rect$

[12] By converse, evaluation of the 2D DFT of a Dirac pulse $\delta(n - n_0, m - m_0)$ located at position n_0, m_0 *in the real space* gives

$$\mathrm{DFT}\{\delta(n - n_0, m - m_0)\} = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} \delta(n - n_0, m - m_0)\, \mathrm{e}^{-j2\pi\left(\frac{hn}{N} + \frac{km}{M}\right)} = \exp\left(-j2\pi\left(\frac{hn_0}{N} + \frac{km_0}{M}\right)\right)$$

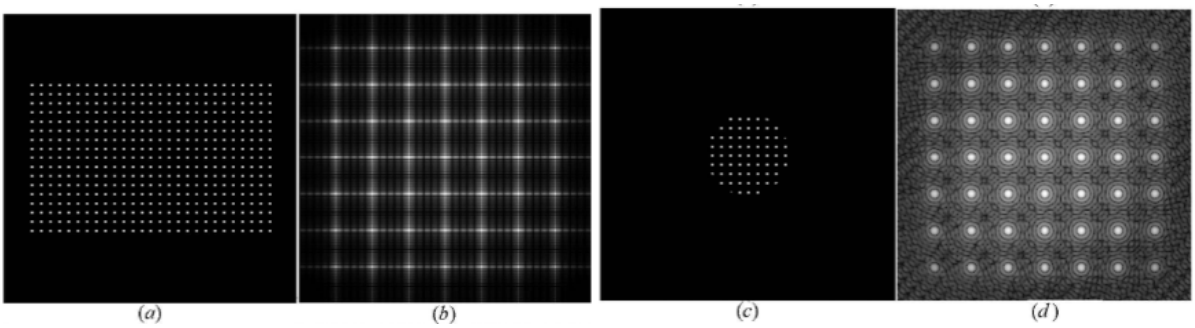
so that the DFT of an image convoluted with a shifted Dirac *inherits* a negative linear phase contribution of $2\pi\frac{n_0}{N}$ radians for every point of the reciprocal lattice in the vertical direction, and of $2\pi\frac{m_0}{M}$ radians in horizontal.

replicated in the pulse positions within the BON: in such cases, the $rect$ is sometimes referred to as *the motif* which *decorates* the BON, where the latter gives the lattice spacing, and the motif gives the shape of the object which is replicated. Finally, if we perform a 2D DFT of (a) we obtain the reciprocal space module shown as (b), which can easily recognized to be the product of (e) times the BON resulting from the transform of the BON used for building (a) (the large spacing Δ used in the real space BON reflects in a narrower period $1/\Delta$ in the reciprocal space, see § 10.2).



Windowing of the BONs Let us now consider the effect of limiting *the extension* of a BON with respect of the *whole period* of the image (always consider that the DFT makes things to be periodic). Such an operation has a similar effect of the *windowing* already analyzed for the 1D case: the convolution theorem (7) also applies in the *reverse direction*, so that *the product* (in direct space) of an image times a spatial mask (the window function), give rise to a 2D DFT which is *the convolution* (in reciprocal space) of the original DFT with the DFT of the mask.

An understanding of this concept can be gained with the aid of the following pictures. In (a) a regular (periodically infinite) BON is windowed by a rectangular mask: its 2D DFT module is shown as (b), where it can be appreciated that, besides an enlarged spacing of the transformed BON, every Dirac pulse of it is convoluted by the DFT of the mask (a bidirectional $sinc$). In (c), the same BON image is now windowed by *a narrower* circular mask, to which corresponds a *wider* (and circularly symmetric) 2D DFT, producing the 2D DFT of (c) shown as (d).



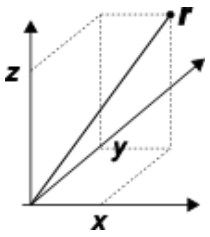
Radon Transform and Tomography Oh well, this goes far beyond the current scope of this document, but it will be developed in the future, since it has very important implication in 2D and 3D *image probing*. The associated theory can be found with some help from the Wikipedia pages Radon transform and Tomography.

4 Tri-dimensional Fourier analysis

The three-dimensions Fourier transform of a function of three variables $f(x,y,z)$ is formally expressed as

$$F(u,v,w) = \int_x \int_y \int_z f(x,y,z)\, e^{-j2\pi(ux+vy+wz)} dxdydz \tag{9}$$

where u, v, w are a triad of volumetric frequencies along the directions of x, y, z.



The notation can be simplified by re-writing the 3D coordinate triads (x,y,z) and (u,v,w) in vector form, i.e. as $\boldsymbol{r} = (x,y,z)^T$ and $\boldsymbol{u} = (u,v,w)^T$. This allows to write the exponent that appears in (9) in the form of a scalar product, i.e. $ux+vy+wz = \boldsymbol{u}\cdot\boldsymbol{r}$, so that (9) can be rewritten in the more compact form

$$F(\boldsymbol{u}) = \int_{\boldsymbol{r}} f(\boldsymbol{r})\, e^{-j2\pi\boldsymbol{u}\cdot\boldsymbol{r}} d\boldsymbol{r} \tag{10}$$

Integrals in three dimensions The simple and double integral utilized in the case of functions of one or two variables have a direct geometric interpretation, as it respectively represents the evaluation of *the area* underlying a function of a single variable, and *the volume* below the surface described by a two-dimensional function. On the contrary, the integral of a function defined in three dimensions seems to have a more subtle interpretation, although it has an extensive use in many physical applications, the simplest of which is the evaluation of the total mass of an *dishomogeneous* object, once its *volumetric density* (for example, measured in grams/mm^3) is known.

In other words, the integral $\int f(\boldsymbol{r})\, d\boldsymbol{r}$ is an abstraction with the meaning of a sum, and the triple integral *sums* the density $f(\boldsymbol{r})$ *times* the infinitesimal volume $d\boldsymbol{r}$, for any position $\boldsymbol{r}$ within the extremes of integration.

Periodic densities Consider now a density $f(\boldsymbol{r})$ which is also *periodic* in all of the three directions x, y and z, with periods Δ_x, Δ_y and Δ_z respectively. This can be the case when $f(\boldsymbol{r})$ represents the *electrons density of charge* in a crystal, which is the periodical repetition of the electron density of a single molecule (more on this later). Then, the more appropriate Fourier description for such a situation is a three-dimensional Fourier series

$$f(\boldsymbol{r}) = \textstyle\sum_h \sum_k \sum_l F(h,k,l)\, e^{j2\pi\left(\frac{hx}{\Delta_x}+\frac{ky}{\Delta_y}+\frac{lz}{\Delta_z}\right)} \tag{11}$$

where $F(h,k,l)$ are the *Fourier coefficients* given by

$$F(h,k,l) = \frac{1}{\Delta_x\Delta_y\Delta_z} \int_{\Delta_x}\int_{\Delta_y}\int_{\Delta_z} f(x,y,z)\, e^{-j2\pi\left(\frac{hx}{\Delta_x}+\frac{ky}{\Delta_y}+\frac{lz}{\Delta_z}\right)} dxdydz \tag{12}$$

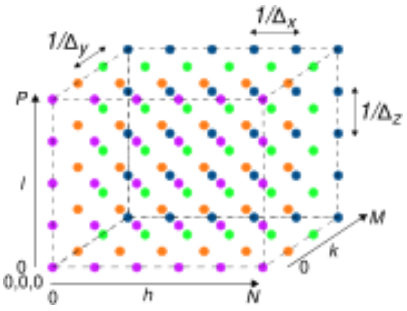
in which $\Delta_x\Delta_y\Delta_z$ is *the volume* V of a single 3D period, and h, k, l are *integers* representing the indexes of the *volumetric harmonic frequencies* in the three directions x, y, z, corresponding to the *continuous* frequencies $u = \frac{h}{\Delta_x}, v = \frac{k}{\Delta_y}$ and $w = \frac{k}{\Delta_z}$ as defined by (9).

3D reciprocal lattice For a periodic density $f(\boldsymbol{r})$ whose values are real, its Fourier coefficients (12) come in complex conjugate pairs, i.e. $F(h,k,l) = F^*(-h,-k,-l)$, so that it is sufficient to know only *half*[13] *of the index space*. Moreover, if the values $F(h,k,l)$ become negligible for indexes

[13] That is, it is sufficient to know the values of $F(h,k,l)$ with only two of the three indices that change sign.

$(h,k,l) \geq (N,M,P)$ then $f(\boldsymbol{r})$ can be considered to have *a limited bandwidth*. In this case all the indexes (h,k,l) for which $F(h,k,l)$ has to be known in order for $f(\boldsymbol{r})$ to be reconstructed by (11) lie in a *parallelepiped*, which is called *the reciprocal lattice*.

The sketch on the right tries to give the idea, where the lower left corner corresponds to $(0,0,0)$, and the maximum values N,M,P for h,k,l have been respectively limited to $5,3,4$. Also, the h,l planes are drawn with different colors. We will return shortly on the meaning of the spacing shown.



Although the drawing represents a lattice whose nodes are identified by indices h,k,l of integer value, in fact (as already described) the indices correspond to the frequencies $u = \frac{h}{\Delta_x}, v = \frac{k}{\Delta_y}$ and $w = \frac{k}{\Delta_z}$. To illustrate the following concept, we refer to the particular (or normalized) case of a *unitary period* in the three directions, that is $\Delta_x = \Delta_y = \Delta_z = 1$, so that (11) turns into

$$f(\boldsymbol{r}) = \sum_h \sum_k \sum_l F(h,k,l)\, e^{j2\pi(hx+ky+lz)} = \sum_{h,k,l} F(\boldsymbol{h})\, e^{j2\pi \boldsymbol{h}\cdot\boldsymbol{r}} \tag{13}$$

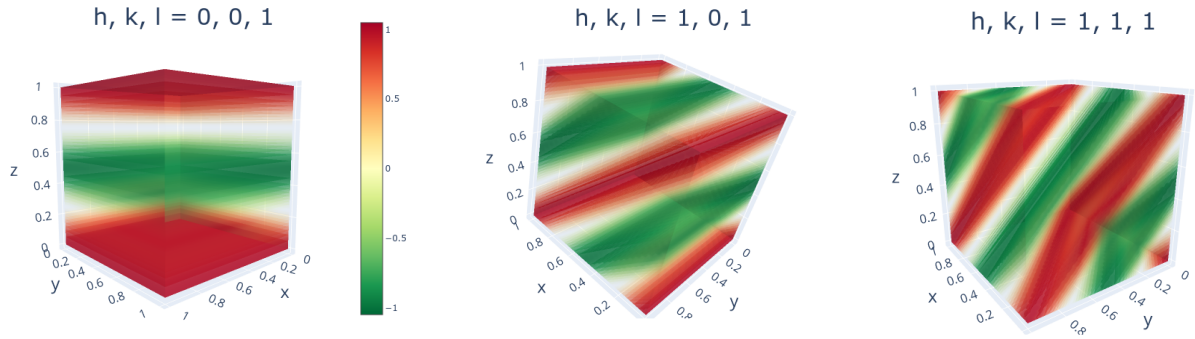
having defined $\boldsymbol{h} = (h,k,l)^T$ as *a vector* with integer elements.

3D basis functions Like it has been done for 2D case, we want to give more meaning to the term

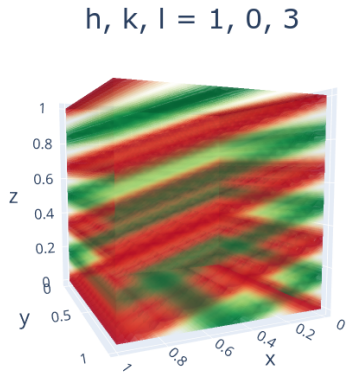
$$B_{h,k,l}(x,y,z) = e^{j2\pi \boldsymbol{h}\cdot\boldsymbol{r}} = e^{j2\pi(hx+ky+lz)} = \cos(2\pi(hx+ky+lz)) + j\sin(2\pi(hx+ky+lz)) \tag{14}$$

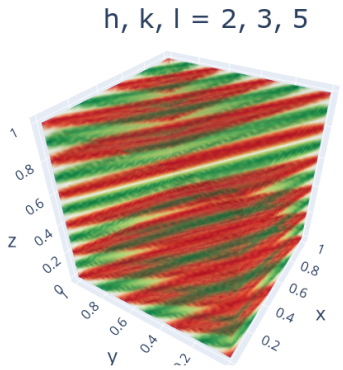
appearing in (13), which expresses the 3D *basis function* related to the reciprocal lattice point h,k,l, and which take values for any point x,y,z in *real space*. Also $B_{h,k,l}(x,y,z)$ is *a density*, and as such, it is not so easy to visualize it. Nevertheless, here below some *static* plots are shown, but for a deeper understanding about their shapes from different points of view, you can run the Python code given in the appendix.

The first image below on the left shows the *real part* of the first vertical harmonic, for which (14) becomes $\cos(2\pi z)$. Its values in between -1 and $+1$ are mapped to colors as given by the scale; we can observe a single period along z, and no variations along the x and y axes. On the contrary, the basis function $B_{1,0,1}$ given in the centre has a first harmonic along both the x and z axes, resulting in a *static density wave* with an *oblique* orientation. The basis function $B_{1,1,1}$ on the right has a first order component also along y, so that for a proper visualization of the density wave the image needs to be rotated.



Two other images are given further at the next page. The first one shows $B_{1,0,3}$ with a 3^{rd} harmonic along z, while the last image shows how the h,k,l triples more distant from the reciprocal lattice origin corresponds to basis functions with an higher volumetric frequency.





As a final remark, we can say that the 3D basis functions are quite similar to those already seen in 2D: they just now *live* in a 3D space!

A box of stars Consider now the 3D extension of the concept introduced as a *bed-of-nails*. A three-dimensional Dirac pulse is written as $\delta(x, y, z)$ or $\delta(\boldsymbol{r})$, and its value is zero everywhere except in $\boldsymbol{r} = 0$, where its value is ∞, but its three dimensional integral $\int \delta(\boldsymbol{r})\, d\boldsymbol{r}$ is equal to one. Think at it as *a star* in the space, if this can help you. Then, consider a 3D lattice in *direct* (or real) space, with 3D Dirac pulses $\delta(\boldsymbol{r} - \boldsymbol{r}_i)$ displaced at every node i of the lattice at position $\boldsymbol{r}_i$. How can you call it, if not *a box of stars* (or BOS)?? In formulas, we can write

$$\mathrm{BOS}(\boldsymbol{r}) = \sum_{i \in lattice} \delta(\boldsymbol{r} - \boldsymbol{r}_i)$$

5 Crystals and their representations

The science of crystallography began its developments starting from 1611, and the study method based on the diffraction of X-rays was born only in 1912. On the other hand, the Dirac delta function was defined only around 1930. It is therefore quite natural that the terminologies used in crystallography and in signal processing diverge significantly, even if the former can be explained in a fairly direct way using the concepts of the latter. Before succeeding, however, it will also be necessary to deal with other aspects, partly of a geometric nature, and partly linked to the propagation of waves. But for the moment, we continue from where we had arrived ...

5.1 Periodical 3D electrons density

First of all, let us define a $\mathrm{BOS}_{crystal}$ in the *real space*, with 3D Dirac pulses located at the lattice nodes $\boldsymbol{r}_i$ where each single molecule belonging to the crystal is located, and name the distance (in the three directions) in between nodes as Δ_x, Δ_y and Δ_z.

Since the convolution of a density $f(\boldsymbol{r})$ with a displaced $\delta(\boldsymbol{r} - \boldsymbol{r}_i)$ moves $f(\boldsymbol{r})$ to the position $\boldsymbol{r}_i$ where the pulse is located, it is possible to describe the overall, periodic electrons charge density $\rho_{crystal}(\boldsymbol{r})$ inside of a crystal, as *the convolution* in between the charge density $\rho_{cell}(\boldsymbol{r})$ of an elementary molecule, and the pulses which make $\mathrm{BOS}_{crystal}$. In other words, $\rho_{cell}(\boldsymbol{r})$ is *the motif* which *decorates* the direct lattice on which $\mathrm{BOS}_{crystal}$ is built. In formulas, it is written as

$$\rho_{cryst}(\boldsymbol{r}) = \rho_{cell}(\boldsymbol{r}) * \mathrm{BOS}_{\mathrm{crystal}}(\boldsymbol{r}) \tag{15}$$

Performing a 3D Fourier transform (10) of both sides of (15) we get

$$F_{cryst}(\boldsymbol{u}) = \int_{\boldsymbol{r} \in crystal} \rho_{cryst}(\boldsymbol{r})\, e^{-j2\pi \boldsymbol{u}\cdot\boldsymbol{r}} d\boldsymbol{r} = \int_{\boldsymbol{r} \in cell} \rho_{cell}(\boldsymbol{r})\, e^{-j2\pi \boldsymbol{u}\cdot\boldsymbol{r}} d\boldsymbol{r} \cdot \frac{1}{V} \cdot \mathrm{BOS}_{\text{reciprocal}}(\boldsymbol{u}) \quad (16)$$

as a consequence of the convolution theorem. Here $\boldsymbol{u} = (u, v, w)^T$ is a vector with real components representing volumetric frequencies, *but:*

- $\rho_{cryst}(\boldsymbol{r})$ is 3D periodic, so its transform $F_{cryst}(\boldsymbol{u})$ is better described by a 3D Fourier series, and
- the 3D Dirac pulses of which the $\mathrm{BOS}_{\text{reciprocal}}(\boldsymbol{u}) = V \cdot \mathcal{F}\{\mathrm{BOS}_{\text{crystal}}(\boldsymbol{r})\}$ in (16) is composed are centered on the nodes of the reciprocal lattice, which have coordinates $\boldsymbol{u} = \left(\frac{h}{\Delta_x}, \frac{k}{\Delta_y}, \frac{l}{\Delta_z}\right)$ with h, k, l integers; in this sense, the reciprocal lattice nodes are spaced (in frequency) by the inverse of the spatial periodicity of $\rho_{cryst}(\boldsymbol{r})$ in the x, y and z directions.

For these reasons $F_{cryst}(\boldsymbol{u}) = \mathcal{F}\{\rho_{cryst}(\boldsymbol{r})\}$ is a collection of 3D Dirac pulses, placed at the nodes of $\mathrm{BOS}_{\text{reciprocal}}(\boldsymbol{u})$, with *masses* given by the *volumetric sampling* (times $\frac{1}{V}$) at frequencies $\boldsymbol{u} = (u, v, w) = \left(\frac{h}{\Delta_x}, \frac{k}{\Delta_y}, \frac{l}{\Delta_z}\right)$ of

$$F_{cell}(\boldsymbol{u}) = \int_{\boldsymbol{r} \in cell} \rho_{cell}(\boldsymbol{r})\, e^{-j2\pi \boldsymbol{u}\cdot\boldsymbol{r}} d\boldsymbol{r} \quad (17)$$

Structure factors The quantity $F_{cell}(\boldsymbol{u}) = \mathcal{F}\{\rho_{cell}(\boldsymbol{r})\}$ given by (17) in crystallography is called the *structure factor*, because it gives *the shape* (or *the mass*) of the pulses located at the $\mathrm{BOS}_{\text{reciprocal}}(\boldsymbol{u})$ nodes. Since the values of $F_{cell}(\boldsymbol{u})$ are *sampled* by $\mathrm{BOS}_{\text{reciprocal}}(\boldsymbol{u})$ only at the discrete frequency values $(u, v, w) = \left(\frac{h}{\Delta_x}, \frac{k}{\Delta_y}, \frac{l}{\Delta_z}\right)$, often happens that *the samples* of $F_{cell}(\boldsymbol{u})$ are referred to as *structure factors* (plural), and given in the form

$$F(h, k, l) = F_{cell}\left(\frac{h}{\Delta_x}, \frac{k}{\Delta_y}, \frac{l}{\Delta_z}\right) \quad (18)$$

or $F_{h,k,l}$. But this notation also applies to *the coefficients* of the 3D Fourier series that could be obtained by applying (12) to $\rho_{cryst}(\boldsymbol{r})$. This clash of terminology sometimes creates confusion about of what we are speaking about, but as we will discuss soon, crystallography is an instrumental technique based on observations, to which the theory has to be wrapped around.

Contribution of individual atoms The density of charge $\rho_{cell}(\boldsymbol{r})$ of electrons inside a crystal cell is that resulting from the sum of the densities $\rho_j(\boldsymbol{r})$ relevant to each of the $j-th$ of the N_a atoms that make up the molecule inside the cell, each one located in a certain position $\boldsymbol{r_j}$, so that it can be written

$$\rho_{cell}(\boldsymbol{r}) = \sum_{j=1}^{N_a} \rho_j(\boldsymbol{r} - \boldsymbol{r}_j)$$

By remembering now that a translation in *direct space* is reflected into an *linear phase* term for the correspondent Fourier transform, the structure factor $F_{cell}(\boldsymbol{u})$ can also be expressed as

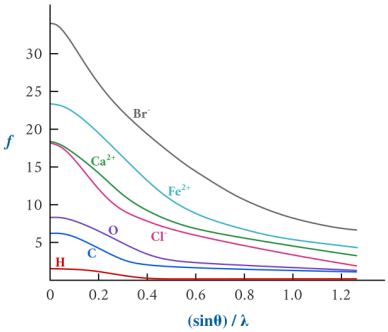
$$F_{cell}(\boldsymbol{u}) = \sum_{j=1}^{N_a} f_j(\boldsymbol{u})\, e^{-j2\pi \boldsymbol{u}\cdot\boldsymbol{r}_j} \quad (19)$$

where

$$f_j(\boldsymbol{u}) = \int_{\boldsymbol{r} \in cell} \rho_j(\boldsymbol{r})\, e^{-j2\pi \boldsymbol{u}\cdot\boldsymbol{r}} d\boldsymbol{r} \tag{20}$$

is the Fourier transform of the electron density $\rho_j(\boldsymbol{r})$ for atom j in the molecule, and it is called the *atomic scattering factor*, for a reason that will be given soon.

Atoms with a bigger nuclei, and a larger number of electrons, occupy a larger portion of the space. Then, according to the reciprocal domain measure which arises after a Fourier transform, bigger atoms possesses an *atomic scattering factor* $f_j(\boldsymbol{u})$ which is *more compact* in the low-frequency region with respect to smaller atoms.

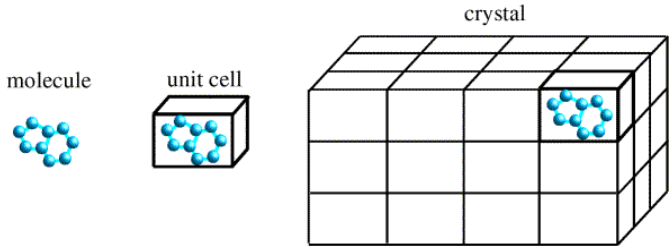


This fact is confirmed by the curves shown aside (source), where $f_j(\boldsymbol{u})$ is compared for different kind of atoms and ions: higher values of $f_j(\boldsymbol{u})$ at $\boldsymbol{u} = 0$ for bigger atoms is justified by the fact that with $\boldsymbol{u} = 0$ (20) is equal to the total number of electrons of the atom. Also, neglecting the fact that the horizontal axis in the figure is labeled in a (by the moment) obscure way, note that due to the spherical symmetry of $\rho_j(\boldsymbol{r})$, also $f_j(\boldsymbol{u})$ has the same symmetry, so that showing only *a section* of it is enough.

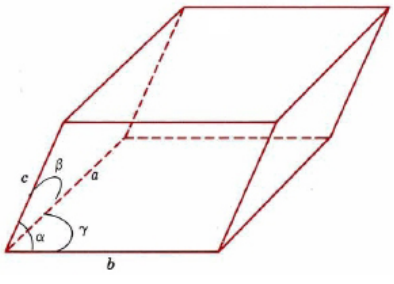
5.2 Geometric aspects

We now move toward a more specific crystallographic point of view. Since the writer was completely unaware about this science before writing these notes, please apologize if you find some inaccuracies.

The picture summarize how a molecule is fit within a *unit cell*, the *smallest* possible volume enclosing it *and filling* the space when repeated. The unit cell is replicated in the order of ten of thousands of times *per dimension* (for macromolecules as proteins, which are around 100 Å big or more) to make a crystal with dimensions of 0.1-0.3 mm. In three dimensions, you will have around 10^{12} molecules, or a thousand of billions. In such a mess, avoiding deviations from a rigorously regular crystal is almost impossible; however we will not worry about that, proceeding as if the crystals were *perfect*.



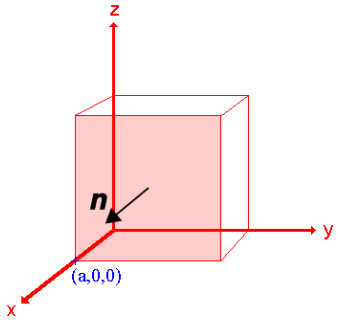
Coordinate system and Bravais lattices Another simplification is about the shape of the unit cell, which can deviate from a parallelepiped: the picture shows the more general form for it, with sides of length a, b, c and angles in between them equal to α, β, γ. By replacing the sides with three linear independent vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ with the same size and orientation, the points *inside* the cell can be identified by three numbers x, y, z, every one comprised in between $[0, 1]$, so that the vector $\boldsymbol{r}_{cell} = x\boldsymbol{a} + y\boldsymbol{b} + z\boldsymbol{c}$ points toward an inner point of the cell. At the same time, the *whole crystal* can be described *as a lattice* by a triple of *integer* numbers m, n, o, so that the vector $\boldsymbol{r}_{cryst} = m\boldsymbol{a} + n\boldsymbol{b} + o\boldsymbol{c}$ points to the lower-left corner of the cell identified by the triple m, n, o. Obviously enough, any point within the crystal has the position which is pointed to by the vector $\boldsymbol{r} = \boldsymbol{r}_{cryst} + \boldsymbol{r}_{cell}$.



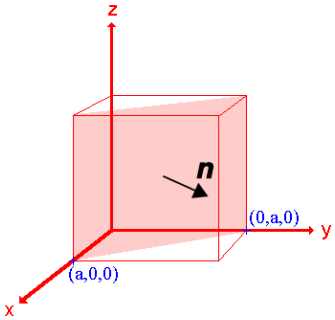
Crystallographic restrictions about the required symmetry[14] limit the shapes (i.e., equality or not in between elements of the triple a, b, c and the values of α, β, γ) that the unit cell can assume to 14 different choices, named Bravais lattices after its definition in 1848 by the homonym scientist. In the following we do not break our heads on these aspects, and we assume that the unitary cell has angles of 90 degrees.

Miller indexing of crystal planes The periodicity in a crystal is a consequence of the fact that each individual atom of the molecule (with its atomic scattering factor) always occupies the same position in each cell. For large molecules, the same atom (e.g. carbonium) appears several times in the unit cell, and this can produce a repetition period which is shorter than the cellular dimension. Also, repetitions have to be considered not only along all the three directions, but also along the diagonals which crosses the crystal. Since any specific direction can be considered as *the normal* to an orthogonal plane, the allowed periodicity inside of a crystal *have been coded* in terms of planes, following the system introduced by W. H. Miller in 1839, and based on the use of three small integers, the *Miller indices*, for identifying crystal faces. To simplify things, let us consider only a cubic crystal system, with all the sides of equal length a; for other situations, please refer elsewhere.

Let us start with the top-left picture (source) in which the surface parallel to the y, z plane is indexed by the 100 triple, and represents a recurrence along the x axis, with period equal to the cell width a, reflecting the recurrence of atoms of the molecule lying on this plane. Consider now three unit vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ oriented as the axes, and indicate the Miller indexes as h, k, l. Then $\boldsymbol{n} = h\boldsymbol{x} + k\boldsymbol{y} + l\boldsymbol{z}$ is a vector *orthogonal* to the plane it represents. In fact, the *oblique* plane parallel to the z axis given on the right, describing a periodicity of $a/\sqrt{2}$[15], is indexed as 110.



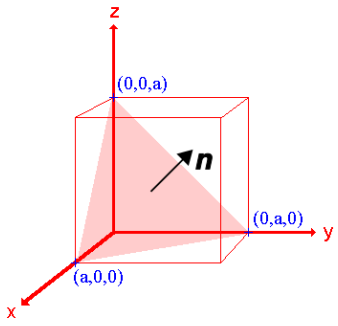
100



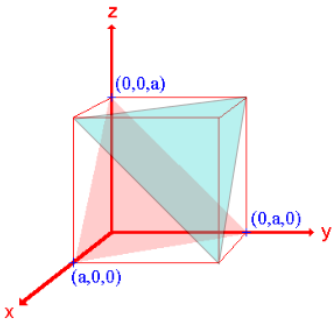
110

Planes 100 and 110 are *unique* inside of an unit cell, and also represent other *equivalent* planes that can be obtained by a rotation[16].

A plane touching three not-contigous vertices of the cube gives an inter-plane distance equal to $a/\sqrt{3}$ and takes Miller indexes 111, so that $\boldsymbol{n} = 1\boldsymbol{x} + 1\boldsymbol{y} + 1\boldsymbol{z}$ has an outgoing direction orthogonal to the plane. In this case, two of such planes enters into the cell, as the other (equivalent) is *upside down*, touching the other three vertices, as represented on the rigth.



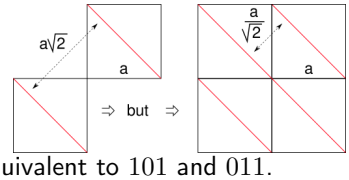
111



111 - equivalent

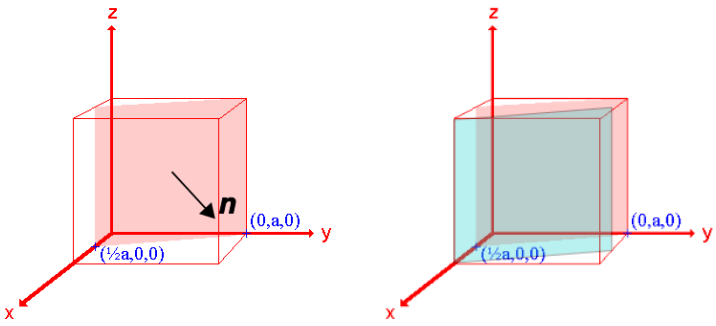
[14] A sequence of rotations by the same angle around one of the crystal axes must put the lattice in the original arrangement, and the number of rotations indicates the symmetry order. For example, a cubic cell has a symmetry order of four, given by rotations of 90 degrees.

[15] It may seem that the distance with respect to a similar plane in the following cell should be $a\sqrt{2}$, but this halves because of the other neighboring cells.



[16] For example, the 100 plane is equivalent to the 010 and the 001 ones, and 110 is equivalent to 101 and 011.

An infinite number of other planes may be defined. The last exemplified one intersects the x axis in *the middle* of the cell, and is identified by the indexes 210, so that the vector $\boldsymbol{n} = 2\boldsymbol{x} + 1\boldsymbol{y}$ is orthogonal to it. It also has an equivalent plane inside the cell, obtained by a 180° rotation, as shown by the last picture, giving an even shorter interplane distance[17]. Finally, consider a plane 300 (not shown): it will be parallel to 100, but repeats *three times* inside the cell.

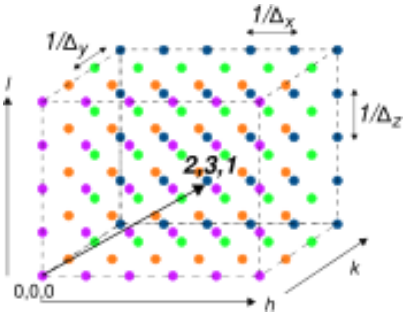


210 210 - equivalent

Not all plans corresponding to any possible triple of $h.k.l$ indices are present in a crystal, but only those for which there is some atom of the molecule, repeating according to the frequency and orientation of the plane. The type of planes that are really present determines many physical properties of the crystal, one of which is the *diffraction*, on which X -ray crystallography is based.

Miller indexes refers to coordinates in the reciprocal lattice It is no coincidence that the letters h, k, l chosen to indicate Miller's indices are the same as those used for the indexes of the volumetric harmonics of a 3D Fourier series, and appearing in (11) - (14). In fact, the value of each Miller index of a triple says how many times the corresponding equivalent planes intercept the axis to which the index refers: in this sense, to say e.g. that $k = 3$ means that a periodicity 3 occurs within the unit cell along the y axis.

If you compare the appearance of the 3D basis function $B_{h,k,l}(x, y, z)$ (14) given at fig. 4 with that of the crystal planes seen before, they look very very similar. So the existence, let's say, of plane 231 means that the 3D Fourier transform $F_{cryst}(\boldsymbol{u})$ of the electron density $\rho_{cryst}(\boldsymbol{r})$ will show the presence of the volumetric harmonic 231, with a 3D frequency $\boldsymbol{u}$ whose components (for a cubic cell) are $(u, v, w) = \left(\frac{2}{a}, \frac{3}{a}, \frac{1}{a}\right)$.



6 The inverse and indirect problem of crystallography

Until now we have supposed to know the electron density $\rho_{cell}(\boldsymbol{r})$ and the crystal lattice, so that the structure factor $F_{cell}(\boldsymbol{u})$ could be evaluated by applying (17), or to know the positions $\boldsymbol{r}_j$ of the atoms inside of the unit cell together with their atomic scattering factor $f_j(\boldsymbol{u})$, to be able to apply (19). *None of these hypotheses is true!* On the contrary, the electron density $\rho_{cell}(\boldsymbol{r})$ must be obtained by means of an *inverse* Fourier transform of *an estimate* of the structure factors $F_{h,k,l}$ defined by (18), so that $\rho_{cell}(\boldsymbol{r})$ can be evaluated as a 3D Fourier series[18]

$$\rho_{cell}(\boldsymbol{r}) = \sum_h \sum_k \sum_l F_{h,k,l} e^{j2\pi(hx+ky+lz)} \tag{21}$$

in which the values of $F_{h,k,l}$ are *indirectly* measured by the technique of X-ray crystallography, starting from a *diffraction pattern*. But, to further complicate things, the diffraction pattern only provides

[17] The general formula for the distance in between planes h, k, l of a cubic cell is $d = \frac{a}{\sqrt{h^2+k^2+l^2}}$

[18] or more rigorously, the exponent in (21) should be substituted by $\frac{hx}{a} + \frac{ky}{b} + \frac{lz}{c}$ in which a, b, c are the lengths of the sides of the unit cell, like the ones depicted at page 16.

information about *the module* of $F_{h,k,l}$, so that in order to apply (21) we must *guess* the information about *the phases* of $F_{h,k,l}$, that is to find a way to solve what is generally indicated as *the phase problem*. At this point *an estimate* of $\rho_{cell}(\boldsymbol{r})$ is obtained, which is checked to see if it adapts well to the other information available, such as the chemical composition of the molecule that fills the unit cell.

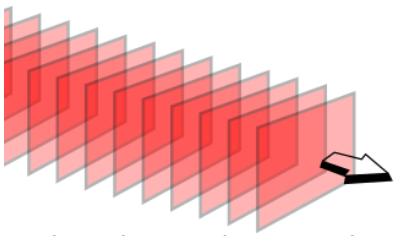
The aspect of how to provide the correct phase will not be treated, considering it beyond the goal that we had given ourselves. But before understanding the wonder of how a 3D signal sequence $F_{h,k,l}$ defined on the reciprocal lattice can be recovered from a 2D image (the diffraction pattern), we have to deepen our knowledge on how the scattering of X-rays works, and how the diffraction pattern is related to the distances between the crystal planes.

X-rays plane waves Objects as small as protein molecules, with dimensions going from 10 to 100 nm or more, cannot be observed by means of optical microscopes, whose wavelength λ is in the order of 500 nm. The λ of X-rays is around $0,1$ nm (or 1 Å), but it is not possible to make lenses capable of allowing them to focus, so that their contribution to imaging is based on the way they interact with the solid matter.

X-rays are a form of electromagnetic wave, propagating in space at the speed of light $c = 3 \cdot 10^8$ m/sec as a traveling plane wave, that is, a sinusoidal wave for which the amplitude A of the electrical field depends on time t and space $\boldsymbol{r}$ as

$$A\cos\left(2\pi\frac{1}{\lambda}(\boldsymbol{r}\cdot\boldsymbol{n} - ct)\right) \tag{22}$$

where $\frac{1}{\lambda}$ is it *spatial* frequency, c is the velocity of propagation, $\boldsymbol{n}$ is a unitary vector that aims towards its direction of propagation, and $\boldsymbol{r}\cdot\boldsymbol{n}$ is the scalar product in between $\boldsymbol{n}$ and a point $\boldsymbol{r}$ in space, so that for points $\boldsymbol{r}_\perp$ which are orthogonal to $\boldsymbol{n}$ there is no variation of the electrical field (at the same instant, or with t fixed), as shown in the Wikipedia picture.



Suppose now that the wave is propagating along the positive x axis, and to observe the wave along it. Equation (22) then becomes $A\cos\left(2\pi\left(\frac{1}{\lambda}x - ft\right)\right)$ as now $\boldsymbol{r}\cdot\boldsymbol{n} = x$ and $\frac{c}{\lambda} = f$, the frequency of oscillation (in time). By substituting the values for c and λ, we get a frequency of 3 *millions of terahertz*, making the measure of its phase absolutely impossible. For this reasons, X-rays detectors are able to measure only *the intensity* of the wave, i.e. its power or $A^2/2$, measuring his presence and strength, or his absence. Therefore, we are not interested in the time dependency, and for our goals (22) is further simplified as

$$A\cos\left(\frac{2\pi}{\lambda}x\right)$$

The physicists give $\boldsymbol{k} = \frac{2\pi}{\lambda}\boldsymbol{n}$ the name of angular wavevector, in which unitary vector $\boldsymbol{n}$ provides orientation towards the direction of propagation. The amplitude $|\boldsymbol{k}| = \frac{2\pi}{\lambda}$ is measured in radians/meter and is the spatial analogue of the angular velocity $\omega = \frac{2\pi}{T}$ for time-dependent trigonometric functions, just like the amplitude of the *wavevector* (or *wave number*) $\boldsymbol{\nu} = \frac{1}{\lambda}\boldsymbol{n}$ is the equivalent of frequency $f = \frac{1}{T}$.

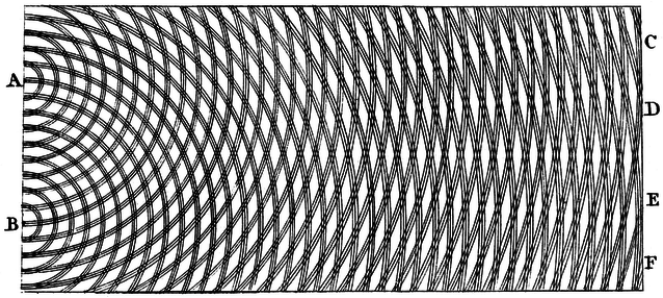
Scattering and diffraction of X-rays When an X-ray hits a particle, its energy is re-emitted *with the same wavelength* in (almost) all directions according to the Thomson (or elastic) *scattering law*[19],

[19] More precisely (or not), the electron oscillates in the direction of the polarization of the incoming wave, but with a phase delay of $\pi/2$, and its acceleration causes the re-emission of electromagnetic radiation with a further phase shift of

which also provides that, with a $\lambda \simeq 1\,\mathring{A}$, the phenomenon is attributable to valence electrons, whose density has a comparable dimensional scale.

Waves which are scattered by different particles *add up* in the surrounding space, and since for any *direction* they can combine in phase or not, the resulting *interference* can be *constructive* or *destructive*, thus originating the phenomenon known as *diffraction*.

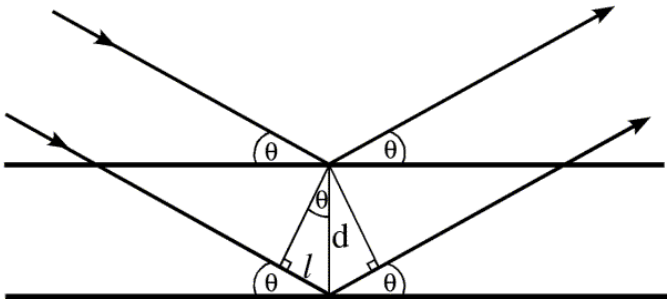
Diffraction occurs in many contexts of real-life, perhaps the most evident of which it is when two stones are launched in calm water, and the consequent circular waves mix by creating nice geometrical patterns. On the side, a historical image (1803) from Wikipedia sketches the diffraction for water waves coming from the left and scattering through the two slits in A and B. The points C, D, E and F then show the first directions along which a constructive interference occurs.



Here comes the *beautiful thing* about diffraction: it *converts the distances into angles!* To be more precise, the points C, D, E and F would become *more distant*, if the slits A and B were *brought closer*[20]. This behavior can be experienced using the interactive tool present at the *Gratz University*, where you can change the position of the scatterers and check the resulting diffraction pattern, as well as you can change the wavelength λ, and verify that it should be smaller than the distance to be measured, so that diffraction can occur.

Diffraction by Bragg planes In X-ray crystallography, the phenomenon of scattering is used to measure the distance d between parallel crystal planes, based on the knowledge of the relationship between this distance and the angle θ for which a constructive interference occurs. The angle θ is in fact measured thanks to the observation, in correspondence of the same, of a significant increase in intensity of the radiation that reaches a screen located at some distance from the irradiated object.

To find the relationship between d and θ, we retrace the reasoning historically due to William Bragg and his son Lawrence (1913). The image[21] shows an incident plane wave of wavelength λ coming from left with an angle θ, that hits two electrons lying on two planes of the crystal at distance d. Of all the rays scattered in any directions, we only focus on the one *reflected* by the same angle θ, wanting to discover what the value of θ should be, in order the outgoing waves (on the right) to have *the same phase*, so that they add up *coherently*, and the intensity of the sum can be detected on the screen. The simple answer is that the difference in path length, equal to $2l$, must be an integer multiple n of the λ (you can experiment with a simulator). From trigonometry it results that $l = d\sin\theta$, which gives the *Bragg condition*



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

$\pi/2$, so that the new wave will be in phase opposition compared to the incoming one.

[20] Incidentally, this inverse relationship is the historical reason for which the name of *reciprocal lattice* was given to the space that is actually a 3D frequency domain

[21] Here and in the following we make a wide use of material from the Protein Crystallography Course of Randy J Read, which did a very good job.

so that, limiting ourselves to only the first order *reflection* for $n = 1$, we have

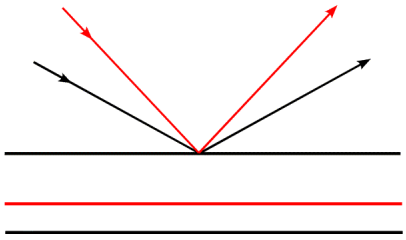
$$d = \frac{\lambda}{2 \sin \theta} \tag{23}$$

or, put in the other way around, the angle θ is related to d as

$$\theta = \arcsin \frac{\lambda}{2d} \tag{24}$$

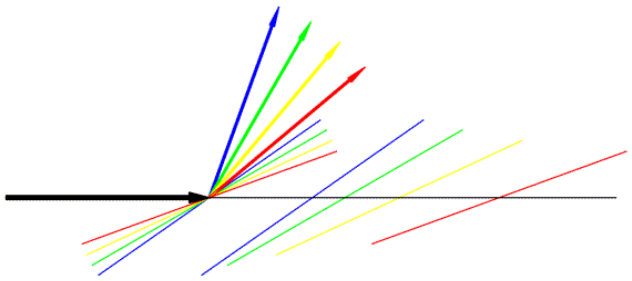
The rays that are scattered with a different angle adds up *incoherently*, and the resulting intensity on the screen is *negligible* for those angles, although... the coherent sum produced only by two planes *also has* a negligible intensity. *But here comes the crystal!* It acts like *an amplifier*: in a crystal there are thousands of planes with the same orientation, and for each of them the Bragg condition applies, so the coherent sum is actually made up of *thousands of waves!*

Minimum distance and maximum frequency We now have the concept, exemplified in the image: the greater the distance d between the planes is, the smaller the angle θ for which the first peak of intensity will be detected on the screen. This puts a practical limit to the minimum intra-plane distance d which can be measured: as $\theta \to \pi/2$, $\sin \theta \to 1$, and $d_{min} = \frac{\lambda}{2}$.



This means that in order to get an higher resolution, you need to use a shorter λ. Or, for a given λ, you cannot observe distances between planes that are shorter than d_{min}, and remembering that the families of closer planes are those associated with higher Miller indexing and consequently to higher volumetric frequencies, the result of having too large a λ is to suffer a bandwidth limitation in $\rho_{cell}(\boldsymbol{r})$ reconstruction by (21).

Varying the angle θ As the incoming X-ray direction is varied[22], one can take note of the angles θ_i for which an intensity peak is detected, and obtain the distances d_i of the different planes by (23). But this is not yet the complete story. If you take both the incoming ray and the crystal fixed, you will still observe several diffraction peaks. These originates from set of planes that are not parallel to each other, as shown in the figure in which the large black arrow is the incoming radiation, and the blue, green, yellow and red arrows represent coherently diffracted rays, generated by planes with the same color code, each arranged with a different orientation, and a different inter-plane distance.

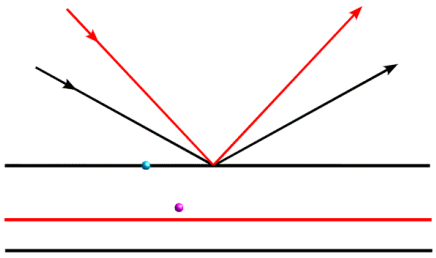


Diffraction from approximately positioned particles What happens if a particles is in a position that does not belong *exactly* to a plan that produces diffraction? After all, the *electrons move!*

Consider the figure at the following page: the black plane produces diffraction at the angle represented by the black arrow. As the purple particle is half-way in between the black planes, its scattered wave will have a $180°$ phase difference with respect to the waves scattered by the blue particle, so its contribution to the total diffracted wave will cancel out. However, when we consider the diffraction

[22] Actually, it is the crystal which is rotated, see the animation on Wikipedia.

event represented by the red arrows, the red particle will be quite near to the red plane, and the scattered waves will add up more nearly in phase.



In general, then, a single spot in the diffraction pattern tells us to what extent the objects are concentrated on the corresponding planes. It tells us about the average position of objects in the direction perpendicular to the planes, but nothing about their position in the directions parallel to the planes. For that, we need to collect diffraction data from otherwise oriented planes.

Intensity of a diffraction spot As anticipated, the observed diffraction spots are measured only by their intensity, but, what does it refer about? As explained above, observing a diffraction spot says nothing about the position (on the corresponding plane) of the particle that originates it: this also means that if many electrons lie on the same diffraction plane, all of them will participate to the constructive interference which generates the spot, whose amplitude (the square root of the intensity) is therefore directly proportional to the number of electrons lingering on the plane.

7 When Fourier met diffraction at the Ewald's sphere

Finally we are quite ready to understand the way by which the observed Bragg diffraction peaks, whose angles θ depend on the distance between crystal planes as given by (24), can be related to the 3D reciprocal lattice indexes h, k, l, which also represent the Miller indices of the plane that produces the peak. This will be done with the aid of a geometric construction known as the Ewald's sphere, from the name of the scientist who introduced it in 1921.

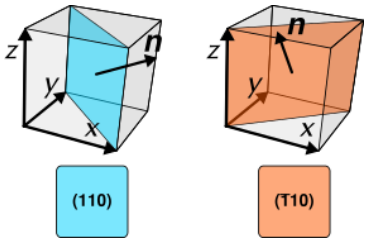
In this way, the amplitude of the structure factors $F_{h,k,l}$ can be estimated, and (after finding how to solve the *phase problem*) eq. (21) can be calculated, thus obtaining the desired crystal electron density $\rho_{cell}(\boldsymbol{r})$. But first, we still need to do some small clarifications regarding the things said so far.

Miller indices multiple personality To recapitulate things about Miller indices (pag. 17 and following), they are three integers h, k, l which

- identify a set of parallel planes inside the crystal, touching the x, y, z axes (within the unit cell) respectively h, k, l number of times
  - let define d_{hkl} as the distance in between these planes;
- express the coordinates of a vector $\boldsymbol{n} = (h, k, l)$ which is *orthogonal* to such planes, and
- express the coordinates of the vector $\boldsymbol{h} = (h, k, l)$ expressing the volumetric frequencies indexes, whose *head* touches the nodes of the *reciprocal lattice*
  - it can be shown[23] that $|\boldsymbol{h}| = \frac{1}{d_{hkl}}$, i.e. the length of $\boldsymbol{h}$ is the reciprocal of the distance in between the Miller planes h, k, l

[23] For example, for a cubic cell with side of length a, the distance between planes (applying the formula in note 17) is $d_{hkl} = \frac{a}{\sqrt{h^2+k^2+l^2}}$, while $|\boldsymbol{h}|$ is the measure of the diagonal of a paralleplped as shown for the reciprocal lattice of pag. 18, or $|\boldsymbol{h}| = \sqrt{\left(\frac{h}{\Delta_x}\right)^2 + \left(\frac{k}{\Delta_y}\right)^2 + \left(\frac{l}{\Delta_z}\right)^2} = \frac{1}{a}\sqrt{k^2+h^2+k^2}$.

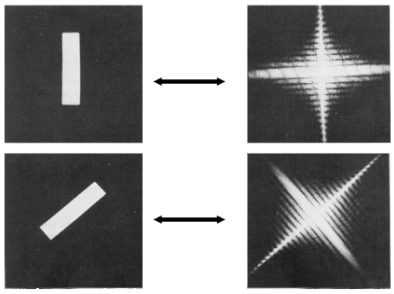
Negative Miller indices and reciprocal lattice extension In reality, each index can also take on a *negative* integer value, indicated by an *overline*, and which describes a different type of orientation. The image (adapted form Wikipedia) compares planes with indexes 110 and $\bar{1}10$, in which the vector $\boldsymbol{n}$ (orthogonal to the plane) in the second case has the first component which is oriented in the negative verse of the x axis.



The negative indices are reported as such in the reciprocal lattice, so in this example we have $\boldsymbol{h} = (-1, 1, 0)$. This means that the extension of the reciprocal lattice is not limited to positive indices as indicated at pag. 18, but it continues all around the origin $h, k, l = 0, 0, 0$.

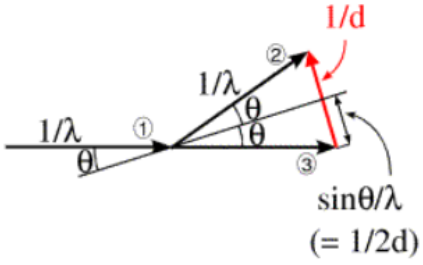
Friedel's law (conjugate symmetry) Since the structure factors F_{hkl} are the 3D Fourier transform of *a real* electrons density, the conjugate symmetry property still applies, so that $F_{h,k,l} = F^*_{-h,-k,-l}$, which in this context is called the Friedel's law, and which substantially means that $F_{h,k,l}$ has not to be collected for all the allowed indices.

Fourier transform of a rotated signal You can easily see this property at work in two dimensions: if you rotate an image by, for example, $90°$ in some direction, the correspondent Fourier transform will rotate by the same quantity. In fact, periodicities that were in the vertical direction, after the rotation will convert to the horizontal direction, and viceversa. Please check also the example image for a 45° rotation of a 2D DFT (only the module is shown).



This has direct consequences on the new orientation of the reciprocal lattice if the crystal is rotated, as in fact it happens in the X-ray crystallography, in order to measure the diffractions caused by further planes.

The Ewald's sphere First of all, consider the scheme represented in illustration, which is more adherent to the instrumental setup used in X-ray crystallography. The incoming ray ① come from the left, hits the crystal with a θ angle with respect of some set of planes, is scattered ② with the same angle, and continues ③ also straight towards the detector screen. The rays ①, ② and ③ can be usefully (for our purposes) be drawn as vectors of length $1/\lambda$ (after all, they all have the same λ)[24].



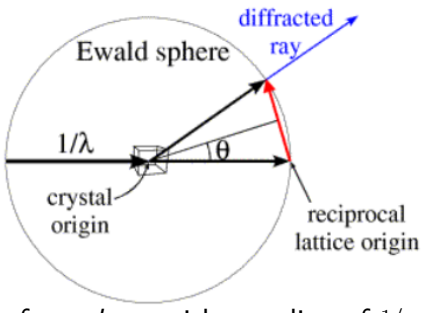
But now put your attention to the red arrow, equal to the difference in between vectors ② and ③, and connecting their heads:

- by construction, it is perpendicular to the depicted set of planes, so it is in the direction of the corresponding reciprocal lattice vector;
- by trigonometry, the red arrow has a length which is equal to $2\frac{1}{\lambda}\sin\theta$, and
- if the angle θ satisfies the Bragg condition (23) for the considered set of planes, then

[24]Here is a more cultured explication. The vectors used to represent the rays ①, ② and ③ are the wavevectors (pag 19) associated with them, whose equal length $\frac{1}{\lambda}$ reflects their equal *relativistic* energy given by $E = hf = \frac{hc}{\lambda}$, as a consequence of elastic scattering, see Wikipedia.

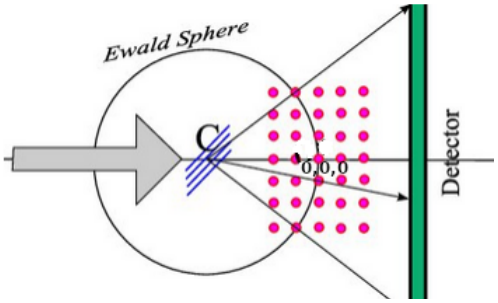
$$\frac{\sin\theta}{\lambda} = \frac{1}{2d}$$

so that when diffraction occurs, the red arrow has length $\frac{1}{d}$, which is exactly the length $\frac{1}{d_{hkl}}$ of the vector $\boldsymbol{u} = (u, v, w) = \left(\frac{h}{\Delta_x}, \frac{k}{\Delta_y}, \frac{l}{\Delta_z}\right)$ in the reciprocal space that reaches the lattice point indexed by the integer triple $\boldsymbol{h} = (h, k, l)$, associated to the set of planes identified by the same triple of Miller indices.



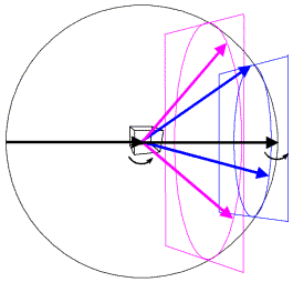
Such a construction could be made for any set of planes in the diffracting condition, and the corresponding reciprocal space vector would be seen to go from the position of the undeflected direct beam to the tip of the vector representing the diffracted ray. Since the latter has a length $1/\lambda$, the red arrow head will move along the *Ewald's circle* with radius $1/\lambda$. Moreover, since the angle 2θ can be anything from 0 to 180 degrees (so θ can be anything from 0 to 90 degrees), and given that the diffracted ray can go in any direction in three dimensions, the latter can have its tip anywhere at the surface of *a sphere* with a radius of $1/\lambda$, which is the **Ewald sphere**.

The meeting Now consider that, because the red arrow is a reciprocal space vector which *starts* at the lattice origin, its base defines *the origin* $(h, k, l) = (0, 0, 0)$ of the reciprocal space. So we are facing with an unexpected reality: in the Ewald sphere there coexists two spaces[25]: *the crystal* which is at the sphere center, diffracting the incoming X-ray, and *the reciprocal lattice* which is centered where the X-ray beam would exit the Ewald sphere, as shown in the figure (source).



The diffraction spots that are observed on the detector screen occur only for angles 2θ at which the reciprocal lattice nodes *intersects* the Ewald sphere. The actual (h, k, l) reciprocal lattice *indexes* of these intersections thus provide *the labeling* for the detected spot intensity, which can therefore be assigned to the relevant structure factor F_{hkl}.

When some of the reciprocal lattice nodes intersect the Ewald sphere, also the reciprocal lattice plane to which the node belongs do intersect the sphere. So, we will see *a circle* of diffraction spots, corresponding to all the reciprocal lattice nodes laying on that plane. This is shown schematically in the new figure, illustrating (with different colors) planes of nodes in the reciprocal lattice that intersect the Ewald sphere, and give a circle of spots surrounding the direct beam position.



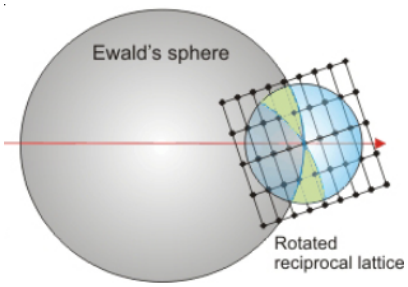
Simultaneous rotations Also if the detector screen could have infinite dimensions, a range of diffraction angles as large as $90°$ would cover only the half ($45°$) of incident angles[26]. Moreover, many of the reciprocal lattice nodes does not intersect the Ewald sphere, letting the amplitudes of the corresponding F_{hkl} structure factors completely unknown. So, in order to complete the data collection the crystal needs to be rotated. As explained one page before, the rotation of the crystal lattice (and of its Bragg

[25] If you feel uncomfortable about that, just realise that it is just a geometrical construction that makes the mathematics of diffraction easy to picture.

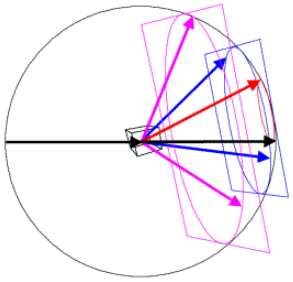
[26] Here please remember that $\theta_{screen} = 2\theta_{Bragg}$

planes) is reflected in the rotation of the reciprocal lattice (in which the Fourier transform lies) by the same amount.

The first picture (source) shows the reciprocal lattice rotation around its center, located at the exit of the incident X-ray beam. In the case of macromolecules such as proteins, a big unit cell may result in very *dense* reciprocal lattices, so that small rotation angles may suffice for the task of sampling a large number of structure factors amplitudes.



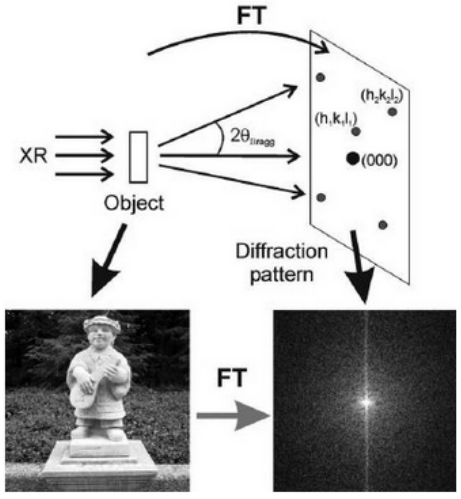
The second image duplicates the one given before, and shows that after (a little) crystal rotation, also the circles made by the intersection in between the reciprocal lattice planes and the Ewald sphere will move by the same extent, giving on the screen the appearance of *lunes* of spots[27]. Moreover, this little rotation may let the reciprocal lattice nodes which fall near the lattice origin to enter the sphere surface, thus producing observable diffraction spots on the screen which would be otherwise missing. This event is represented in figure by a new red vector, together with a small red circle. This offers the opportunity to measure the low (volumetric) frequencies of electron density, which provide its large scale spatial distribution.



Go interactive The concepts given so far deserve adequate visualization tools. You can start by jumping to this video of a tilted reciprocal lattice, rotating around its origin, and *shooting* diffracted rays all around when its nodes touch the Ewald sphere. Then you can play with The Interactive Ewald Sphere Tutorial by Kevin Cowtan, and enjoy tilting the crystal and see the diffraction pattern appearing while the reciprocal lattice dances with you. But if you still need more, maybe the EPFL website is the one good for you, where you can find a reciprocal lattice generator, a CrystalOgraph for drawing crystal structures, a calculator for the 2D Fourier transform of a density function ρ(x) drawn by hand, and more.

8 Sometimes size matters

Here we develop some considerations on the reason why the maximum diffraction angle recorded determines the obtainable spatial resolution, and why the value of the λ wavelength of the X-rays can produce some misunderstanding on the true geometry of the diffractometer. Finally, we will discuss the consequences of an imperfect geometry of the crystal in analysis.



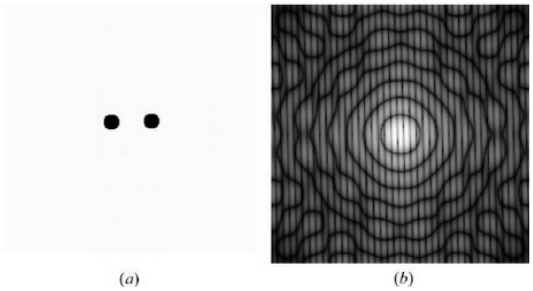
Analogy in between images ad crystals As it should be already evident, most of the concepts given at § 3 for images are directly applicable the the X-ray cristallography, in the light of the analogy described by the illustration. In fact, as we discussed, the *intensities* $|F_{hkl}|^2$ of the structure factors observed on the

[27] This can be explained by thinking that the diffraction spot do not have a *negligible volume*, as it will be clear after discussing the impact of the crystal shape, at page 27.

diffractometer screen are linked to the electronic density of a unitary cell of the crystal by means of a 3D Fourier transform. Therefore many situations that may occur for crystallography can be better understood by considering an image and its 2D Fourier transform.

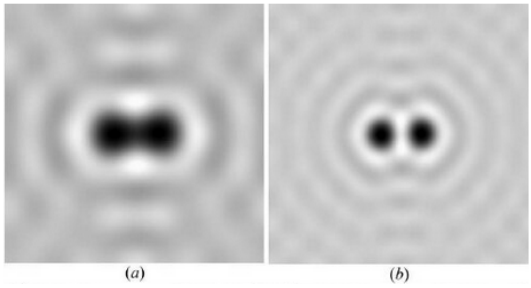
Limiting the diffraction angle range As discussed, higher θ angles correspond to planes that are separated by a shorter distance. So we can ask ourselves if, knowing the value of the shortest distance we want to measure, we can limit the amount of angles for which to collect the diffraction intensities. Or, said in other words, we wonder how many Fourier coefficients we need.

For this purpose we can consider the image (a) with size 256x256 pixels, in which the two circles (with diameter 15 pixels and separated by 40) may represent two atoms that we want to distinguish from each other. Since (a) can be considered as the convolution between a single circle and two Dirac pulses moved to the left and right with respect to the center of the image, the module of its 2D Fourier's transform (b) is the product of the transform of a circle (producing *oscillations* with radial symmetry), multiplied by the transform of the two Diracs, which produces the dense vertical lines.


(a) (b)

One can assume that the lowest frequency necessary to distinguish the two circles corresponds to a sine wave having a periodicity equal to the separation between the circles. Since the first Fourier index corresponds to a sine wave of one period on the image, i.e. 256 pixels, then a sine wave with a period of 40 pixels corresponds to the sixth or seventh Fourier coefficient ($256/40 = 6.4$).

In fact, since the image (a) on the right is reconstructed by an IDFT of the (b) above by using only the central area of the first five Fourier coefficients in radius, the two discs are not resolved. This is because limiting the number of coefficients acts as *a window* in the frequency domain, with a low-pass filter effect in the space domain. Using instead eight Fourier coefficient around the center as done for fig. (b) on the right, the two discs are clearly shown.


(a) (b)

Returning to crystals, the shortest interatomic distance that one can expect in a compound is, for example, a C-H bond, ($\simeq 1$ Å). By applying the Bragg law (23) one can write $\frac{\sin\theta}{\lambda} = \frac{1}{2d} = 0.5\,\text{Å}^{-1}$, and considering a Mo K$\alpha$ radiation of 0.771 Å, an angle $\theta_{Bragg} = \arcsin(0.5\,\lambda) \simeq 20.8°$ is derived, so that to distinguish H atoms from their neighbors we need to collect the $|F_{hkl}|^2$ values up to an $\theta_{screen} = 2\theta_{Bragg} = 42°$ angle.

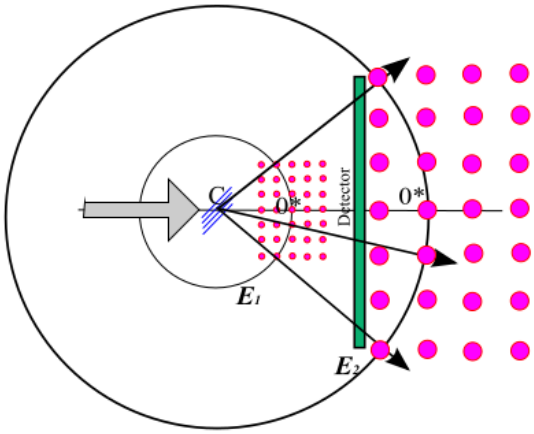
The giant sphere and the embedded screen Up to now we draw the Ewald sphere with its center on the crystal and the detector screen lying outside of the sphere, see pag. 24. But since the wavelength λ is of the order of the Angstrom, or 10^{-10} meters, and given that the Ewald sphere has a radius of $1/\lambda$, it seems that the screen should be placed more than *10 millions of kilometers far!* But here we forgot that $1/\lambda$ has to be measured in meters^{-1}, and the distance from the screen to the crystal usually is a few centimeters, so we are simply drawing on the same piece of paper and scale, objects with dimensions that refers to different units.

For what regards the reciprocal lattice dimensions, its nodes are spaced (in the three directions) by *the inverse* of length of the sides of the *unit cell*, so their inter-distance is expressed in the same unit of measure of the Ewald sphere. But unit cells are rather *larger* than λ, so the distance between the

nodes of the reciprocal lattice is much smaller than $1/\lambda$, and ultimately many planes of the reciprocal lattice can be *housed* within the space defined by the radius of the Ewald sphere.

As for the position of the diffraction spots on the screen, Bragg's law imposes $\sin\theta_{hkl} = \frac{\lambda}{2d_{hkl}}$, where the $\frac{\lambda}{d_{hkl}}$ ratio is adimensional as between homogeneous quantities. For a screen at distance d_s from the crystal, the angle $2\theta_{hkl}$ corresponds to a distance D of the hkl spot from the center of the screen equal to $D = d_s \tan 2\theta_{hkl}$. As if to say, while the angle θ_{hkl} do exists in real space, sphere and mutual lattice exist in Fourier's space. At this point we can think to *re-scale* the representation of both the sphere and the lattice, so that the screen can be drawn *out of them* as it is usually done, noting also that the re-scaling operation does not change the scattering angles.

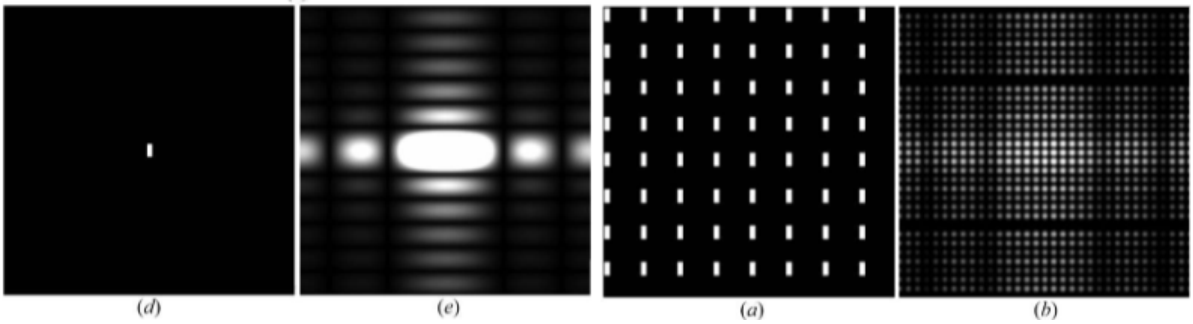
Since we are now changing the scale, nothing prohibits us from modifying it so that the screen *enters* the Ewald sphere, as shown in the picture (source), where sphere E2 has been drawn at a scale three times larger than sphere E1, and the reciprocal lattice is expanded accordingly, so that the diffraction directions are preserved. Also the distance between the crystal and the screen remains the same, so that now the screen is depicted *inside* of the Ewald's sphere. Hopefully, the concepts of the Ewald sphere and the reciprocal lattice could now be a little demystified.



Effect of the crystal shape We observed at page 15 that the structure factors F_{hkl} observed on the screen as diffraction spots are the samples at the volumetric frequencies $\boldsymbol{u} = (u, v, w) = \left(\frac{h}{\Delta_x}, \frac{k}{\Delta_y}, \frac{l}{\Delta_z}\right)$ of the 3D function

$$F_{cryst}(\boldsymbol{u}) = F_{cell}(\boldsymbol{u}) \cdot \frac{1}{V} \mathrm{BOS}_{\mathrm{reciprocal}}(\boldsymbol{u})$$

in which $\mathrm{BOS}_{\mathrm{reciprocal}}(\boldsymbol{u})$ is a *Box-of-Stars*, see also the figures (and the discussion) at page 11 that are proposed again below.



But this is true only for *perfectly periodic* crystals, whose dimension is assumed to be *infinite!*

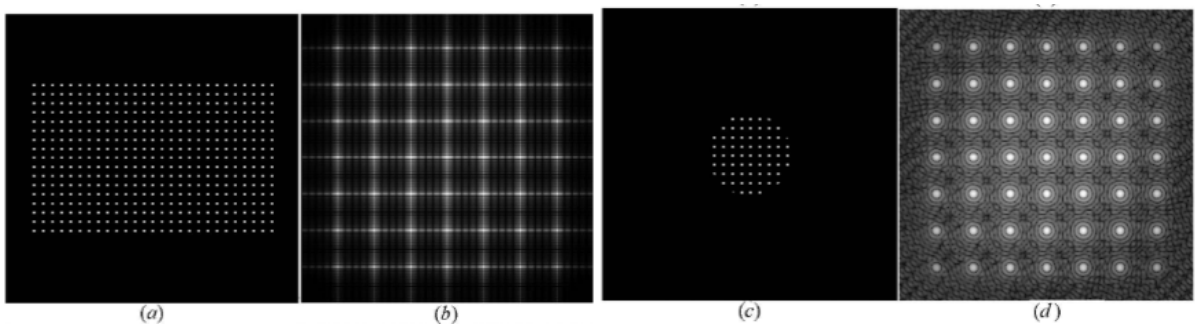
On the contrary, if the crystal is limited in space (with typical dimensions in the order of 0.1 mm), then the same takes place for the *box* which encloses *the stars* in the $\mathrm{BOS}_{\mathrm{crystal}}(\boldsymbol{r})$, or in Signal Processing parlance, the direct lattice has been *windowed*, that is, it has been multiplied for a limited-extension $\mathrm{W}(\boldsymbol{r})$. The effect of limited size for the crystal is that the stars (3D Dirac pulses) in

$$\mathrm{BOS}_{\mathrm{reciprocal}}(\boldsymbol{u}) = V \cdot \mathcal{F}\{\mathrm{BOS}_{\mathrm{crystal}}(\boldsymbol{r}) \cdot \mathrm{W}(\boldsymbol{r})\}$$

begin to *widen,* since every 3D Dirac pulse in the (unlimited crystal) $\mathrm{BOS}_{\mathrm{reciprocal}}(\boldsymbol{u})$ has now been convoluted with the transform of the 3D window $\mathrm{W}(\boldsymbol{r})$ that delimits the crystal. For this reason

$F_{cryst}(\boldsymbol{u})$ does not more exists only *at* the lattice nodes in the reciprocal space, but *occupy* some volume *around them*.

This can be hopefully clarified by looking at the pictures at pag. 11 and reported below



where a lattice in real space (the crystal equivalent) is windowed by a mask of rectangular (a) or circular (c) shape, simulating a crystal with limited extension. The figs. (b) and (d) show the respective 2D DFT (the reciprocal lattice equivalent), highlighting the predicted nodes widening. In particular, (d) exemplifies the case in which the overall shape of the crystal does not reproduce that of the unit cell.

9 Conclusions

The science of crystallography developed at the beginning of the last century, when the Dirac pulse had not yet been defined, thus finding itself forced to borrow the notations of solid-state physics, optics, electromagnetism, geology, and the science of materials that was developing.

The interpretation of the phenomena involved based on a 3D Fourier transform was therefore adopted with unconcern only in later times, when one could no longer do without it, without wondering if it would not have been possible to find a convergence somewhere earlier in time.

Here it has been shown that the description of a crystal starting from a signal processing point of view from the beginning is a very intimately related approach to the description of the crystal planes based on Miller's indices, allowing a direct interpretation of the reciprocal space such as that of volumetric frequencies, without the need to submit to purely geometric notations.

This involves a more direct interpretation of the diffraction patterns obtainable by X-rays, also on the basis of the analogies that can be highlighted with the case of image processing.

10 Appendix

10.1 2D Fourier basis functions plot

This is the Octave/Matlab code used for the drawings of the 2D Fourier basis functions $A_{h,k}(n,m)$. Matlab users may need to change the comment sign from # to %

```
dim = 128; M = N = dim;     # number of columns and rows
iM = iN = linspace (0, dim-1, dim); # two vectors with columns and row indexes
[m, n] = meshgrid (iM, iN); # two arrays: m has equal rows, n has columns

# Here the vertical (h) and horizontal (k) harmonic indexes are set
h = 20;                     # index of the row harmonic (vertical)
k = 120;                    # index of the column harmonic (horizontal)

basis = 0.99*cos(2*pi*(n*h/N + m*k/M)); # only the real part
surf (iM, iN, basis);

xlabel ("m");
ylabel ("n");
zlabel ("z");
nome = sprintf ("2D Fourier basis function (h,k) = (%d,%d)", h, k);
title (nome);

colormap copper;            # gives the colors for different heights
view (2);                   # view from top
shading("interp");          # interpolate luminances
axis ([0 M 0 N -1.05 1.05]);

nomefile = strcat (nome, ".jpg");
print ("-djpg", nomefile);
```

10.2 DFT of a space-limited bed of nails

At page 10 it is stated that *the transform of a BON still is a BON in the reciprocal space* without a proof, which is given here for the 2D discrete and space-limited sequence

$$\mathrm{BON}_{\Delta_n,\Delta_m,N,M}(n,m) = \sum_{p=0}^{\frac{N}{\Delta_n}-1}\sum_{q=0}^{\frac{M}{\Delta_m}-1}\delta(n-p\Delta_n, m-q\Delta_m)$$

First of all, remember (note 12) that $\mathrm{DFT}\{\delta(n-n_0,m-m_0)\} = \exp\left(-j2\pi\frac{hn_0}{N}\right)\exp\left(-j2\pi\frac{km_0}{M}\right)$, and, as the DFT operator is linear, we have

$$DFT\{\mathrm{BON}_{\Delta_n,\Delta_m,N,M}(n,m)\} = \sum_{p=0}^{\frac{N}{\Delta_n}-1}\exp\left(-j2\pi hp\frac{\Delta_n}{N}\right)\sum_{q=0}^{\frac{M}{\Delta_m}-1}\exp\left(-j2\pi kq\frac{\Delta_m}{N}\right) \quad (25)$$

since the translated pulse transform is factorized. We can then proceed analyzing just one of the factors, for instance the first one. By applying the Euler relation, we have

$$\sum_{p=0}^{\frac{N}{\Delta}-1}\exp\left(-j2\pi hp\frac{\Delta}{N}\right) = \sum_{p=0}^{\frac{N}{\Delta}-1}\cos\left(2\pi hp\frac{\Delta}{N}\right) - j\sum_{p=0}^{\frac{N}{\Delta}-1}\sin\left(2\pi hp\frac{\Delta}{N}\right) \quad (26)$$

The imaginary part is zero for any choice of h, in fact

- with $h = 0$ we have $\sin\left(2\pi \mathbf{0} p \frac{\Delta}{N}\right) = 0$ for any p
- with $h = 1$ the term $\sin\left(2\pi \mathbf{1} p \frac{\Delta}{N}\right)$ with $p = 0, 1, \cdots, \frac{N}{\Delta} - 1$ samples an entire period of the sine, so that the sum of the samples is zero;
- with $h > 1$ the same reasoning above applies, extended over h periods.

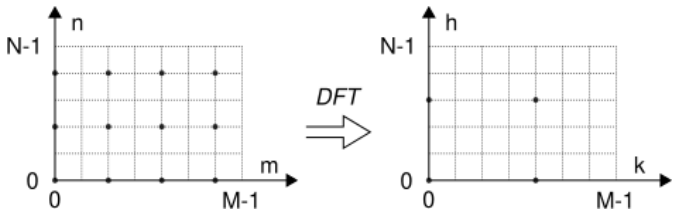
For what regards the real part, we have

- with $h = 0$ we have $\cos\left(2\pi \mathbf{0} p \frac{\Delta}{N}\right) = 1$ for any p, so the sum is equal to the number of its terms, $\frac{N}{\Delta}$;
- with $h = 1$ the term $\cos\left(2\pi \mathbf{1} p \frac{\Delta}{N}\right)$ with $p = 0, 1, \cdots, \frac{N}{\Delta} - 1$ samples an entire period of the cosine, so that the sum of the samples is zero;
- when $h = \frac{N}{\Delta_n}$ the term $\cos\left(2\pi h p \frac{\Delta_n}{N}\right)$ becomes $\cos(2\pi p)$, which for $p = 0, 1, \cdots, \frac{N}{\Delta_n} - 1$ is always one, so the sum is $\frac{N}{\Delta_n}$;
- when h is an integer multiple of $\frac{N}{\Delta_n}$ the same of the above applies, and the sum is $\frac{N}{\Delta_n}$;
- in all of the other cases for h, the second consideration applies to multiple periods of cosine, so the sum is zero.

For these reasons, (26) results in a sequence in h of regularly spaced pulses with period $\frac{N}{\Delta_n}$ and amplitude $\frac{N}{\Delta_n}$. Since identical considerations can be made in regard of the second factor appearing in (25), with the only substitution of $\frac{M}{\Delta_m}$ instead of $\frac{N}{\Delta_n}$, it can be concluded that

$$\begin{aligned} DFT\left\{\mathsf{BON}_{\Delta_n,\Delta_m,N,M}(n,m)\right\} &= \frac{NM}{\Delta_n \Delta_m} \mathsf{BON}_{\frac{N}{\Delta_n},\frac{M}{\Delta_m},N,M}(h,k) \\ &= \frac{NM}{\Delta_n \Delta_m} \sum_{p=0}^{\frac{N}{\Delta_n}-1} \sum_{q=0}^{\frac{M}{\Delta_m}-1} \delta\left(h - p\frac{N}{\Delta_n}, k - q\frac{M}{\Delta_m}\right) \end{aligned}$$

Example By choosing $\Delta_n = \Delta_m = 2$ and $N, M = 6, 8$ as in figure, the DFT of a $\mathsf{BON}_{2,2,6,8}(n,m)$ is $12 \cdot \mathsf{BON}_{3,4,6,8}(n,m)$



Square images In this case it will be $\Delta_n = \Delta_m$ and $N = M$, so that we can write

$$DFT\left\{\mathsf{BON}_{\Delta,N}(n,m)\right\} = \left(\frac{N}{\Delta}\right)^2 \mathsf{BON}_{\frac{N}{\Delta},N}(h,k)$$

10.3 3D Fourier basis functions plot

This other listing is python code for the drawing of the 3D Fourier series basis functions $B_{h,k,l}(x, y, z)$

```
# Drawing of 3D Fourier basis functions
# inspired by https://plotly.com/python/3d-volume-plots/
# Run the code by typing into the terminal    python3 "3D Fourier basis.py"

import plotly.graph_objects as go
import numpy as np

X, Y, Z = np.mgrid[-1:1:30j, -1:1:30j, -1:1:30j]

# harmonics number along directions x, y and z
h = 7;
k = 3;
l = 6;

# build the real part of the 3D basis density
values = np.cos(2*np.pi*(h*X + k*Y + l*Z))

# build the volumetric picture object
fig = go.Figure(data=go.Volume(
    x=X.flatten(),
    y=Y.flatten(),
    z=Z.flatten(),
    value=values.flatten(),
    isomin=-1.05,
    isomax=1.05,
    opacity=0.7,
    surface_count=21,
    opacityscale=[[0, 1], [0.3, 0.5], [0.4, 0.4], [0.5, 0.1],
                          [0.6, 0.4], [0.7, 0.5], [1, 1]],
    colorscale="RdYlGn_r",
    spaceframe_fill=0.9,
    ))

# Set the title
title_string = "h, k, l = %d, %d, %d" % (h, k, l)
fig.update_layout(title_text=title_string)
fig.update_layout(title_x=0.5)
fig.update_layout(title_y=0.85)
fig.update_layout(title_xanchor="center")
fig.update_layout(title_font_size=32)

fig.update_layout(      # font size for the axes
    scene=dict(
        xaxis=dict(tickfont=dict(size=14),title_font=dict(size=24)),
        yaxis=dict(tickfont=dict(size=14),title_font=dict(size=24)),
        zaxis=dict(tickfont=dict(size=14),title_font=dict(size=24))
    ))

fig.show()              # the resulting picture will be drawn in your web browser
```

References

- Aubert, E., Lecomte, C. (2007) Illustrated Fourier Transforms for crystallography. Journal of Applied Crystallography. 1153. 10.1107/S0021889807043622 and Fourier Transform Lab software
- Discrete Fourier transform in Wikipedia
- 2D Discrete Fourier Transform at Verona University
- Discrete Fourier Transform at Worcester Polytechnic Institute
- 2D Fourier transforms and applications at Oxford University
- Brad Osgood The Fourier Transform and its Applications, Lecture Notes for EE 261, Electrical Engineering Department Stanford University. At Cap. 8, n-dimensional Fourier Transform applied to crystallography
- Protein Crystallography Course at University of Cambridge
- Crystallography Classnotes at the University of Toledo
- S. V. Smaalen, (2013), Structure Determination of Single Crystals in Modern Diffraction Methods, First Edition, Ed. by E. J. Mittemeijer and U. Welzel, 2013 Wiley-VCH Verlag GmbH & Co. KGaA
- F., James, E., Gwyndaf. (2008). Elucidations on the reciprocal lattice and the Ewald sphere European Journal of Physics. 29, 1059.

Wide coverage of arguments

- Basics of image processing at University of Strasbourg
- The Scientist and Engineer's Guide to Digital Signal Processing an on-line book about DSP
- Solids and Surface Chemistry a collaborative LibreText
- Welcome to the world of Crystallography at Dept. of Crystallography & Struc. Biol., Consejo Superior de Investigaciones Científicas, Madrid
- 250 questions to a crystallographer by Gervais Chapuis at École Polytechnique Fédérale de Lausanne
- Crystallography 101 by Bernhard Rupp
- Online Dictionary of Crystallography maintained by the Commission for Crystallographic Nomenclature (CCN) of the International Union of Crystallography (IUCr).

Interactive tools

- Fourier Transform Filtering Techniques in Optical Microscopy Primer - noise removal
- 2D Inverse Fourier Transform Playground lets to draw 2D filters by hand, also shows 2D basis functions in reciprocal space
- http://lampz.tugraz.at/∼hadley/ss1/crystaldiffraction/interference2d.php from Gratz University, animated 2D interference by two scatterers
- Kevin Cowtan at the University of York, has many interactive demos about 2D Fourier transform, Miller indices and structure factor, Ewald sphere, and more