united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

ICTP 40th Anniversary

SCHOOL ON SYNCHROTRON RADIATION AND APPLICATIONS
In memory of J.C. Fuggle & L. Fonda

19 April - 21 May 2004

Miramare - Trieste, Italy

1561/9

Optical Path Function

Anna Bianco

strada costiera, 11 - 34014 trieste italy - tel. +39 040 2240111 fax +39 040 224163 - sci_info@ictp.trieste.it - www.ictp.trieste.it

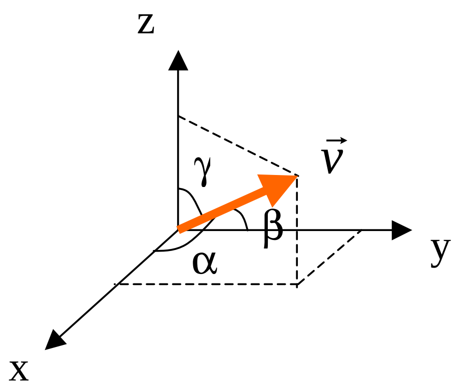
The electron beam

The electron beam circulating in a synchrotron radiation source is an ensemble of particles moving in such a way that the velocity in one direction is much greater than the velocity in the other directions.
Let us define a Cartesian coordinate system in which the y axis is the instantaneous beam direction and x and z are respectively in and perpendicular to the electron ring plane.
The state of each electron can be specified by a point in the 6-dimensional phase space $\{x,y,z,v_x,v_y,v_z\}$. Usually the longitudinal motion along the beam axis y is decoupled from the motion in the transverse xz plane and the 6-dimensional space can be split into a longitudinal 2-dim phase space and a transverse 4-dim phase space. We will consider only the transverse phase space $\{x,z,v_x,v_z\}$. The transverse velocities of the electrons are usually expressed in terms of the direction cosines:

$$v_x = v\cos\alpha$$

$$v_z = v\cos\gamma$$

where α and γ are the angles between the instantaneous electron velocity $\vec{v}$ and the reference axes x and z.



Introducing the new coordinates:

$$x' = \cos\alpha$$

$$z' = \cos\gamma$$

and using the hypothesis:

$$v_y >> v_x, v_z$$

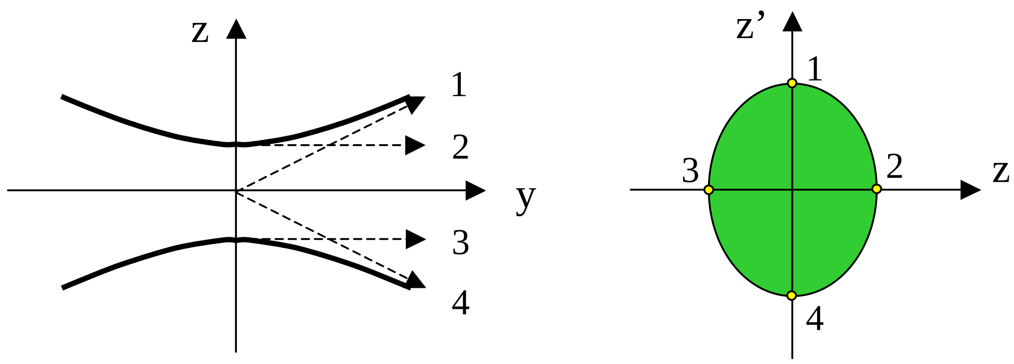
the transverse velocities are proportional to x',z':

$$v_x \cong v_y x'$$

$$v_z \cong v_y z'$$

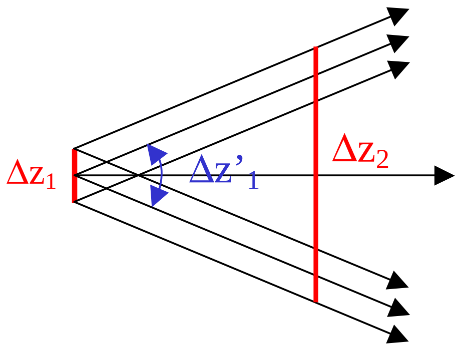
In first approximation, x' and z' have a very simple meaning: x' and z' are the angles between the y axis and the projections of the instantaneous electron velocity respectively in the xy and yz plane.
One is usually more interested to know about the angles x' and z', therefore the transverse phase space is considered to be $\{x,z,x',z'\}$ instead of $\{x,z,v_x,v_z\}$.

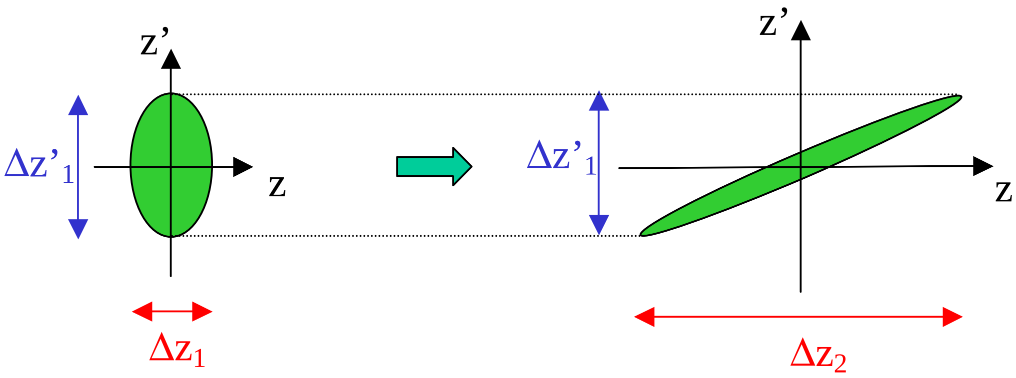
An electron beam can be represented in a phase space diagram as a cloud of points within a close contour, usually an ellipse. The example of an instantaneous cross section of a beam at a waist is shown in the following sketch.



The area within the close contour is called the emittance of the beam. Emittance is a very important parameter for the electron beam. Liouville theorem states that if the forces acting on the electrons are conservative, emittance is conserved. This is what happens to the electron beam: along the ring, the shape of the occupied phase space changes but the volume remains constant.
If the ellipse is upright, its area, which is the product of the two semi-axes, is the beam size multiplied by the beam divergence. In this case Liouville's theorem states that the product beam size multiplied by beam divergence is conserved.

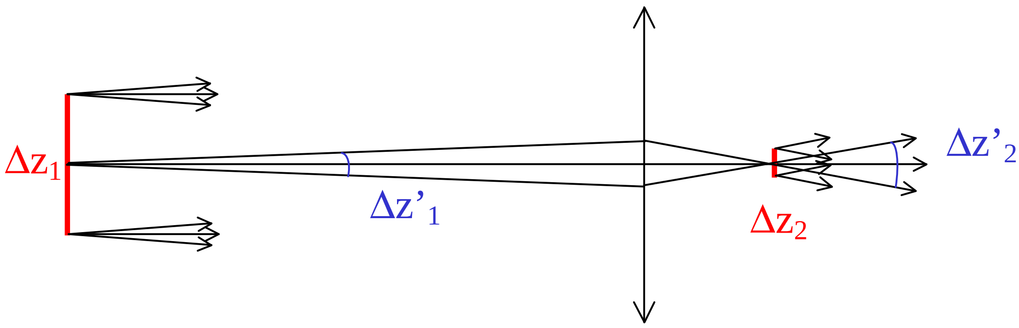
Example 1: a diverging beam.

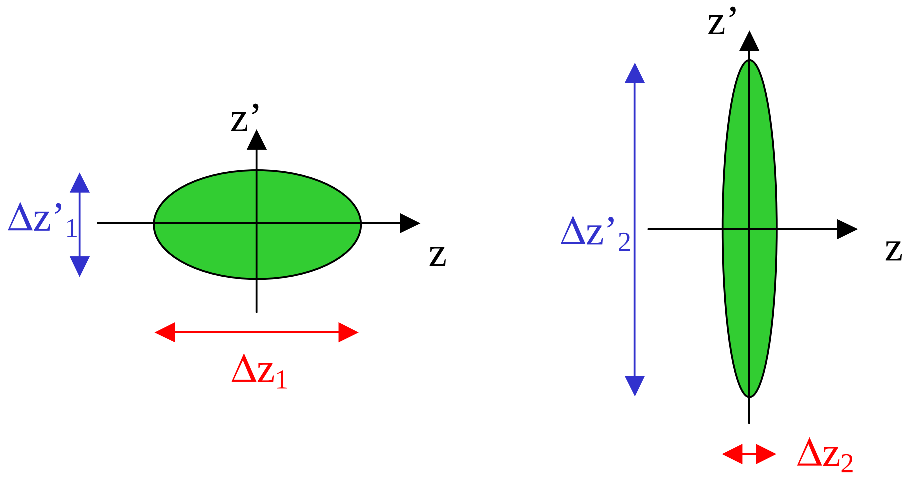




In the first position, electrons are crowded in physical space and spread out in angles
In the second position, electrons are diffuse in physical space and angles: the corresponding ellipse is tilted, but its area remains constant

Example 2: a focusing beam.
After the focusing action of a quadrupole magnet, the beam will be more confined in space and more divergent in angles.





The area of the ellipse representing the occupied phase space is always the same: if you confine the beam in space you make it more divergent, if you make it more parallel the beam diameter increases.
For a 4 dimensional transverse phase space, the electron beam emittance is**:**
$\varepsilon = \sigma_e \sigma'_e$
where:
σ_e is the lateral extent of the electron beam
σ'_e is the solid angle of the electrons trajectories around the ideal trajectory.
Since the two orthogonal lateral directions are only weakly coupled, it is useful to define a horizontal emittance ε_x and a vertical emittance ε_z:
$\varepsilon_z = C\, \varepsilon_x$
where C is the coupling factor, which, like the emittance, is an invariant of the system. The coupling factor depends on the “goodness” of the alignment of the magnet fields in the ring.

The third generation synchrotron radiation sources are optimized for low-emittance electron beams.
Case of ELETTRA:
Electron energy=2GeV
Horizontal emittance ε_X = 7 micron mrad

C = 1%
In the straight sections the electron beam has:
Horizontal/vertical dimensions = 241/15 micron
Horizontal/vertical divergence = 29/6 microrad

Basic properties of synchrotron radiation

The main distinguishing properties of synchrotron radiation (SR) that make it so advantageous depend on the characteristics of the storage ring and are summarized in the following:

1. photon energy can be selected in a very broad and continuous spectral range, from infrared through vacuum-ultraviolet, extreme ultraviolet, soft x-rays up to hard x-ray region.
2. SR has high intensity
3. the emitted radiation is highly collimated and emanates from a very small source, the electron beam
4. the electrons in the storage ring are grouped in bunches, therefore SR is emitted as pulses from each bunch
5. SR is highly polarized: SR emitted from bending magnets and conventional ID in the plane of the ring is linear polarized (the electric vector lies in the plane of the ring). Above and below the plane of the ring the radiation is elliptically polarized.

Property number 3 can be well described introducing a parameter called "brilliance", which is a combination of the emitted flux and of two geometric characteristics: the source size and the angular spreads of the emitted beam:

$$Brilliance = \frac{photon\ flux}{I}\frac{1}{\sigma_x \sigma_z \sigma_{x'} \sigma_{z'} BW}$$

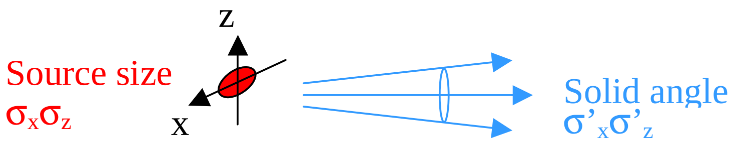
I = electron current in the storage ring

$\sigma_x \sigma_z$ = transverse area from which the SR is emitted

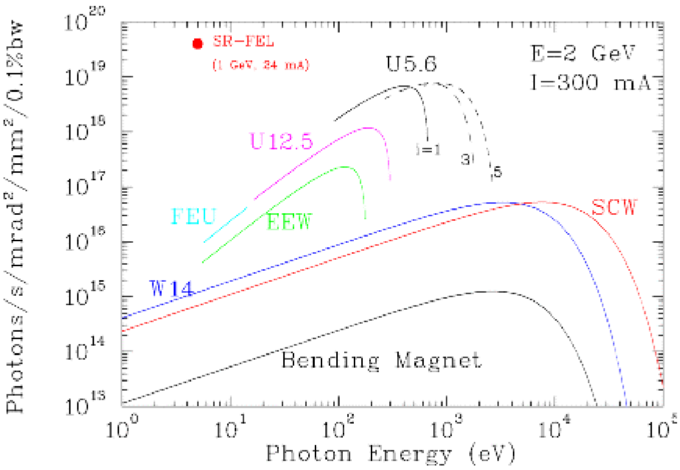
$\sigma_{x'} \sigma_{z'}$ = solid angle into which the SR is emitted

BW = spectral bandwidth, usually $\frac{\Delta E}{E} = 0.1\ \%$

The Cartesian reference system is the same as that introduced for the electron beam: the y axis is the central direction of the photon beam while x and z are in the transverse plane, in and perpendicular to the orbit plane.



Brilliance is the main figure of merit of SR sources.
As an example, the following figure shows the photon brilliance (photons/s/0.1%BW/mrad) as a function of photon energy for different ID's at ELETTRA, electron energy 2GeV, electron current 300mA.



The figure illustrates how wiggler and undulator sources can provide greatly improved synchrotron characteristics over more conventional bending magnet radiation.

Why is brilliance important?
On one hand, the importance of a higher flux is evident: more flux means more signal for the experiment. Moreover, having a lot of photons in a narrow energy band allows to reach a good energy resolution with a sufficient signal, that is without increasing the acquisition time.
But why combining the flux with geometrical factors? The answer is provided by Liouville's theorem, that we have already applied to the electron beam. Applied to optics it states that the occupied phase space volume is constant along the optical path for an optical system without losses. More simply, Liouville's theorem states that the product (beam size x beam divergence) cannot be decreased without loosing photons. This implies that to focus a beam in a small spot (which is needed for achieving high energy and/or spatial resolution) one must accept an increase in the beam divergence. Along the beamline this requires the installation of large size optical devices, which are expensive and have lower optical qualities. But also at the sample the divergence cannot be made arbitrarily large: it is impossible to concentrate the whole radiation in an arbitrarily small spot. With a not brilliant source the spot size at the sample can be made small only reducing the photon flux.
The answer to the question "why is brilliance important?" is therefore the following: the high brilliance of the radiation source allows the development of monochromators with high energy resolution and high throughput and gives also the possibility to image a beam down to a very small spot on the sample with high intensity.

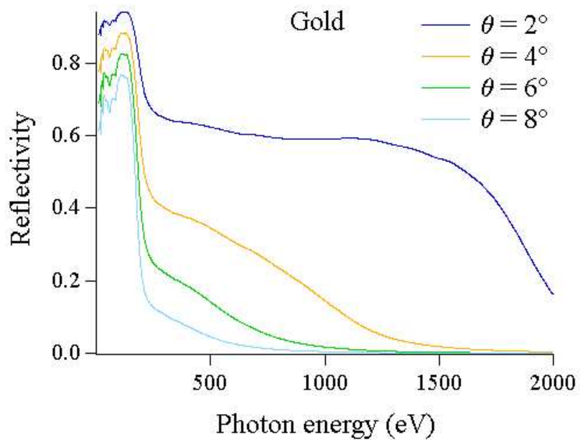
The beamline

The researcher needs at his experiment a certain number of photons/second into a phase volume of some particular characteristics. Each kind of experiment imposes different demands, for example scanning microscopy needs a high flux of photons focused in a sub-micron spot. Moreover, these photons have to be monochromatized to a greater or less extent.
This is the job of the beamline: the beamline is the means of bringing the radiation from the source to the experiment transforming the phase volume in a controlled way, that is: demagnifying and monochromatizing the source and refocusing it onto a sample.
In doing this, the beam-line must preserve the excellent qualities of the radiation up to the experiment. We have seen that third generation synchrotron radiation sources are optimized for low-emittance electron beams and high-brilliance radiation: the beam-line scientist and engineer has to design a beam-line that transfers this brilliance to the experiment. We will come back to this issue.

We restrict ourselves to photon energies from 10 to 2000eV. This energy range includes Vacuum Ultra Violet (VUV,which extends from 4eV to 30eV), Extreme Ultra Violet (EUV, from 30eV to 250eV) and soft x-ray regions (up to several keV). These limits are not precisely defined and can vary from author to author.
VUV, EUV and soft x-ray regions are very interesting because they are characterized by the presence of the absorption edges of most low and intermediate Z elements (Z is the atomic number). Thus photons with these energies are a very sensitive tool for elemental and chemical identification. But from the point of view of the optics, these absorption edges make these regions difficult to access. What makes VUV and soft x-ray radiation so different from neighboring spectral regions is the high degree of absorption in all materials. At lower photon energies, in the visible and ultraviolet, and at higher photon energies, in the hard x-ray region, many materials become transparent, but at VUV and soft x-rays all materials are absorbing. Because no windows are available, the entire optical system must be kept under vacuum in order not to disturb the storage ring and the experiment. Moreover, ultrahigh vacuum conditions ($P=1\text{-}2\times10^{-9}$ mbar) are required to avoid photon absorption in air and to protect the optical surfaces from contamination (especially from carbon).
The other consequence of the strong absorption of radiation by all materials in this spectral region is that no refractive optics are available. The only[1] optical elements which can be used in the VUV and soft x-rays regions are mirrors and diffraction gratings, used in reflection.
Moreover, the reflectivity of ordinary materials[2] are very small and, at a given energy, drop down fast with the increasing of the grazing incidence angle. Therefore only mirrors and gratings at grazing incidence angles (1-2 degrees) are used. As an example, the following figure shows the reflectivity of gold as a function of photon energy for different incidence angles.

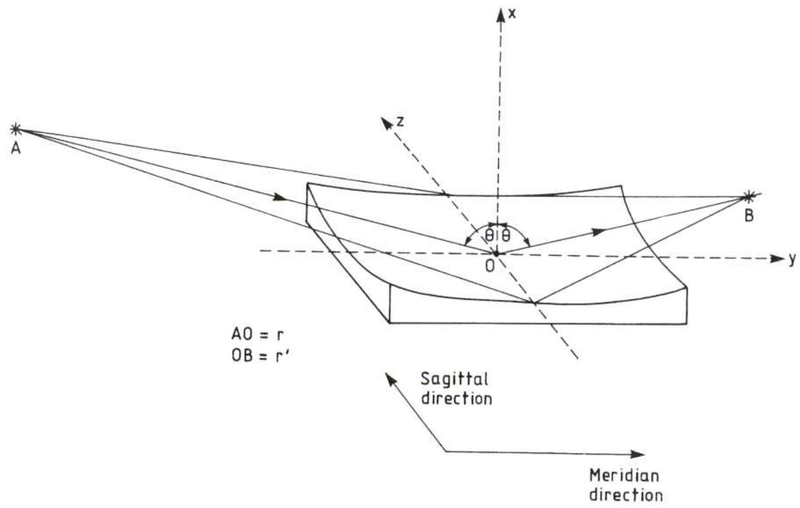
---

[1] There are few exceptions, for example zone plates, which are transmission optical devices composed of nanometer-scale concentric metal rings which are used to focus x-rays.
[2] Special materials called multilayers can be built with high reflectivity at particular given energies.



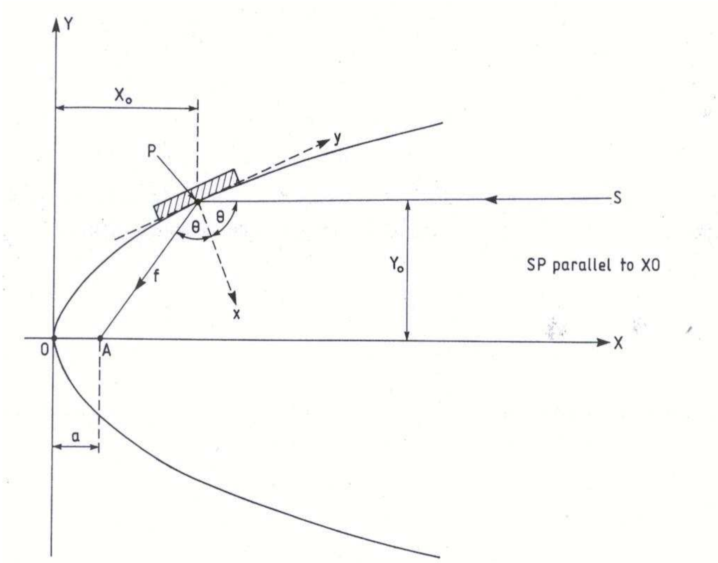
Focusing properties and surface equations of common optical elements

X-ray optics can have different kinds of geometrical shapes: they vary from plane to more exotic aspherical forms. We will review the main focusing properties of the more common geometrical surfaces: paraboloid, ellipsoid, toroid.
First, let us consider a generic concave mirror and a point source A which is focused to the point B. The plane defined by the central ray AO and the normal to the mirror in O is called tangential or meridional plane. The plane perpendicular to the tangential plane and containing the normal to the surface in O is called sagittal plane. We'll use a Cartesian reference system with origin in O, x axis normal to the surface in O , y and z axis belonging to the plane tangent to the mirror in O, y belongs to the tangential plane, x to the sagittal plane.



Paraboloid

The parabola has the property that rays traveling parallel to the symmetry axis OX are all focused to a point at A. Conversely, the parabola will collimate rays emanating from a point situated at the focus of the parabola A.



If focusing in one plane only is required, the mirror surface would be plane in the direction perpendicular to the plane of the diagram. If the angle of incidence on the off-axis segment is ϑ, f is the distance from the pole of the segment to the focus and a is the distance from the focus to the pole of the parabola, then the line equation for the parabola is:

$$Y^2 = 4aX$$

where

$$a = f\cos^2\vartheta$$

The position of the pole can be given by

$$X_o = a\tan^2\vartheta$$
$$Y_o = 2a\tan\vartheta$$

If double focusing is required, a paraboloidal shape could be used and can be obtained by rotating a parabola around its axis of symmetry OX. The equation of the surface of the paraboloid would then be

$$Y^2 + Z^2 = 4aX$$

The above equations use a Cartesian coordinate system centered on the pole of the parabola (X,Y,Z). It is often necessary, however, to use the reference coordinate system (xyz) introduced before, centered on the center of the optical segment under consideration. In this coordinate system, the equation of the paraboloid becomes

$$x^2 \sin^2 \vartheta + y^2 \cos^2 \vartheta + z^2 - 2xy \sin\vartheta \cos\vartheta - 4ax \sec\vartheta = 0$$

Ellipsoid

The ellipse has the property that rays from one point focus F_1 will always be perfectly focused to the second point focus F_2. In a Cartesian system (X,Y,Z) centered on the center of the ellipse O, the equation of the ellipse is given by

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1$$

The parameters of the ellipse *a* and *b* can be defined in terms of the object and image distances *r* and *r'* and the angle of incidence ϑ by

$$a = \frac{r + r'}{2}$$
$$b = a\sqrt{1 - e^2}$$

where *e* is the eccentricity and is given by

$$e = \frac{1}{2a}\sqrt{r^2 + r'^2 - 2rr'\cos(2\vartheta)}$$

The position of the pole of the element is given by

$$Y_0 = \frac{rr'\sin(2\vartheta)}{2ae}$$
$$X_0 = a\sqrt{1 - \frac{Y_0^2}{b^2}}$$

If double focusing is required, an ellipsoidal shape could be used. This would be obtained by rotating the ellipse around the major axis OX. The equation of the surface formed would be given

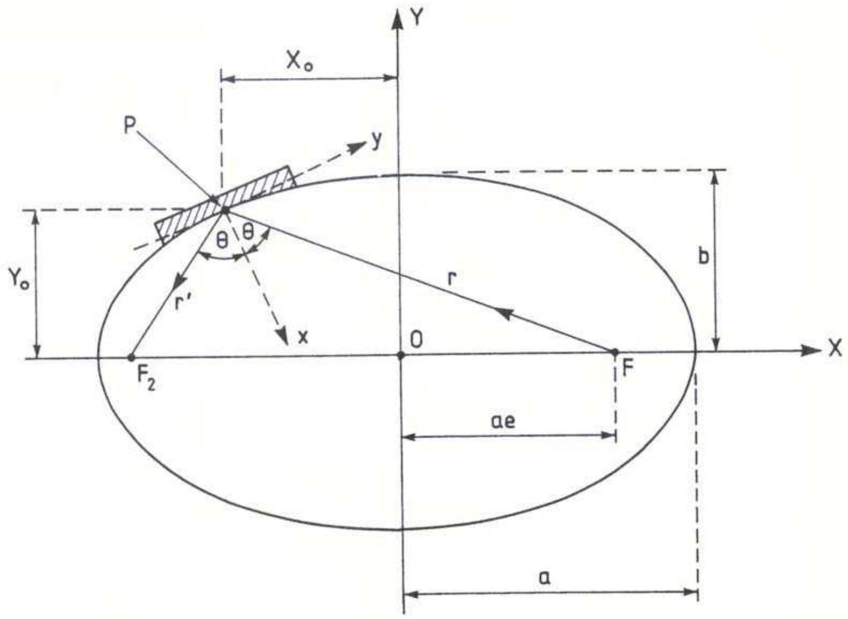
$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} + \frac{Z^2}{b^2} = 1$$

In the coordinate system x,y,z with the origin at the center of the optical, the surface equation of the ellipsoid is given by

$$x^2\left(\frac{\sin^2\vartheta}{b^2}+\frac{1}{a^2}\right)+y^2\left(\frac{\cos^2\vartheta}{b^2}\right)+\frac{z^2}{b^2}-x\left(\frac{4f\cos\vartheta}{b^2}\right)-xy\left[\frac{2\sin\vartheta\sqrt{e^2-\sin^2\vartheta}}{b^2}\right]=0$$

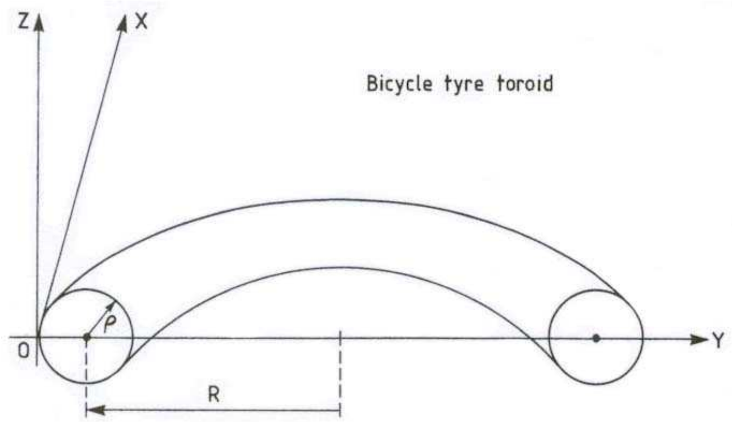
where

$$f=\left(\frac{1}{r}+\frac{1}{r'}\right)^{-1}$$



Toroids

The bicycle tyre toroid is generated by rotating a circle of radius ρ in an arc of radius R. (ρ is the minor radius, R is the major radius.).



In the coordinate system x,y,z with the origin at the center of the optical element, the surface equation can be given as

$$x^2 + y^2 + z^2 = 2Rx - 2R(R-\rho) + 2(R-\rho)\sqrt{(R-x)^2 + y^2}$$

The conditions for a meridian focus are

$$\left(\frac{1}{r} + \frac{1}{r'}\right)\frac{\cos\vartheta}{2} = \frac{1}{R}$$

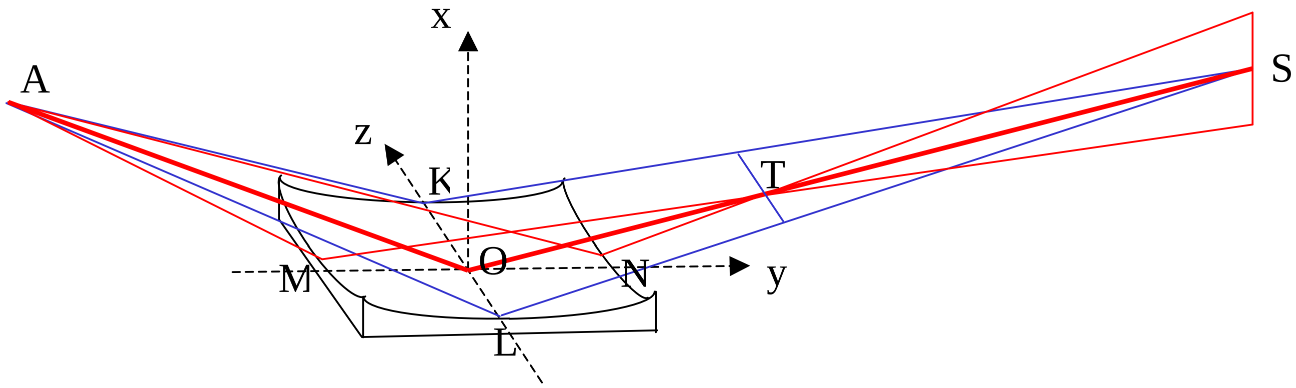
and for a sagittal focus are

$$\left(\frac{1}{r} + \frac{1}{r'}\right)\frac{1}{2\cos\vartheta} = \frac{1}{\rho}$$

To obtain a stigmatic image of a point source we therefore have the relationship between R and ρ:

$$\frac{\rho}{R} = \cos^2\vartheta$$

The following sketch illustrates a vertical deflecting toroidal mirror: the point source A has the tangential focus T and the sagittal focus S in two different positions. The bundle of rays in the meridian plane MON, which also contains the source point A, forms an image of A at the vertical meridional focus T. Bundles of meridional rays like AK and AL will be focused each in their own plane of incidence, so the meridional image of the point A is a horizontal line. The sagittal bundles reflected from KOL will be focused at the sagittal focus S. The sagittal bundles reflected from zones of the mirror before and after KOL will be focused to points above and below S. The sagittal image of the point A is therefore a vertical line.



A special case of toroid is a sphere: ρ=R. The relation $\frac{\rho}{R} = \cos^2 \vartheta$ implies that a stigmatic image can be obtained only if ϑ =0, that is: for a spherical mirror a stigmatic image can only be obtained at normal incidence. For a vertical deflecting spherical mirror at grazing incidence the horizontal sagittal focus is always further away from the mirror than the vertical tangential focus. The spherical mirror only weakly focusses in the sagittal direction.

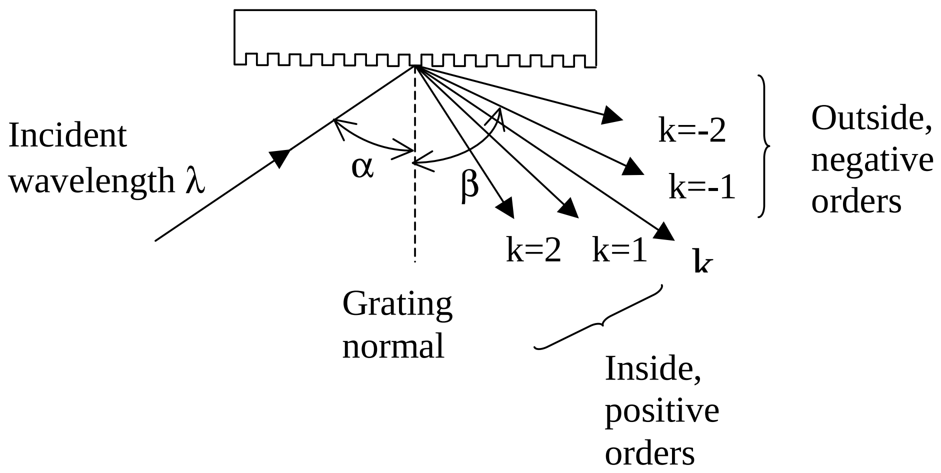
Diffraction gratings

A diffraction grating is the heart of a soft x-rays beamline: it separates the different components of the spectrum by redirecting the radiation by an amount which depends upon the wavelength. A slit after a grating is therefore able to select and trasmit a given narrow band of wavelengths. The most common grating consists of a suitable rigid substrate with an optical surface upon which are produced a series of equispaced parallel grooves. The standard notation used is shown in the following sketch.
Radiation of wavelength λ incident on the grating at an angle α, measured with respect to the surface normal, is either reflected (zero order) or diffracted at an angle β. The diffraction angle β is related to the angle of incidence α, the groove density N and the wavelength λ by the grating equation:

$$\sin\alpha + \sin\beta = Nk\lambda$$

α and β are of opposite sign if on opposite sides of the surface normal. In the figure, α is positive and β is negative. Diffraction into orders whose angles β are greater than that of

the zero-order beam are called “negative” or outside orders, and for these k takes negative values. In the figure, β is shown for the first negative order.
The relative amount of light of wavelength λ reflected or diffracted into a particular order k is determined by the shape of the grooves and by the material on the grating surface.

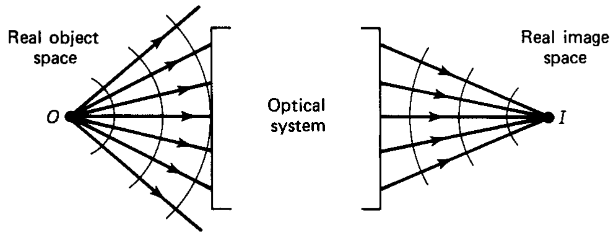


Conserving brilliance

The basic elements of a beam-line working in the photon energy range from 10 to 2000eV are mirrors, gratings and slits.
Total reflection X-ray mirrors operating at grazing angles of incidence can efficiently focus a wide beam of x-rays. Mirrors can also be employed as filters to effectively reflect or pass x-rays below a certain cut-off energy and to absorb or block x-rays above this energy. Gratings diffract radiation as a function of the wavelength and slits spatially select the dispersed radiation.
In doing these jobs, the optical elements have to preserve the high quality of the radiation. We have seen that the parameter that well describes the quality of the emitted radiation is brilliance. Conserving brilliance along a beam-line requires special attention.
Brilliance decreases because of different reasons:

- Micro-roughness and slope errors on optical surfaces
- Thermal deformations of optical elements due to heat load produced by the high power radiation
- Aberrations of optical elements

We will not consider the effects of real optical surfaces, such as slope errors, roughness, and thermal induced deformations; we will study the focusing properties of optical elements, mirrors and gratings, with theoretical surface shapes. These optical elements produce aberrated images. Let us understand what aberrations are.
An ideal optical element is able to perform perfect imaging if all the rays originating from a single object point cross at a single image point, or, equivalently, the wave front in image space has a spherical shape centered on the image point.



Deviations from perfect imaging are called aberrations.
Image quality is essential for achieving high energy and spatial resolution: knowledge of the aberration theory is therefore necessary for the design of a beamline.
Aberration theory allows to treat mathematically in general terms the focusing properties of a concave optical element. We will study the case of a grating, a mirror can be considered a particular case of grating with zero density of lines. The general theory of aberrations of diffraction gratings applies Fermat's principle to derive expressions for the aberrations coefficients.

Principle of Fermat

Fermat's principle asserts that light rays choose their paths to minimize the optical length. The optical length is defined by the actual length of the path times the index of refraction of the medium. Assuming the index of refraction to vary as a function of the position, $n(\vec{r})$, the optical path length from point A to point B is given by

$$\int_A^B n(\vec{r})dl$$

where $dl = \sqrt{dx^2 + dy^2 + dz^2}$ is the line segment along the path.
If a medium has a constant refraction index, Fermat's principle states that light will travel along straight lines.
Since

$$n(\vec{r})dl = \frac{c}{v}dl = cdt$$

where c is the light speed in vacuum and v is the light speed in the medium, the optical path length is equal to $c\int_A^B dt$. That is why the principle of Fermat is also known as the principle of least time: the actual path between two points taken by a beam of light is the one which is traversed in the least time. The laws of reflection and refraction directly follow from Fermat'principle.

Theory of conventional diffraction gratings

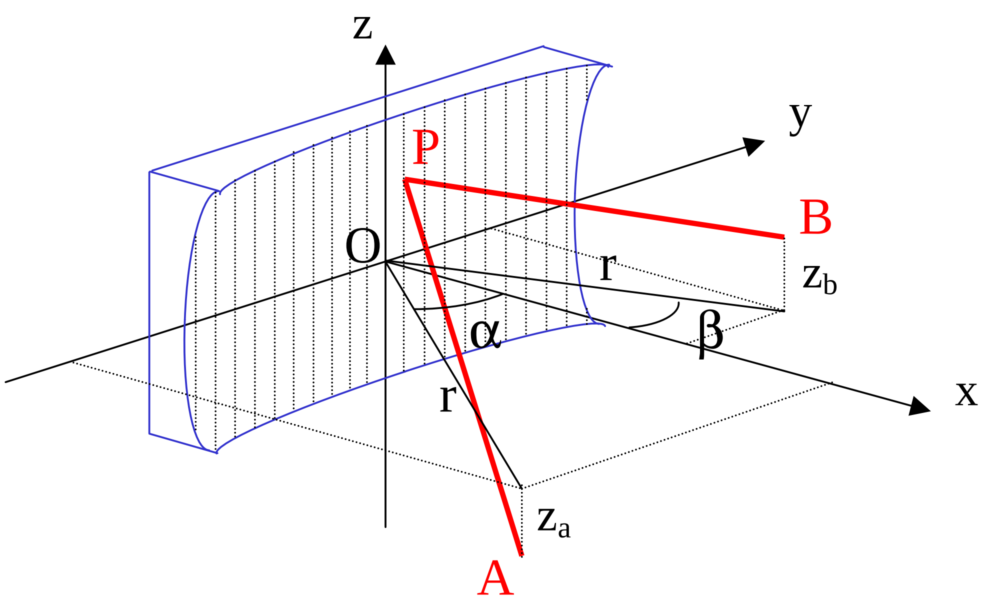
The following figure shows a classical concave grating with parallel rectilinear grooves and constant spacing d. The Cartesian coordinate system has origin O in the grating pole, x-axis along the normal to the surface at O and y- and z-axes belonging to the plane tangent to the grating at O. The z-axis is parallel to the groove which passes through O. Let us consider a light ray which leaves the generic point $A(x_a, y_a, z_a)$, hits the mirror/grating surface at an arbitrary point P(x,y,z) and is reflected/diffracted to the point $B(x_b, y_b, z_b)$.
Let assume that the refraction index is constant, therefore light travels along straight lines. The optical path length F for that ray in the case of a mirror is simply:

$$F = \overline{AP} + \overline{PB}$$

For a grating an additional term has to be inserted to cope with the physical optics of diffraction. For a classical grating with rectilinear grooves with constant spacing d, this additional term is:

$$F = \overline{AP} + \overline{PB} + kN\lambda\, y$$

where λ is the wavelength of the diffracted light, k is the order of diffraction (±1,±2,...), N=1/d is the grating groove density. To better understand this additional term, let us consider two points on the grating surface, P_1 and P_2 , separated by an integer number of grooves. The last term accounts for the fact that the distances AP_1B and AP_2B may be different and the light along both rays may be in phase at B if the difference between these two distances contains an integer number of wavelengths. If $P_1$and P_2 are one groove apart, the number of wavelengths in the difference is the order of diffraction.



In the above drawing the positions of A and B are specified by the cylindrical coordinates r, α, z_a and r',β, z_b; the angles of incidence and diffraction, α and β, are measured in the xy plane and the signs of α and β are opposite if the incident and diffracted rays are on opposite sides of the xy plane (in the figure α is negative and β is positive.)

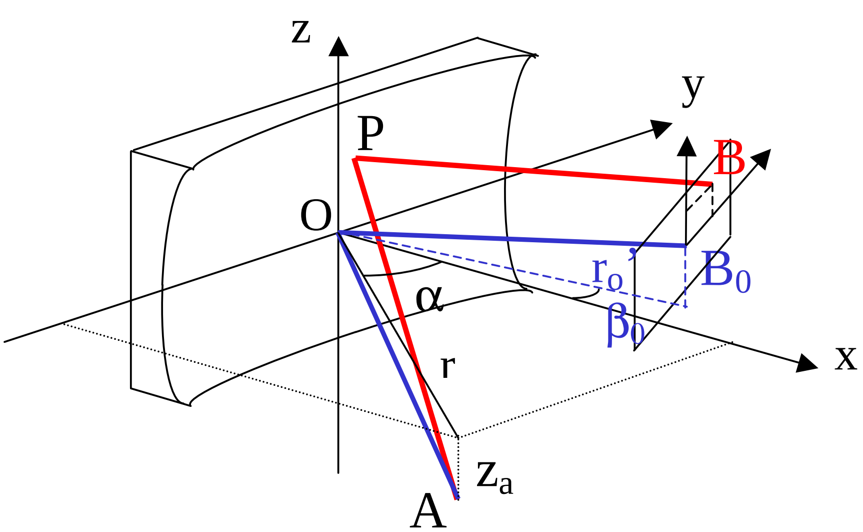
The groove profile is ignored in this geometrical treatment of grating aberrations, but it is of course extremely important in the calculation of intensity distribution.

Now, let us consider some number of light rays starting from A and impinging on the grating at different points P. Fermat's principle states that if the point A is to be imaged at the point B, then all the optical path lengths from A via the grating surface to B will be the same. Thus B will be the point of a perfect focus if the two equations:

$$\frac{\partial F}{\partial y} = 0 \quad \frac{\partial F}{\partial z} = 0$$

are satisfied simultaneously by any pair of (y,z). For a mirror, this is accomplished if the surface has the correct geometry. For a grating, the shape of the surface and the groove density N decide if the desired result will be achieved.

In general, $\frac{\partial F}{\partial y}$ and $\frac{\partial F}{\partial z}$ are functions of y and z and can not be made zero for any (y,z). Therefore when the point P wanders over the grating surface, diffracted rays fall on slightly different points on the focal plane and an aberrated image is formed. The image produced by the central ray (which is the ray passing through the origin O) is called the Gaussian image (point B_0 in the following figure).



The ray diffracted by the generic point P on the grating surface will arrive at a different point B: the displacement of B with respect to B_0 is called the ray aberration. The goal of an aberration theory is to produce simple expressions for these intersection points B in the image plane produced by the rays diffracted from different points on the gratings surface. We will see that these expressions can be reduced to a form consisting of a sum in powers of the aperture coordinates y and z, each term representing a particular geometric aberration. By adjusting the values of the different parameters, such as substrate shape, groove density, object and image distances, the sum of the aberrations can be reduced.

The mirror/grating surface may be described by its analytical equation or may in general be described by a series expansion:

$$x = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{ij} y^i z^j$$

where $a_{00} = a_{10} = a_{01} = 0$ because of the choice of origin and j=even if the xy plane is a symmetry plane.
Giving suitable values to the coefficients a_{ij}'s, we obtain the expressions for the various possible surfaces. The first coefficients for toroidal, paraboloidal and ellipsoidal surfaces are reported in the following. Values for the toroidal and ellipsoidal coefficients are given to sixth order at http://xdb.lbl.gov/Section4/Sec_4-3Extended.pdf. The geometry and the parameters for the three different cases have been introduced before, when we described these three geometrical surfaces.
Note that sphere, cylinder and plane are special cases of the toroid:
R=ρ → sphere
R= ∞ → cylinder
R=ρ= ∞ → plane.
For gratings only toroids, sphere and plane surfaces are of interest.

Toroid

$$a_{02} = \frac{1}{2\rho}; \quad a_{20} = \frac{1}{2R}; \quad a_{22} = \frac{1}{4R^2\rho}; \quad a_{40} = \frac{1}{8R^3};$$
$$a_{04} = \frac{1}{8\rho^3}; \quad a_{12} = 0; \quad a_{30} = 0$$

Paraboloid

$$a_{02} = \frac{1}{4f\cos\vartheta}; \quad a_{20} = \frac{\cos\vartheta}{4f}; \quad a_{22} = \frac{3\sin^2\vartheta}{32f^3\cos\vartheta};$$
$$a_{12} = -\frac{\tan\vartheta}{8f^2}; \quad a_{30} = -\frac{\sin\vartheta\cos\vartheta}{8f^2}$$
$$a_{40} = \frac{5\sin^2\vartheta\cos\vartheta}{64f^3}; \quad a_{04} = \frac{\sin^2\vartheta}{64f^3\cos^3\vartheta}$$

Ellipsoid

$$a_{02} = \frac{1}{4f\cos\vartheta}; \quad a_{20} = \frac{\cos\vartheta}{4f}; \quad a_{04} = \frac{b^2}{64f^3\cos^3\vartheta}\left[\frac{\sin^2\vartheta}{b^2} + \frac{1}{a^2}\right];$$

$$a_{12} = \frac{\tan\vartheta}{8f^2\cos\vartheta}\left(e^2 - \sin^2\vartheta\right)^{1/2}; \quad a_{30} = \frac{\sin\vartheta}{8f^2}\left(e^2 - \sin^2\vartheta\right)^{1/2};$$

$$a_{40} = \frac{b^2}{64f^3\cos^3\vartheta}\left[\frac{5\sin^2\vartheta\cos^2\vartheta}{b^2} - \frac{5\sin^2\vartheta}{a^2} + \frac{1}{a^2}\right];$$

$$a_{22} = \frac{\sin^2\vartheta}{16f^3\cos^3\vartheta}\left[\frac{3}{2}\cos^2\vartheta - \frac{b^2}{a^2}\left(1 - \frac{\cos^2\vartheta}{2}\right)\right]$$

$$where \quad f = \left[\frac{1}{r} + \frac{1}{r'}\right]^{-1}$$

Now, let us consider the optical path function F. The path lengths $\overline{AP}$ and $\overline{PB}$ can be expressed as:

$$\overline{AP} = \sqrt{(x_a - x)^2 + (y_a - y)^2 + (z_a - z)^2}$$

$$\overline{PB} = \sqrt{(x_b - x)^2 + (y_b - y)^2 + (z_b - z)^2}$$

and x can be substituted with the power series in y and z:

$$x = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_{ij} y^i z^j$$

It is also convenient to express the position of A and B in cylindrical coordinates:

$$x_a = r\cos\alpha \qquad y_a = r\sin\alpha$$
$$x_b = r'\cos\beta \qquad y_b = r'\sin\beta$$

After these substitutions, the resulting expression for F can be expressed as a power series in the aperture coordinates y and z:

$$
\begin{aligned}
F &= \sum_{ijk} F_{ijk} y^i z^j \\
&= F_{000} + yF_{100} + zF_{011} + \frac{1}{2}y^2 F_{200} + \frac{1}{2}z^2 F_{020} + \frac{1}{2}y^3 F_{300} + \frac{1}{2}yz^2 F_{120} + \frac{1}{8}y^4 F_{400} + \frac{1}{4}y^2 z^2 F_{220} + \\
&\quad + \frac{1}{8}z^4 F_{040} + yzF_{111} + \frac{1}{2}yF_{102} + \frac{1}{4}y^2 F_{202} + \frac{1}{2}y^2 zF_{211} + \ldots
\end{aligned}
$$

It can be proved that the coefficient F_{ijk} has the general structure:

$$F_{ijk} = z_a^k C_{ijk}(\alpha, r) + z_b^k C_{ijk}(\beta, r') + N\,k\,\lambda\, f_{ijk}$$

$$f_{ijk} = \begin{cases} 1 & \text{when } \text{ijk} = 100 \\ 0 & \text{otherwise} \end{cases}$$

The first two subscripts i,j refer to the exponents of y and z, the third subscript k refers to the exponent of z_a and z_b .

We have seen that the condition for having a perfect focus is that the two partial derivatives of F are zero for each pair (y,z). This means that all (except F_{000}) the coefficients F_{ijk} must be zero. No optical system can be made to satisfy all these conditions at once: the operative approach is to understand which are the most significant terms and set them to zero. In this way the two derivatives, also if not equal to zero, are small enough.
Each term in the series (except the first two terms F_{000} and F_{100}) represents a particular type of aberration and the corresponding coefficient F_{ijk} is related to the strength of that aberration. This will be clearer in the following.

The expressions of the most important terms, for r,r' >> z_a,z_b, are reported in the following.

$$F_{000} = r + r'$$

$$F_{100} = Nk\lambda - (\sin\alpha + \sin\beta)$$

$$F_{200} = \left(\frac{\cos^2\alpha}{r} + \frac{\cos^2\beta}{r'}\right) - 2a_{20}\left(\cos\alpha + \cos\beta\right)$$

$$F_{020} = \frac{1}{r} + \frac{1}{r'} - 2a_{02}\left(\cos\alpha + \cos\beta\right)$$

$$F_{300} = \left[\frac{T(r,\alpha)}{r}\right]\sin\alpha + \left[\frac{T(r',\beta)}{r'}\right]\sin\beta - 2a_{30}\left(\cos\alpha + \cos\beta\right)$$

$$F_{120} = \left[\frac{S(r,\alpha)}{r}\right]\sin\alpha + \left[\frac{S(r',\beta)}{r'}\right]\sin\beta - 2a_{12}\left(\cos\alpha + \cos\beta\right)$$

$$F_{400} = \left[\frac{4T(r,\alpha)}{r^2}\right]\sin^2\alpha - \left[\frac{T^2(r,\alpha)}{r}\right] + \left[\frac{4T(r',\beta)}{r'^2}\right]\sin^2\beta - \left[\frac{T^2(r',\beta)}{r'}\right]$$

$$\text{-}8a_{30}\left[\frac{\sin\alpha\cos\alpha}{r} + \frac{\sin\beta\cos\beta}{r'}\right] - 8a_{40}\left(\cos\alpha + \cos\beta\right) + 4a_{20}^2\left[\frac{1}{r} + \frac{1}{r'}\right]$$

$$F_{220} = \left[\frac{2S(r,\alpha)}{r^2}\right]\sin^2\alpha + \left[\frac{2S(r',\beta)}{r'^2}\right]\sin^2\beta - \left[\frac{T(r,\alpha)S(r,\alpha)}{r}\right] - \left[\frac{T(r',\beta)S(r',\beta)}{r'}\right]$$

$$+4a_{20}a_{02}\left[\frac{1}{r} + \frac{1}{r'}\right] - 4a_{22}\left(\cos\alpha + \cos\beta\right) - 4a_{12}\left[\frac{\sin\alpha\cos\alpha}{r} + \frac{\sin\beta\cos\beta}{r'}\right]$$

$$F_{040} = 4a_{02}^2\left[\frac{1}{r} + \frac{1}{r'}\right] - 8a_{04}\left(\cos\alpha + \cos\beta\right) - \left[\frac{S^2(r,\alpha)}{r}\right] - \left[\frac{S^2(r',\beta)}{r'}\right]$$

$$F_{011} = -\frac{z_a}{r} - \frac{z_b}{r'}$$

$$F_{111} = -\frac{z_a\sin\alpha}{r^2} - \frac{z_b\sin\beta}{r'^2}$$

$$F_{102} = \frac{{z_a}^2\sin\alpha}{r^2} + \frac{{z_b}^2\sin\beta}{r'^2}$$

$$F_{202} = \left(\frac{z_a}{r}\right)^2\left[\frac{2\sin^2\alpha}{r} - T\left(r,\alpha\right)\right] + \left(\frac{z_b}{r'}\right)^2\left[\frac{2\sin^2\beta}{r'} - T(r',\beta)\right]$$

$$F_{211} = \frac{z_a}{r^2}\left[T(r,\alpha) - \frac{2\sin^2\alpha}{r}\right] + \frac{z_b}{r'^2}\left[T(r',\beta) - \frac{2\sin^2\beta}{r'}\right]$$

where $T(r,\alpha) = \dfrac{\cos^2\alpha}{r} - 2a_{20}\cos\alpha$ and $S(r,\alpha) = \dfrac{1}{r} - 2a_{02}\cos\alpha$

and analogous expressions for $T(r',\beta)$ and $S(r',\beta)$

Now, let us apply Fermat's principle to the central ray, that is the ray which passes via the pole of the grating where x=y=z=0. Imposing:

$$\left(\frac{\partial F}{\partial y}\right)_{y=z=0} = 0 \qquad \left(\frac{\partial F}{\partial z}\right)_{y=z=0} = 0$$

we obtain the following expressions:

$F_{100}=0$ → $\sin\alpha + \sin\beta_0 = Nk\lambda$

$F_{011}=0$ → $\dfrac{z_a}{r} = -\dfrac{z_{b0}}{r_0'}$

The first equation, $F_{100}=0$, is the grating equation. For N→0 (no grating) it simply gives the law of reflection in the plane xy: $\alpha=-\beta$.
The second equation, $F_{011}=0$, is the law of magnification in the sagittal direction. Let us remind that z_a ,z_a are the out-of-plane heights at the object and image points. The equation says that in the vertical plane the grating acts as a mirror: the angle of incidence in the sagittal plane, $\arctan\dfrac{z_a}{r}$, is equal to the angle of reflection, $\arctan\dfrac{z_{b0}}{r_0'}$. The sagittal magnification $\dfrac{z_{b0}}{z_b}$ is equal to the ratio $\dfrac{r_0'}{r}$.

The tangential focal distance r'_0 of the Gaussian image is obtained by setting the focusing term F_{200} equal to zero:

$F_{200}=0$ → $\left(\dfrac{\cos^2\alpha}{r} + \dfrac{\cos^2\beta_0}{r_0'}\right) - 2a_{20}(\cos\alpha + \cos\beta_0) = 0$

The three equations: $F_{100}=0$, $F_{011}=0$, $F_{200}=0$, determine the Gaussian image point $B_0(r'_0,\beta_0,z_{b0})$, that is the perfect image point.

While the second order aberration term F_{200} governs the tangential focusing, the second order term F_{020} governs the sagittal focusing.
The sagittal focus condition is obtained by setting F_{020} equal to zero:

$F_{020}=0$ → $\dfrac{1}{r} + \dfrac{1}{r'} - 2a_{02}(\cos\alpha + \cos\beta) = 0$

The associated aberrations are called defocus and astigmatism, respectively.

The two aberrations describe the extent of an image, defocus describes the blurring of the image in the tangential plane, astigmatism pertains to the extent of the image in the sagittal direction. In a soft x-ray monochromator, the tangential plane is usually vertical and the sagittal plane is horizontal. In a storage ring the vertical source size is smaller than the horizontal source size and therefore the dispersion plane is usually chosen to be the vertical plane in order to reach a higher energy resolution.
Actually astigmatism defines the condition in which the tangential and sagittal foci are not coincident, which implies a line image at the tangential focus.

Toroidal mirror example.
Substituting the coefficients a_{20} and a_{02} with their expressions corresponding to a toroidal surface:

$$a_{02} = \frac{1}{2\rho}; \quad a_{20} = \frac{1}{2R}$$

and assuming that $\alpha=-\beta=-\theta$, the tangential and sagittal focusing conditions become the relations already introduced for the tangential and sagittal foci of a toroidal mirror:

$$\left(\frac{1}{r} + \frac{1}{r'}\right)\frac{\cos\vartheta}{2} = \frac{1}{R}$$

$$\left(\frac{1}{r} + \frac{1}{r'}\right)\frac{1}{2\cos\vartheta} = \frac{1}{\rho}$$

While higher order aberrations are usually of less importance than defocus and astigmatism, they can be significant. For example it can be important to minimize the third order aberrations primary coma F_{300} and astigmatic coma F_{120}. The fourth order terms F_{400} F_{220} F_{040} can be regarded as generalized types of spherical aberration.
Other higher order terms contribute to secondary coma, secondary spherical aberrations, field curvature and distortion.

Calculation of ray aberrations

By definition, the central ray AO arrives at the Gaussian image point B_0 , which coordinates (r_0',β_0,z_{b0}) are determined by the equations: $F_{100}=0$**,** $F_{011}=0$**,** $F_{200}=0$. In general, the ray starting from A will arrive at the focal plane at a point B displaced from the Gaussian image point B_0 by the ray aberrations Δy_b and Δz_b .
It can be shown that:

$$\Delta y_b = \frac{r_0'}{\cos\beta_0}\frac{\partial F}{\partial y}$$

$$\Delta z_b = r_0'\frac{\partial F}{\partial z}$$

The meaning of the non-zero differentials $\frac{\partial F}{\partial y} \neq 0, \frac{\partial F}{\partial z} \neq 0$, is that of a change in the appropriate direction cosine of the emergent ray from the Gaussian value. The division by cos β_0 comes from the coordinate transformation to the image plane.
If we substitute the power series expansion of F in the aperture coordinate,
$F = \sum_{ijk} F_{ijk} y^i z^j$, we can calculate the ray aberrations for each aberration type:

$$\Delta y_b^{\ ijk} = \frac{r_0'}{\cos \beta_0} F_{ijk} \ i \ y^{i-1} \ z^j$$

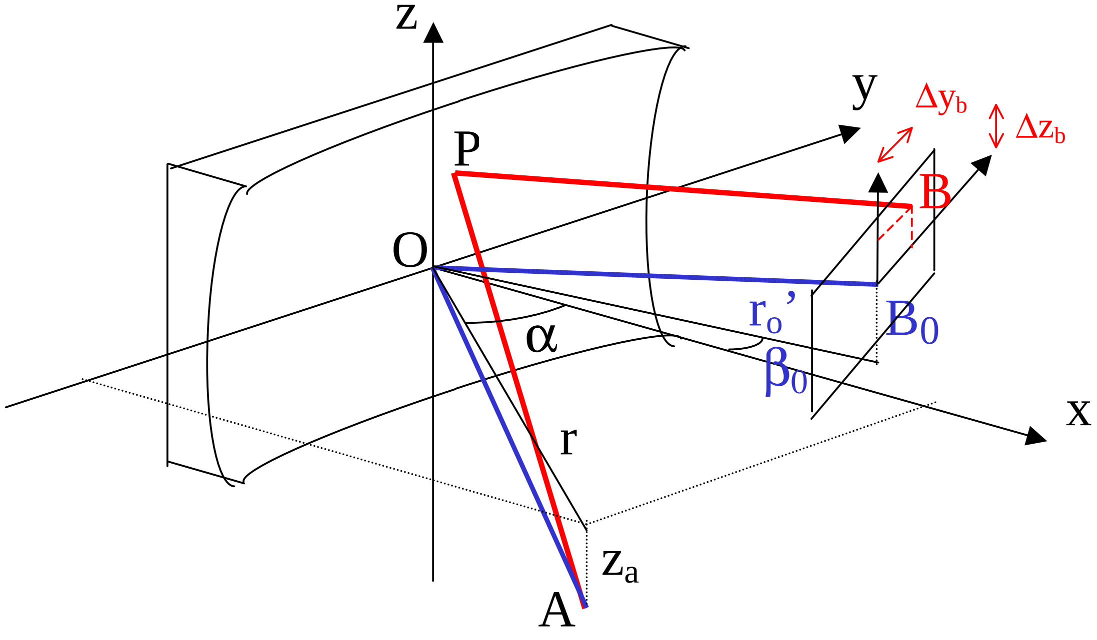
$$\Delta z_b^{\ ijk} = r_0' \ F_{ijk} \ y^i \ j \ z^{j-1}$$

The above equations describe the impact of each aberration term on the final image.
The final ray displacement is given by summing all the individual aberrations:

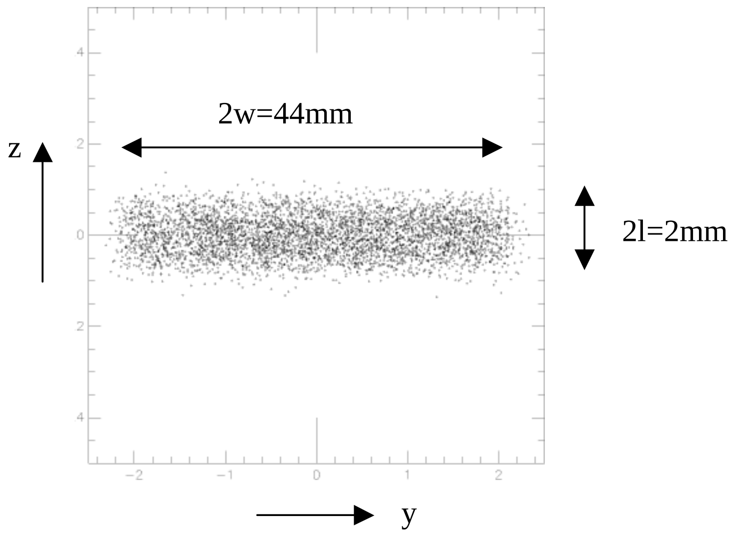
$$\Delta y_b = \sum_{ijk} \Delta y_b^{\ ijk}$$
$$\Delta z_b = \sum_{ijk} \Delta z_b^{\ ijk}$$

Each coefficient F_{ijk} represents a particular form of aberration and is related to the strength of that aberration. An aberration is absent from an image (of a given wavelength in a given order of diffraction) if its associated coefficient F_{ijk} is zero.

Inserting the geometrical parameters a_{ij} corresponding to the grating surface shape, we obtain the F_{ijk} coefficients and with them we can calculate the size of the aberrated image that we can expect from a given (perfectly made) grating. The parameters y and z have to be chosen half the dimension of the photon spot on the grating surface:
y=w (half ruled lenght of the grating); z=l (half ruled width of the grating)
In this way we evaluate the contributions of the rays which are more distant from the pole of the grating.



Example grating footprint:



Let us consider in more details the defocus and coma contributions.
The defocus term is linear in the ruled length of the grating (±w) and gives an error (in the dispersive direction) which is symmetric about the Gaussian image point, corresponding to rays from the top and bottom of the grating :

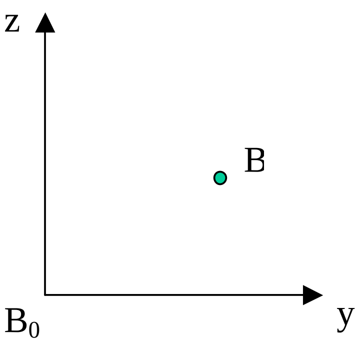
$$\Delta y_b^{\ 200}(\pm w) = \pm \frac{r_0'}{\cos\beta_0} F_{200}\ 2\,w$$
$$\Delta z_b^{\ 200} = 0 \qquad (14)$$

If a monochromator is only limited by defocus, a decrease in the dispersive aperture will cause a corresponding linear decrease in the dispersive error.
In contrast, the coma term is proportional to w^2 giving a dispersive error which only occurs on one side of the Gaussian image point for rays from both the top and the bottom of the grating ($y=\pm w$):

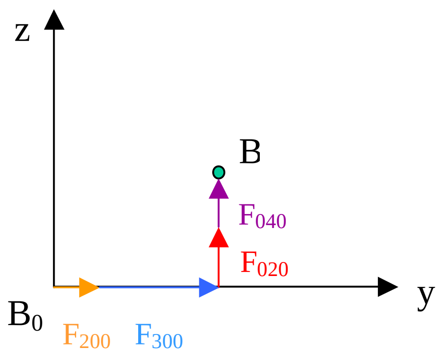
$$\Delta y_b^{\ 300}(\pm w) = \frac{r_0'}{\cos\beta_0} F_{300}\ 3\,w^2$$
$$\Delta z_b^{\ 300} = 0$$

If the monochromator is only limited by coma, a reduction in dispersive aperture by a factor of 2 will cause the error from the most aberrated rays to reduce by a factor of 4.
These conclusions have been reached without the need to consider the form of the substrate. The shape of the substrate will alter the magnitude of the F_{ijk} terms and their wavelength dependence but not their aperture dependence.

Let us compare what can be learnt from a ray trace and from the aberration theory, considering the simple example in the following figure. Ray trace simple tells us that the ray arrives in a certain point, the aberration theory shows that the ray arrives off the perfect focus in the tangential direction because of F_{200}, which is defocus (the image is imaged but not where you want it), and F_{300} which is primary coma. It arrives off the perfect focus in the sagittal direction by F_{020}, which is astigmatism, and F_{040}. Knowing this, the designer can have at least a beginning idea of how to combat the aberrated performance.



Ray tracing



Aberration-based calculations

Calculation of aberration contribution to resolution

Let us calculate the contribution to resolution caused by aberrations.

$$\Delta\lambda = \left(\frac{\partial\lambda}{\partial\beta}\right)_{\alpha=const}\Delta\beta$$

Deriving the grating equation:

$$\left(\frac{\partial\lambda}{\partial\beta}\right)_{\alpha=const} = \frac{\cos\beta}{Nk}$$

we have:

$$\Delta\lambda = \frac{\cos\beta}{Nk}\Delta\beta$$

Substituting:

$$\Delta\beta = \frac{\Delta y_b}{r'}$$

we get:

$$\Delta\lambda = \frac{\cos\beta}{Nk}\frac{\Delta y_b}{r'}$$

We have already seen that:

$$\frac{\partial F}{\partial y} = \frac{\cos\beta}{r'}\Delta y_b$$

Then:

$$\Delta\lambda = \frac{1}{Nk}\frac{\partial F}{\partial y}$$

and therefore:

$$\Delta\lambda = \frac{1}{Nk}\sum_{ijk} F_{ijk}\ i\ y^{i-1}\ z^{j}$$

Inserting the geometrical parameters a_{ij} corresponding to the grating surface shape we obtain the F_{ijk} coefficients and with them we can calculate the aberration-limited wavelength resolution that we can expect from a given (perfectly made) grating.

In conclusion:

We have seen that the condition for having a perfect focus is that the two partial derivatives must zero for each pair (y,z). This means that all the coefficients F_{ijk} (ijk#000) must be zero. Non-zero values for the coefficients F_{ijk} lead to displacements of the rays arriving in the image plane from the ideal Gaussian image point. We have found the expressions for these rays displacements and the corresponding contributions to energy resolution. In this way the impact on the imaging and energy resolution properties of a given grating can be evaluated. By a proper choice of the grating shape, groove density, object and image distances, the sum of the aberrations may be reduced to a minimum.

These notes have been taken from the following references:

D.Attwood, “Soft x-rays and extreme ultraviolet radiation”, Cambridge University Press, 1999

D.Attwood, “Challenges for utilization of the new Synchrotron facilities”, Nucl. Instr.and Meth.A291, 1-7, 1990

B.W.Batterman and D.H.Bilderback, “X-Ray Monocromators and Mirrors” in “Handbook on Synchrotron Radiation”, Vol.3, G.S.Brown and D.E.Moncton, Editors, North Holland, 1991, chapter 4

W.Gudat and C.Kunz, “Instrumentation for Spectroscopy and Other Applications”, in “Syncrotron Radiation”, “Topics in Current Physics”, Vol.10, C.Kunz, Editor, Springer-Verlag, 1979, chapter 3

M.Howells, “Gratings and monochromators”, Section 4.3 in “X-Ray Data Booklet”, Lawrence Berkeley National Laboratory, Berkeley, 2001

M.Howells, “Vacuum Ultra Violet Monochromators”, Nucl. Instrum. and Meth. 172, 123-131, 1980

M.C. Hutley, “Diffraction Gratings”, Academic Press, 1982

R.L. Johnson, “Grating Monochromators and Optics for the VUV and Soft-X-Ray Region” in “Handbook on Synchrotron Radiation”, Vol.1, E.E.Koch, Editor, North Holland, 1983, chapter 3

C.Kunz and J.Voss, “Scientific progress and improvement of optics in the VUV range”, Rev. Sci.Instrum. 66 (2), 1995

G.Margaritondo, Y.Hwu and G.Tromba, “Synchrotron light: from basics to coherence and coherence related applications”, in: “Syncrotron Radiation:Fundamentals, Methodologies and Applications”, Conference Proceedings, Vol. 82, S.Mobilio and G.Vlaic, Editors, S.Margherita di Pula, 2001

W.R.McKinney, M.Howells, H.Padmore, “Aberration analysis calculations for Synchrotron radiation beamline design”, in “Gratings and Grating Monochromators for Synchrotron Radiation”, W.R. McKinney, C. A. Palmer, Eds., Proc. SPIE Vol.3150, 1997

A.G. Michette, “Optical Systems for X Rays”, Plenum Press, 1986

W.B.Peatman, “Gratings, mirrors and slits”, Gordon and Breach Science Publishers, 1997

J.B. West and H.A. Padmore, “Optical Engineering” in “Handbook on Synchrotron Radiation”, Vol.2, G.V.Marr, Editor, North Holland, 1987, chapter 2

G.P.Williams, “Monocromator Systems”, in “Synchrotron Radiation Research: Advances in Surface and Interface Science”,Vol.2, R.Z.Bachrach, Editor, Plenum Press, 1992, chapter 9