WAVE OPTICS
(FOURIER OPTICS)

ARNAUD DUBOIS

October 2012
References


Light is an electromagnetic wave phenomenon. Electromagnetic radiations are classified in different categories, depending on their frequency. Light occupies a limited region of the electromagnetic spectrum. The range of optical wavelengths contains three bands: ultraviolet (10 to 390 nm), visible (390 to 760 nm), and infrared (760 nm to 1 mm). The corresponding range of optical frequencies stretches from $3 \times 10^{11}$ Hz to $3 \times 10^{16}$ Hz.
Since light is an electromagnetic wave, light can be described by the same theoretical principles that govern all forms of electromagnetic radiation. Electromagnetic radiation propagates in the form of two coupled vector waves, an electric-field wave and a magnetic-field wave.

Although light is described by two wave vectors, it is possible to describe many optical phenomena using a scalar wave theory. In this theory, light is described by a single scalar wavefunction. This approximate way of treating light is called scalar wave optics, or simply wave optics.

When light waves propagate through and around objects whose dimensions are much greater than the wavelength, the wave nature of light is not discernable. In that case, the behaviour of light can be described by rays. This model of light, called ray optics, obeys a set of geometrical rules. Theoretically, ray optics is the limit of wave optics when the wavelength is zero. However, the wavelength need not actually be equal to zero for the ray-optics theory to be useful. As long as the light waves propagate through and around objects whose dimensions are much greater than the wavelength, the ray theory suffices for describing most phenomena. Because the wavelength of visible light is much shorter than the dimensions of the visible objects encountered in our daily lives, manifestations of the wave nature of light are not apparent without careful observation.

Thus, the electromagnetic theory of light encompasses wave optics, which in turn, encompasses ray optics. Ray optics and wave optics provide approximate models of light which derive their validity from their successes in producing results that approximate those based on rigorous electromagnetic theory.

Although electromagnetic optics provides the most complete treatment of light within the confines of classical optics, there are certain optical phenomena that are characteristically quantum mechanical in nature and cannot be explained classically (spectrum of blackbody radiation, absorption and emission of light by matter at specific wavelengths, etc...). These phenomena are described by a quantum electromagnetic theory known as quantum electrodynamics. For optical phenomena, this theory is also referred to as quantum optics.

Historically, optical theory developed roughly in the following sequence: (1) ray optics; (2) wave optics; (3) electromagnetic optics; (4) quantum optics. Not surprisingly, these models are progressively more difficult and sophisticated, having been developed to provide explanations for the outcomes of successively more complex and precise optical experiments.
Theory of quantum optics provides an explanation of virtually all optical phenomena. The electromagnetic theory of light provides the most complete treatment of light within the confines of classical optics. Wave optics is a scalar approximation of electromagnetic optics. Ray optics is the limit of wave optics when the wavelength is very short.
Chapter 1: INTRODUCTION TO WAVE OPTICS

Wave optics is a scalar approximation of electromagnetic optics. The vectorial nature of light is therefore ignored in the wave optics theory. Wave optics can be used for describing many optical phenomena that cannot be described by ray optics, including interference and diffraction. However, wave optics cannot describe optical phenomena that require a vector formulation, such as polarization effects. The wave optics theory is also not capable of providing a complete picture of reflection and refraction of light at the boundaries between dielectric materials.

1. POSTULATES OF WAVE OPTICS

The theory of wave optics can be built from a set of postulates. These postulates can be justified by the electromagnetic theory of light:

- Light propagates in the form of waves.
- In free space, light waves travel with speed $c_0$.
- A homogeneous transparent medium (such as glass) is characterized by a single constant, its refractive index $n \ (\geq 1)$. In a medium of refractive index $n$, light waves travel with a reduced speed $c = c_0 / n$.
- An optical wave is described mathematically by a real function of position $\mathbf{r} = (x, y, z)$ and time $t$, denoted $u(\mathbf{r}, t)$ and known as the wavefunction. It satisfies the wave equation:

$$\nabla^2 u - \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} = 0$$

where $\nabla^2$ is the Laplacian operator. It is a scalar differential operator. In Cartesian coordinates, it is defined as $\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}$.

Any function satisfying the wave equation represents a possible optical wave.

Because the wave equation is linear, the principle of superposition applies; i.e., if $u_1(\mathbf{r}, t)$ and $u_2(\mathbf{r}, t)$ represent optical waves, then $u(\mathbf{r}, t) = u_1(\mathbf{r}, t) + u_2(\mathbf{r}, t)$ also represents a possible optical wave.
The wave equation is approximately applicable to media with position-dependent refractive indices, provided that the variation is slow within distances of a wavelength. The medium is then said to be locally homogeneous. For such media, $n$ and $c$ are simply replaced by position-dependent functions $n(\mathbf{r})$ and $c(\mathbf{r})$, respectively.

2. MONOCROMATIC WAVES

In the wave optics theory, a monochromatic wave is represented by a wavefunction with harmonic time dependence,

$$u(\mathbf{r},t) = a(\mathbf{r}) \cos[2\pi vt + \varphi(\mathbf{r})],$$

where

- $a(\mathbf{r}) = \text{amplitude}$
- $\varphi(\mathbf{r}) = \text{phase}$
- $\nu = \text{frequency (cycles/s or Hz)}$
- $\omega = 2\pi \nu = \text{angular frequency (radians/s)}$.

Both the amplitude and the phase are generally position dependent, but the wavefunction is a harmonic function of time with frequency $\nu$ at all positions. As seen in the introduction, the frequency of optical waves lies in the range $3 \times 10^{11}$ to $3 \times 10^{16}$ Hz.

2.1 Complex Wavefunction

It is usual and convenient to represent the real wavefunction $u(\mathbf{r},t)$ in terms of a complex function

$$U(\mathbf{r},t) = a(\mathbf{r}) \exp[j\varphi(\mathbf{r})] \exp(j2\pi vt),$$

so that

$$u(\mathbf{r},t) = \text{Re}\{U(\mathbf{r},t)\} = 1/2\left[U(\mathbf{r},t) + U'(\mathbf{r},t)\right].$$

The function $U(\mathbf{r},t)$, known as the complex wavefunction, describes the wave completely; the wavefunction $u(\mathbf{r},t)$ is simply its real part. Like the wavefunction $u(\mathbf{r},t)$, the complex wavefunction $U(\mathbf{r},t)$ must also satisfy the wave equation

$$\nabla^2 U - \frac{1}{c^2} \frac{\partial^2 U}{\partial t^2} = 0.$$

2.2 Complex Amplitude

The complex wavefunction can be written in the form
\[ U(\mathbf{r}, t) = U(\mathbf{r}) \exp(j2\pi vt), \]

where the time-independent factor \( U(\mathbf{r}) = a(\mathbf{r}) \exp[j \varphi(\mathbf{r})] \) is referred to as the complex amplitude. The wavefunction \( u(\mathbf{r}, t) \) is therefore related to the complex amplitude by

\[ u(\mathbf{r}, t) = \text{Re}\{U(\mathbf{r}) \exp(j2\pi vt)\} \]

At a given position \( \mathbf{r} \), the complex amplitude \( U(\mathbf{r}) \) is a complex variable whose magnitude \( |U(\mathbf{r})| = a(\mathbf{r}) \) is the amplitude of the wave and whose argument \( \arg\{U(\mathbf{r})\} = \varphi(\mathbf{r}) \) is the phase.

### 2.3 The Helmholtz Equation

The monochromatic wave must satisfy the wave equation. Substituting \( U(\mathbf{r}, t) = U(\mathbf{r}) \exp(j2\pi vt) \) into the wave equation, we obtain the differential equation

\[ \left( \nabla^2 + k^2 \right) U(\mathbf{r}) = 0 \]

called the Helmholtz equation, where

\[ k = 2\pi v/c = \omega/c \]

is referred to as the wavenumber.

### 2.4 Optical Intensity

The optical intensity of a monochromatic wave is defined as the absolute square of its complex amplitude

\[ I(\mathbf{r}) = |U(\mathbf{r})|^2 \]

The intensity of a monochromatic wave does not vary with time.

### 2.5 Wavefronts

The wavefronts are the surfaces of equal phase, \( \varphi(\mathbf{r}) = \) constant. The constants are often taken to be multiples of \( 2\pi \), \( \varphi(\mathbf{r}) = 2\pi q \), where \( q \) is an integer.
Summary

- A monochromatic wave of frequency $\nu$ is described by a complex wavefunction $U(\mathbf{r}, t) = U(\mathbf{r}) \exp(j2\pi\nu t)$, which satisfies the wave equation.

- The complex amplitude $U(\mathbf{r})$ satisfies the Helmholtz equation; its magnitude $|U(\mathbf{r})|$ and argument $\arg\{U(\mathbf{r})\}$ are the amplitude and phase of the wave, respectively. The optical intensity is $I(\mathbf{r}) = |U(\mathbf{r})|^2$. The wavefronts are the surfaces of constant phase, $\varphi(\mathbf{r}) = \arg\{U(\mathbf{r})\} = 2\pi q$ ($q$ integer).

- The wavefunction $u(\mathbf{r}, t)$ is the real part of the complex wavefunction, $u(\mathbf{r}, t) = \text{Re}\{U(\mathbf{r}, t)\}$. The wavefunction also satisfies the wave equation.

3. ELEMENTARY WAVES

We consider here two special monochromatic waves: the plane wave and the spherical wave. These waves are the simplest solutions to the Helmholtz equation in a homogeneous medium.

3.1 The Plane Wave

The plane wave has a complex amplitude

$$U(\mathbf{r}) = A \exp(-j\mathbf{k} \cdot \mathbf{r}) = A \exp\left[-j(k_x x + k_y y + k_z z)\right],$$

where $A$ is a constant called the complex envelope, and $\mathbf{k} = (k_x, k_y, k_z)$ is called the wavevector.

For the plane wave to satisfy the Helmholtz equation, $k_x^2 + k_y^2 + k_z^2 = k^2$, so that the magnitude of the wavevector $\mathbf{k}$ is the wavenumber $k$.

Since the phase $\arg\{U(\mathbf{r})\} = \arg\{A\} - \mathbf{k} \cdot \mathbf{r}$, the wavefronts obey $\mathbf{k} \cdot \mathbf{r} = k_x x + k_y y + k_z z = 2\pi q + \arg\{A\}$ ($q$ integer). This is the equation describing parallel planes perpendicular to the wavevector $\mathbf{k}$ (hence the name "plane wave"). These planes are separated by a distance $\lambda = 2\pi / k$, so that

$$\lambda = c / \nu$$

where $\lambda$ is called the wavelength.
The plane wave has a constant intensity $I(r) = |A|^2$ everywhere in space so that it carries infinite power. This wave is clearly an idealization since it exists everywhere and at all times.

If the $z$ axis is taken in the direction of the wavevector $\mathbf{k}$, then $U(r) = A \exp(-j k z)$ and the corresponding wavefunction is

$$u(r, t) = |A| \cos \left[ 2\pi \nu t - k z + \arg \{A\} \right] = |A| \cos \left[ 2\pi \nu (t - z/c) + \arg \{A\} \right]$$

The wavefunction is therefore periodic in time with period $1/\nu$, and periodic in space with period $2\pi/k$, which is equal to the wavelength $\lambda$. Since the phase of the complex wavefunction, $\arg \{U(r, t)\} = 2\pi \nu (t - z/c) + \arg \{A\}$, varies with time and position as a function of the variable $(t - z/c)$ (see Fig), $c$ is called the phase velocity of the wave. $u(z, t)$

A plane wave travelling in the $z$ direction is a periodic function of $z$ with spatial period $\lambda$ and a periodic function of $t$ with temporal period $1/\nu$.

In a medium of refractive index $n$:

- the phase velocity is $c = c_0 / n$

- the wavelength is $\lambda = c / \nu = c_0 / n \nu$, so that $\lambda = \lambda_0 / n$ where $\lambda_0 = c_0 / \nu$ is the wavelength in free space. For a given frequency $\nu$, the wavelength in the medium is reduced relative to that in free space by the factor $n$.

- as a consequence, the wavenumber $k = 2\pi / \lambda$ increases relative to that in free space ($k_0 = 2\pi / \lambda_0$) by the factor $n$.

In summary: as a monochromatic wave propagates through media of different refractive indices its frequency remains the same, but its velocity, wavelength, and wavenumber are modified.

### 3.2 The Spherical Wave

Another simple solution of the Helmholtz equation is the spherical wave. The complex amplitude of a spherical wave is

$$U(\mathbf{r}) = (A/r) \exp(-jkr)$$
where $r$ is the distance from the origin and $k = 2\pi v / c = \omega / c$ is the wavenumber.

The intensity of a spherical wave, $I(r) = |A|^2 / r^2$, is inversely proportional to the square of the distance. Taking $\arg \{A\} = 0$ for simplicity, the wavefronts are the surfaces $kr = 2\pi q$ or $r = q\lambda$, where $q$ is an integer. These are concentric spheres separated by a radial distance $\lambda = 2\pi / k$ that advance radially at the phase velocity $c$ (see Figure).

A spherical wave originating at the position $r_0$ has a complex amplitude $U(r) = (A/|r - r_0|)\exp[-jk|r - r_0|]$. Its wavefronts are spheres centered about $r_0$.

Note that a wave with complex amplitude $U(r) = (A/r)\exp(+jkr)$ is a spherical wave traveling toward the origin instead of away from the origin.

### 3.4 Fresnel Approximation of the Spherical Wave; The Paraboloidal Wave

Let us consider a spherical wave at points $r = (x, y, z)$ sufficiently close to the $z$ axis but far from the origin, so that $(x^2 + y^2)^{1/2} \ll z$.

We note $\theta^2 = (x^2 + y^2) / z^2$. By using an approximation based on the Taylor series expansion:

$$r = (x^2 + y^2 + z^2)^{1/2} = z(1 + \theta^2)^{1/2} = z\left(1 + \frac{\theta^2}{2} + \frac{\theta^4}{8} + \ldots\right) \approx z\left(1 + \frac{\theta^2}{2}\right) = z + \frac{x^2 + y^2}{2z}$$

Substituting $r = z + \frac{x^2 + y^2}{2z}$ into the phase, and $r = z$ into the magnitude of $U(r)$, we obtain
A more accurate value of \( r \) was used in the phase since the sensitivity to errors of the phase is greater.

This is called the **Fresnel approximation**. This approximation plays an important role in simplifying the theory of transmission of optical waves through apertures (theory of diffraction).

With the Fresnel approximation, the complex amplitude of a spherical wave may be viewed as representing a plane wave \( A\exp(-jkz) \) modulated by the factor \( \left( \frac{1}{z} \right) \exp\left(-jk \frac{x^2 + y^2}{2z} \right) \), which involves a phase \( k \left( \frac{x^2 + y^2}{2z} \right) \). This phase factor serves to bend the planar wavefronts of the plane wave into paraboloidal surfaces (see Fig.), since the equation of a parabola of revolution is \( \frac{x^2 + y^2}{z} = \text{constant} \). Thus the spherical wave is approximated by a **paraboloidal wave**. When \( z \) becomes very large, the phase approaches \( kz \) and the magnitude varies slowly with \( z \), so that the spherical wave eventually resembles the plane wave \( \exp(-jkz) \), as illustrated below.

A spherical wave can be approximated at points near the \( z \) axis and sufficiently far from the origin by a paraboloidal wave. For very far points, the spherical wave approaches the plane wave.

The condition of validity of the Fresnel approximation is not simply that \( \theta^2 / 2 \ll 1 \). Although the third term of the series expansion, \( \theta^4 / 8 \) may be very small in comparison with the second and first terms, when multiplied by \( kz \) it may become comparable to \( \pi \). The approximation is therefore valid when \( kz\theta^4 / 8 \ll \pi \), or \( (x^2 + y^2)^2 \ll 4z^3\lambda \). For points \((x, y)\) lying within a circle of radius \( a \) centered about the \( z \) axis, the validity condition is \( a^4 \ll 4z^3\lambda \) or

\[
\frac{N \theta_m^2}{4} \ll 1
\]

where \( \theta_m = a / z \) is the maximum angle and
is known as the Fresnel number.

**EXERCISE 1:** Validity of the Fresnel Approximation.

Determine the radius of a circle within which a spherical wave of wavelength $\lambda = 633$ nm, originating at a distance 1 m away, may be approximated by a paraboloidal wave. Determine the maximum angle $\theta_m$ and the Fresnel number $N_F$.

4. TRANSMISSION THROUGH OPTICAL COMPONENTS

We now proceed to examine the transmission of optical waves through transparent optical components such as plates and lenses. The effect of reflection at the surfaces of these components will be ignored, since it cannot be properly described using the scalar wave optics model of light. The effect of absorption in the material is also ignored. The main emphasis here is on the phase shift introduced by these components and on the associated wavefront bending.

4.1 Transparent Plate of constant thickness

Consider first the transmission of a plane wave through a transparent plate of refractive index $n$ and thickness $d$ surrounded by free space. The surfaces of the plate are the planes $z = 0$ and $z = d$ and the incident wave travels in the $z$ direction. Let $U(x, y, z)$ be the complex amplitude of the wave. Since external and internal reflections are ignored, $U(x, y, z)$ is assumed to be continuous at the boundaries. The ratio $t(x, y) = U(x, y, d)/U(x, y, 0)$ therefore represents the complex amplitude transmittance of the plate. The incident plane wave continues to propagate inside the plate as a plane wave with wavenumber $nk_0$, so that $U(x, y, z)$ is proportional to $\exp(-jnk_0d)$. Thus $U(x, y, d)/U(x, y, 0) = \exp(-jnk_0d)$, so that

$$t(x, y) = \exp(-jnk_0d)$$

i.e., the plate introduces a phase shift $nk_0d = 2\pi (d/\lambda)$.
4.2 Thin Transparent Plate of Varying Thickness

We consider now a thin transparent plate whose thickness $d(x, y)$ varies smoothly as a function of $x$ and $y$. The incident wave is a paraxial wave. The plate lies between the planes $z = 0$ and $z = d_0$, which are regarded as the boundaries of the optical component.

The wave crosses a thin plate of width $d(x, y)$ surrounded by thin layers of air of total width $d_0 - d(x, y)$. The local transmittance is the product of the transmittances of a thin layer of air of thickness $d_0 - d(x, y)$ and a thin layer of material of thickness $d(x, y)$ so that

$$t(x, y) = \exp[-jk_0d(x, y)]\exp[-jk_0\{d_0 - d(x, y)\}],$$

from which

$$t(x, y) = h_0 \exp[-j(n-1)k_0d(x, y)]$$

where $h_0 = \exp[-jk_0d_0]$ is a constant phase factor. This relation is valid in the paraxial approximation (all angles $\theta$ are small) and when the thickness $d_0$ is sufficiently small.

A transparent plate of varying thickness.

4.3 Thin Lens

The general expression for the complex amplitude transmittance of a thin transparent plate of variable thickness is now applied to the plano-convex thin lens.
Since the lens is the cap of a sphere of radius \( R \), the thickness at the point \((x, y)\) is \( d(x, y) = d_0 - PQ = d_0 - (R - QC) \) or

\[
d(x, y) = d_0 - \left( R - \left[ R^2 - (x^2 + y^2) \right]^{1/2} \right)
\]

This expression can be simplified by considering only points for which \( x \) and \( y \) are sufficiently small in comparison with \( R \), so that \( x^2 + y^2 \ll R^2 \). In that case

\[
\left[ R^2 - (x^2 + y^2) \right]^{1/2} = R \left( 1 - \frac{x^2 + y^2}{R^2} \right)^{1/2} \approx R \left( 1 - \frac{x^2 + y^2}{2R^2} \right)
\]

and

\[
d(x, y) \approx d_0 - \frac{x^2 + y^2}{2R}
\]

We obtain then

\[
t(x, y) \approx h_0 \exp \left( jk_0 \frac{x^2 + y^2}{2f} \right)
\]

where

\[
f = \frac{R}{n - 1}
\]

is the focal length of the lens and \( h_0 = \exp[-jnk_0d_0] \) is a constant phase factor that is usually of no significance.

The lens imparts to the incident wave a phase proportional to \( x^2 + y^2 \); it bends the planar wavefronts of a plane wave, transforming it into a paraboloidal wave centered at a distance \( f \) from the lens.
Double-Convex Lens: the complex amplitude transmittance of the double-convex lens is given by

\[ t(x, y) \approx \exp \left( jk_0 (n-1) \frac{x^2 + y^2}{2R_1} \right) \times \exp \left( jk_0 (n-1) \frac{x^2 + y^2}{-2R_2} \right) \]

\[ t(x, y) \approx b_0 \exp \left( jk_0 \frac{x^2 + y^2}{2f} \right) \quad \text{with} \quad \frac{1}{f} = (n-1) \left( \frac{1}{R_1} - \frac{1}{R_2} \right) \]

Recall that, by convention, the radius of a convex/concave surface is positive/negative, i.e., \( R_1 \) is positive and \( R_2 \) is negative for the lens in Figure. The parameter \( f \) is recognized as the focal length of the lens.

**General case:** In the Fresnel approximation, the transmittance of a thin lens can be simply written:

\[ t(x, y) = \exp \left( jk_0 \frac{x^2 + y^2}{2f} \right) \]

where \( f \) is the focal length of the lens. A lens is therefore characterized by its focal length only, regardless of its geometrical shape.

**EXERCISE 2:** Focusing of a plane wave by a thin lens.

Show that when a plane wave is transmitted through a thin lens of focal length \( f \) in a direction parallel to the axis of the lens, it is converted into a paraboloidal wave (the Fresnel approximation of a spherical wave) centered about a point at a distance \( f \) from the lens.

**EXERCISE 3:** Imaging property of a Lens.

We consider a point source \( P_1 \) emitting a paraboloidal wave. What is the wave just after the lens of focal length \( f \)? Where is the image point \( P_2 \)? Check the Descartes’s law of Ray Optics.
EXERCISE 4: Show that a thin plate of uniform thickness $d_0$ and quadratically graded refractive index

$$n(x, y) = n_0 \left[ 1 - \frac{1}{2} \alpha^2 \left( x^2 + y^2 \right) \right],$$

where $\alpha d_0 \ll 1$, acts as a converging lens. Express the focal length $f$. 

![Diagram of a converging lens with focal length $f$ and thickness $d_0$.]
Chapter 2: FOURIER OPTICS

1. INTRODUCTION

Fourier optics provides a description of the propagation of light waves based on harmonic analysis and linear systems.

Harmonic analysis is based on the expansion of an arbitrary function of time as a sum of harmonic functions of time of different frequencies and different amplitudes:

\[
f(t) = \int_{-\infty}^{\infty} \tilde{f}(\nu) \exp(-j2\pi\nu t) d\nu
\]

An arbitrary function \( f(t) \) can be analyzed as a sum of harmonic functions of different frequencies and complex amplitudes.

The function \( \tilde{f}(\nu) \exp(-j2\pi\nu t) \) is a harmonic function with frequency \( \nu \) and complex amplitude \( \tilde{f}(\nu) \).

The complex amplitude \( \tilde{f}(\nu) \), as a function of frequency, is called the Fourier transform of \( f(t) \):

\[
\tilde{f}(\nu) = \int_{-\infty}^{\infty} f(t) \exp(j2\pi\nu t) dt
\]

This approach is useful for the description of linear systems. If the response of the system to each harmonic function is known, the response to an arbitrary input function can be determined (principle of superposition).

The response of a linear system can therefore be determined by the use of harmonic analysis at the input and superposition at the output.
Consider now an arbitrary function $f(x, y)$ of the two variables $x$ and $y$. We suppose that $x$ and $y$ represent the spatial coordinates in a plane. $f(x, y)$ can also be written as a sum of harmonic functions of $x$ and $y$:

$$f(x, y) = \int \int \tilde{f}(\nu_x, \nu_y) \exp[-j2\pi(\nu_x x + \nu_y y)] d\nu_x d\nu_y$$

$\tilde{f}(\nu_x, \nu_y)$ is the complex amplitude of the harmonic functions, $\nu_x$ and $\nu_y$ are the spatial frequencies in the $x$ and $y$ directions, respectively.

The harmonic function $\tilde{f}(\nu_x, \nu_y) \exp[-j2\pi(\nu_x x + \nu_y y)]$ is the building block of the Fourier Optics theory. It can be used to generate an arbitrary function of two variables $f(x, y)$, as illustrated below. The weights $\tilde{f}(\nu_x, \nu_y)$ of the different harmonic functions are determined by use of the two-dimensional Fourier transform

$$\tilde{f}(\nu_x, \nu_y) = \int \int f(x, y) \exp[j2\pi(\nu_x x + \nu_y y)] dx dy$$

An arbitrary function $f(x, y)$ can be analyzed as a sum of harmonic functions of different frequencies and complex amplitudes.

Consider a plane wave $U(x, y, z) = A \exp[-j(k_x x + k_y y + k_z z)]$. In an arbitrary plane $z$, $U(x, y, z)$ is a harmonic function. In the $z = 0$ plane, for example, $U(x, y, 0)$ is the harmonic function $A \exp[-j2\pi(\nu_x x + \nu_y y)]$ where $\nu_x = k_x / 2\pi$ and $\nu_y = k_y / 2\pi$ are the spatial frequencies.

Principle of Fourier optics: an arbitrary wave can be analyzed as a superposition of planes waves.
Since an arbitrary function \( f(x, y) \) can be analyzed as a sum of harmonic functions, an arbitrary optical wave \( U(x, y, z) \) may be analyzed as a sum of plane waves.

If it is known how a linear optical system modifies plane waves, the principle of superposition can be used to determine the effect of the system on an arbitrary wave.

It is therefore useful to describe the propagation of light through linear optical components, including free space (free space is a linear system because the wave equation is linear). The complex amplitudes in two planes normal to the optic (z) axis are regarded as the input and output of the system.

The transmission of an optical wave \( U(x, y, z) \) through an optical system between an input plane \( z = 0 \) and an output plane \( z = d \). This is regarded as a linear system whose input and output are the functions \( f(x, y) = U(x, y, 0) \) and \( g(x, y) = U(x, y, d) \), respectively.

## 2. PROPAGATION OF LIGHT IN FREE SPACE

### 2.1 Correspondence Spatial Harmonic Function / Plane Wave

Consider a plane wave of complex amplitude \( U(x, y, z) = A \exp \left[ -j \left( k_x x + k_y y + k_z z \right) \right] \) with wavevector \( k = (k_x, k_y, k_z) \), wavelength \( \lambda \), wavenumber \( k = (k_x^2 + k_y^2 + k_z^2)^{1/2} = 2\pi/\lambda \), and complex envelope \( A \). The vector \( k \) makes angles \( \theta_x = \sin^{-1}(k_x/k) \) and \( \theta_z = \sin^{-1}(k_z/k) \) with the y-z and x-z planes, respectively, as illustrated below.
The complex amplitude in the $z = 0$ plane, $U(x,y,0)$, is a spatial harmonic function $f(x,y) = A \exp[-j2\pi(n_x x + n_y y)]$ with spatial frequencies $n_x = k_x / 2\pi$ and $n_y = k_y / 2\pi$ (cycles/mm). The angles of the wavevector are therefore related to the spatial frequencies of the harmonic function by

$$
\begin{align*}
\theta_x &= \sin^{-1} \lambda n_x \\
\theta_y &= \sin^{-1} \lambda n_y
\end{align*}
$$

$\Lambda_x = 1/n_x$ and $\Lambda_y = 1/n_y$ are the periods of the harmonic function in the $x$ and $y$ directions. We see that the angles $\theta_x = \sin^{-1}(\lambda / \Lambda_x)$ and $\theta_y = \sin^{-1}(\lambda / \Lambda_y)$ are governed by the ratios of the wavelength of light to the period of the harmonic function in each direction. The wavefronts of the wave match to the periodic pattern of the harmonic function in the $z = 0$ plane.

If $k_x \ll k$ and $k_y \ll k$, so that the wavevector $k$ is paraxial, the angles $\theta_x$ and $\theta_y$ are small and

$$
\begin{align*}
\theta_x &= \lambda n_x \\
\theta_y &= \lambda n_y
\end{align*}
$$

Thus the angles of inclination of the wavevector are directly proportional to the spatial frequencies of the corresponding harmonic function.

### 2.2 Transmission through optical components

Consider a thin optical element with complex amplitude transmittance $f(x,y) = \exp[-j2\pi(n_x x + n_y y)]$. When a plane wave of unity amplitude traveling in the $z$ direction is transmitted through this optical element, the wave is modulated by the harmonic function, so that $U(x,y,0) = f(x,y)$. The incident wave is then converted into a plane wave with a wavevector at angles $\theta_x = \sin^{-1} \lambda n_x$, and $\theta_y = \sin^{-1} \lambda n_y$ (see Figure). The optical element is a diffraction grating which acts like a prism.
An arbitrary transmittance of an optical element $f(x, y)$ can be decomposed as a sum of several harmonic functions of different spatial frequencies. The transmitted optical wave is therefore also the sum of plane waves dispersed into different directions; each spatial frequency is mapped into a corresponding direction. The amplitude of each wave is proportional to the amplitude of the corresponding harmonic component of $f(x, y)$.

More generally, if $f(x, y)$ is a superposition integral of harmonic functions,

$$f(x, y) = \iint F(v_x, v_y) \exp\left[-j2\pi(v_x x + v_y y)\right] dv_x dv_y,$$

with frequencies $(v_x, v_y)$ and amplitudes $F(v_x, v_y)$, the transmitted wave $U(x, y, z)$ is the superposition of plane waves,

$$U(x, y, z) = \iint F(v_x, v_y) \exp\left[-j2\pi(v_x x + v_y y)\right] \exp(-jkz) dv_x dv_y,$$

with complex envelopes $F(v_x, v_y)$, where $k_z = \left(k_x^2 + k_y^2 - k_z^2\right)^{1/2} = 2\pi \left(1/\lambda^2 - v_x^2 - v_y^2\right)^{1/2}$. Note that $F(v_x, v_y)$ is the Fourier transform of $f(x, y)$.

Since an arbitrary function may be Fourier analyzed as a superposition integral, the light transmitted through an optical element of arbitrary transmittance may be written as a superposition of plane waves, provided that $v_x^2 + v_y^2 \leq \frac{1}{\lambda^2}$.

**EXERCISE 5: Diffraction grating**

Consider a diffraction grating with a sinusoidal amplitude transmittance: $t(x, y) = \frac{1}{2} \left[1 + \cos\left(2\pi x / p\right)\right]$.

Show that an incident plane wave (normal incidence) of wavelength $\lambda$ is converted into 3 plane waves. In which directions?
2.3 Transfer function of free space

As seen before, the complex amplitude of a plane wave can be written as

\[ U(x, y, z) = A \exp[-j(k_x x + k_y y + k_z z)] \]
\[ = A \exp[-j(k_x x + k_y y)] \exp[-jk_z z] \]
\[ = A \exp[-j2\pi(v_x x + v_y y)] \exp[-jk_z z] \]

where

\[ k_x = 2\pi v_x, k_y = 2\pi v_y. \]

A plane wave at \( z = d \) is equal to this plane wave at \( z = 0 \) (harmonic function \( A \exp[-j2\pi(v_x x + v_y y)] \)) multiplied by the phase factor \( \exp[-jk_z d] \).

Since \( k^2 = k_x^2 + k_y^2 + k_z^2 \) (Helmholtz)

\[ k_z = \pm \left(k^2 - k_x^2 - k_y^2 \right)^{1/2} = \pm 2\pi \left(1/\lambda^2 - v_x^2 - v_y^2 \right)^{1/2} \]

- If \( v_x^2 + v_y^2 \leq 1/\lambda^2 \), then \( k_z \) is real. The sign of \( k_z \) depends on the direction of propagation. Here we consider a propagation in the direction of increasing \( z \) (+ sign).

We can then write

\[ U(x, y, d) = U(x, y, 0) \times H(v_x, v_y) \]

where

\[ H(v_x, v_y) = \exp\left[-j2\pi\left(1/\lambda^2 - v_x^2 - v_y^2 \right)^{1/2} d\right] \]

\( H \) is called the transfer function of free space.

The magnitude of the transfer function \( |H(v_x, v_y)| = 1 \). The intensity of a plane wave is not modified during propagation, but it undergoes a phase change.

- If \( v_x^2 + v_y^2 > 1/\lambda^2 \), then \( k_z = \pm j2\pi \left(v_x^2 + v_y^2 - 1/\lambda^2 \right)^{1/2} \). The transfer function is now real: \( H(v_x, v_y) = \exp\left[\pm 2\pi \left(v_x^2 + v_y^2 - 1/\lambda^2 \right)^{1/2} d\right] \). The + sign gives an exponentially growing function, which is physically unacceptable. We have therefore
\[ H(v_x, v_y) = \exp \left[ -2\pi \left( v_x^2 + v_y^2 - \frac{1}{\lambda^2} \right)^{1/2} d \right] \], which represents an attenuation factor.

The wave is then called an evanescent wave. We may therefore regard \(1/\lambda\) as the cutoff frequency of the system. Features contained in spatial frequencies greater than \(1/\lambda\) (details smaller than \(\lambda\)) cannot be transmitted by an optical wave over distances much greater than the wavelength \(\lambda\). The spatial resolution of imaging systems in the “far field” is limited by the wavelength. Near-field optical imaging was demonstrated recently to achieve super-resolution (not limited by the wavelength).

**Fresnel approximation**

The expression for the transfer function may be simplified if the spatial frequencies are such that \(v_x^2 + v_y^2 \ll 1/\lambda^2\).

\[
\left( \frac{1}{\lambda^2} - v_x^2 - v_y^2 \right)^{1/2} = \frac{1}{\lambda} \left[ 1 - \lambda^2 \left( v_x^2 + v_y^2 \right) \right]^{1/2}
\]

\[
= \frac{1}{\lambda} \left[ 1 - \frac{\lambda^2 (v_x^2 + v_y^2)}{2} + \frac{\lambda^4 (v_x^2 + v_y^2)^2}{8} - \ldots \right]
\]

We can approximate the transfer function by

\[
H(v_x, v_y) \approx \exp(-jkd) \exp\left[ j\pi d \left( v_x^2 + v_y^2 \right) \right]
\]

In this approximation, the phase is a quadratic function of the spatial frequencies. This approximation is known as the **Fresnel approximation**.

The condition of validity of the Fresnel approximation is that \(2\pi \frac{1}{8} \lambda^3 (v_x^2 + v_y^2)^2 d \ll \pi\). Noting \(\theta^2 = \theta_x^2 + \theta_y^2 \approx \lambda^2 (v_x^2 + v_y^2)\), the condition of validity can be rewritten as:

\[
\frac{\theta^2 d}{4\lambda} \ll 1
\]

If \(a\) is the largest radial distance in the output plane, the largest angle is \(\theta_m = a / d\). The condition of validity of the Fresnel approximation may be written in the form

\[
\frac{N_f \theta_m^2}{4} \ll 1
\]

where \(N_f = a^2 / \lambda d\) is the Fresnel number.
Input-output relation

Given an arbitrary monochromatic wave with complex amplitude \(U(x, y, z)\) at \(z = 0\). It can be decomposed into a sum of plane waves:

\[
U(x, y, 0) = \iint \tilde{U}(v_x, v_y, 0) \exp\left[-j2\pi(v_x x + v_y y)\right] dv_x dv_y
\]

\(\tilde{U}(v_x, v_y, 0)\) is the Fourier transform of \(U(x, y, 0)\). It represents the complex envelop of plane-wave components at \(z = 0\):

\[
\tilde{U}(v_x, v_y, 0) = FT[U(x, y)]
\]

\[
\tilde{U}(v_x, v_y, 0) = \iint U(x, y) \exp\left[j2\pi(v_x x + v_y y)\right] dx dy
\]

The complex envelop of plane-wave components at \(z = d\) is:

\[
\tilde{U}(v_x, v_y, d) = \tilde{U}(v_x, v_y, 0) \times H(v_x, v_y)
\]

The complex amplitude of the wave \(U(x, y, z)\) at \(z = d\) is the sum of the contribution of these plane waves:

\[
U(x, y, d) = \iint \tilde{U}(v_x, v_y, d) \exp\left[-j2\pi(v_x x + v_y y)\right] dv_x dv_y
\]

\[
U(d) = FT^{-1}\{FT[U(0)] \times H\}
\]

\[
H(v_x, v_y) = \exp(-jkd) \exp\left[j\pi\lambda d \left(v_x^2 + v_y^2\right)\right]
\]
### 2.4 Impulse-response function of free space

From \( U(d) = FT^{-1}\{FT[U(0)] \times H}\), we see that \( U(d) = U(0) \ast FT^{-1}(H) \) (convolution). The inverse Fourier transform of the Transfer function is called the impulse-response function:\(^*\)

\[
h(x, y) = h_0 \exp\left(-j k \frac{x^2 + y^2}{2d}\right)\quad \text{with}\quad h_0 = \left(\frac{j}{\lambda d}\right) \exp(-jkd)
\]

The impulse-response function of free space is proportional to the complex amplitude at \( z = d \) of a paraboloidal wave centered about the origin \((0, 0)\).

An alternative procedure for relating the complex amplitude of an arbitrary monochromatic wave between \( z = 0 \) and \( z = d \) is to regard the wave at \( z = 0 \) as a superposition of points, each producing a paraboloidal wave. The wave originating at the point \((x', y', 0)\) has an amplitude \( U(x', y', 0) \) and is centered about \((x', y', 0)\) so that it generates a wave with amplitude \( U(x', y', 0) \times h(x - x', y - y') \). The sum of these contributions is the two-dimensional convolution

\[
U(x, y, d) = \iint U(x', y', 0) h(x - x', y - y') dx' dy'
\]

\[
U(x, y, d) = h_0 \iint U(x', y', 0) \exp\left[-j\pi \frac{(x - x')^2 + (y - y')^2}{\lambda d}\right] dx' dy'
\]

This relation is also called the Fresnel integral. Note that it is possible to obtain this relation by simplifying the Rayleigh and Sommerfeld integral (using the Fresnel approximation), this integral being derived from the Helmholtz and Kirchhoff theorem based on the Helmholtz equation and the Green theorem (see appendix)

---

\(^*\) Using the following Fourier transform relation: \( \exp[-\pi a(v_z^2 + v_y^2)] \to \frac{1}{a} \exp[-\frac{\pi}{a} (x^2 + y^2)] \)
This result is consistent with the **Huygens-Fresnel principle** which states that each point on a wavefront generates an elementary spherical wave. The envelope of these elementary waves constitutes a new wavefront. Their superposition constitutes the wave in another plane. In the Fresnel approximation, a spherical wave centered at \((0, 0)\) is approximated by the paraboloidal wave:

\[
h(x, y) = \exp\left(-jk \frac{x^2 + y^2}{2d}\right)
\]

### 2.5 Summary

In summary, within the Fresnel approximation, there are two approaches to determining the complex amplitude of an arbitrary wave at \(z = d\), given the amplitude at \(z = 0\). In the **space domain approach**, in which the wave is expanded in terms of paraboloidal elementary waves, and in the **frequency-domain approach** in which the wave is expanded as a sum of plane waves. The procedure of calculation is described in the following table:

\[
\begin{align*}
U(x, y, 0) &\xrightarrow{\text{FT}} \hat{U}(\nu_x, \nu_y, 0) &\xrightarrow{\times H(\nu_x, \nu_y)} \hat{U}(\nu_x, \nu_y, d) \\
\hat{U}(\nu_x, \nu_y, 0) &\xrightarrow{\text{FT}^{-1}} U(x', y', d)
\end{align*}
\]

\[
h(x, y) = \frac{j\exp(-jkd)}{\lambda d} \exp\left[-j\pi \frac{(x^2 + y^2)}{\lambda d}\right]
\]

\[
H(\nu_x, \nu_y) = \exp(-jkd) \exp\left[j\pi \lambda d (\nu_x^2 + \nu_y^2)\right]
\]
3. OPTICAL FOURIER TRANSFORM

3.1 Amplitude in the far-field

Consider the relation between $U(x, y, d)$ and $U(x', y', 0)$ in the space-domain approach. We have

$$U(x, y, d) = h_0 \iint U(x', y', 0) \exp \left[- j \pi \frac{(x-x')^2 + (y-y')^2}{\lambda d}\right] dx' dy'.$$

The phase in the exponent is:

$$(\pi / \lambda d) \left[ (x-x')^2 + (y-y')^2 \right] = (\pi / \lambda d) \left[ (x^2 + y^2) + (x'^2 + y'^2) - 2(xx' + yy') \right]$$

Suppose that $U(x', y', 0)$ is confined to a circle of radius $b$, and that the distance $d$ is sufficiently large so that the Fresnel number $N_F = b^2 / \lambda d \ll 1$.

$$(\pi / \lambda d) \left[ (x-x')^2 + (y-y')^2 \right] \approx (\pi / \lambda d) (x^2 + y^2) - (2\pi / \lambda d)(xx' + yy')$$

We have then

$$U(x, y, d) = h_0 \exp \left[- j \pi \frac{x^2 + y^2}{\lambda d}\right] \iint U(x', y', 0) \exp \left[ j 2\pi \frac{xx' + yy'}{\lambda d} \right] dx' dy'.$$

Suppose that $U(x, y, d)$ is confined to a circle of radius $a$, and that the distance $d$ is sufficiently large so that the Fresnel number $N_F = a^2 / \lambda d \ll 1$. We have then

$$U(x, y, d) \approx h_0 \iint U(x', y', 0) \exp \left[ j 2\pi \frac{xx' + yy'}{\lambda d} \right] dx' dy'.$$

We can recognize a Fourier transform:

$$U(x, y, d) \approx \hat{U}(\frac{x}{\lambda d}, \frac{y}{\lambda d}, 0)$$

This approximation is called the Fraunhofer approximation. It is valid when the Fresnel numbers $N_F$ and $N_F'$ are small. The Fraunhofer approximation is more difficult to satisfy than the Fresnel approximation, which requires that $N_F \theta_m^2 / 4 \ll 1$. Since $\theta_m \ll 1$ in the paraxial approximation, it is possible to satisfy the Fresnel condition $N_F \theta_m^2 / 4 \ll 1$ for Fresnel numbers $N_F$ not necessarily $\ll 1$. 
In the **Fraunhofer** approximation, the complex amplitude \( U(x,y,d) \) of a monochromatic wave of wavelength \( \lambda \), at \( z = d \), is proportional to the Fourier transform of the complex amplitude \( U(x,y,0) \) at \( z = 0 \), evaluated at the spatial frequencies \( \nu_x = x/\lambda d \) and \( \nu_y = y/\lambda d \). The approximation is valid if \( U(x,y,0) \) is confined to a circle of radius \( b \) satisfying \( b^2/\lambda d \ll 1 \) and if \( U(x,y,d) \) is confined to a circle of radius \( a \) satisfying \( a^2/\lambda d \ll 1 \).

**EXERCISE 6:** Conditions of validity of the Fresnel and Fraunhofer approximations.

Demonstrate that the Fraunhofer approximation is more restrictive than the Fresnel approximation by taking \( \lambda = 0.5 \) µm; assuming that the objects points \((x, y)\) lie within a circle of radius \( a = 1 \) cm, and determining the range of distances \( d \) for which the two approximations are applicable. The wave is observed on the axis \( z \).

### 3.2 Amplitude in the back focal plane of a lens

An arbitrary wave can be decomposed into a sum of plane waves. The plane wave-components may be separated by use of a converging lens as illustrated in the Figure below. The spherical lens transforms each plane wave into a paraboloidal (spherical) wave focused in the back focal plane of the lens (see Exercice 2). If a plane wave arrives at small angles \((\theta_x, \theta_y)\), it is transformed into a paraboloidal wave centered about the point \((x, f, y, \theta_x)\), where \( f \) is the focal length of the lens. The lens therefore maps each direction \((\theta_x, \theta_y)\), determined by the spatial frequencies \((\nu_x = \theta_x/\lambda, \nu_y = \theta_y/\lambda)\) into a single point \((x, y)\) in the focal plane and thus separates the different plane waves components.


Let \( f(x, y) \) be the complex amplitude of the optical wave in the plane \( z = 0 \), located at an arbitrary distance \( d \) from the lens. Light is decomposed into plane waves with complex amplitudes proportional to \( \tilde{f}(\nu_x, \nu_y) \). The planes waves are focused in the focal plane at points \((x, y)\) defined by \( x = \theta_x f = \lambda \nu_x f \) and \( y = \theta_y f = \lambda \nu_y f \). The complex amplitudes at points
\((x, y)\) in the focal plane are therefore proportional to the Fourier transform of \(f(x, y)\) evaluated at \(v_x = x/\lambda f, v_y = y/\lambda f\).

\[
g(x, y) \propto \mathcal{F}\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)
\]

To determine the proportionality factor, we analyse the incident wave into its Fourier components (plane wave decomposition) and propagate them into the optical system. We then superimpose the contributions of these plane waves in the focal plane. We assume that the Fresnel approximation is valid.

In the plane \(z = 0\); the plane wave with angles \((\theta_x, \theta_y)\) has a complex amplitude

\[
U(x, y, 0) = \tilde{f}(v_x, v_y) \exp[-j2\pi(v_x x + v_y y)].
\]

In the plane \(z = d^-\) (just before crossing the lens), the amplitude of the plane wave is

\[
U(x, y, d^-) = \tilde{f}(v_x, v_y) \exp[-j2\pi(v_x x + v_y y)] \times H_d(v_x, v_y),
\]

with \(H_d(v_x, v_y) = \exp(-jkd)\exp\left[j\pi\lambda d\left(v_x^2 + v_y^2\right)\right].\)

Upon crossing the lens, the complex amplitude is multiplied by the transmittance of the lens. We suppose that the lens is very thin (\(\Delta = 0\)). Within the Fresnel approximation, the complex transmittance of the lens is

\[
t(x, y) = \exp\left(\frac{jk_0 x^2 + y^2}{2f}\right) = \exp\left(\frac{j\pi x^2 + y^2}{\lambda f}\right).
\]

Upon crossing the lens, the amplitude of the plane wave is therefore

\[
U(x, y, d^+) = \tilde{f}(v_x, v_y) \exp[-j2\pi(v_x x + v_y y)] \times H_d(v_x, v_y) \times t(x, y)
\]

By noting

\[
-2v_x, x^2/\lambda f = (x^2 - 2v_x, \lambda f x)/\lambda f = [(x - x_0)^2 - x_0^2]/\lambda f
\]

\[
-2v_y, y^2/\lambda f = (y^2 - 2v_y, \lambda f y)/\lambda f = [(y - y_0)^2 - y_0^2]/\lambda f
\]

with \(x_0 = \lambda v_x f\) and \(y_0 = \lambda v_y f\),

the amplitude of the wave can be rewritten

\[
U(x, y, d^+) = A(v_x, v_y) \exp\left[j\pi \frac{(x-x_0)^2 + (y-y_0)^2}{\lambda f}\right]
\]

where

\[
A(v_x, v_y) = \exp(-jkd)\exp\left[j\pi\lambda (d-f)(v_x^2 + v_y^2)\right]\tilde{f}(v_x, v_y)
\]
We recognize here the complex amplitude of a paraboloidal wave (spherical wave) converging toward the point \((x_0, y_0)\) in the lens focal plane.

We now have to examine the propagation in free space between the lens and its focal plane. By using the Fresnel integral, we obtain finally

\[
U(x, y, d + f) = h_0 \left(\lambda f\right)^2 A(v_x, v_y) \delta(x - x_0) \delta(y - y_0)
\]

where

\[
h_0 = \left(\frac{j}{\lambda f}\right) \exp(- jkf)
\]

The plane wave is focused into a single point at \(x_0 = \lambda v_x f, \ y_0 = \lambda v_y f\).

The last step is to integrate over all the plane waves (all values of \(v_x\) and \(v_y\)). We finally obtain

\[
g(x, y) = h_0 A(x/\lambda f, y/\lambda f)
\]

or

\[
g(x, y) = h_i \exp\left[j\pi \left(\frac{x^2 + y^2}{\lambda f^2}\right)(d - f)\right] \tilde{f}\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)
\]

where

\[
h_i = \exp(-jkd) \exp\left[j\pi \lambda (d - f) \left(v_x^2 + v_y^2\right)\right] \tilde{f}(v_x, v_y)
\]

The intensity in the lens focal plane is

\[
I(x, y) = \frac{1}{(\lambda f)^2} \left| \tilde{f}\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right) \right|^2
\]

The intensity of light in the back focal plane of the lens is proportional to the squared absolute value of the Fourier Transform of the amplitude of the wave at the input plane, regardless of the distance \(d\).

If \(d = f\) (front focal plane of the lens), we have then

\[
g(x, y) = h_i \tilde{f}\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)
\]

where

\[
h_i = (j/\lambda f) \exp(-j2kf).
\]
The complex amplitude of light at a point \((x, y)\) in the back focal plane of a lens of focal length \(f\) is proportional to the Fourier transform of the complex amplitude of the complex amplitude in the front focal plane evaluated at the spatial frequencies \(\nu_x = x/\lambda f\) and \(\nu_y = y/\lambda f\). This relation is valid in the Fresnel approximation. That is why the back focal plane of a lens is called the Fourier plane. Without lens, the Fourier transformation is obtained only in the Fraunhofer approximation, which is more restrictive.

### 4. DIFFRACTION

When an optical wave is transmitted through an aperture and travel some distances in free space, its intensity is called the diffraction pattern. If light were treated as rays, the diffraction pattern would be a shadow of the aperture. Because of the wave nature of light, however, the diffraction pattern may deviate from the aperture shadow, depending on the distance between the aperture and the observation plane, the wavelength, and dimensions of the aperture.

The aperture is characterized by a transmittance, or aperture function:

\[
p(x, y) = \begin{cases} 
1 & \text{inside the aperture} \\
0 & \text{outside the aperture}
\end{cases}
\]

The incident wave is multiplied by the aperture function and then propagates in free space. The diffraction pattern is the intensity distribution of the light at a distance \(d\) from the aperture. The diffraction pattern is known as the Fresnel diffraction or Fraunhofer diffraction, depending on whether free space propagation is described using the Fresnel approximation or the Fraunhofer diffraction, respectively.

#### 4.1 Fresnel diffraction

The theory of Fresnel diffraction is based on the assumption that the incident wave is multiplied by the aperture function \(p(x, y)\) and propagates in free space in accordance with the Fresnel approximation (valid at any distance).

Explicit analytical expression of Fresnel diffraction patterns are often difficult to establish (it was however easy for a Gaussian wave, see exercise). Numerical simulation are generally performed (see examples).
Numerical simulations of Fresnel diffraction patterns resulting from diffraction by a circular aperture (diameter = 218 µm) illuminated with a plane wave at $\lambda = 632.8$ nm. The plane $z = 0$ mm corresponds to the plane of the aperture.

Numerical simulations of Fresnel diffraction patterns resulting from diffraction by a circular aperture (diameter = 2 cm) illuminated with a plane wave at $\lambda = 500$ nm. The propagation distance is 1 m (left) and 2 m (right).

Numerical simulations of Fresnel diffraction. The aperture (left) is illuminated with a plane wave at $\lambda = 500$ nm. The diffraction pattern (right) is calculated after propagation over a distance of 2.56 m.
4.2 Fraunhofer diffraction

The theory of Fraunhofer diffraction is based on the assumption that the incident wave is multiplied by the aperture function $p(x, y)$ and propagates in free space in accordance with the Fraunhofer approximation (valid only in far-field).

Assuming that the incident wave is a plane wave of intensity $I_0$ travelling in the $z$ direction, the complex amplitude in the plane of the aperture is $\sqrt{I_0} p(x, y)$. As seen earlier, the complex amplitude in the far field is $U(x, y, d) = \sqrt{I_0} h_0 \hat{p} \left( \frac{x}{\lambda d}, \frac{y}{\lambda d} \right) \text{ with } h_0 = \left( \frac{j}{\lambda d} \right) \exp(-jkd)$.

The diffraction pattern is therefore:

$$I(x, y) \propto I_0 \left| \hat{p} \left( \frac{x}{\lambda d}, \frac{y}{\lambda d} \right) \right|^2$$
The Fraunhofer diffraction pattern at the point \((x, y)\) is proportional to the squared magnitude of the Fourier transform of the aperture function \(p(x, y)\) evaluated at the spatial frequencies 

\[ \nu_x = \frac{x}{\lambda d} \text{ and } \nu_y = \frac{y}{\lambda d}. \]

The Fraunhofer approximation is valid for distances \(d\) that are usually extremely large. They are satisfied in applications of long-distance free-space optical communication such as laser radar (lidar) and satellite communication. However, if a lens of focal length \(f\) is used to focus the diffracted light, the intensity pattern in the focal plane is proportional to the squared magnitude of the Fourier transform of \(p(x, y)\) evaluated at 

\[ \nu_x = \frac{x}{\lambda f} \text{ and } \nu_y = \frac{y}{\lambda f}. \]

**Diffraction from a rectangular aperture:** The Fraunhofer diffraction pattern from a rectangular aperture, of height and width \(D_x\) and \(D_y\), observed at a distance \(d\) is:

\[
I(x, y) = I_0 \left| \tilde{p} \left( \frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \right|^2
\]

Fraunhofer diffraction from a rectangular aperture (1.6 × 3.2 mm) (left), illuminated with a plane wave at \(\lambda = 500\) nm. The diffraction pattern (right) is calculated at a distance of 30 m.
Diffraction from a circular aperture: the Fraunhofer diffraction pattern from a circular aperture of diameter \( D \) is:

\[
I(x, y) = I_0 \left[ \frac{2J_1(\pi Dr / \lambda d)}{\pi Dr / \lambda d} \right]^2 \quad \text{with} \quad r = \left( x^2 + y^2 \right)^{1/2}
\]

The diffraction pattern consists of a central disk surrounded by rings. It is known as the Airy pattern. The radius of the central disk is

\[
\begin{cases}
\rho_0 = 1.22 \lambda d / D \\
\theta_0 = 1.22 \lambda / D
\end{cases}
\]

Fraunhofer diffraction from a circular aperture (diameter = 160 µm) (left), illuminated with a plane wave at \( \lambda = 500 \) nm. The diffraction pattern (right) is calculated at a distance of 1 m.

5. SPATIAL FILTERING

Consider a two-lens imaging system as the one illustrated below:

This system, called 4-f imaging system, has a magnification of -1.
This optical system can be recognized as an association of two Fourier-transforming subsystems. The first subsystem performs a Fourier Transform, and the second another Fourier transform. As a result, in the absence of an aperture the image is the perfect replica of the object (with a magnification -1).

Let \( f(x, y) \) be the complex amplitude in the object plane and \( g(x, y) \) the complex amplitude in the image plane. The first lens analyzes \( f(x, y) \) into its spatial Fourier transform and separates its Fourier components so that each point in the Fourier plane corresponds to a single spatial frequency. These components are then recombined by the second lens system and the object distribution is perfectly reconstructed.

The 4-f imaging system can be used as a spatial filter in which the image \( g(x, y) \) is a filtered version of the object \( f(x, y) \). Since the Fourier components of \( f(x, y) \) are available in the Fourier plane, a mask may be used to adjust them selectively, blocking some components and transmitting others, as illustrated.

Examples of spatial filters

The ideal low-pass filter has a transmittance represented by the “Disk” function. For example, if the radius of the disk is $r_0$ and the optical wavelength $\lambda$, the filter eliminates the spatial frequencies greater than $\nu_c = \frac{r_0}{\lambda f}$.

- The high-pass filter is the complement of the low-pass filter. It blocks low frequencies and transmits high frequencies. The mask is a clear transparency with an opaque central disk. The filter output is high at regions of large rate of change and small at regions of smooth or slow variation of the object. This filter is therefore useful for edge enhancement in image-processing applications.

- The vertical-pass filter blocks horizontal frequencies and transmits vertical frequencies. Only variations in the $x$ direction are transmitted. If the mask is a vertical slit of width $D$, the highest transmitted frequency is $\nu_y = \frac{(D/2)}{\lambda f}$.
Composit photo of the moon (a). Filtered photo without horizontal lines (b). Fourier transform (c). Filtered Fourier transform (d).


6. Image formation

We consider image formation using the single lens imaging system shown below:

6.1 Impulse response function

By definition, the impulse response of the complex amplitude in the object plane of the wave emitted by the point (an impulse) on the optical axis in the object plane. This point emits a spherical wave, whose complex amplitude in the aperture plane (just before the lens) is

\[
U(x, y) = h_1 \exp\left( -jk \frac{x^2 + y^2}{2d_1} \right) \quad \text{where} \quad h_1 = \left( \frac{j}{\lambda d_1} \right) \exp(-jkd_1)
\]

Upon crossing the aperture and the lens, the complex amplitude becomes

\[
U_1(x, y) = U(x, y) \exp\left( jk \frac{x^2 + y^2}{2f} \right) p(x, y)
\]

The resultant field \( U_1(x, y) \) then propagates in free space a distance \( d_2 \). The complex amplitude in the image plane is then obtained by a convolution, according to

\[
h(x, y) = h_2 \int U_1(x', y') \exp\left[ -j\pi \left( \frac{(x-x')^2 + (y-y')^2}{\lambda d_2} \right) \right] \, dx' \, dy' \quad \text{where} \quad h_2 = \left( \frac{j}{\lambda d_2} \right) \exp(-jkd_2)
\]

We have finally:

\[
h(x, y) = h_0 h_2 \exp\left[ -j\pi \frac{x^2 + y^2}{\lambda d_2} \right] \tilde{p}_1 \left( \frac{x}{\lambda d_2}, \frac{y}{\lambda d_2} \right),
\]

where \( \tilde{p}_1 \) is the Fourier transform of the function \( p_1(x, y) = p(x, y) \exp\left[ -j\pi \frac{x^2 + y^2}{\lambda} \right] \), and

\[ e = \frac{1}{d_2} + \frac{1}{d_1} - \frac{1}{f}. \]

If the system is focused (\( e = 0 \)), \( p_1 = p \). We have then

\[
h(x, y) = h_0 \exp\left[ -j\pi \frac{x^2 + y^2}{\lambda d_2} \right] \tilde{p}_1 \left( \frac{x}{\lambda d_2}, \frac{y}{\lambda d_2} \right)
\]

where \( h_0 = h_0 h_2 \).

The function \( h(x, y) \) is called the impulse response function of the imaging system.

If the phase factor \( \pi \frac{x^2 + y^2}{\lambda d_2} \) is much smaller than 1, it can be neglected, so that

\[
h(x, y) = h_0 \tilde{p} \left( \frac{x}{\lambda d_2}, \frac{y}{\lambda d_2} \right)
\]

*The system’s impulse response is proportional to the Fourier transform of the pupil function \( p \).*
6.2 Image formation with coherent illumination

In the ray optics theory, each point $A$ of the object, located at $(x_0, y_0)$, has an image which is a point $A'$ located at $(x'_0, y'_0)$. In the wave optics theory, the image of the point $A$ is proportional to the impulse response function centered on $A'$: $h(x' - x'_0, y' - y'_0)$. The complex amplitude $U_{\text{image (wave optics)}}$ in the image plane is therefore:

$$U_{\text{image (wave optics)}}(x', y') = \iint U_{\text{image (ray optics)}}(x_0, y_0)h(x' - x'_0, y' - y'_0)dx_0'dy_0'$$

where $U_{\text{image (ray optics)}}$ is the image obtained by ray tracing. This result can be written in a condensed manner as using the convolution operator:

$$U_{\text{image (wave optics)}} = U_{\text{image (ray optics)}} * h$$

The image in amplitude is the perfect image (in amplitude, according to the ray optics theory) convoluted by the impulse response function.

The spatial frequencies of the wave in the image plane are given by

$$\tilde{U}_{\text{image (wave optics)}} = \tilde{U}_{\text{image (ray optics)}} \times H$$

where $H(\nu_x, \nu_y)$ is the transfer function of the imaging system, defined as the Fourier transform of the impulse response function.

$$H(\nu_x, \nu_y) = p(\lambda d_2 \nu_x, \lambda d_3 \nu_y)$$

where $p(x, y)$ is the pupil function.

The image in intensity (which is what we can observe on a screen or a camera) is then:

$$|U_{\text{image (wave optics)}}|^2 = |U_{\text{image (ray optics)}}|^2 h^2$$

The description of image formation considered here is valid if we consider coherent light.

The concept of coherence will be presented in the next chapter.

6.3 Image formation with incoherent illumination

In the previous studies, the light was supposed to be spatially coherent. This means that when several optical waves are superimposed, the amplitude of the resulting wave is the sum of the amplitude of each wave. If the light is supposed to be spatially incoherent, this result is not true. The superposition of several incoherent waves gives a wave whose intensity (and not amplitude) is the sum of the intensity (and not amplitude) of each wave.
The image of a point given by an optical system (the impulse response) has an intensity \( |h(x' - x_0', y' - y_0')|^2 \). The image formation with incoherent illumination is therefore given by

\[
I_{\text{image (wave optics)}}(x', y') = \iint I_{\text{image (ray optics)}}(x_0', y_0') h_{\text{inc}}(x' - x_0', y' - y_0') \, dx_0' \, dy_0',
\]

with \( h_{\text{inc}}(x', y') = |h(x', y')|^2 \).

Using condensed notation, we write:

\[
I_{\text{image (wave optics)}} = I_{\text{image (ray optics)}} \ast h_{\text{inc}}.
\]

By taking the Fourier transform of the previous relation, we have

\[
\tilde{I}_{\text{image (wave optics)}} = \tilde{I}_{\text{image (ray optics)}} \times H_{\text{inc}},
\]

where \( H_{\text{inc}} = \tilde{h}_{\text{inc}} \) is referred to as the incoherent transfer function of the optical system.

The modulation transfer function is defined as \( |H_{\text{inc}}| \).
Summary

Image given by an optical system

Coherent illumination

amplitude  \[ U_{\text{wave-optic}} = U_{\text{nat-optic}} \ast h \]

\[ h(x, y) \approx \hat{p} \left( \frac{x}{\lambda d}, \frac{y}{\lambda d} \right) \]

intensity  \[ I_{\text{wave-optic}} = \left| U_{\text{wave-optic}} \right|^2 \]

\[ \hat{U}_{\text{wave-optic}} = \hat{U}_{\text{nat-optic}} \times H \]

\[ H(v_x, v_y) = p(\lambda d v_x, \lambda d v_y) \]

Incoherent illumination

intensity  \[ I_{\text{wave-optic}} = I_{\text{nat-optic}} \ast h_{\text{incoherent}} \]

\[ h_{\text{incoherent}} = |h|^2 \]

\[ \hat{I}_{\text{wave-optic}} = \hat{I}_{\text{nat-optic}} \times H_{\text{incoherent}} \]

\[ H_{\text{incoherent}} = \hat{I}_{\text{incoherent}} \]

\[ p = \text{pupil function (unity within the aperture, zero elsewhere)} \]
\[ h = \text{impulse response function in amplitude} \]
\[ H = \text{transfer function (with coherent illumination)} \]
\[ h_{\text{incoherent}} = \text{incoherent impulse response function, point spread function (PSF)} \]
\[ H_{\text{incoherent}} = \text{incoherent transfer function} \]
\[ |H_{\text{incoherent}}| = \text{modulation transfer function} \]
\[ d = \text{distance from the optical system to the image plane} \]
**Example: Imaging system with a circular aperture**

If the pupil is a circle of radius \( a \), the pupil function \( p(x, y) = 1 \) for points inside the circle and zero elsewhere. The pupil function is described by \( p(x, y) = \text{dis}\left(\frac{r}{a}\right) \) where \( r = \sqrt{x^2 + y^2} \).

The coherent impulse response function of the imaging system is then

\[
h(x, y) \propto \frac{J_1(2\pi \nu r)}{2\pi \nu r} \text{ where } \nu = \frac{a}{\lambda d}
\]

The incoherent impulse response function is

\[
h_{\text{inc}}(x, y) = \left|h(x, y)\right|^2 \propto \left[\frac{J_1(2\pi \nu r)}{2\pi \nu r}\right]^2 \text{ (Airy function)}
\]

The first zeros of \( h \) and \( h_{\text{inc}} \) are located at \( r_i = 1,22 \frac{\lambda d}{a} \).

With incoherent illumination, each point in the object plane has an image that is proportional to \( h_{\text{inc}} \). When two points of the object plane are close enough so that their images are distant of the quantity \( r_i \), then it becomes impossible to resolve them. The quantity \( r_i = 1,22 \frac{\lambda d}{a} \) is a measure of the resolution of the imaging system.

In the case of **coherent** illumination, the coherent transfer function of the imaging system is

\[
H(\nu_x, \nu_y) = \text{dis}\left(\frac{\nu}{\nu_x}\right) \text{ where } \nu = \sqrt{\nu_x^2 + \nu_y^2}
\]

In the case of **incoherent** illumination, the incoherent transfer function of the imaging system is

\[
H_{\text{inc}}(\nu_x, \nu_y) = \begin{cases} 
\frac{2}{\pi} \left\{ \cos^{-1} \frac{\nu}{2\nu_x} - \frac{\nu}{2\nu_x} \left[1 - \left(\frac{\nu}{2\nu_x}\right)^2\right]^{1/2} \right\} & \text{if } \nu < 2\nu_x \\
0 & \text{otherwise}
\end{cases}
\]

The cutoff frequency is \( \nu_c = \begin{cases} \frac{a}{\lambda d} & \text{with coherent illumination} \\
\frac{2a}{\lambda d} & \text{with incoherent illumination}
\end{cases} \)

The impulse response functions and transfer functions with coherent or incoherent illumination are shown below:
Impulse-response functions and transfer functions of a diffraction-limited single lens under coherent (a) and incoherent (b) illumination. The pupil is supposed to be a circle of radius $a$. 
Chapter 3: COHERENCE

In the previous chapters, light was assumed to be deterministic. The dependence of the wave function on time and position was perfectly predictable. In reality, optical waves have random fluctuations. For example, light radiated by a hot object (thermal light) is random because it is a superposition of emissions from a large number of atoms radiating independently and at different frequencies, amplitudes and phases. Randomness in light may also be a result of scattering from rough surfaces, diffused glass or turbulent fluids.

The theory of the random fluctuations of light is known as statistical optics or also as the theory of optical coherence. This theory is based on statistical averaging to define a number of non-random quantitative measures.

1. Statistical properties of random light
An arbitrary optical wave is defined by a wavefunction \(u(r,t) = \text{Re}\{U(r,t)\}\), where \(U(r,t)\) is the complex wavefunction. For random light, \(U(r,t)\) is a random function that can be characterized by a number of statistical averages introduced in this section.

1.1 Optical intensity
The intensity of coherent (deterministic) light is defined as

\[ I(r,t) = |U(r,t)|^2 \]
For monochromatic deterministic light, the intensity is independent of time.

In the case of random light, the intensity \( |U(r,t)|^2 \) is also a random function of position and time. It is then called the instantaneous intensity.

The average intensity is defined as

\[
I(r,t) = \langle |U(r,t)|^2 \rangle
\]

The symbol \( \langle \rangle \) denotes an ensemble average over many realizations. The average intensity may be time independent or not (see figure below). When the optical wave is statistically stationary, its statistical average does not depend on time. Stationary does not mean constancy; it means constancy of the average properties (example = halogen lamp with constant electric current supply).

(a) Statistically stationary wave. The average intensity does not vary with time.
(b) A statistically non-stationary wave has a time-varying intensity.

When the light is stationary, the statistical averaging operation can be determined by time averaging over a long duration (instead of averaging over many realizations), so that:

\[
I(r) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} |U(r,t)|^2 dt
\]
1.2 Temporal coherence and spectrum

We consider the fluctuation of stationary light as a function of time, at a fixed position. The random fluctuation of $U(t)$ can be characterized by a “memory time”. Fluctuations separated by a time longer than the “memory time” are independent, so that the process “forgets” itself. The function $U(t)$ appears to be smooth within the memory time, but “rough” when examined over a longer time scale. A quantitative measure of this temporal behaviour can be established by defining a statistical behaviour known as the autocorrelation function.

Temporal coherence function

The autocorrelation function of a stationary complex random function $U(t)$ is

$$G(\tau) = \langle U^*(t)U(t+\tau) \rangle$$

or

$$G(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} U^*(t)U(t+\tau)dt$$

In the theory of optical coherence, the autocorrelation function $G(\tau)$ is called the temporal coherence function. This function has a Hermitian symmetry: $G(-\tau) = G^*(\tau)$. The intensity of the wave is equal to $G(\tau)$ when $\tau = 0$ : $G(0) = I$. The phase of the function $G(\tau)$ is equal to the average of the phase difference between $U(t)$ and $U(t+\tau)$. If there is no correlation between $U(t)$ and $U(t+\tau)$, $G(\tau) = 0$.

Degree of temporal coherence

The temporal coherence function $G(\tau)$ carries information about the degree of correlation (coherence) of stationary light and the intensity. A measure of the coherence that is insensitive to the intensity is given by the normalized autocorrelation function, known as the degree of temporal coherence:

$$g(\tau) = \frac{G(\tau)}{G(0)} = \frac{\langle U^*(t)U(t+\tau) \rangle}{\langle U^*(t)U(t) \rangle}$$

$|g(0)| = 1$. The absolute value of $g(\tau)$ cannot be greater than one. $0 \leq |g(\tau)| \leq 1$

The value of $g(\tau)$ is a measure of the degree of correlation between $U(t)$ and $U(t+\tau)$. In the case of deterministic and monochromatic light, i.e. $U(t) = A\exp( j2\pi \nu t )$, $|g(\tau)| = 1$. $U(t)$ and $U(t+\tau)$ are completely correlated for all time delays $\tau$.

Usually, $|g(\tau)|$ drops from 1 for $\tau = 0$ to 0 for sufficiently large value of $\tau$. 

48
Coherence time / coherence length

The width of \( |g(\tau)| \) denoted \( \tau_c \), is a measure of the memory time of the fluctuations known as the coherence time. For \( \tau \ll \tau_c \), the fluctuations are strongly correlated, whereas for \( \tau \gg \tau_c \) they are weakly correlated. Although a definition of the width of a function is arbitrary, the usual definition of the coherence time is the power-equivalent width:

\[
\tau_c = \int_{-\infty}^{\infty} |g(\tau)|^2 d\tau
\]

The coherence time of monochromatic light is infinite.

The coherence length is defined as

\[
l_c = c\tau_c
\]

Power spectral density

The power spectral density is defined as \( S(\nu) = |\hat{U}(\nu)|^2 \)

The power spectral density is the Fourier Transform of the autocorrelation function. This relation is known as the Wiener-Khinchin theorem

\[
S(\nu) = \int_{-\infty}^{\infty} G(\tau) \exp(-j2\pi\nu\tau) d\tau
\]

Demonstration of the Wiener-Khinchin theorem

\[
G(\tau) = \int_{-\infty}^{\infty} U^*(t)U(t+\tau)dt
\]

\[
U(t) = \int_{-\infty}^{\infty} \hat{U}(\nu)\exp(-j2\pi\nu t)d\nu \quad \text{et} \quad U^*(t) = \int_{-\infty}^{\infty} \hat{U}^*(\nu)\exp(+j2\pi\nu t)d\nu
\]

\[
G(\tau) = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \hat{U}^*(\nu)\exp(+j2\pi\nu t)d\nu \right] \left[ \int_{-\infty}^{\infty} \hat{U}(\nu)\exp(-j2\pi\nu (t+\tau))d\nu \right] dt
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{U}^*(\nu)\hat{U}(\nu)\exp\left(+j2\pi(t+\tau)\left(\nu - \nu'\right)\right)\exp(-j2\pi\nu \tau)d\nu d\nu' dt
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{U}^*(\nu)\hat{U}(\nu)\delta\left(\nu - \nu'\right)\exp(-j2\pi\nu \tau)d\nu d\nu
\]

\[
= \int_{-\infty}^{\infty} \hat{U}^*(\nu)\hat{U}(\nu)\exp(-j2\pi\nu \tau)d\nu
\]

\[
= \int_{-\infty}^{\infty} |\hat{U}(\nu)|^2 \exp(-j2\pi\nu \tau)d\nu
\]
We have \( G(\tau) = \int_{-\infty}^{\infty} S(\nu) \exp(-j2\pi\nu\tau) d\nu \)

Spectral width

The spectral width of light is defined as the width of the spectral density \( S(\nu) \). Because of the Fourier-transform relation between \( S(\nu) \) and \( G(\tau) \), their widths are inversely related. A light source of broad spectrum has a short coherence time (and short coherence length).

\[
\Delta \nu_c = \frac{1}{\tau_c} = \frac{c}{l_c}
\]

<table>
<thead>
<tr>
<th>Source</th>
<th>( \tau_c )</th>
<th>( l_c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filtered sunlight ((\lambda_0 = 0.4 - 0.8 , \mu\text{m}))</td>
<td>2.67 fs</td>
<td>800 nm</td>
</tr>
<tr>
<td>LED ((\lambda_0 = 1 , \mu\text{m}, \Delta\lambda = 50 , \text{nm}))</td>
<td>67 fs</td>
<td>20 ( \mu\text{m} )</td>
</tr>
<tr>
<td>Low-pressure Sodium lamp</td>
<td>2 ps</td>
<td>600 ( \mu\text{m} )</td>
</tr>
<tr>
<td>Multimode He-Ne laser ((\lambda_0 = 633 , \text{nm}))</td>
<td>0.67 ns</td>
<td>20 cm</td>
</tr>
<tr>
<td>Single-mode He-Ne laser ((\lambda_0 = 633 , \text{nm}))</td>
<td>1 ( \mu\text{s} )</td>
<td>300 m</td>
</tr>
</tbody>
</table>

1.3 Spatial coherence

Mutual coherence function

A descriptor of the spatial and temporal fluctuations of the random function \( U(\mathbf{r}, t) \) is the cross-correlation function of \( U(\mathbf{r}_1, t) \) and \( U(\mathbf{r}_2, t) \):

\[
G(\mathbf{r}_1, \mathbf{r}_2, \tau) = \langle U^*(\mathbf{r}_1, t)U(\mathbf{r}_2, t + \tau) \rangle
\]

This function of the time delay \( \tau \) is known as the mutual coherence function. Its normalized form is called the complex degree of mutual coherence:

\[
g(\mathbf{r}_1, \mathbf{r}_2, \tau) = \frac{G(\mathbf{r}_1, \mathbf{r}_2, \tau)}{[I(\mathbf{r}_1)I(\mathbf{r}_2)]^{1/2}}.
\]
When the two points coincide \( \mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r} \), \( g(\mathbf{r}, \mathbf{r}, \tau) \) is the degree of temporal coherence.

The absolute value of the degree of coherence is bounded between zero and unity. It is a measure of the degree of correlation between the fluctuations at \( \mathbf{r}_1 \) and those at \( \mathbf{r}_2 \) at a time \( \tau \) later.

**Degree of spatial coherence**

The spatial correlation of light may be assessed by examining the dependence of the mutual function on position for a fixed time delay \( \tau = 0 \) usually. The mutual coherence function at \( \tau = 0 \) is

\[
G(\mathbf{r}_1, \mathbf{r}_2, 0) = \left\langle U^*(\mathbf{r}_1, t) U(\mathbf{r}_2, t) \right\rangle
\]

This function is denoted \( G(\mathbf{r}_1, \mathbf{r}_2) \) for simplicity.

Its normalized form called the degree of spatial coherence:

\[
g(\mathbf{r}_1, \mathbf{r}_2) = \frac{G(\mathbf{r}_1, \mathbf{r}_2)}{\left[ I(\mathbf{r}_1) I(\mathbf{r}_2) \right]^{1/2}}
\]

**Coherence area**

For a given point at \( \mathbf{r}_2 \), the area scanned by the point \( \mathbf{r}_1 \) within which \( |g(\mathbf{r}_1, \mathbf{r}_2)| \geq 1/2 \) (for example) is called the coherence area around the point \( \mathbf{r}_2 \). In the case of perfectly spatially coherent light the coherence area is infinite. In the case of totally spatially incoherent light the coherence area is reduced to a point.

It is important to consider the coherence area when using an optical system. If the area of coherence is greater than the size of the aperture through which light is transmitted, the light may be regarded as coherent. If the coherence area is smaller than the resolution of the optical system (size of the system’s incoherent impulse response), the light can be considered as incoherent.

Light emitted by a hot object (thermal light) with an extended surface has an area of coherence on the order of \( \lambda^2 \), where \( \lambda \) is the central wavelength. It can therefore be considered as spatially incoherent is most cases. However, complete coherence or incoherence are only idealizations (two limits of partial coherence).

In a microscope using Köhler illumination, the spatial coherence of illumination can be adjusted with the aperture diaphragm (D1). The primary advantage of Köhler illumination is to provide even illumination of the sample.
1.4 Gain of spatial coherence by propagation: Zernike and Van Cittert theorem

We consider a spatially incoherent light source in the plane \( z = 0 \). We propose to calculate the degree of, spatial coherence in an arbitrary plane \( z \) after propagation over a distance \( d \).

\[ U'(r_1) = -\frac{j}{\lambda d} \exp(-jkd) \iint \overline{U'(s)} \exp \left[ j \pi \frac{(r_1 - s)^2}{\lambda d} \right] d^2s \]

\[ U(r_2) = \frac{j}{\lambda d} \exp(-jkd) \iint U(s) \exp \left[ -j \pi \frac{(r_2 - s)^2}{\lambda d} \right] d^2s \]

The (spatial) mutual coherence function in the plane \( z \) is
\[ G_\lambda(r, r') = \left\langle U(r) U^*(r') \right\rangle = \frac{1}{(\lambda d)^2} \int \int \int U(s) U^*(s) \exp \left[ j \pi \frac{(r_1 - s)^2 - (r_2 - s)^2}{\lambda d} \right] d^2 s \]

Since the light source is supposed to be totally spatially incoherent, the mutual coherence function in the plane of the source is

\[ \left\langle U(s) U^*(s) \right\rangle = G_\lambda(s, s) = \sqrt{I(s)} I(s) \delta(s - s'), \delta(s - s') \text{ being the Dirac distribution (= 0 for } s \neq s'). \]

The mutual coherence function after propagation over a distance \( d \) becomes

\[ G_\lambda(r, r') = \frac{1}{(\lambda d)^2} \int I(s) \exp \left[ j \pi \frac{(r_1 - s)^2 - (r_2 - s)^2}{\lambda d} \right] d^2 s \]

Its normalized form is

\[ g(r, r') = \exp \left[ j \pi \frac{r_1^2 - r_2^2}{\lambda d} \right] \frac{\mathcal{F} \left( \frac{r_1 - r_2}{\lambda d} \right)}{\mathcal{F}(0)} \]

and

\[ |g(r, r')| = \left| \frac{\mathcal{F} \left( \frac{r_1 - r_2}{\lambda d} \right)}{\mathcal{F}(0)} \right| \]

The degree of spatial coherence in the plane \( z \) after propagation over a distance \( d \) is related to the Fourier transform of the intensity distribution of the incoherent light source. This result is known as the Zernike and Van Cittert theorem. Although light is totally spatially incoherent in the plane of the source, after propagation light is no longer totally incoherent (the light source having a finite size). In general, light gains spatial coherence by the mere act of propagation. This result is not surprising. Although light fluctuations at different points in the plan of the source are uncorrelated, the radiation from each point spreads and overlaps with that from the
neighboring points. The light reaching two points in the output plane come from many points in
the source plane, some of which are common. These common contributions create partial
correlation between fluctuations at the output points.

If the area of the source is small, the area of coherence after propagation becomes large. In the
extreme limit of a point source, the area of coherence after propagation is infinite: the wave is
then spatially coherent.

2. Interference of partially coherent light

2.1 Interference and temporal coherence

In a Michelson interferometer (see figure below), a beam splitter generates two identical waves,
one of which having travelled a longer optical path before the two waves are recombined.

A interferometer such as a Michelson interferometer may therefore be used to add a wave to a
 replica of a time-delayed replica of itself.

The intensity of light detected at the interferometer output is

\[ I = \langle |U(t) + U(t+\tau)|^2 \rangle = \langle |U(t)|^2 \rangle + \langle |U(t+\tau)|^2 \rangle + \langle U^*(t)U(t+\tau) \rangle + \langle U(t)U^*(t+\tau) \rangle \]

We can write

\[ I = 2I_0 \left[ 1 + \text{Re}\{g(\tau)\} \right] \]
\[ = 2I_0 \left[ 1 + |g(\tau)| \cos \phi(\tau) \right] \]

where \( g(\tau) \) is the degree of temporal coherence of the wave.

The ability of a wave to interfere with a time delayed replica of itself is governed by the degree
of temporal coherence. The visibility of the interference fringes is determined by \( |g(\tau)| \).

The function \( I(\tau) \) is called interferogram. It is proportional to the real part of the temporal
autocorrelation of the wave. A Michelson interferometer is thus sometimes called optical
autocorrelator. The width of the interferogram is the coherence length. This quantity can be very
easily measured by using a Michelson-type interferometer. By computing the Fourier transform
of the interferogram, the whole spectrum of the light can be obtained. This technique is known as Fourier-transform spectroscopy.

Since \( G(\tau) = I_{oG}(\tau) = \int_0^{\infty} S(\nu) \exp(j2\pi\nu \tau) d\nu \) (Wiener-Khinchin theorem)

and \( I_o = \int_0^{\infty} S(\nu) d\nu \)

we can write

\[
I = 2\left[ \int_0^{\infty} S(\nu) d\nu + \text{Re}\left\{ \int_0^{\infty} S(\nu) \exp(j2\pi\nu \tau) d\nu \right\} \right]
\]

or

\[
I = 2 \int_0^{\infty} S(\nu) \left[ 1 + \cos(2\pi\nu \tau) \right] d\nu
\]

The interferogram can be interpreted as a weighted superposition of interferograms produced by each of the monochromatic components of the wave. Each component of frequency and weight \( S(\nu) \) produces an interferogram with period \( 1/\nu \) and unity visibility.

---

Example of interferogram obtained with a halogen lamp of effective spectral width \( \Delta \sigma (\text{um}^{-1}) \); the horizontal axis is \( \delta = c \tau \).

#### 2.2 Interference and spatial coherence

The effect of spatial coherence on interference can be demonstrated by considering the Young’s double pinhole interference experiment. A partially coherent wave \( U(\mathbf{r}, \tau) \) illuminates an opaque screen with two pinholes located at positions \( \mathbf{r}_1 \) and \( \mathbf{r}_2 \).
Light is diffracted by the two pinholes in the form of spherical waves centered at the pinholes. The two waves interfere and the intensity $I$ is:

$$I = \left| \left( |U_1|^2 + |U_2|^2 \right) \right| = \left( |U_1|^2 \right) + \left( |U_2|^2 \right) + \left( U_1^* U_2 \right) + \left( U_1 U_2^* \right)$$

The intensity is observed as a function of $x$.

We have $r_1 = (-a, 0, 0)$, $r_2 = (a, 0, 0)$ and $r = (x, 0, d)$.

Within the Fresnel approximation, the two diffracted spherical waves can be written as paraboloidal waves:

$$U_1(r, t) \approx \frac{|r - r_1|}{c} U_1 \left( r, t - \frac{d + (x + a)^2/2d}{c} \right)$$

$$U_2(r, t) \approx \frac{|r - r_2|}{c} U_2 \left( r, t - \frac{d + (x - a)^2/2d}{c} \right)$$

The two waves have equal intensity equal to $I_0$. The normalized cross-correlation between the two waves at $r$ is

$$g_{12} = \frac{\langle U_1^*(r, t) U_2(r, t) \rangle}{I_0} = g(r_1, r_2, \tau_x)$$

where

$$\tau_x = \frac{|r - r_1| - |r - r_2|}{c} = \frac{(x + a)^2 - (x - a)^2}{2dc} = \frac{2ax}{2dc} = \frac{\theta x}{c}$$

is the time delay between the two waves.

If we consider monochromatic light with frequency $\nu_0$, $g(r_1, r_2, \tau_x) = g(r, r_c) \exp\left( j2\pi\nu_0\tau_x \right)$. The intensity at the interferometer output is then

$$I(x) = 2I_0 \left[ 1 + |g(r_1, r_2)| \cos \left( \frac{2\pi \theta}{\lambda} x + \phi \right) \right]$$
where \( \lambda = c / \nu_0 \), \( \tau_s = \theta \lambda / c \) and \( \varphi = \arg \{ g(\mathbf{r}_1, \mathbf{r}_2) \} \). The interference fringe pattern is sinusoidal with spatial period \( \lambda / \theta \) and visibility \( |g(\mathbf{r}_1, \mathbf{r}_2)| \).

The visibility of the interference pattern is equal to the magnitude of the degree of spatial coherence at the two pinholes.