Physical Optics

Lecture notes on chapters 1, 3, 5, 6 by Pierre Chavel and Anthony Berdeu. This Physical Optics course is being offered in English for the first time in 2018-2019. It is part of the syllabus for the first year (M1) of the "Optics, Image, Vision, Multimedia" Master Course for the Université Jean Monnet de Saint-Etienne and Institut d'Optique – Graduate School students. It is a requisite for the "Optics in Surface and Interface Science and Engineering" (SISE) and "Advanced Imaging and Material Appearance metrology and modelling" (AIMA) study tracks.

With the inauguration, in the fall of 2018, of the "Surface Light Engineering Health and Society" (SLEIGHT) joint graduate school (École universitaire de Recherche MANUTECH SLEIGHT), it is part of the SLEIGHT course offer.

Contents

INTRODUCTION TO THE PHYSICAL OPTICS COURSE

The goal of this course is to offer a description of light propagation in free space and in optical systems, in particular optical imaging systems, which is both more comprehensive and more accurate than that provided by geometrical optics, at least when the latter is restricted to tracing light rays that are considered independent of each other. Indeed, the mere laws of reflection and refraction of light rays on a locally plane interface do not fully account for the propagation of light, except in the limiting case of vanishingly small wavelengths. Physical Optics rests mainly on diffraction as the main concept to describe the propagation of light in homogeneous media, optical refractive and reflective components, including wavelength-scale structures that interact with light.

The course has been designed to offer practical tools useful for experimental work, for the design of optical components and instruments, as well as for modelling purposes. At the same time, it directly rests on solid, if not always rigorous, physical principles. It is therefore a relevant part of the curriculum both for physicists and for engineers manipulating light. It is composed of

- 30h lectures including exercises (10 times 3h, taught by Corinne Fournier for Chapters 2 and 4, Pierre Chavel and Anthony Berdeu for the other chapters)
- 16h practical training on Matlab (4 times 4h, Corinne Fournier, Emmanuel Marin).

It is organized in six chapters:

1. The Physics of Light Propagation: a reminder on diffraction, including electromagnetic diffraction, “scalar” diffraction, simple analytical tools for calculating diffraction and for inspiring an intuition about how diffraction behaves in practical situations. The basic concept here is the complex amplitude of light, and how it propagates. Coherence is assumed throughout the chapter. One parameter of major interest to describe light propagation is spatial frequency, the spatial analog of time frequency for time signals. The special propagation properties of pure spatial frequencies both in free space and in imaging systems make this concept an indispensable tool for this course, in relation with the (prerequisite) concepts of Fourier Transform and Fourier Series.
2. The Sampling of Optical Signals: to effectively model light propagation situations, care must be taken to properly sample the wavefront: how many samples (“pixels”), and what distance between pixels are required to avoid modelling artefacts, how does sampling in real space relate to sampling in the plane of spatial frequencies.
3. Imaging in the Approximation of Spatial Frequency Filtering: a naïve requirement for a “perfect” optical imaging instrument would be that the image be in some sense “identical” to the object. In particular, the image of a “point”, in the geometrical sense, should likewise be a point. In geometrical optics, a stigmatic instrument satisfies those conditions. How does diffraction modify this situation? What are the relevant physical quantities to describe an image, how do they relate to those that describe the object? The answer is strongly dependent of the source properties that determine the coherence of light, i.e. the possibility for light beams emerging from two points in the object to interfere with each other. Spatial frequency filtering is a powerful heuristic tool, based on the concept of spatial frequency, to appropriately account for the resolution limit of an optical instrument. It describes the relation between an object and its image as a convolution, in a way that depends on coherence.
4. Holography: in coherent illumination, an object is characterized by its influence on the electric field of an illuminating beam, which may be transmitted or reflected, or both. The full information on optical behavior of the object is contained in the modulus and phase of the transmitted or reflected electromagnetic field as diffracted by the object. Holography is the main technique to record that field on photographic film or on a opto-electronic sensor (typically a 2D array of photodiodes).

Computer generated holography allows to create an otherwise unavailable object from its diffracted field computed and fabricated by computer controlled processes.

5. Diffractive optics: an extension of computer generated holography, diffractive optics creates an arbitrary wavefront from microstructures, or even nanostructures. It offers record low volumes and shows a distinctive and possibly useful spectral behavior. Among such “diffractive optical elements”, the simplest case, and probably still the most important, is the diffraction grating, which from an incoming plane wave diffracts a set of strongly chromatic plane waves. The chapter covers a subset of the growing family of diffractive structures: spherical wave generation by diffractive lenses, “Fresnel zone plates”, or arbitrary wavefronts. This field is currently merging into the broader field of nanophotonics, where the behavior of light in structures smaller than the wavelength involves phenomena that are not available otherwise.
6. Speckle: while optical components must accurately control the phase profile of their diffracted wavefronts, in many cases plane or spherical, most objects and translucent media show a completely opposite behavior and randomly perturb the phase of any incoming wavefront: light is said to be scattered *(in French diffusée)*. A coherent beam of light scattered by such an object shows a specific phenomenon known as “speckle”.

References:

Standard textbooks covering mostly the same content as this course are many. They include J.W. Goodman’s classic Introduction to Fourier Optics. The latest edition is the fourth, published by W.H. Freeman and Co, 2017, ISBN 978-1319119164. The first edition (1968) was translated into French by Christian Durou and José Philippe Pérez, 1972, éditions Masson, Paris. Pioneers include A. Maréchal and M. Françon’s Diffraction, structure des images, second edition published by éditions Masson, Paris 1970, while the real first book on the subject is P.M. Duffieux’ l’intégrale de Fourier et ses applications à l’Optique, chez l’auteur, Besançon, 1946.

Chapter 1 – The Physics of Light Propagation

1.1 – Light diffraction from a known distribution of amplitude on a plane:

1.1.1 – Electromagnetic diffraction

In this course, only macroscopic fields are considered, i.e. fields averaged over a large number of atoms. Since the wavelength of visible light is on the order of hundredths of nanometers and the typical size of an atom is smaller than one nanometer, it is a generally acceptable approximation to assimilate the electric and magnetic fields, $\mathbf{E}$ and $\mathbf{B}$, with their spatial average over some atoms or molecules (care should be taken, in fact, for large polymer molecules). The average of bound charges and local currents then allows to define the $\mathbf{D}$ and $\mathbf{H}$ fields. The macroscopic form of Maxwell's equations, which are assumed to be known to the reader, relate the $\mathbf{E}$, $\mathbf{B}$, $\mathbf{D}$ and $\mathbf{H}$ fields and the electric permittivity *ε*, the magnetic susceptibility *μ* and the macroscopic charge and current densities ρ_c and $\mathbf{j_c}$. We shall consider here only isotropic materials, where *ε* and *μ* are scalars (the square bracket will therefore be omitted). Moreover, we consider "non magnetic media" (extremely few media are magnetic at optical frequencies), therefore $\mu = \mu_o$. Light is emitted by time varying charge and current densities, but we shall be interested in the propagation of light outside the light sources, therefore in regions of space where they both vanish. Classical optics is completely described by the macroscopic Maxwell's equations. In nonlinear optics – a situation that arises with intense local fields – the relation between $\mathbf{E}$ and $\mathbf{D}$ is nonlinear, which can be expressed by saying that *ε* depends on the field. This course considers linear optics, where *ε* is therefore a constant field, which at each point in space fully characterizes the local behavior of the medium. The refractive index is defined here as the positive solution of $n^2 = \varepsilon$. Complex values of *ε* and *n* represent absorption or, possibly, amplification in a medium. The real part of *n* is always positive.

Notations: here and in the following, boldface characters denote vector fields, vector $\mathbf{R}$ designates the running point in 3D space, *t* designates time. Lower case space vectors such as $\mathbf{r}$, $\boldsymbol{\rho}$ designate 2D vector in an identified plane. At angular frequency ω, the frequency is ν. With *c* the speed ("celerity") of light in a vacuum, the vacuum wavelength is $\lambda = c/\nu$, the wavelength in a medium of refractive index *n* is $\lambda_n = \lambda/n$. The inverse of the wavelength, called "wave number", is denoted by *σ*.

In optics, fields vary at frequencies in the range of hundredths of terahertz: considering the boundaries of the visible spectrum, vacuum wavelength 400 nm corresponds to 750 THz and vacuum wavelength 750 nm, to 400 THz. However, electromagnetic phenomena are essentially identical to those in the visible over a broad range of frequencies, typically any frequency where no detector is able to follow the time variation of the fields, but only time averages are accessible. Therefore, the domain of optics can be extended from the thermal infrared (vacuum wavelength 10 or even 20 µm, frequency 300 or even 150 GHz) to the X or even γ ray frequency ranges.

To solve an electromagnetic problem is to find the full electromagnetic field ($\mathbf{E}$, $\mathbf{B}$, $\mathbf{D}$ and $\mathbf{H}$) at every point $\mathbf{R}$ in space and every time *t* given the geometry, i.e. the distribution of *ε*, and given complete information about the sources and their time dependencies. When the time variation of fields is restricted to optical frequencies, this is called an electromagnetic diffraction problem. Nevertheless, following Sommerfeld, it is customary to speak about diffraction when the difference between the description of light propagation in the terms of geometric optics rays and the electromagnetic solution are conspicuous. That is what typically happens in the vicinity of refractive index discontinuities, which is particularly obvious when the refractive index changes over scales comparable to a wavelength. The word “scattering” is usually associated with diffraction by rapidly varying random variations of the refractive index.

As we shall see, some diffraction problems can be treated in an approximate but acceptable way without explicitly solving for the full electromagnetic field. Most of this course will indeed relate to situations where those approximate solutions are valid. Because they are simpler, they can also help gain intuition about diffraction. “Electromagnetic diffraction” is therefore the more general case, but “scalar diffraction”, defined in the next section, is a heuristically powerful special case.

1.1.2 – Scalar optics

Throughout this course, we consider scalar optics, When the dependency in space and time of all components of a vector field are proportional, one scalar field is sufficient to represent them all, to within a multiplicative constant. That happens in electromagnetism whenever the following three conditions are all met:

- the refractive index is a scalar: there is no birefringence, optical activity (sometimes called circular birefringence) or other anisotropic effect in matter;
- the beam aperture is small. As a typical counter-example, consider two plane waves interfering at right angles: the three components of the total $\mathbf{E}$ field are easily expressed and seen to be non-proportional to each other. We exclude that case;
- the incidence angles on the various diopters are small enough for the Fresnel coefficients to be identical for all polarization directions.

In this chapter, we only consider monochromatic fields, with a time dependence denoted in the complex notation as $\exp(-i\omega t) = \exp(-2i\pi\nu t)$, using a complex notation under which the real physical quantities are the real part of their complex extension, mathematically known as the “analytical signal” associated to the real quantity. Note that throughout this course, the minus sign has been selected in front of the time dependence. We shall denote by $U(\mathbf{R},t)$ the scalar light amplitude, a complex quantity which we shall call “the complex amplitude”. We call $\mathrm{e}(\mathbf{R},t) = |U(\mathbf{R},t)|^2$ the “instantaneous intensity” at point $\mathbf{R}$ and time *t*, and $\mathfrak{E}(\mathbf{R},t) = \left\langle |U(\mathbf{R},t)|^2 \right\rangle$ “the observed intensity”, which is the time average of the former over many periods of the complex amplitude (in most cases, the word “observed” will be omitted).

The link between those quantities and photometry determines the multiplicative constant in the complex amplitude. One may therefore call $\mathfrak{E}$ “intensity” or “illumination”: the difference will affect only the constant. For example, one may choose to claim that $\mathfrak{E}$ is the detected illumination for a detector placed perpendicular to the Poynting vector.

1.1.3 – Plane wave expansion

One simple and general way to deal with scalar diffraction is to consider light propagation in a volume devoid of any obstacle, i.e. in a portion of space filled with a homogeneous medium which, in practice, is often air or a vacuum – we shall consider the case of a vacuum. Diffracting slits and other structures **inside** that volume are therefore excluded – a strong limitation indeed. However, if the complex amplitude $U(\mathbf{R},t)$ is known on a closed surface whose interior is free space, it is indeed possible to find it everywhere within that volume using a plane wave expansion.

Indeed, plane waves are solutions to Maxwell’s equations in free space. In a linear medium, any sum of solutions is a solution, and conversely, any solution to Maxwell’s equation that has a Fourier transform is an integral of plane waves, as is obvious from the very definition of the Fourier transform.

Such an approach is valid for any closed surface in electromagnetic optics. We shall consider it only for a plane and in the case of scalar optics. Extension of a closed surface to a plane can mathematically be shown to be possible, for example by starting from a sphere, and letting its center move to infinity in a direction perpendicular to the plane to be considered: it is valid as long as no source or other non-homogeneity is encountered in that half space. All sources or obstacles in the problem are therefore restricted to one side of the plane.

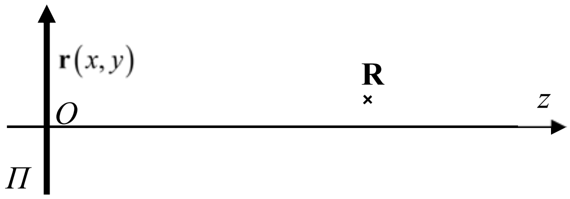
We therefore seek to solve the following problem: space is divided into two half spaces by plane *Π*. The complex amplitude U, monochromatic of frequency *ν*, is assumed to be known at every point $\mathbf{R}(\mathbf{r},0)$ on that plane. In half-space $z > 0$, U has a Fourier transform with respect to variable $\mathbf{r}$. In that half-space, a classic vector calculation on Maxwell's equations shows that all vector fields and their components, and therefore also the complex amplitude, obey the Helmholtz equation:

$$\Delta U + k_o^2 U = 0 \text{, with } k_o = \omega/c = 2\pi/\lambda = 2\pi\sigma . \tag{1.1}$$

All sources and obstacles being on the negative z half-space, one seeks solutions with only "entering" plane waves, i.e. plane waves that propagate in the positive z direction or decrease exponentially in the positive z direction:

$$U_{plane}(\mathbf{R},t) = U_o \exp(i\mathbf{k}\bullet\mathbf{R} - i\omega t) = U_o \exp 2i\pi(\boldsymbol{\sigma}\bullet\mathbf{R} - \nu t) \tag{1.2}$$

Here, $\boldsymbol{\Sigma}$ is the 3D spatial frequency. Of more common use is its 2D counterpart $\boldsymbol{\mu}$. It is easy to check that such a plane wave obeys the Helmholtz equation if and only if $|\mathbf{k}| = k_o$, and the propagation or exponential decay in the positive z direction condition corresponds to a positive real part of the wave vector, and, for evanescent waves, with a positive imaginary part of the wave vector $\mathbf{k}$. Throughout the rest of this chapter, the time dependency, being always the same, will be omitted.



We can now solve the problem at hand: given $U(\mathbf{r},0)$, express $U(\mathbf{R})$ throughout the positive z half-space. Let us firstly express the definition of the 2D Fourier transform (FT) of $U(\mathbf{r},0)$:

$$U(\mathbf{r},0) = \iint_{\mathbb{R}^2} \tilde{U}(\boldsymbol{\mu}) \exp(2i\pi\mathbf{r}\bullet\boldsymbol{\mu})\, d^2\boldsymbol{\mu} \tag{1.3}$$

Based on the linearity of Helmholtz's equation and the unicity of the solution (admitted here without a demonstration), we can identify the solution sought to the integral of all "entering" plane waves that have the proper expression in plane *Π:* $\exp(2i\pi\mathbf{r}\bullet\boldsymbol{\mu})$ is the expression for $z = 0$ of the "entering" plane wave $\exp(2i\pi\boldsymbol{\sigma}\bullet\mathbf{R})$, solution of the Helmholtz equation, with

$$\boldsymbol{\sigma}^2 = \boldsymbol{\mu}^2 + \mu_z^2 = \frac{\nu^2}{c^2} = \frac{1}{\lambda^2} \text{, whence}$$

$$U(\mathbf{R}) = \iint_{\mathbb{R}^2} \tilde{U}(\boldsymbol{\mu}) \exp 2i\pi\left(\mathbf{r}\bullet\boldsymbol{\mu} + z\sqrt{\boldsymbol{\sigma}^2 - \boldsymbol{\mu}^2}\right) d^2\boldsymbol{\mu} \tag{1.4}$$

Here, if the radicand is negative, the imaginary square root with a positive imaginary part should be selected. *(Exercise 1.1: the Fresnel approximation. That exercise is part of the course.)*

1.1.4 – The Huygens-Fresnel "principle"

Consider an arbitrary distribution of complex amplitude $U(\mathbf{r},0)$. If all spatial frequencies are much smaller than σ, express Eqn (1.4) can be simplified using the Fresnel approximation on the spatial frequencies (in contrast with Exercise 1.1, where the Fresnel approximation was derived on space variables). The result is:

$$U(\mathbf{R}) = \exp\frac{2i\pi z}{\lambda}\iint_{\mathbb{R}^2}\tilde{U}(\boldsymbol{\mu})\exp\left(-i\pi\lambda z\boldsymbol{\mu}^2\right)\exp\left(2i\pi\mathbf{r}\bullet\boldsymbol{\mu}\right)d^2\boldsymbol{\mu}\,, \qquad (1.4\text{bis})$$

which is identified as the inverse FT of a product, i.e. the convolution of the two inverse FTs. The former is known to be $U(\mathbf{r},0)$. The latter is obtained from the FT of a quadratic phase, a mathematical result that we shall admit here: if α is a non-negative real and β an arbitrary real, then $\mathrm{FT}\left[\exp-\pi(\alpha+i\beta)x^2\right] = \frac{1}{\sqrt{\alpha+i\beta}}\exp\frac{-\pi\mu^2}{\alpha+i\beta}$, where μ is the conjugate variable to x and where the complex square root designates a number with a positive real part. In particular, $\mathrm{FT}\left[\exp i\pi x^2\right] = \exp\left(-i\pi\mu^2 + i\,\pi/4\right)$.

From there,

$$U(\mathbf{R}) = \frac{-i\exp(2i\pi z/\lambda)}{\lambda z}U(\mathbf{r},0) * \exp\left(i\pi\mathbf{r}^2/\lambda z\right) \qquad (1.5)$$

Where the convolution is applied over variables x and y. In Eqn 1.5, it appears that every point x, y in plane Π produces a spherical wave that reaches plane z. That is the analytical expression of the Huygens-Fresnel "principle", which since the time of Maxwell (fifty years after Fresnel, nearly two centuries after Huygens) is no more a principle, but a theorem, expressed here under the Fresnel approximation: "each point in the diffracting plane behaves as a secondary source of light emitting a spherical wave proportional to the incoming wave, and in quadrature ahead of phase." As can be seen, it applies to any arbitrary surface in free space (we expressed it for a plane, but it can be extended to other surface shapes).

1.1.5 – Complex amplitude transmittance:

While Eqn (1.4) is simple and general, it suffers from an obvious major drawback: how can $U(\mathbf{r},0)$ be known? The only general answer is to solve the Maxwell's equations in the negative z half-space, were sources and diffractive obstacles are located – obviously a much trickier question! Nevertheless, that equation is fairly useful and practical when it is a valid approximation to assume that the expression of $U(\mathbf{r},0)$ is simply provided by geometrical optics: for example, if plane Π "contains" (or, practically, is located in the immediate vicinity behind) a slit illuminated in normal incidence by a plane wave of complex amplitude U_o, it is an intuitively appealing approximation to declare that the complex amplitude is U_o just behind the slit, and zero elsewhere in plane Π. For a sufficiently broad slit, the approximation gives an adequate description of diffraction by the slit, as is usually presented in elementary courses. That approximation is known as the "given field approximation", meaning that the field in plane Π is assumed to be given by geometrical optics. In opposition to that case, in nanophotonics, slits may behave quite differently.

More generally, that approximation is convenient way to express the complex amplitude immediately after a diffracting obstacle in the form:

$$U(\mathbf{r},0)=t(\mathbf{r})U_{illum}(\mathbf{r},0) \quad (1.6)$$

In that case, if an object placed in the negative z half-space but close to plane *Π* is illuminated by wave $U_{illum}(\mathbf{r},0)$, one considers that quantity $t(\mathbf{r})$, called "complex amplitude transmittance" (or in short transmittance, or sometimes just transmission) **is independent of the illuminating wave and can be empirically described by geometrical optics**. While that assumption does not generally hold and is obviously wrong for thick objects, where shadowing effects and more generally diffraction occurs within the object itself. Diffracting objects for which Eqn (1.6) applies with $t(\mathbf{r})$ independent of the illuminating wave are called **"thin objects"**. *(Exercise 1.2, the Talbot effect).*

1.2 Fraunhofer diffraction and Fresnel diffraction

1.2.1 – Fraunhofer diffraction, "far field diffraction":

Simply considering that Eqn (1.3) is an integral of plane waves shows that diffraction, from its physical principles, forms the Fourier transform of the diffracting complex amplitude "at infinity" (each plane wave corresponds to one point at infinity).

Since one essential property of a lens is to optically conjugate a plane wave with one point in its focal plane in the image space, one therefore reaches the conclusion that a lens placed just behind plane *Π* will form the Fourier transform of the initial complex amplitude distribution in the focal plane. *(Exercise 3: the lens as a Fourier transformer.)*

Fraunhofer diffraction is also called "far field diffraction" or "diffraction at infinity". Note that the "far field" here refers to "infinity" or to a distance large enough that the pattern observed is essentially the same as would be seen at infinity. In many cases, it is more convenient to use a lens and to visualize the Fourier transform in its focal plane, whence the other denomination "diffraction around a focus".

1.2.2 – Fresnel diffraction:

Consider a thin object normally illuminated by a plane wave of complex amplitude U_o. The complex amplitude just behind the object is therefore $U(\mathbf{r},0)=U_o t(\mathbf{r})$. Equation (1.5) gives the expression of diffraction at an arbitrary distance, usually known as « Fresnel diffraction ». *Question 4 of Exercise 3 shows that Fraunhofer diffraction is really a particular case of Fresnel diffraction.*

A note of caution on vocabulary is on order here. In the study of diffraction "near field" should not be used as the opposite to "far field". "Far field" diffraction is a usual denomination for Fraunhofer diffraction, where the FT of a given object can be observed. Current instrumentation allows to observe, using a so-called "near field optical microscope", the field so close to a diffracting object that is evanescent waves are sensed. That cannot be achieved using a lens, but by scanning the immediate vicinity of the object (nanometric distances) using a finely pointed tip. Those instruments have now been available for over thirty years. Before that time, it was customary to oppose "near field" and "far field" in diffraction phenomena, a practice which nowadays would be confusing and should be avoided: the denomination "near field" optics should be left for instruments observing evanescent waves.

1.3 – A summary:

This summary covers the lecture itself and the exercises, including in fact exercise 3.1.

- For a scalar field satisfying the Helmholtz equation, a general monochromatic solution given the field in the *z=0* plane in the homogeneous *z>0* half-space, all sources and obstacles being in the *z<0* half space, is the "plane wave expansion" of Eqn (1.4):

$$U(\mathbf{R}) = \iint_{\mathbb{R}^2} \tilde{U}(\boldsymbol{\mu}) \exp 2i\pi \left(\mathbf{r} \bullet \boldsymbol{\mu} + z\sqrt{\boldsymbol{\sigma}^2 - \boldsymbol{\mu}^2} \right) d^2\boldsymbol{\mu} \tag{1.4}$$

- It is however impractical to solve that equation. In the domain of validity of the order 2 expansion of the square root in the exponent, the propagation law can be written as the following plane wave expansion,

$$U(\mathbf{R}) = \exp\frac{2i\pi z}{\lambda} \iint_{\mathbb{R}^2} \tilde{U}(\boldsymbol{\mu}) \exp\left(-i\pi\lambda z \boldsymbol{\mu}^2\right) \exp\left(2i\pi \mathbf{r} \bullet \boldsymbol{\mu}\right) d^2\boldsymbol{\mu} \quad , \tag{1.4bis}$$

or equivalently as the following convolution in the space domain,

$$U(\mathbf{R}) = \frac{-i \exp(2i\pi z/\lambda)}{\lambda z} U(\mathbf{r}, 0) * \exp\left(i\pi \mathbf{r}^2/\lambda z\right) \tag{1.5}$$

which can be expressed heuristically in terms of the Huygens-Fresnel “principle”: each point in the *z=0* plane emits a spherical “wavelet” proportional to the field it has received but in phase quadrature, and all wavelets interfere.

- In that context, a spherical wave is expressed as a quadratic phase, as a result of the aforementioned second order expansion, called in this case the “Fresnel approximation”. Specifically, a spherical wave centered at $(0,0,z_o)$ produces in plane z=0 a complex amplitude

$$A_{sph,Fresnel}(\mathbf{r}, 0) = A_o \exp\frac{-i\pi \mathbf{r}^2}{\lambda z_o} \tag{1.7}$$

From Eqn (1.7), expressions at an arbitrary point plane *z* for a spherical wave centered at an arbitrary point $\boldsymbol{R}_o$ are easily retrieved.

- In particular, a spherical wave such as that of Eqn (1.7) illuminating an object of transmittance $t(\boldsymbol{r})$ and converging at a distance z_o of the object plane produces in plane $z=z_o$ the Fraunhofer diffraction pattern of that object, expressed as a consequence of Eqn (1.5) by the Fourier transform of the object:

$$A_{Fraunh}(\boldsymbol{\rho}, z_o) = \frac{-iA_o \exp ikz_o}{\lambda z_o} \tilde{t}\left(\frac{\boldsymbol{\rho}}{\lambda z_o}\right) \exp\frac{i\pi \boldsymbol{\rho}^2}{\lambda z_o} \tag{1.8}$$

In that expression, the first phase term obviously accounts for the phase delay along the axis. The last term is a spherical phase that is not observed on a screen, but participates in the diffraction phenomenon.

- In any plane other than $z=z_o$, Eqn (1.5) has no easy expression as a Fourier transform and the pattern is known as Fresnel diffraction.
- In a few situations, it is possible to cancel the spherical phase term in Eqn (1.8), which makes life easier for the calculation but not necessarily in practical implementations (because of field limitations and aberrations). The best known case is the “f-f” configuration where the object is illuminated by a plane wave in the front focal plane of a lens and the Fourier transform is observed in the back focal plane.

Chapter 3 – Spatial Frequency Filtering

It is the purpose of this chapter to provide a powerful, albeit approximate, tool for expressing the illuminance in an image, i.e. the amount of light that reaches point **r'** in the image plane as a function of the amount of light coming from the object. In an ideal imaging system, there would be a one to one correspondence between an object point **r** and its image **r'**, with a scale factor which is simply the magnification. However, imaging is not perfect, therefore light from several object points **r** gets scrambled into point **r'**. We want to address, using the appropriate physical quantities, the relation between the whole image, a function of **r'**, and the whole object, a function of **r**. As we shall see, those quantities depend on coherence.

Objects are usually volumes, or curved surfaces in a volume, and it is common for images to be plane because they are detected on planar sensors. For the sake of simplification, only plane objects will be considered. In the first section, the image illuminance is directly derived from light diffraction and from the behavior of diffracting objects and lenses, based on Chapter 1. The result is a tool of practical interest: imaging in spatially coherent light as a spatial frequency filtering process has applications in microscopy and imaging under laser illumination. Section 2 is devoted to introducing a more global viewpoint, that of linear filtering, which is common to many branches of theoretical and applied physics, including electrical engineering, and is a major concept for understanding imaging. Section 3 uses that viewpoint to describe imaging in spatially incoherent light as a spatial frequency filtering process, which corresponds to by far the most common situation.

3.1 – Coherent imaging, spatial frequency filtering in coherent light (a reminder)

This lecture follows, with some intentional overlap, on material taught at Palaiseau by H. Benisty and at the beginning of M1 SISE by N. Destouches.

Let us consider a thin object illuminated by a spherical wave. Figure 3.1 directly follows from Chap 1, Eqn 1.6, which expresses the complex amplitude in plane *z=-d* immediately behind a thin object, and from the expression of a spherical wave in Exercise 1.1. If the (complex amplitude) transmission of that object is *t(**r**)*, and the spherical wave converges at the coordinates origin *O*, an algebraic distance - *d* from the object center, then the complex amplitude leaving the object is

$$U(\mathbf{r},d) = \exp\frac{i\pi\mathbf{r}^2}{\lambda d} \quad U_o t(\mathbf{r}) \tag{3.1}$$

with notations directly deriving from Chap 1 and U_o the complex amplitude at the center *(0,0,d)* of the object. Note that *d* is a negative quantity.

Because the object is illuminated by a spherical wave, Fraunhofer diffraction arises at plane *z=0* to yield a complex amplitude

$$U(\boldsymbol{\rho},0_-) = \frac{iU_o}{\lambda d}\exp\frac{2i\pi}{\lambda}\left(-d-\frac{\boldsymbol{\rho}^2}{2d}\right)\tilde{t}\left(\frac{-\boldsymbol{\rho}}{\lambda d}\right). \tag{3.2}$$

The coordinates in the Fraunhofer diffraction plane *z=0* are commonly denoted by $\boldsymbol{\rho}(\xi,\eta)$. As can be seen, spatial frequency µ comes to a focus at point $\boldsymbol{\rho} = -\boldsymbol{\mu}\lambda d$. The minus sign below the 0 on the left hand sign will be explained shortly.

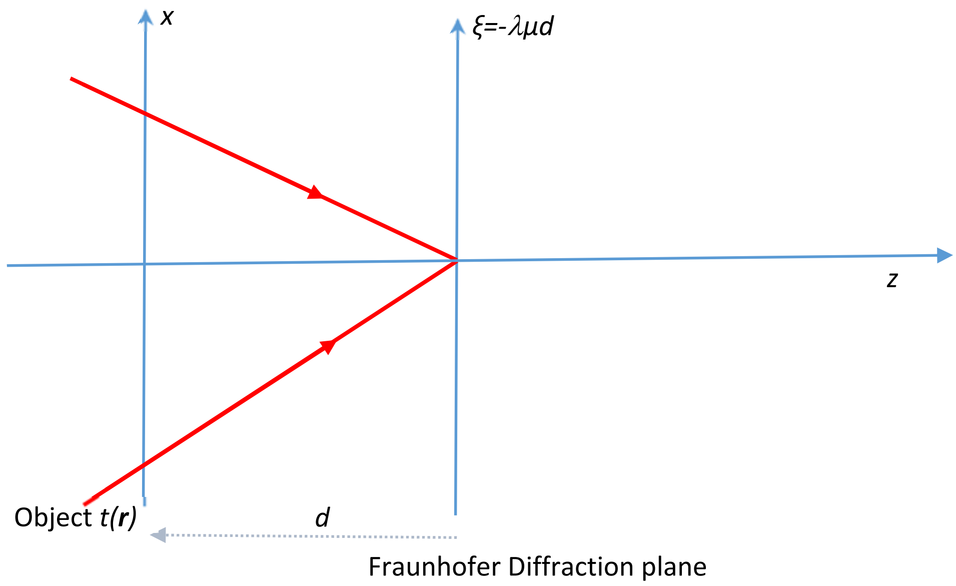
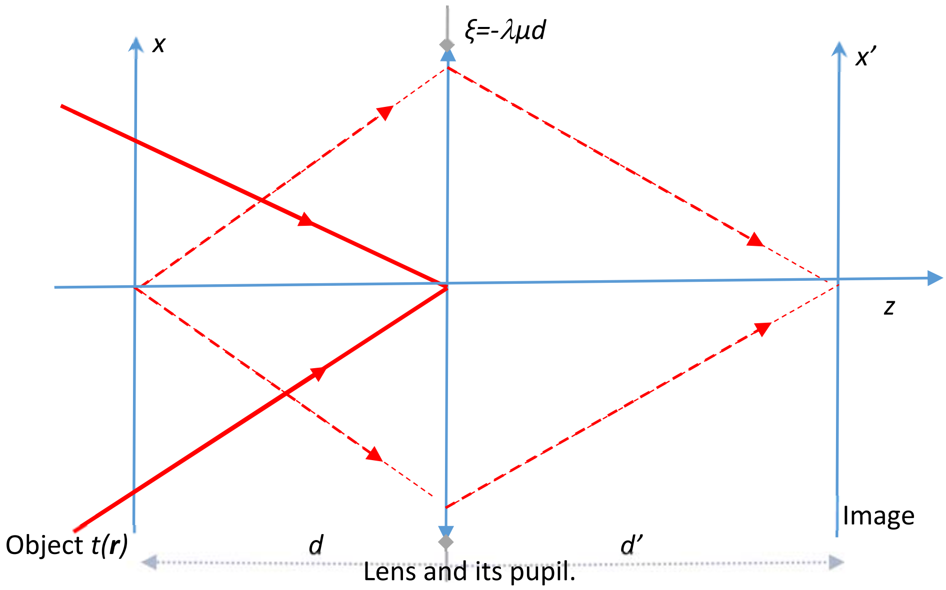


Fig 3.1 – The first diffraction.

3.1.1 – Geometrical imaging as a limiting case of double diffraction

We now consider the following setup, where a lens of image focal length f has been placed in plane z=0 of Fig. 3.1. The lens is limited by a "pupil" (so called by a straight analogy with the pupil of the eye). It is called a "double diffraction setup" because, as we shall now examine in detail, it will be considered as a cascade of one diffraction from the object plane to the pupil plane followed by a diffraction which occurs when light that has gone through the lens propagates to the image plane.



The pupil *p(**ρ**)* is depicted by two grey segments ending with a diamond

Fig. 3.2 – the two diffractions: one from the object to the lens, one from the lens to the image.

The pupil that limits the lens will be seen to play a major role because of its diffraction. Nevertheless, in a first stage, we shall consider that the pupil does not diffract, which is equivalent to saying that it is infinite. In practice, the situation does occur when essentially all light diffracted from the object reaches the lens, and is therefore not affected by a pupil, a relatively common situation. In that case, geometrical optics accounts for imaging. In this subsection, we shall show that the framework of diffraction does include geometrical optical imaging as a special case. In the next subsection, we shall take the exact shape of the pupil into account and therefore see how diffraction affects imaging.

Here, therefore, function *p(**ρ**)* which describes the pupil (see Fig 3.2) is just equal to unity everywhere. Let us analyze how the complex amplitude that has already been diffracted from the object to the pupil plane (Fig 3.1 and Eqn 3.2) is first transformed by going through the lens and then diffracted from the lens to the image plane.

Equation (3.2) describes the **first Fraunhofer diffraction** in this specific imaging process.

With the lens present and considered thin, we have to distinguish between the light just before the lens ($z=0_-$, Eqn 3.2) and the light just after the lens ($z=0_+$, Eqn 3.3), which is the product of the former by the transmittance of the thin lens. The latter (Exercise 1.3) reads $t_{lens}(\boldsymbol{\rho}) = \exp\frac{-i\pi\boldsymbol{\rho}^2}{\lambda f'}$, whence

$U(\boldsymbol{\rho},0_+) = \frac{iU_o}{\lambda d}\exp\frac{2i\pi}{\lambda}\left(-d-\frac{\boldsymbol{\rho}^2}{2d}-\frac{\boldsymbol{\rho}^2}{2f'}\right)\tilde{t}\left(\frac{-\boldsymbol{\rho}}{\lambda d}\right)$, which in turn is simplified using the law of thin lenses:

$$U(\boldsymbol{\rho},0_+) = \exp\frac{-i\pi\boldsymbol{\rho}^2}{\lambda d'}\quad\frac{iU_o}{\lambda d}\exp\frac{-2i\pi d}{\lambda}\tilde{t}\left(\frac{-\boldsymbol{\rho}}{\lambda d}\right) \tag{3.3}$$

The quadratic phase term from the lens combines the phase terms in such a way that Eqn (3.3) is formally equivalent to Eqn (3.1), with the amplitude of the spherical wave converging a distance *–d* behind the object being formally replaced by that of a spherical wave converging at a distance *d'* behind the pupil. Therefore, the principle of Fraunhofer diffraction operates again and the propagation between the pupil plane and the image plane is really a **second Fraunhofer diffraction**, whence the name "double diffraction" for this setup. Note that the object term of the first diffraction, *t(**r**)*, is replaced by the slightly trickier expression $\tilde{t}\left(\frac{-\boldsymbol{\rho}}{\lambda d}\right)$. The amplitude in the image plane then reads

$$U(\mathbf{r}',d') = \frac{-i}{\lambda d'}\exp\frac{i2\pi}{\lambda}\left(d'+\frac{\mathbf{r}'^2}{2d'}\right)\mathrm{FT}\left[\frac{iU_o}{\lambda d}\exp\frac{-2i\pi d}{\lambda}\tilde{t}\left(\frac{-\boldsymbol{\rho}}{\lambda d}\right)\right]_{\mathbf{r}'/\lambda d'},$$

with the last subscript indicating at which point the Fourier transform should be calculated. After a little algebra, that expression further yields to

$$U(\mathbf{r}',d') = \frac{U_o d}{d'}\exp\frac{i2\pi}{\lambda}\left(d'-d+\frac{\mathbf{r}'^2}{2d'}\right)t\left(\frac{d\mathbf{r}'}{d'}\right). \tag{3.4}$$

Eqn (3.5), although deriving from the modelling of diffraction, is in fact a geometrical optics law: it describes the complex amplitude of the stigmatic geometrical image of the initial object. That is because no limitation of the beams by any pupil has introduced a discrepancy between geometrical optics and diffraction. Compared to the expression of the complex amplitude in the object, the image amplitude is affected by three terms:

- the image is a homothetic version of the object, with a magnification ratio *d'/d*, as expected,
- the illuminating beam amplitude is globally affected by the same amount, *d'/d*, which accounts for energy conservation,

- the phase is delayed by a quantity that accounts for propagation over distance *d′-d* along the axis, and by a quadratic (i.e. spherical) term of radius *d′* which straightforwardly corresponds to the illuminating wave.

In what follows, we shall investigate how diffraction by a limited or intentionally obstructed pupil affects the expected geometrical image $t'_g(\mathbf{r}') = t\left(\frac{d\mathbf{r}'}{d'}\right)$. Notation t'_g stands for “geometrical image”.

3.1.2 – Double diffraction with an arbitrary pupil:

Obviously, every system has in fact a bounded pupil. In addition, one may, intentionally or not, introduce inside the pupil phase variations departing from the wanted spherical phase term. Similarly, the complex amplitude modulus (often called “the amplitude” by oversimplification) can be modified, either through absorption or even, in a gain medium, through amplification. This is why the lens, with its pupil, will now be decomposed as follows: $p(\boldsymbol{\rho})t_{lens}(\boldsymbol{\rho})$, with t_{lens} already introduced earlier, and p(**ρ**) called the “pupil function”, which induces the effects whereby diffraction modifies geometrical optics. In the basic situation, the pupil function is just the characteristic function of the pupil, in the mathematical sense, i.e; unity inside the pupil and zero everywhere else. But pupil plane engineering can be used to introduce various other modifications and effectively perform analog image processing. That principle can be emphatically be presented as “massively parallel processing at the speed of light” because all points in the object are processed in parallel and the result is available as soon as light has reached the image (well, in fact, as soon as a sufficient quantity of energy has been detected in the image).

The calculation leading from Eqn (3.3) to Eqn (3.4) needs to be changed only by introducing the pupil function: after the lens and its pupil,

$$U(\boldsymbol{\rho},0_+) = \exp\frac{-i\pi\boldsymbol{\rho}^2}{\lambda d'} \quad \frac{iU_o}{\lambda d}\exp\frac{-2i\pi d}{\lambda}\tilde{t}\left(\frac{-\boldsymbol{\rho}}{\lambda d}\right)p(\boldsymbol{\rho}) \qquad \text{(3.3bis)}$$

and in the image, remembering that the Fourier transform of a product is the convolution of the Fourier transform:

$$U(\mathbf{r}',d') = \frac{U_o d}{d'}\exp\frac{i2\pi}{\lambda}\left(d'-d+\frac{\mathbf{r}'^2}{2d'}\right)\left[t\left(\frac{d\mathbf{r}'}{d'}\right)*\frac{\tilde{p}\left(\frac{\mathbf{r}'}{\lambda d'}\right)}{\lambda^2 d'^2}\right]. \qquad \text{(3.4bis)}$$

In Eqn (3.4bis), the star denotes 2D convolution over variable ***r′***.

3.1.3 – The coherent point spread function and the coherent modulation transfer function:

The main result here is contained in the square bracket: instead of the geometrical image t'_g, the real image t'_d is affected by diffraction, which can be understood as **a spatial frequency filtering** operation, as will be explained now. The “impulse response” or “point spread function” is the image of a point in the object (a Dirac delta peak at the origin of the object plane)[1]:

[1] En français, « réponse percussionnelle », mais évitons « fonction d’étalement de point » puisqu’il existe une expression française, d’ailleurs antérieure à l’anglaise !

$$P(\mathbf{r}') = \frac{\tilde{p}\left(\frac{\mathbf{r}'}{\lambda d'}\right)}{\lambda^2 d^2} \tag{3.5}$$

It is just the Fraunhofer diffraction pattern by the pupil, as can be straightforwardly understood by considering a single point on axis as the object. In the spatial frequency filtering operation, the input is **the geometrical image** t'_g, the output is **the image affected by diffraction** t'_d, both expressed as part of the complex amplitude at the image plane, and the relation reads:

$$t'_d = t'_g * P \text{, a convolution in the image space coordinates } \boldsymbol{r'}. \tag{3.6}$$

The "gain" function of the filter, specifically called here the **coherent modulation transfer function**, CTF[2], is the Fourier transform of the impulse response, and is identical to the pupil function itself except that it is expressed in units of spatial frequency:

$$\tilde{t}'_d(\boldsymbol{\mu}') = \tilde{t}'_g(\boldsymbol{\mu}')\mathrm{CTF}(\boldsymbol{\mu}') \text{, with } \mathrm{CTF}(\boldsymbol{\mu}') = p(-\lambda d'\boldsymbol{\mu}'). \tag{3.7}$$

That fairly simple relation, where the 2D spatial frequency $\boldsymbol{\mu'}$ is the conjugate variable to the space variable $\boldsymbol{r'}$ in the image space, shows the direct and straightforward effect of introducing a pupil to affect the diffraction phenomenon. In the pupil plane, where the object FT is observed, one can individually change the phase and modulus of each and every point (i.e., spatial frequency of the image), by inserting a phase, absorbing or amplifying device.

This simple model of a convolution relation between the object and the image is restricted to linear optics and, mainly, by aberrations that occur in almost every realistic imaging process: lenses and oblique propagation introduce aberrations. Nevertheless, a local convolution, depending on the local behavior of aberrations in a region of the image, is still a valid model.

Nevertheless, the simple model valid under the Fresnel approximation is useful to provide a fairly practical physical intuition of the imaging process, in this case under coherent illumination. This is summarized in the table below.

Summary of coherent spatial frequency filtering (double diffraction experiment)

$$\text{Object transmittance } t \xrightarrow[\text{geometrical imaging}]{} t'_g \xrightarrow[\text{diffraction}]{} t'_d = t'_g * P$$

where P is the Fraunhofer diffraction pattern by the pupil

and in Fourier space (the 2D spatial frequency space, physically observable in the pupil itself):

$$\tilde{t}'_d = \tilde{t}'_g p,$$

where p is simply the pupil function itself, expressed in suitable units.

Exercises will illustrate the usefulness of this theory. Let it be reminded that temporal coherence will not need to be considered here, just because "two different optical (time) frequencies never interfere", the image in the presence of an extended (time) spectrum is just the integral of all monochromatic images.

3.2 – Shift invariant, linear filtering (a reminder)

The Fourier transform allows to introduce the concept of spatial frequency: practically every object described as a function of *x* and *y* has a Fourier transform that describes its behavior in terms of spatial frequencies. One major interest of the concept is to provide an intuitive understanding of how images

[2] En français, « fonction de transfert cohérente ».

are formed. Indeed, imaging optical instruments can usually be modelled as shift-invariant, linear filters. Generalizing section 3.1 on coherent spatial frequency filtering, this section is a short with a reminder of the mathematical properties of shift invariant, linear filters, a concept found in many domains of physics, both theoretical and applied, and in particular in electronics, which are described both in the time (or space) domain and in the dual frequency (or spatial frequency) domain. Further, we turn to the application of those concepts to imaging under spatially incoherent illumination. We shall proceed as we did for spatial frequency filtering in coherent light, starting with the description of the ideal geometrical image, in the absence of any diffraction – as would be obtained from a stigmatic instrument with an infinitely extended pupil, – and then proceeding to finite pupils and applications.

3.2.1 – Definitions:

Consider a set of functions f defined over $\mathbb{R}^n$ and whose values belong to some set E where addition and multiplication are defined. In practice, n is commonly equal to unity in electronics, and to 2 or sometimes 3 in optics, and E is either $\mathbb{R}, \mathbb{R}^+$ or $\mathbb{C}$:

$$\begin{array}{ccc} \mathbb{R}^n & & E \\ \mathbf{x}(x_1,\cdots,x_n) & \xrightarrow{f} & f(\mathbf{x}) \end{array}.$$

An application *F* from set $\{f\}$ into itself $\begin{array}{ccc} \{f\} & & \{f\} \\ f & \xrightarrow{F} & F(f) \end{array}$ is said to be linear if for any vectors ***x₁***, ***x₂*** and constants *α₁*, *α₂*:

$$F(\alpha_1\mathbf{x}_1+\alpha_2\mathbf{x}_2)=\alpha_1 F(\mathbf{x}_1)+\alpha_2 F(\mathbf{x}_2). \tag{3.8}$$

For an function f in $\{f\}$ and any vector x_o, let us denote by f_{x_o} the function defined by:

$$\text{for any } \mathbf{x}\in\mathbb{R}^n,\ f_{\mathbf{x}_o}(\mathbf{x})=f(\mathbf{x}-\mathbf{x}_o).$$

Application *F* is said to be shift-invariant (or "homogeneous") if for any vector x_o,

$$F(f)=g \Rightarrow F(f_{\mathbf{x}_o})=g_{\mathbf{x}_o} \tag{3.9}$$

Applications that are both linear and shift invariant, often called (linear, shift-invariant) filters, have remarkable properties in relation to their Fourier transform.

Examples include

- R, L, C electronic circuits ($E=\mathbb{C}, n=1$, ***x*** is time, *f* is the voltage applied to the circuit, *F(f)* is the voltage between two given points in the circuit),
- image formation in optics ($E=\mathbb{C}$ or $\mathbb{R}^+, n=2$, ***x*** are the image coordinates, *f* is the ideal geometrical image, *F(f)* is the image in the presence of diffraction).

3.2.2 – Fundamental property:

Exponential functions of the form $\exp(2i\pi\boldsymbol{\mu}\cdot\mathbf{x})$ are eigenfunctions of any linear, shift-invariant filter. Let us demonstrate that property. For an arbitrary frequency $\boldsymbol{\mu}\in\mathbb{R}^n$, we define function ${}^{\boldsymbol{\mu}}f$ (a particular member of set $\{f\}$) by

$${}^{\mu}f(\mathbf{x})=\exp(2i\pi\boldsymbol{\mu}\cdot\mathbf{x}) \text{ for all } \mathbf{x} \text{ in } \mathbb{R}^n.$$

Besides, consider an arbitrary linear, shift-invariant filter *F* and let us define function $^{\mu}g$ as $^{\mu}g = F\left({}^{\boldsymbol{\mu}}f\right)$.

Using known properties of exponentials:

$\forall \mathbf{x}_o \in \mathbb{R}^n, \forall \mathbf{x} \in \mathbb{R}^n, {}^{\boldsymbol{\mu}}f_{\mathbf{x}_o}(\mathbf{x}) = \exp\left[2i\pi\boldsymbol{\mu}\bullet(\mathbf{x}-\mathbf{x}_o)\right] = \exp\left(-2i\pi\boldsymbol{\mu}\bullet\mathbf{x}_o\right){}^{\boldsymbol{\mu}}f(\mathbf{x})$,

so that ${}^{\boldsymbol{\mu}}f_{\mathbf{x}_o} = \exp\left(-2i\pi\boldsymbol{\mu}\bullet\mathbf{x}_o\right){}^{\boldsymbol{\mu}}f$. Then, using the property of homogeneity:

$F\left({}^{\boldsymbol{\mu}}f_{\mathbf{x}_o}\right) = {}^{\boldsymbol{\mu}}g_{\mathbf{x}_o}$ and using now the property of linearity,

$F\left({}^{\boldsymbol{\mu}}f_{\mathbf{x}_o}\right) = F\left[\exp\left(-2i\pi\boldsymbol{\mu}\bullet\mathbf{x}_o\right){}^{\boldsymbol{\mu}}f\right] = \exp\left(-2i\pi\boldsymbol{\mu}\bullet\mathbf{x}_o\right)F\left({}^{\boldsymbol{\mu}}f\right)$, whence

$\forall \mathbf{x}_o \in \mathbb{R}^n, \forall \mathbf{x} \in \mathbb{R}^n, {}^{\boldsymbol{\mu}}g(\mathbf{x}-\mathbf{x}_o) = \exp\left(-2i\pi\boldsymbol{\mu}\bullet\mathbf{x}_o\right){}^{\boldsymbol{\mu}}g(\mathbf{x})$, which is true in particular for $\mathbf{x} = \mathbf{x}_o$, therefore : $\forall \mathbf{x} \in \mathbb{R}^n, {}^{\boldsymbol{\mu}}g(\mathbf{x}) = {}^{\boldsymbol{\mu}}g(\mathbf{0})\exp\left(2i\pi\boldsymbol{\mu}\bullet\mathbf{x}_o\right)$ and therefore

$${}^{\boldsymbol{\mu}}g = F\left({}^{\boldsymbol{\mu}}f\right) = {}^{\boldsymbol{\mu}}g(\mathbf{0})\,{}^{\boldsymbol{\mu}}f. \text{ QED} \tag{3.10}$$

The quantity ${}^{\boldsymbol{\mu}}g(\mathbf{0})$ is just a complex number depending on parameter ***μ***. It is a function of ***μ***, and can be denoted as *G(**μ**)*. It is called the gain of filter F, or sometimes the filter function. In optics, as introduced specifically for coherent and incoherent illumination in this chapter, other names are used based on the concept of "transferring" the contrast and phase – modulus and phase of *G(**μ**)* – of a sinusoid from the input (the ideal geometrical image) to the output (the image affected by diffraction or various other effects).

3.2.3 – The convolution theorem:

In the common situation where the input of a filter has a Fourier transform:

$$f(\mathbf{x}) = \int_{\mathbb{R}^n} \tilde{f}(\boldsymbol{\mu})\exp\left(2i\pi\boldsymbol{\mu}\bullet\mathbf{x}\right)d^n\boldsymbol{\mu},$$

it is straightforward to express the output using the definition of the gain function and the linearity:

$$\begin{aligned} g(\mathbf{x}) = F\left[f(\mathbf{x})\right] &= F\left[\int_{\mathbb{R}^n} \tilde{f}(\boldsymbol{\mu})\exp\left(2i\pi\boldsymbol{\mu}\bullet\mathbf{x}\right)d^n\boldsymbol{\mu}\right] \\ &= \int_{\mathbb{R}^n} \tilde{f}(\boldsymbol{\mu})F\left[\exp\left(2i\pi\boldsymbol{\mu}\bullet\mathbf{x}\right)\right]d^n\boldsymbol{\mu} \\ &= \int_{\mathbb{R}^n} \tilde{f}(\boldsymbol{\mu})G(\boldsymbol{\mu})\exp\left(2i\pi\boldsymbol{\mu}\bullet\mathbf{x}\right)d^n\boldsymbol{\mu} \end{aligned},$$

which says that g has a Fourier transform, which is

$$\boxed{\tilde{g}(\boldsymbol{\mu}) = \tilde{f}(\boldsymbol{\mu})G(\boldsymbol{\mu})} \tag{3.11}$$

in words, that equation says that the effect of the linear shift-invariant filter is to multiply the amplitude of each frequency in the input by the gain. By inverse Fourier transformation, and defining the inverse Fourier transform of the gain function as $P(\mathbf{x}) = \int_{\mathbb{R}^n} G(\boldsymbol{\mu})\exp\left(2i\pi\boldsymbol{\mu}\bullet\mathbf{x}\right)d^n\boldsymbol{\mu}$, the **impulse response** of the filter, it straightforwardly follows that:

$$g(\mathbf{x}) = f(\mathbf{x}) * P(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{x}_1) P(\mathbf{x} - \mathbf{x}_1) d^n\mathbf{x}_1 \quad (3.12)$$

which implies that the impulse response is the output to a delta peak as input.

3.3 – Spatial frequency filtering in spatially incoherent illuminance

3.3.1 – Illuminance of the stigmatic geometric optical image

a) Reminders

Solid angle:

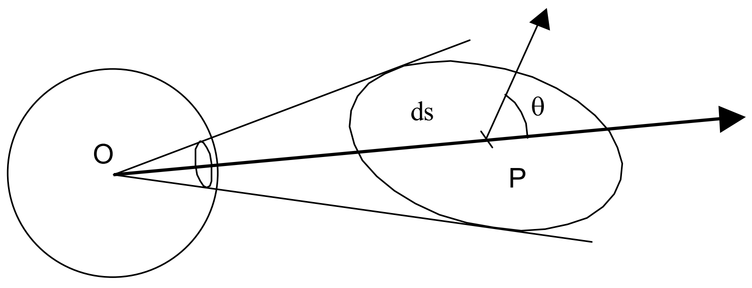
Consider a cone of apex O based on an arbitrary closed curve. The solid angle is the area of the intersection of this cone with the unit sphere centered at *O*. It is therefore a dimensionless quantity and the solid angle under which one point sees the whole space is 4π. We define solid angle as an arithmetic, not algebraic, quantity.

Applications:

- the solid angle *Ω* under which a given point *O* “sees” a disk of center *C* and radius *a*, with *CO* perpendicular to the disk plane, can be calculated by integrating from 0 to α the area of the elementary annulus of the unit sphere comprised between *ϑ* and *ϑ+dϑ*.

The result is $\Omega = 2\pi(1-\cos\alpha)$ with $\alpha = \arctan\dfrac{a}{OC}$. (3.13)

- The differential solid angle $d^2\Omega$ under which *O* “sees” an differential area element d^2s centered at *P* and such that the segment *OP* is inclined by *ϑ* from the normal to the area element at *P* is obtained by projecting the area ds on the plane perpendicular to *OP* at *P*, and considering that, since *ds* is an differential area element, the element of the sphere (*O*,*OP*) cut by the cone (*O*,*ds*) is equal to its projection. Therefore, $d^2\Omega = \dfrac{d^2 s\cos\theta}{OP^2}$.



- Geometric étendue (the English word here is just the French): consider now two differential area elements d^2s centered at *P* and $d^2\sigma$ centered at *O* such their normal make the respective angles *ϑ* and *ψ* with respect to *OP*. Calling respectively $d^2\Omega$ the solid angle under which *O* sees *ds* and $d^2\Omega_P$ the solid angle under which *P* sees *dσ*, the quantity

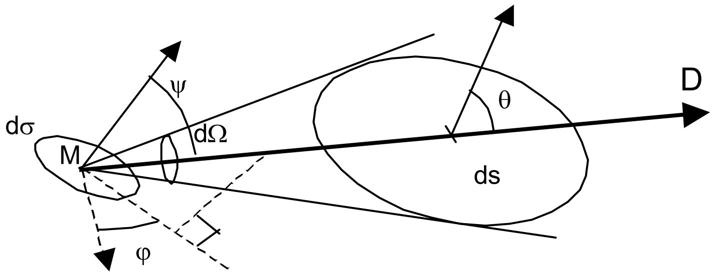
$\dfrac{d^2 s\,\cos\theta\, d^2\sigma\,\cos\psi}{OP^2} = d^2\Omega\, d^2\sigma\,\cos\psi = d^2\Omega_p\, d^2 s\cos\theta = d^4G$ is called the differential geometric étendue spanned by d^2s and $d^2\sigma$.

Photometry

The word “flux” used for light is just the power of light traversing a given surface. Flux, a physical quantity, is measured in watts. But if photometric units, that have been introduced in the international system of units specifically to account for the spectral sensitivity of the "average" human eye, the unit is the lumen. At the maximal spectral sensitivity (λ=550 nm), 1 W corresponds to 683 lumens[3].

The light energy emitted by a source or propagating through space is usually not isotropic, but depends on the propagation direction. The quantity named "luminance"[4] has been introduced to account for that fact. The luminance of a source is the flux emitted at point *M* of its surface in a given direction (*D*) defined by angles φ and ψ (see figure) by unit solid angle and by unit area of the source projected normally to (*D*):

$$d^4\phi = L(M,\varphi,\psi)\cos\psi \; d^2\sigma \; d^2\Omega = L(M,\varphi,\psi)\, d^4G \quad (3.14)$$



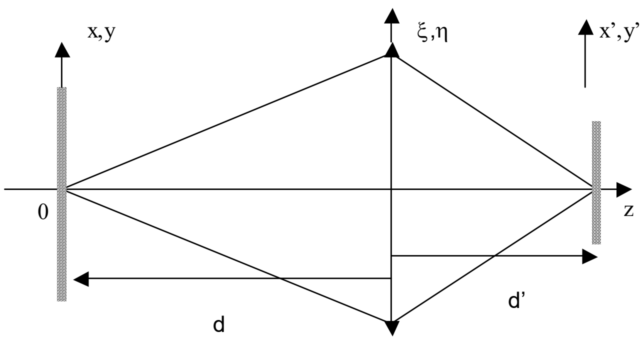
For a spatially incoherent source, the flux due to various elementary areas just add (on an incoherent basis): they do not interfere. That is the very definition of the word "incoherent", coherence being the ability to interfere.

A source whose luminance does not depend on the direction D is said to be "Lambertian" (from Jean-Henri Lambert, a precursor of photometry in the 18th century).

The concept of luminance, defined for sources as said, is in fact applicable to any surface in space, be it a source or not. In the latter case, the flux is not emitted by propagated from the sources to the surface considered in direction (*D*). Just like in the Huygens-Fresnel principle, any surface illuminated by a primary source behaves in photometric terms as a "secondary source": an arbitrary surface in isotropic space just transmits the flux that it has received. Physical surfaces such as scattering objects may absorb the energy and/or redistribute it in various directions.

Illuminance[4] E is the flux by unit area for a given surface: $d^2\phi = E\, d^2 s$.

b) Expression of the illuminance



Consider a differential area d^2s on a Lambertian object of luminance L_o located on the axis of a thin lens. As indicated on the figure, the notations are the same as in section 3.1 for the object, pupil and

[3] See www.bipm.org for more detail.

[4] En français, « luminance », à ne pas confondre avec l'anglais « illuminance » qui désigne l'éclairement.

image plane and their respective distances. In the limit of small angles (Gaussian optics), the cosine in equation 3.13 reduces to $1-\alpha^2/2$ and more generally (for an arbitrary pupil shape) the pupil area can be approximated to the area of the sphere in the definition of solid angle. If the image is stigmatic, with *Σ* the lens pupil area, and neglecting diffraction, the flux on the pupil is

$$d^2\phi = L_o \frac{\Sigma\, d^2 s}{d^2}.$$

The image magnification is d'/d . Because of losses that may occur due to reflection, absorption or scattering and a possible partial intentional attenuation of the pupil, the image illuminance is then

$$\mathcal{E}'_g = \frac{\tau\, d^2\phi}{\left(\frac{d'}{d}\right)^2 d^2 s} = \frac{\tau L_o \Sigma}{d'^2}. \qquad (3.15)$$

More generally, but still in the limit of Gaussian optics and assuming stigmatism, for an extended object, letting $\mathbf{r}' = \frac{d'}{d}\mathbf{r}$ be the position in the image plane of the image of object point **r**,

$$\mathcal{E}'_g(\mathbf{r}') = \frac{\tau\, d^2\phi}{\left(\frac{d'}{d}\right)^2 d^2 s} = \frac{\tau L_o(\mathbf{r})\Sigma}{d'^2} = \frac{L'(\mathbf{r}')\Sigma}{d'^2}, \text{ with } L'(\mathbf{r}') = \tau L(\mathbf{r}) \qquad (3.16)$$

L'(***r'***) is the luminance in the image, which is just the luminance of the object corrected for losses.

3.3.2. – Illuminance in the diffraction limited image

a) Illuminance of the image of the object center in the presence of diffraction

Consider again the Lambertian object *L*(**r**), this time taking into account diffraction by the pupil It is convenient to start with a differential object zone around the center, small enough that all images of its points are essentially identical. The local illuminance is *L*(**0**). The incoherent addition of all those identical images is just the image of the central point with an illuminance proportional to the object luminiance and the image area. The image of the central point itself has been calculated in section 3.1: if there is only one point in the object, it does not make sense to distinguish between “coherent” and “incoherent”. The calculation is therefore the same as in the coherent case. With the same notations as earlier, the image of a differential area $d^2\boldsymbol{r}$ of the object is:

$d^2\mathcal{E}'_d = \alpha_1 \left|\tilde{p}\left(\frac{\mathbf{r}'}{\lambda d'}\right)\right|^2 d^2\mathbf{r}$, with α_1 a constant that will be calculated now. The flux on the differential area of the image is $d^4\phi = d^2\mathcal{E}' dx' dy' = d^2\mathcal{E}' d^2\mathbf{r}'$. The total flux on the diffraction pattern is the flux from $d^2\boldsymbol{r}$ that has reached the image, i.e., using Parseval’s theorem and taking account that parameter τ and the pupil function *p* together account for all possible causes of attenuation, voluntary or not:

$$\begin{aligned}
d^2\phi = L(\mathbf{0})\frac{\tau\Sigma\, d^2\mathbf{r}}{d^2} &= \iint_{\text{image}} d^2\mathcal{E}_d d^2\mathbf{r}' = \iint_{\mathbb{R}^2} \alpha_1 \left|\tilde{p}\left(\frac{\mathbf{r}'}{\lambda d'}\right)\right|^2 d^2\mathbf{r}\, d^2\mathbf{r}' \\
&= \alpha_1 d^2\mathbf{r}(\lambda d')^2 \iint_{\mathbb{R}^2} |\tilde{p}(\mathbf{u})|^2 d^2\mathbf{u} = \alpha_1 d^2\mathbf{r}(\lambda d')^2 \iint_{\mathbb{R}^2} |p(\boldsymbol{\rho})|^2 d^2\boldsymbol{\rho} \\
&= \alpha_1 d^2\mathbf{r}(\lambda d')^2 \tau\Sigma
\end{aligned}$$

whence finally

$$d^2\mathcal{E}'_d = \frac{\tau L(\mathbf{0})}{(\lambda d d')^2}\left|\tilde{p}\left(\frac{\mathbf{r}'}{\lambda d'}\right)\right|^2 d^2\mathbf{r} = \frac{L'(\mathbf{0})}{(\lambda d'^2)^2}\left|\tilde{p}\left(\frac{\mathbf{r}'}{\lambda d'}\right)\right|^2 d^2\mathbf{r}' \tag{3.17}$$

Note of caution: while in the calculation $d^2\mathbf{r}'$ was a differential element in an integral, in Eqn (3.17) it is the geometrical image of $d^2\mathbf{r}$ – still an arbitrary differential element, but the specific element around the center of coordinates considered in this subsection.

b) Extension to the full image:

The above expression for the illuminance of the center of coordinates directly gives the point spread function of a spatially incoherent imaging system, as soon as linearity and shift invariance are satisfied. Linearity by addition of illuminances due to the various source points is guaranteed as soon as they are really mutually incoherent and that the threshold for nonlinear optical phenomena is not reached. Shift invariance, on the other hand, is really guaranteed only in the absence of field aberrations, which is never strictly the case. As in the coherent case (end of Section 3.1), the image of off-axis object point $\mathbf{r}$ is to a good approximation a $\mathbf{r}\,d'/d$ shifted version of the image of the central point only in a limited part of the field. In that part, linear shift invariant filtering is achieved. Outside it, it is customary to segment the complete image field into regions where shift-invariance is approximately satisfied and to talk about a “local point spread function” and a “local modulation transfer function”. In the following, we do however consider the case where shift invariance is achieved indeed.

In that case, it is straightforward to add the illuminances of the various object points according to Eqn (3.17):

$$\mathcal{E}'_d(\mathbf{r}') = \iint_{\mathbb{R}^2} \frac{L'(\mathbf{r}'_1)}{\Sigma d'^2}\frac{1}{(\lambda d')^2}\left|\tilde{p}\left(\frac{\mathbf{r}'-\mathbf{r}'_1}{\lambda d'}\right)\right|^2 d\mathbf{r}'_1 = \mathcal{E}'_g(\mathbf{r}') * P_i(\mathbf{r}') \tag{3.18}$$

with

$$P_i(\mathbf{r}') = \frac{1}{\Sigma(\lambda d')^2}\left|\tilde{p}\left(\frac{\mathbf{r}'}{\lambda d'}\right)\right|^2 \tag{3.19}$$

P_i is the **incoherent point spread function** (français réponse percussionnelle incohérente) of the imaging system. Again “incoherent” here refers to spatial incoherence.

c) Incoherent transfer function:

The incoherent point spread function has a Fourier transform, named **incoherent transfer function**:

$$\mathrm{ITF}(\boldsymbol{\mu}') = \frac{p \otimes p(\lambda d'\boldsymbol{\mu}')}{\Sigma} = \frac{p \otimes p(\lambda d'\boldsymbol{\mu}')}{p \otimes p(\mathbf{0})}. \tag{3.20}$$

In this expression, the sign $\otimes$ stands for the correlation in the sense of functions, an operation similar to but not quite identical to convolution: $f \otimes g(v) = \int_{\mathbb{R}} f(u)\, g^*(u-v)\, du$ (with straightforward extension to higher dimensions – please note the complex conjugate sign on the second operand). It is immediate to show that $f \otimes g(v) = \int_{\mathbb{R}} f(u+v)\, g^*(u)\, du$ and that moreover,

$$\widetilde{f \otimes g} = \tilde{f}\,\tilde{g}^*.$$

Because our introduction of the incoherent point spread function requested that the energy be conserved between the geometrical image and the diffraction limited image, a noteworthy consequence is:

$$\mathrm{ITF}(\mathbf{0}) = 1. \tag{3.21}$$

One should be aware of a peculiar definition of the expression "Modulation Transfer Function" in English: $\mathrm{MTF}(\boldsymbol{\mu}') = |\mathrm{ITF}(\boldsymbol{\mu}')|$, while the quantity $\exp i \,\mathrm{Arg}\left[\mathrm{ITF}(\boldsymbol{\mu}')\right]$ is known as the phase transfer function.

Summary of incoherent spatial frequency filtering

$$\text{Object luminance } L \xrightarrow[\text{geometrical imaging}]{} \mathcal{E}'_g \xrightarrow[\text{diffraction}]{} \mathcal{E}'_d = \mathcal{E}'_g * P_i$$

where P_i is the **squared modulus** of the Fraunhofer diffraction pattern by the pupil

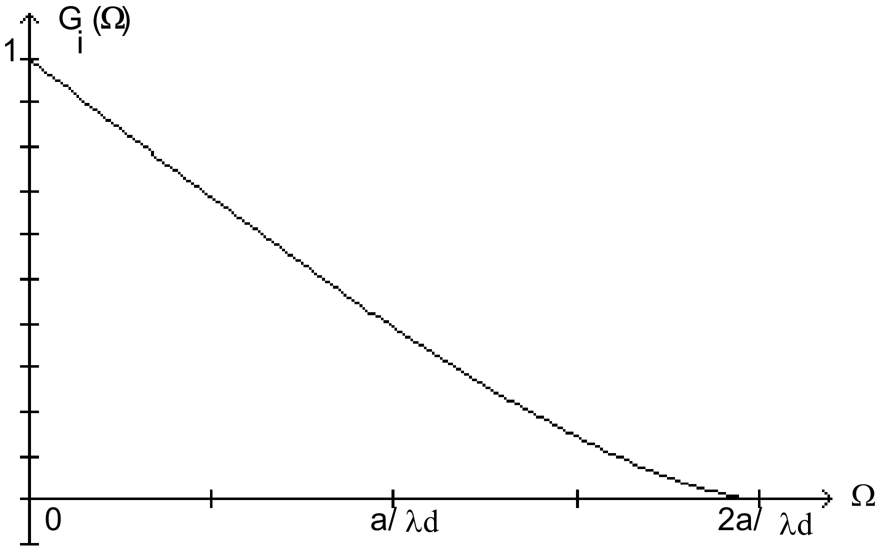
and in Fourier space (the 2D spatial frequency space, which is not physically accessible in incoherent illumination):

$$\tilde{\mathcal{E}}'_d = \tilde{\mathcal{E}}'_g \ \mathrm{ITF}(\boldsymbol{\mu}),$$

with $\mathrm{ITF}(\boldsymbol{\mu}') = \dfrac{p \otimes p(\lambda d' \boldsymbol{\mu}')}{p \otimes p(\mathbf{0})}$ the suitably scaled pupil autocorrelation.

3.3.3. Cutoff frequency, ITF of the stigmatic circular pupil instrument

a) Incoherent transfer function and point spread function:



From the above development, it results

1) That the point spread function is just a suitably normalized "Airy disk" intensity:

$$P_i(\mathbf{r}') = \frac{\pi a^2}{(\lambda d')}\left[\frac{2J_1(v)}{v}\right]^2 \text{ with } v = 2\pi \frac{a|\mathbf{r}'|}{\lambda d'} \qquad (3.22)$$

2) that the ITF of a stigmatic instrument with a circular pupil is the normalized intersection area of two disks of equal radius:

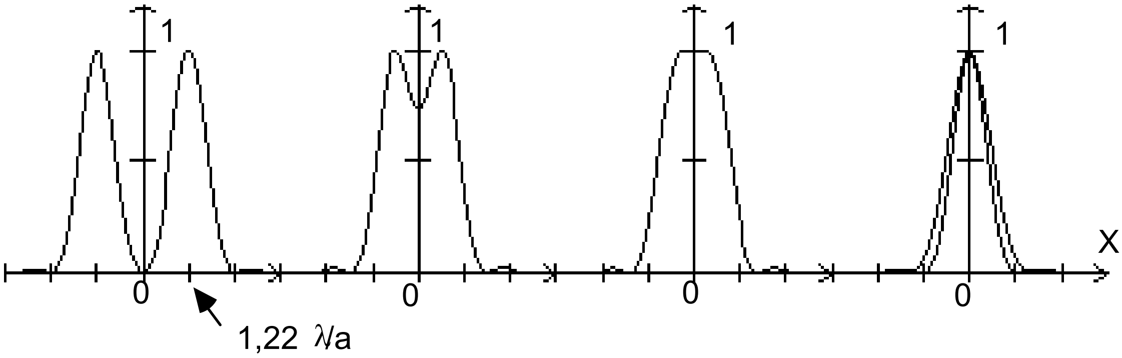
$$\mathrm{ITF}(\boldsymbol{\mu}') = \frac{2}{\pi}\left(\mathrm{Arccos}\, u - u\sqrt{1-u^2}\right) \text{ with } u = \frac{\lambda d' |\boldsymbol{\mu}'|}{2a}. \qquad (3.23)$$

It is a circularly symmetric function, strictly positive within the support $|\boldsymbol{\mu}'| \le \mu'_{\max}$, meaning that there is a strict cutoff frequency $\mu_{\max} = \dfrac{2a}{\lambda d'}$.

b) Discussion on the diffraction limited resolution of optical instruments

Image of two bright spots next to each other:

Consider an object composed of only two bright points with the same intensity and with their geometrical images separated by distance $1.22\,\varepsilon l d'/2a$, i.e. a fraction ε of the classic Rayleigh limit. The following shows the curves corresponding to ε=2, 1, 0.8 and 0.5 respectively, the latter two being plotted on the same diagram. All are normalized to unity at their own maximum.



It appears that the resolution limit, arbitrarily assessed from these curves, is slightly better than the Rayleigh criterion as usually expressed. However, mathematically speaking, there is no minimal distance between two points such that their combined image reduces to that of one single point: a suitable measurement, integrating over noise with an excellent detector and perhaps using so-called "super-resolution" techniques such as pupil mask, is always in principle capable of telling that there is more than a single point in the image.

Periodic object:

Considering the ITF, it is seen that if the period is larger than the inverse of the cutoff frequency, some modulation reaches the image. If however the period is smaller than $\lambda d'/2a$, strictly no modulation reaches the image[5].

In summary, the concept of the diffraction limited resolution in imaging is an ill-defined concept in the object space, but has a clear answer in the spatial frequency space.

[5] One may note that a similar situation arises in coherent illumination, but the cutoff period is $\lambda d'/a$, i.e. twice as large, with a sharp cutoff instead of a progressive loss of contrast as shown by the ITF shape. Note that spatial frequency filtering in coherent light applies to the complex amplitude transmittance t(r') while in incoherent light it applies to the illuminance. Those are two distinct physical quantities except in the case of binary objects.

CHAPTER 5: DIFFRACTIVE OPTICS

5.1. Diffraction is not always an obstacle:

It is known that every optical instrument is subject to diffraction. Geometrical optics is just an asymptotic limit when the wavelength tends to zero and is useful whenever the effect of diffraction is small enough compared to the needs of the application to be neglected. But if the user pushes the instrument to its limits, diffraction will always show up.

Diffraction is therefore mostly seen as a limitation to the performance of optical instruments. We have seen in Chapter 3 that it sets a limit to the magnitude of the spatial frequencies that are transferred from an object to its image, and similar effects arise in spectroscopy. There are however some situations when diffraction can be brought to use and create new functions: the term "diffractive optics" has been coined for that case. They will be considered in the present chapter, where phenomena that appear at scales close to the optical wavelength, or smaller, will be investigated.

The most common case of diffractive optics, and that has also been historically the subject of most investigations, is the diffraction grating. It is used in spectroscopy, in optical telecommunications to separate the wavelength channels carrying different data streams, and in a wealth of other applications, often to metrology. It will be covered in the following section. The other sections will cover other phenomena and components, generally of lesser practical importance so far, but currently the subjects of intense research.

It should be noted here that holography, covered in Chapter 4 of this course, is in a way an extension of diffraction gratings, and definitely belongs to diffractive optics. While in a grating, the individual grooves are the diffracting objects, in a hologram, they are replaced by the interference fringes between the carrier wave and the object wave. Generally, in diffractive optics, diffraction occurs at a set of neighboring regions (grooves, zones) and the combined effect of interference between the diffracted beams of all those regions determines the specific diffractive function, including wavelength separation, focusing (in a diffractive lens or holographic lens), aberration correction, or other.

5.2. The diffraction grating

5.2.1. Introduction: what is a diffraction grating?

A diffraction grating is a periodic modulation of refractive index in space. There exist in fact several kings of diffractive gratings:

- transmission gratings and reflection gratings (or gratings that are used both in reflection and in transmission),
- thin gratings (described by a "complex amplitude transmittance" independent of the specific illuminating wave, see Chapter 1), and thick gratings,
- gratings with one dimensional, two dimensional or three dimensional periodicities,
- gratings where only the real part of the refractive index is modulated, called phase gratings, and gratings where the imaginary part (or both the real and the imaginary parts) are modulated, somewhat deceptively termed "amplitude gratings",
- the index modulation may be purely scalar or may involve anisotropy and therefore rest on polarization.

Depending whether the period, technically called the pitch (en français "pas" du réseau) is much larger or not than the wavelength, the grating may or may not be approximatively studied in the paraxial and scalar formalism. If that is not the case, the much heavier electromagnetic theory based on Maxwell's equations must be resorted to – the two situations will be considered next. In 1D gratings, each period is called a "groove".

5.2.2. 1D gratings in scalar optics

A thin grating, by definition, is described by its complex amplitude transmittance $t(x)$. Since it is periodic, let us denote its pitch by Λ (the upper case of letter λ) and expand the transmittance into a Fourier series:

$$t(x)=\sum_{l\in\mathbb{Z}} t_l \exp\left(2i\pi l \frac{x}{\Lambda}\right). \tag{5.1}$$

In the Fraunhofer diffraction regime, this translates into a set of equidistant dots at infinity, each corresponding to one "order" of the grating. If illuminated by a plane monochromatic wave at a small angle α with respect to the main propagation axis z, each order is a plane wave and focusses in the Fraunhofer diffraction pattern at the focal plane of a lens at spots centered at points

$$\left(\alpha+\frac{l\lambda}{\Lambda}\right)f', \tag{5.2}$$

where *f'* is the image focal length of the lens and λ the wavelength in a vacuum (here, the grating is considered to be placed in a vacuum. To a fair approximation, the refractive index of air is the same as that of vacuum). The spot width, measured between its maximum and its first zero, is determined by the diffraction by the pupil. Typically, the pupil limiting the grating is rectangular shaped and therefore forms a cardinal sine Fraunhofer diffraction pattern of width

$$\frac{\lambda f'}{A}, \tag{5.3}$$

where A is the rectangle dimension across the grating grooves. That means that the spacing between diffracted orders is *N* times the width of one order, with *N* the number of grating grooves illuminated.

Diffraction efficiency in order *l* is defined as the fraction of the incoming light which gets diffracted into order l, i.e. the ratio between the flux by unit area at the exit of the grating in that order and the flux of the incident wave on the illumination side. Sometimes, the ratio between the flux in order *l* and the sum of fluxes in all diffracted orders is mistakenly (cheatingly!) called diffraction efficiency.

5.2.3. Grating orders in electromagnetic optics: the Floquet theorem

We now consider a general grating, still one-dimensional in the *x* direction and with a pitch Λ, but not necessarily thin, and possibly with a pitch too small for the scalar theory of Fraunhofer diffraction to yield good results. Consider that the grating index modulation is confined between planes *z*=0 and *z*=*h*, and that the grating is illuminated by one unique monochromatic plane wave with its wave vector ***k*** in the x,*z* plane with a positive *z*-component and its electric field polarized perpendicular to that plane, i.e. along *y* (TE polarization). A mathematical result known as the Floquet theorem, which will be taken here as a given result using the fact that the electromagnetic field obeys the Maxwell's equation, states that if the incident wave vector component along *x* is $k_x = k\sin\alpha_o$, then $E(x,z)\exp(-ik_x x)$ is a periodic function of *x* with period Λ. $k=\frac{2\pi}{\lambda}$ is the length of the wave vector in a vacuum.

From there, based on a Fourier series expansion, the electromagnetic field assumes above ($z>h$) as well as below ($z<0$) the grating the form of a sum of plane waves, where the incident field can be distinguished from the "diffracted field" (everything else). Above the grating, the field writes:

$$E_{diff}(x,z)=\sum_{l\in\mathbb{Z}} E_{l+}\exp\left[ik\left(\sin\alpha_o+\frac{l\lambda}{\Lambda}\right)x+ik_{z,l+}z\right]. \tag{5.4}$$

Relative integer parameter l stands for the grating order. Orders for which $-k < k_{x,l} = k\left(\sin\alpha_o + \frac{l\lambda}{\Lambda}\right) < k$ are homogeneous monochromatic plane waves leaving the grating at angle α_l such that

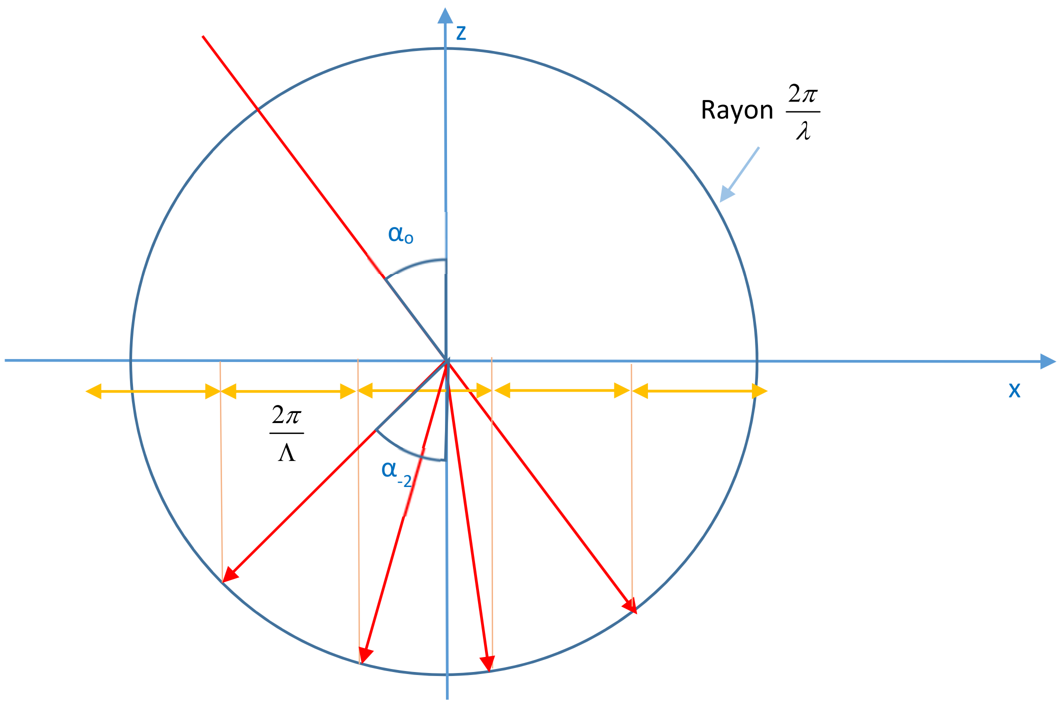
$$\boxed{\sin\alpha_l - \sin\alpha_o = \frac{l\lambda}{\Lambda}} \qquad (5.5)$$

and their z component is the solution $k_{z,l+}$ of equation $k^2 = k_{x,l}^2 + k_{z,l+}^2$ for which $k_{z,l+} > 0$, i.e. it propagates in the direction of increasing *z*. The other orders are evanescent, which are essentially localized at distances from the grating on the order of the wavelength and are exponentially attenuated as *z* increases. For them, $k_{z,l+}$ is therefore a complex number with a positive imaginary part.

Similarly, "below" the grating

$$E_{diff}(x,z) = \sum_{l\in\mathbb{Z}} E_{l-} \exp\left[ik\left(\sin\alpha_o + \frac{l\lambda}{\Lambda}\right)x + ik_{z,l-}z\right], \qquad (5.4bis)$$

which shows that (if the refractive indices on either side of the grating are the same), the reflected and transmitted orders l are symmetrically oriented about the grating. Their z components are opposed (both for homogeneous waves and for evanescent waves).



Geometrical construction of the orders of a grating, from the wave vector. In this example, orders -3 through 0 are homogeneous waves, all others are evanescent.

Between the two, in the region where the refractive index is modulated, the calculation is heavy and requires in practice a computer to be solved. Such calculation is required to find the E_{l+} and E_{l-} order amplitude values, and therefore the diffraction efficiencies.

Floquet's theorem generalizes straightforwardly to the other main polarization state (TM), to oblique incidence ($k_y \neq 0$, the so called conical geometry), and to two-dimensional gratings.

Comment: Eqn (5.5) expresses "the grating law". It is general for 1D gratings illuminated in the *x*,*z* plane. It is useful to visualize it as shown on the figure below for the transmitted orders. The construction relies on the fact that the wave vector for all orders has always the same modulus and that the x component of each order is obtained by adding $\frac{2\pi}{\Lambda}$. For the fairly general case of a grating etched on a plane diopter, the construction generalizes by remembering that the wave vector scales with the refractive index: $|\mathbf{k}| = \frac{2\pi n}{\lambda}$

5.2.4. Thick gratings

Diffraction by a thick grating does obey the grating law. However, the order efficiency, as calculated by solving the Maxwell's equations in the index modulated region of space, depends strongly of the grating thickness *h*. A thick grating that consists of parallel, equidistant planes of equal index, which is a sort of artificial crystal, the only orders diffracted with a non-negligible efficiency are those that obey, in addition to the grating law, the condition that the diffracted order corresponds to reflection on the planes of equal index. This constraint is called Bragg's law and writes

$$\sin\alpha = \sin\alpha_l = \frac{l\lambda_o}{2\Lambda} . \qquad (5.6)$$

It is important to stress that Eqn (5.5) is a property which gives the position of all orders, while Eqn (5.6) is a condition which indicates which order, if any, is diffracted by a thick grating with a non-negligible efficiency.

Bragg's law is particularly important in crystallography, natural crystals being 3D periodic gratings with pitches on the order of X ray wavelengths (tens of nanometers to a few nanometers in most cases). In that case, the "planes" of equal index are replaced by the crystallographic planes (see an introduction on crystallography).

Specific Bragg gratings are those fabricated inside an optical fiber, that act as mirror for a specific wavelength or wavelength range for optical routing or switching applications in telecommunications, leaving all other wavelengths unaffected. Another extension of index modulation, accessible only if the index modulation range is large enough, is photonic bandgap materials, where the behavior of a Bragg mirror occurs over several non-parallel Bragg mirrors at the same time. Light can be trapped in a cavity bounded by Bragg mirrors on all sides, and in a photonic bandgap material the same periodic index modulation plays the role of all the Bragg mirrors.

5.3 Segmented optical components:

5.3.1 – Zone plates:

Introduction in the form of an exercise:

The exercise is treated in the Fresnel approximation. An object located in plane z=0 has a complex amplitude transmittance $t(r)=\frac{1}{2}\left(1+\cos\frac{2\pi r^2}{r_o^2}\right)$*, with* $r=\sqrt{x^2+y^2}$ *. It is bound by a pupil of radius a. Plot the transmittance against r, then against* r^2*. Expanding the cosine, show that the amplitude after this object can be regarded as the coherent sum of three waves, of which two would be produced respectively by a positive and by a negative lens. Give the focal length of those lenses.*

Generalization

The transmission of the previous object is not a periodic function of a space variable, like in a diffraction grating. However, it is periodic in the variable r^2, square of the distance to the center. More generally, any object that is periodic in r^2 is called a “zone plate” (or often “Fresnel zone plate”, although they were invented after Fresnel: the annular zones, however, were used by Fresnel as a concept in his pioneering study of diffraction. En français, réseau zoné). If it is thin, its transmission can be expanded in a Fourier series of variable r^2 of the form $t(r)=\sum_{m\in\mathbb{Z}} t_m \exp\frac{2i\pi r^2}{r_o^2}$. These objects can be thought of as coherent superposition of overlapping lenses.

From there, many properties of the diffraction gratings are retrieved. Swiss physicist Jacques-Louis Soret introduced in 1875 the zone plate equivalent of a Ronchi ruling. At the time, it was intriguing and enlightening to see a China ink drawing on paper, reproduced photographically on a glass plate, behave as alens! The equivalent of a blazed grating, which sends almost all of its light in one given order, is a succession of transparent annuli with the profile of a lens but cut into thin slices of thickness $\frac{\lambda}{n-1}$. If antireflection coated, it can theoretically reach a diffraction efficiency of 100%. With suitable technology, this is an interesting optical component for some applications.

5.3.2 – Arbitrary shaped zones:

One can go one step further in generalizing gratings. Here, we shall only consider the most useful case of “blazed” diffractive elements, i.e. where only phase is modulated. As we have seen (exercise 5.2), the phase of a blazed grating varies linearly from 0 to 2π across one period, and then falls back to zero. It consists of parallel equidistant grooves shape like little sections of prisms. Similarly, the phase of a blazed zone plate varies linearly in the variable r^2 from 0 to 2π across one annular zone and then falls back to zero. Its zone shapes are small fractions of a sphere.

More generally, if one needs an optical component whose phase in plane *z*=0 at a given wavelength λ is a given function $\phi(x,y)$, it may be implemented by slicing a surface of the suitable shape into small segments or zones where phase varies from 0 to 2π as $\phi(x,y)$ modulo 2π. Specifically, surface $h(x,y)=\frac{\lambda\phi(x,y)}{2\pi(n-1)}$ modulo 2π is the thinnest conceivable component that will provide the desired phase in plane *z*=0. It plays the same role as the initial non segmented components, albeit only at one specific wavelength. At another wavelength, the phase is the same but the diffraction efficiency drops (remember however that the same phase at two different wavelengths does not correspond to the same wavefront shape). Such a segmented component is light and compact, but expensive to fabricate although replication may be very cheap. Such components are currently investigated by many research projects. Such components may be useful to correct for aberrations, shape wavefronts, or a number of specific applications.

As a final note, components mentioned in this paragraph are blazed in order 1: phase varies from 0 to 2π across one segment. Diffracted components may also be blazed in order *l*, with phase varying from 0 to $2l\pi$.

5.4 Metasurfaces

A recent extension of the above concepts has been enabled by using microelectronics technology to fabricate optical components. In that case, an idea still at the research stage consists in changing the local index rather than the surface height. To continuously change the local index, one procedure is to implement an “effective index” by a juxtaposition of nanostructures, i.e. features much smaller than the wavelength such that light diffracted outside the evanescent wave region essentially experiences an average index over the local nanostructures.

That goal can be achieved for example by periodically etching a surface into a set of nanopillars of identical height, but modulating the diameter of the nanopillars across the component surface. Locally, the nanopillars behave as a grating, but with such a small pitch Λ that all orders, except order zero, are evanescent. The diffraction efficiency is 100% into order zero, but its phase is modulated by the pillar shape. As can be seen, such a component involves two effects of diffraction: one is the decomposition of the desired wavefront into segments, as explained in § 5.3. The other is the modulation of the effective index by the nanopillars.

Such diffractive elements are known as metasurfaces. They may some day tunr important in photonic nanosystems that are at this stage still to be invented. This paragraph will be covered in the form of a seminar based on a presentation at a 2017 scientific conference (and published in a peer reviewed journal the same year).

Chapter 6 – Fundamentals on Speckle

6.1 – Diffraction by a random distribution of index:

When the refractive index, including the profile of a surface between two homogeneous media, is randomly modulated at a scale close to the wavelength, diffraction by that medium is itself random and produces, under coherent illumination, a characteristic phenomenon known as "speckle" (en français "tavelure", peu usité). Speckle is always generated by a "diffuser". The word scattering is used in this context (rather than "diffusion", which refers to other phenomena such as one of the modes for heat transfer). It designates random interferences or, which is essentially the same, diffraction by a set of random structures. (En français, on utilise le mot "diffusion", mot sujet à confusion en raison de ses diverses significations).

While speckle is an interferences phenomenon, its appearance differs strongly from interferences in an interferometer due to its random nature. Indeed, the usual interference fringes are replaced by a set of more or less bright spots on a dark background (as we shall see, the most common case is really bright spots on a dark background as opposed to dark spots on a bright background)

Speckle may be observed in transmission (through ground glass, through fog) or reflection (on most everyday life objects when illuminated by a laser). It is practical to introduce the following properties of speckle.

- A diffuser may be "strong", when all the light received is scattered, or "weak", when part can go through or be reflected unperturbed, i.e. the "zero order" in the sense of diffractive optics, usually called "specular light" (Latin speculum, mirror) in the present context, can be observed. Speckle created by a strong diffuser has specific properties and is termed "fully developed".
- "Fine" and "coarse" diffusers differ by the spatial scale of index fluctuations. We shall see that this has a strong effect on the speckle patterns.
- Speckle may be observed in Fresnel diffraction, in Fraunhofer diffraction ("at infinity" or in the vicinity of the convergence point of a spherical illuminating wave), or in the image of a diffuser when the diffuser is so fine that its details are not individually resolved in the image.

This brief list hints at the fact that speckle "grain" size and the grain size of the diffuser that generates it are two different things. Our analysis below will highlight the distinction.

A small number of random index fluctuation diffracting light are sufficient to create a speckle pattern. We shall nevertheless be mostly interested in the more common case of rather large numbers. From the central limit theorem (français "théorème de la limite centrale"), it then results that the complex amplitude in the speckle pattern obeys Gaussian statistics, a rather strong statement that induces well identified properties, as we shall see.

Everyday life provides many cases of situations generating speckle under sufficiently coherent illumination: atmospheric turbulence, mist on eyeglasses when entering a humid, warm room from outside by cold weather, ground glass, paper, diffusers placed around luminaries to avoid dazzle, a dirty windshield at night. The observer often does not pay attention because partial coherence restricts the speckle contrast and because he is interested in features of the scene observed that are meaningful to him/her, while speckle is not. The interference of modes in a multimode optic fiber, likewise, creates speckle. In spite of a few early observations, speckle has been a subject of interest to scientists mostly after the advent of the laser.

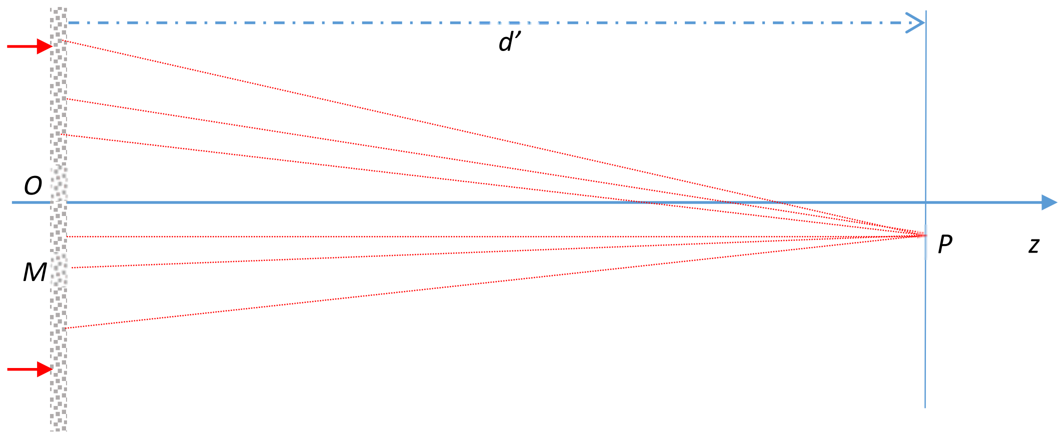
Speckle is a rich and rather complex phenomenon, as can be seen from the excellent books by Dainty (Laser Speckle and Related Phenomena, Springer, 1975, an early synthesis that remains completely relevant) and by Goodman (Speckle Phenomena in Optics, Roberts Publishing, 2006). The present chapter is appropriate for a short introduction. A more expanded course, but still short compared to those books, taught in French a few years ago by the same instructor is available for download on Libres Savoirs and Claroline.

An introductory exercise to Gaussian statistics: calculate and plot the probability distribution of the sum of two unbiased dice, three unbiased dice, four unbiased dice. Compared to the Gaussian distribution that has the same mean and variance. The central limit theorem claims that the sum of dice converges towards the Gaussian when the number of dice tends to infinity. This short exercise shows that convergence is in fact rather fast.

In the following, the main properties of speckle patterns by a stationary diffuser illuminated by a coherent wave on a given surface *S* are described. By "stationary", one should qualitatively understand that the diffuser shows the same properties at every point of surface *S*: for example, there is no region in S where the diffuser grains are more densely packed or coarser than elsewhere: for the observer, the diffuser "looks the same" on the whole surface *S*.

6.2 – Statistics of the illuminance at one point in a speckle pattern

Let us consider a laser beam illuminating a diffuser in transmission, and investigate the statistics of the complex amplitude and illuminance at an arbitrary point *P* on a screen some distance *d'* away from the diffuser. The basic effect is that *P* receives light diffracted by the whole object, or at least the whole part of the diffuser that contributes to some extend to light arriving at *P*. Because the very process of scattering involves a random phase relation between any two points on the diffuser, those are random interferences. Obviously, from previous chapters, the laws of diffraction determine the amplitude at *P* through some integral. But the main properties of the illuminance statistics at the given point *P* can be found without writing the integral by application of the central limit theorem, that states that the statistics at *P* are Gaussian.



If the complex amplitude obeys Gaussian statistics, there may or may not be a privileged phase – as in the obvious particular case of Fraunhofer diffraction by a diffuser where only the modulus, not the phase, is modulated. If that is not the case, the statistics are Gaussian circular, i.e. the real and imaginary part of the complex amplitude on the screen obey identical statistics <u>and</u> are uncorrelated. That case is remarkable and simple. In this section, the statistics of the complex amplitude *A* and of the illuminance $E = |A|^2$ <u>at one point</u> will be given.

$$\mathrm{ddp}_A(z) = \frac{1}{\pi A_o^2} \exp \frac{-|z|^2}{A_o^2} \qquad (7.1)$$

$$\mathrm{ddp}_{\mathcal{E}}(E) = \frac{1}{\mathcal{E}_o} \exp \frac{-E}{\mathcal{E}_o} \tag{7.2}$$

with $\mathcal{E}_o = \left\langle |A_o|^2 \right\rangle$. (7.3)

From there, the illuminance variance follows straightforwardly:

$$\left\langle (E - \mathcal{E}_o)^2 \right\rangle = \mathcal{E}_o^2 . \tag{7.4}$$

In words, that implies that the "signal to noise ratio" is unity, where the "signal" is the average (mathematical expectation) of the illuminance and the noise its rms fluctuation. This quite strong kind of speckle occurs when there is no direct light transmitted, i.e. for "fully developed" speckle.

Exercise: write the integral in the following case: the diffuser is a glass plate of refractive index n that has in a first stage been completely coated with an absorbing layer. Next, fairly small holes ("delta peaks") have been punched in the layer, eliminating the absorbing layer and going a random depth h into the glass, where the statistics of h are uniformly distributed between 0 and $\frac{\lambda}{n-1}$*. Using the central limit theorem, calculate the probability density function for the complex amplitude on the screen. This fairly simple but artificial model gives the correct law for the complex amplitude but is unable to provide an estimate for the speckle grain size.*

6.3 – Two-points statistics in a speckle pattern:

The most striking characteristics of a speckle pattern is that it consists of small bright dots on a dark background. The size of those points can be estimated by comparing the complex amplitude at one point *P* with the complex amplitude at another point *P'* nearby. Here we designate by $\mathbf{r} - \delta\mathbf{r}/2$ and by $\mathbf{r} + \delta\mathbf{r}/2$, respectively, the coordinates of *P* and *P'* on the screen plane. That "comparison" takes preferably the form of a correlation: "how correlated is complex amplitude at those two points?" Obviously, if the distance ***δr*** between the two points tends to zero, then they are fully correlated, and if it tends to infinity, they are independent. The correlation curve width versus ***δr*** is an appropriate definition for the "grain size".

Writing the Fresnel diffraction integral leads to:

$$\left\langle A^*(\mathbf{r} - \delta\mathbf{r}/2, d')\, A(\mathbf{r} + \delta\mathbf{r}/2, d') \right\rangle = \mathcal{E}_o(\mathbf{r}) \frac{\tilde{P}\left(\frac{\delta\mathbf{r}}{\lambda d'}\right)}{\tilde{P}(0)}, \tag{7.5}$$

where P designates the pupil limiting the diffuser. In words, **the autocovariance of a speckle pattern complex amplitude is identical to the Fraunhofer diffraction pattern, in the observation plane, that would be created by the empty pupil suitably illuminated** (i.e., illuminated by a spherical wave converging on the screen). It is appropriate to stress that this result does not assume that the initial laser beam is such a spherical wave.

An important associated result is the behavior of the average illuminance $\mathcal{E}_o$. The finer the diffuser grain, the broader the area illuminated by diffusion (i.e., the broader the speckle pattern taken globally, as opposed to the speckle grains) and conversely. Analytically, for a diffuser illuminated by a laser beam converging on the screen (an assumption that was not required in the previous paragraph!), the law is given by the autocorrelation of the complex amplitude on the exit face of the diffuser (plane *z*=0):

$$\mathcal{E}_o(\mathbf{r}, d') = \mathrm{TF}\left(\left\langle A^*(\boldsymbol{\rho} - \delta\boldsymbol{\rho}/2, 0)\, A(\boldsymbol{\rho} + \delta\boldsymbol{\rho}/2, 0) \right\rangle\right)\left(\frac{\mathbf{r}}{\lambda d'}\right), \tag{7.6}$$

where the Fourier transform is with respect to variable ***δr*** and it has been assumed that the diffuser is stationary (see the introduction). $\mathcal{E}_o\left(\mathbf{r},d'\right)$ is often called the speckle pattern "envelope" (although that vocabulary is not mathematically strictly appropriate). One extreme case is when the diffuser grain is too small to be resolved and can be assimilated to a Dirac delta peak. Its Fourier transform, the power spectrum, is then uniform and, by analogy with the spectrum of white light, is called "white noise". Mathematically, white noise extending to infinity in the Fourier domain (here, variable ***r***) is not possible, but physically, noise can be white at the scale of the area observed. A "white" diffuser therefore creates a speckle pattern which is statistically uniform on the observed area, i.e. "stationary".

Comments:

- in statistics courses, the Fourier transform of an autocorrelation is called "power spectrum", therefore the speckle "envelope" is the power spectrum of the diffuser complex amplitude.
- Indeed, for a thin diffuser, the complex amplitude at the exit face $A\left(\boldsymbol{\rho},0\right)$ is just the product of the transmittance or reflectance of the diffuser and the illuminating wave.

The properties that have been summarize above show that the speckle is essential to understand the appearance of scattering objects. That is true even in incoherent illumination. Besides, speckle is a powerful measurement method.

Summary

1. "Diffusers" consist of a set of randomly distributed "grains". Those grains may in some cases be all identical, but they are usually described by an autocovariance that allows to define the diffuser "average grain".
2. The diffuser "average grain" has a Fraunhofer diffraction pattern, which is its Fourier transform. In Fraunhofer diffraction (spherical wave illumination), the speckle pattern follows the shape of that diffraction pattern, its "envelope".
3. The speckle grain size is statistically given through the complex amplitude autocovariance on the screen and is the Fraunhofer diffraction pattern of the diffuser pupil.

The following diagram recapitulates those properties.

Diffuser consisting of grains of size g and illuminated through a pupil of size a.	Fraunhofer diffraction →	Speckle grains with an average angular size $\frac{\lambda}{a}$ in a speckle pattern of global size $\frac{\lambda}{g}$

ANNEXES

Some useful Fourier transforms

Definition: $f(x)$	$\tilde{f}(\mu) = \int_{\mathbb{R}} f(x) \exp(-2i\pi\mu x)\, dx$	(A.1)
$\exp(-\pi x^2)$	$\exp(-\pi\mu^2)$	(A.2)
$\exp(i\pi x^2)$	$\exp(-i\pi\mu^2 + i\pi/4)$	(A.3)

Index