SPIE PRESS | Field Guide

SPIE

Field Guide to

Physical Optics

Daniel G. Smith

Field Guide to

Physical Optics

Daniel G. Smith

SPIE Field Guides
Volume FG17

John E. Greivenkamp, Series Editor

SPIE
PRESS

Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Smith, Daniel Gene.
Field guide to physical optics / Daniel G. Smith.
pages cm. – (The field guide series ; FG17)
Includes bibliographical references and index.
ISBN 978-0-8194-8548-9
1. Physical optics. I. Title.
QC395.2.S65 2013
535′.2–dc23

2013000915

Published by

SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1.360.676.3290
Fax: +1.360.647.1445
Email: books@spie.org
Web: www.spie.org

Printed in the United States of America.
Second printing

Introduction to the Series

Welcome to the *SPIE Field Guides*—a series of publications written directly for the practicing engineer or scientist. Many textbooks and professional reference books cover optical principles and techniques in depth. The aim of the *SPIE Field Guides* is to distill this information, providing readers with a handy desk or briefcase reference that provides basic, essential information about optical principles, techniques, or phenomena, including definitions and descriptions, key equations, illustrations, application examples, design considerations, and additional resources. A significant effort will be made to provide a consistent notation and style between volumes in the series.

Each *SPIE Field Guide* addresses a major field of optical science and technology. The concept of these Field Guides is a format-intensive presentation based on figures and equations supplemented by concise explanations. In most cases, this modular approach places a single topic on a page, and provides full coverage of that topic on that page. Highlights, insights, and rules of thumb are displayed in sidebars to the main text. The appendices at the end of each Field Guide provide additional information such as related material outside the main scope of the volume, key mathematical relationships, and alternative methods. While complete in their coverage, the concise presentation may not be appropriate for those new to the field.

The *SPIE Field Guides* are intended to be living documents. The modular page-based presentation format allows them to be easily updated and expanded. We are interested in your suggestions for new Field Guide topics as well as what material should be added to an individual volume to make these Field Guides more useful to you. Please contact us at **fieldguides@SPIE.org.**

John E. Greivenkamp, *Series Editor*
College of Optical Sciences
The University of Arizona

The Field Guide Series

Keep information at your fingertips with all of the titles in the Field Guide Series:

Adaptive Optics, Second Edition, Robert Tyson & Benjamin Frazier

Atmospheric Optics, Larry Andrews

Binoculars and Scopes, Paul Yoder, Jr. & Daniel Vukobratovich

Diffractive Optics, Yakov Soskind

Geometrical Optics, John Greivenkamp

Illumination, Angelo Arecchi, Tahar Messadi, & John Koshel

Image Processing, Khan M. Iftekharuddin & Abdul Awwal

Infrared Systems, Detectors, and FPAs, *Second Edition*, Arnold Daniels

Interferometric Optical Testing, Eric Goodwin & Jim Wyant

Laser Pulse Generation, Rüdiger Paschotta

Lasers, Rüdiger Paschotta

Lens Design, Julie Bentley & Craig Olson

Microscopy, Tomasz Tkaczyk

Optical Fabrication, Ray Williamson

Optical Fiber Technology, Rüdiger Paschotta

Optical Lithography, Chris Mack

Optical Thin Films, Ronald Willey

Optomechanical Design and Analysis, Katie Schwertz & James Burge

Physical Optics, Daniel Smith

Polarization, Edward Collett

Probability, Random Processes, and Random Data Analysis, Larry Andrews

Radiometry, Barbara Grant

Special Functions for Engineers, Larry Andrews

Spectroscopy, David Ball

Terahertz Sources, Detectors, and Optics, Créidhe O'Sullivan & J. Anthony Murphy

Visual and Ophthalmic Optics, Jim Schwiegerling

Physical optics is a broad subject that has been in vigorous and continuous development for more than a century. It can be thought of as encompassing all optics, except possibly ray optics, but it may also be regarded as a subset of physical phenomena described by electromagnetic optics.

This Field Guide is a practical overview of the subject area, with specific emphasis on information most useful in the field of optical engineering. Within this Field Guide, the reader will find formulae and descriptions of basic electromagnetic wave phenomena that are fundamental to a wave theory of light. Tools are provided for describing polarization. And, although vector diffraction theory and electromagnetic methods (e.g., FDTD and RCWA) are not treated here, emphasis is placed on scalar diffraction and imaging theory, which are essential in solving most practical optical engineering problems.

I owe thanks to the various professors who first taught me these subjects: Jack Kasher and Daniel Wilkins at the University of Nebraska at Omaha; and then later at the University of Arizona College of Optical Sciences: Arvind S. Marathay, Jack D. Gaskill, Roland V. Shack, James C. Wyant, and John E. Greivenkamp. I also owe thanks to Tom D. Milster, who graciously allowed me to review his soon-to-be published text and adopt portions of his notation. Thanks also go to Kerry McManus Eastwood for her extensive help in preparing this volume and to Eric P. Goodwin for his help in reviewing the material.

Finally, I dedicate this Field Guide to my wife, Jenny, who is always there to keep me coherent and in phase.

Daniel G. Smith
Nikon Research Corporation of America
January 2013

Table of Contents

Table of Contents

Table of Contents

Glossary of Terms and Acronyms

a_x,a_y,a_z	Field amplitudes in x, y, and z directions
A	Scalar optical field amplitude
$\mathbf{A}$	Vector optical field amplitude
A_c	Coherence area
A,B,C,D	Elements of a lens system matrix
$\mathbf{B}$	Magnetic induction
c	Speed of light
CTF	Coherent transfer function
$C(x),S(x)$	Fresnel cosine and sine integrals
D	Diattenuation
$\mathbf{D}$	Electric displacement
DOP	Degree of polarization
e	Base of the natural logarithm
e	Eccentricity
E	Electric field magnitude (scalar quantity)
$\mathbf{E}$	Electric field
$\mathbf{E}$	Jones vector
f	Lens focal length
F	Coefficient of finesse
$\mathfrak{F}$	Finesse
F/#	F-number
FWHM	Full-width at half-max
G	Green's function
h	Point spread function
h	Gaussian beam half-width at half-max
h_z^F	Fresnel's wavelet
h_z^H	Huygens' wavelet
H	Transfer function
$\mathbf{H}$	Magnetic field strength
HLP	Horizontal linear polarization
HWP	Half wave plate
i	Imaginary number $\sqrt{-1}$
I	Irradiance
$\mathbf{j}$	Electric current density
$\mathbf{J}$	Jones matrix
J_n	N^{th} order Bessel function of the first kind
k	Wavenumber
$\mathbf{k}$	Wavevector
K	Obliquity factor

Glossary of Terms and Acronyms

$\mathbf{L}$	Linear operator
LCP	Left-hand circular polarization
LP	Linear polarization/polarizer
LSI	Linear and shift invariant
m	Order of diffraction
$\mathbf{M}$	Mueller matrix/characteristic matrix
$\mathbf{M}_j$	Characteristic matrix of the j^{th} film layer
MTF	Modulation transfer function
n	Index of refraction (real)
$\mathbf{n}$	Surface normal
N	Complex refractive index
NA	Numerical aperture
N_j	Number densities
N_f	Fresnel number
OFZ	Open Fresnel zone
OPD	Optical path difference
OTF	Optical transfer function
P	Pupil function
PER	Polarization extinction ratio
PSF	Point spread function
q	Complex curvature of a Gaussian beam
$\mathbf{q}$	Grating vector
QWP	Quarter wave plate
r	Scalar position coordinate
r_s,r_p	Coefficient of reflection for s and p polarization
$\mathbf{r}$	Vector position coordinate
R	Reflectance
$\Re$	Resolving power
RCP	Right-hand circular polarization
s_1,s_2,s_3	Object and image distance
S_0	Magnitude of the Stokes vector
$\mathbf{S}$	Stokes vector
$\mathbf{S}$	Poynting vector
SR	Strehl ratio
t_s,t_p	Coefficient of transmission for s and p polarization
t_A	Amplitude transparency function
T	Transmittance

Glossary of Terms and Acronyms

TE	Transverse electric
TIR	Total internal reflection
TM	Transverse magnetic
U	Scalar optical field in complex notation
$\mathbf{U}$	Vector optical field in complex notation
$\mathbf{v}$	Fringe direction
V	Fringe visibility
VLP	Vertical linear polarization
w	Gaussian beam $1/e^2$ radius
w_0	Gaussian beam waist radius
x,y,z	Position coordinates
$\hat{x},\hat{y},\hat{z}$	Unit vectors in the x, y, and z directions
y	Optical admittance
z_0	Rayleigh range
α	Absorption coefficient
α,β,γ	Direction cosines
$\gamma(r_1,r_2,\tau)$	Complex degree of coherence
$\Gamma(r_1,r_2,\tau)$	Mutual coherence function
δ	Path difference or a small angle
∇	Laplacian operator
Δl	Coherence length
$\Delta\lambda$	Range of or change in wavelengths
$\Delta\nu$	Range of or change in frequency
Δt	Coherence time
$\varepsilon, \varepsilon_0$	Permittivity and permittivity of free space
η	Tilted or effective optical admittance
θ	An angle
κ	Extinction coefficient
λ	Wavelength
$\overline{\lambda}$	Mean wavelength
λ_1,λ_2	Amplitude transmission of eigenpolarizations
Λ	Grating period
μ, μ_0	Permeability and permeability of free space
ν	Optical frequency
ξ,η	Spatial frequency components
ρ	Electric charge density
σ	Electric conductivity
Σ	Aperture transmission function

Glossary of Terms and Acronyms

τ	Delay time
ϕ	Lens power
ϕ	Phase angle
Φ	Radiant flux
ψ_1,ψ_2	Eigenpolarizations
ω	Angular optical frequency
$\overline{\omega}$	Mean angular optical frequency

Maxwell's Equations and the Wave Equations

Maxwell's equations (SI units) in an unbounded, linear, isotropic, and homogeneous medium:

$$\nabla \cdot \mathbf{D} = \rho, \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$
$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{H} = \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}$$

The **material equations** are $\mathbf{D} = \varepsilon \mathbf{E}$, $\mathbf{B} = \mu \mathbf{H}$, $\mathbf{j} = \sigma \mathbf{E}$

$\mathbf{D}$	Electric displacement	$\mathrm{C/m^2}$
$\mathbf{E}$	Electric field	V/m
$\mathbf{B}$	Magnetic induction	T
$\mathbf{H}$	Magnetic field strength	A/m
$\mathbf{j}$	Electric current density	$\mathrm{A/m^2}$
ρ	Electric charge density	$\mathrm{C/m^3}$
ε	Permittivity	F/m
μ	Permeability	H/m
σ	Electric conductivity	$(\Omega \mathrm{m})^{-1}$

The vacuum permeability and permittivity are

$$\mu_0 = 4\pi \times 10^{-7} \text{ H/m}$$
$$\varepsilon_0 = 8.8541835 \times 10^{-12} \text{ F/m}$$

These seven equations can be manipulated into the following **vector wave equations**:

$$\nabla^2 \mathbf{E} = \varepsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2} + \mu\sigma \frac{\partial \mathbf{E}}{\partial t}, \quad \nabla^2 \mathbf{H} = \varepsilon\mu \frac{\partial^2 \mathbf{H}}{\partial t^2} + \mu\sigma \frac{\partial \mathbf{H}}{\partial t}$$

The **Laplacian** ∇^2 in Cartesian, cylindrical, and spherical coordinates are

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial z^2} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2}$$
$$= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial}{\partial \theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial \phi^2}$$

Since $\mathbf{E}$ and $\mathbf{H}$ are vector quantities, each component must satisfy the **scalar wave equation**, where V is a scalar electric or magnetic field:

$$\nabla^2 V = \varepsilon\mu \frac{\partial^2 V}{\partial t^2} + \mu\sigma \frac{\partial V}{\partial t}$$

Principle of Linear Superposition and Complex Notation

The **principle of linear superposition** (or superposition principle) states that the optical field at each point is the algebraic sum of the overlapping fields. This can be seen as a result of the linearity of the wave equation; if E_1 and E_2 are both solutions of the wave equation then

$$\nabla^2\mathbf{E}_1 = \varepsilon\mu\frac{\partial^2\mathbf{E}_1}{\partial t^2} + \mu\sigma\frac{\partial\mathbf{E}_1}{\partial t}, \quad \nabla^2\mathbf{E}_2 = \varepsilon\mu\frac{\partial^2\mathbf{E}_2}{\partial t^2} + \mu\sigma\frac{\partial\mathbf{E}_2}{\partial t}$$

The sum of the two solutions is another solution.

$$\nabla^2(\mathbf{E}_1 + \mathbf{E}_2) = \varepsilon\mu\frac{\partial^2}{\partial t^2}(\mathbf{E}_1 + \mathbf{E}_2) + \mu\sigma\frac{\partial}{\partial t}(\mathbf{E}_1 + \mathbf{E}_2)$$

This assumes, as is often the case, that the medium is linear or that the field strengths are not so large as to modify the optical characteristics of the material. Under this condition, the fields can overlap without affecting each other.

It is both common and highly convenient to use a complex representation of the optical field.

$$U = Ae^{i(kz - \omega t + \phi)}$$

Although the **complex notation** is used throughout this text, all optical fields are real quantities, and the actual field is the real part of the complex field:

$$\mathbf{E} = \mathrm{Re}[U] = \frac{1}{2}(U + U^*) = A\cos(kz - \omega t)$$

This is particularly important for nonlinear operations where the complex notation leads to incorrect results.

The Helmholtz Equation

When possible, it can be convenient to express the scalar field U as the product of two functions separately depending on the position $\mathbf{r}$ and time t.

$$U(\mathbf{r},t) = F(\mathbf{r})G(t)$$

Choosing $-k^2$ as the constant of separation, the wave equation can be separated into spatial and temporal wave equations, generally known as the **Helmholtz equations**:

$$\nabla^2 F(\mathbf{r}) + k^2 F(\mathbf{r}) = 0, \quad \frac{\partial^2 G(t)}{\partial t^2} + v^2 k^2 G(t) = 0$$

Although both of these equations are sometimes referred to as Helmholtz equations, in the field of physical optics, the term **Helmholtz equation** is usually reserved for the spatial wave equation:

$$\nabla^2 F(\mathbf{r}) + k^2 F(\mathbf{r}) = 0$$

An implication of these second-order homogeneous partial differential equations is that both F and G can be decomposed into harmonic functions, opening a vast mathematical toolbox for analyzing optical systems.

If F has the form $F(\mathbf{r}) = u(\mathbf{r})e^{ikz}$, and the second derivative of u with respect to z can be neglected, the result is the **paraxial Helmholtz equation**:

$$\nabla_{xy}^2 u + 2ik\frac{\partial u}{\partial z} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + 2ik\frac{\partial u}{\partial z} = 0$$

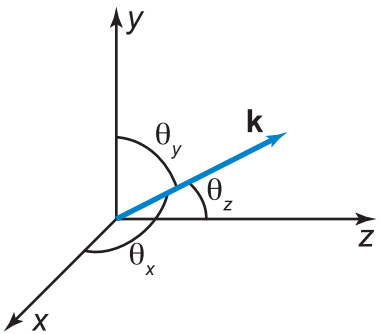
The paraxial Helmholtz equation is useful, for instance, in deriving the form of the **Gaussian beam**.

Plane Wave

A **plane wave** can be any periodic function where the surfaces of equal phase are planar. However, this term usually refers to a function that, satisfying the Helmholtz equation, varies harmonically in space and time as

$$\mathbf{E} = \mathbf{A}\cos(\mathbf{k}\cdot\mathbf{r} - \omega t + \phi)$$

A, in this case, is a constant vector describing the direction of the electric field, and **k** is the **wavevector**, which depends on the wavelength λ and the direction of propagation, which is expressed here in terms of the **direction cosines** α, β, and γ:



$$\mathbf{k} = k\hat{\mathbf{k}} = \frac{2\pi}{\lambda}(\alpha\hat{x} + \beta\hat{y} + \gamma\hat{\mathbf{z}}) = \frac{2\pi}{\lambda}(\hat{x}\cos\theta_x + \hat{y}\cos\theta_y + \hat{z}\cos\theta_z)$$

The magnitude of the wavevector $2\pi/\lambda$ is known as the **wavenumber**. Care must be taken with this term since the field of spectroscopy defines the wavenumber as $1/\lambda$ and typically expresses it in units of 1/cm.

Although actual fields are always real quantities, it is convenient and ubiquitous to specify the plane wave in **complex notation**.

$$\mathbf{U} = \mathbf{A}e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t + \phi)}$$

Beware that using the complex notation in nonlinear expressions can lead to incorrect results. Although ubiquitous, this convenient notation should be used with some caution and with the understanding that fields are actually real quantities.

Spherical and Cylindrical Waves

The idealized point source radiates energy isotropically, and symmetry dictates that the wavefronts form concentric spheres known as **spherical waves**. In a spherical coordinate system centered on the source point, the functional form of the spherical wave can only depend on the radial coordinate r and time t. With the **Laplacian** in spherical coordinates, the scalar wave equation becomes a 1D wave equation where the wave function is the product rU.

$$\frac{\partial^2}{\partial r^2}(rU) = \frac{1}{\nu^2}\frac{\partial^2}{\partial t^2}(rU)$$

The **harmonic spherical wave** is a solution to this equation and may be expressed in terms of a source at some source point location $\mathbf{r}_0 = x_0\hat{\mathbf{x}} + y_0\hat{\mathbf{y}} + z_0\hat{\mathbf{z}}$ and initial phase ϕ_0, and traveling away from the source.

$$U(r,t) = \frac{A}{|r - r_0|}e^{i(k|r - r_0| - \omega t + \phi_0)}$$

True point sources are not found in nature, but the spherical wave is still extremely useful in practical engineering problems, where it can form a good approximation to many types of point-like sources.

Also of immense practical importance is the paraxial approximation to the spherical wave, or the **Fresnel approximation**:

$$U(r,t) \approx \frac{A}{z - z_0}e^{\frac{ik}{2(z - z_0)}\left[(x - x_0)^2 + (y - y_0)^2\right]}e^{i\left(k(z - z_0) - \omega t + \phi_0\right)}$$

The **harmonic cylindrical wave** is also a solution to the wave equation and is most easily expressed in cylindrical coordinates:

$$U(r,t) = \frac{A}{\sqrt{r}}e^{i(kr - \omega t + \phi_0)}$$

Speed of Light

Substituting the complex representation of a plane wave V with frequency ω, traveling along the z axis, into the wave equation provides an expression relating the speed ν to frequency ω, permittivity, permeability, and conductivity:

$$V = Ae^{i\omega\left(\frac{z}{v}-t\right)} \rightarrow \frac{\omega^2}{\nu^2} = \varepsilon\mu\omega^2 + i\mu\sigma\omega$$

Most problems in optics are set in nonabsorbing media where the imaginary part is very close to zero, and $\nu = c/n = 1/\sqrt{\varepsilon\mu}$. In vacuum, $n = 1$ and conductivity is zero, the **speed of light** is the vacuum value c, and

$$c = 1/\sqrt{\varepsilon_0\mu_0} = 2.99792458 \times 10^8 \text{m/s}$$

More generally, the speed of light is modified by the **complex refractive index**, which has real and imaginary parts n and κ, respectively. The imaginary part is known as the **extinction coefficient**, and the real part is often referred to as the **index of refraction** or **refractive index**.

$$N = \frac{c}{\nu} = n + i\kappa$$

Note that the sign of the imaginary part depends on the choice of increasing or decreasing phase.

Decreasing phase	$e^{i(kz-\omega t)} \rightarrow N = n + i\kappa$
Increasing phase	$e^{-i(kz-\omega t)} \rightarrow N = n - i\kappa$

Below are some other useful relationships where the subscript r indicates a value *relative* to the vacuum value.

$$n^2 - \kappa^2 = \frac{\varepsilon\mu}{\varepsilon_0\mu_0} = \varepsilon_r\mu_r, \quad 2n\kappa = \frac{\mu\sigma}{\varepsilon_0\mu_0\omega} = \frac{\mu_r\sigma}{\varepsilon_0\omega}$$

The Beer–Lambert Law

The **transmittance** T through a material is the ratio of the input-to-output power or flux. A complex refractive index in the exponential form of a plane wave produces an exponentially decaying transmittance, which is the essence of the **Beer–Lambert law**, or **Beer's law**. This describes the exponential decay with propagation through a medium with the **extinction coefficient** κ or, equivalently, the **absorption coefficient** α:

$$T = \|U/U_0\|^2 = e^{-\frac{2\omega}{c}\kappa z} = e^{-\frac{4\pi}{\lambda}\kappa z} = e^{-\alpha z}$$

$$\alpha = \frac{2\omega}{c}\kappa = \frac{4\pi}{\lambda}\kappa$$

If an absorbing medium consists of several absorbers, the absorption coefficients add in proportion to their concentrations. For example, if M absorbers have absorption cross sections σ_j and number densities N_j, then the Beer–Lambert law becomes

$$T = e^{-z\sum_{j=1}^{M}\sigma_j N_j}$$

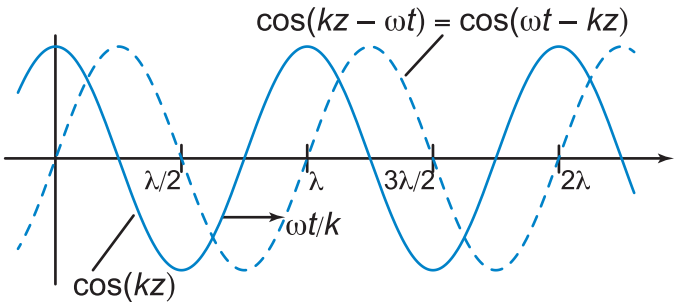
Similar expressions can be formed in terms of other quantities such as molar concentration and mass density absorption. It is also common to see absorption defined in base 10 rather than base e, in which case,

$$T = e^{-\alpha z} = 10^{-\alpha' z} \quad \rightarrow \quad \alpha' = \alpha/\ln 10$$

Beer's law assumes that the medium is linear, homogenous, isotropic, and nonscattering, and, strictly speaking, applies only to collimated beams of a single frequency. The transmission of a more general beam can be estimated by integrating Beer's law over angles and wavelengths of the rays (or components of an **angular spectrum of plane waves**) in the beam.

Increasing versus Decreasing Phase Convention

Both $\cos(kz - \omega t)$ and $\cos(\omega t - kz)$ represent a monochromatic wave traveling in the positive z direction.



In complex notation, both are commonly used and are valid. However, practitioners of physical optics should become skilled at spotting the differences because the resulting errors can be subtle and sometimes devastating.

The most obvious implication of the choice between increasing and decreasing phase conventions is the sign of the imaginary part of the complex refractive index:

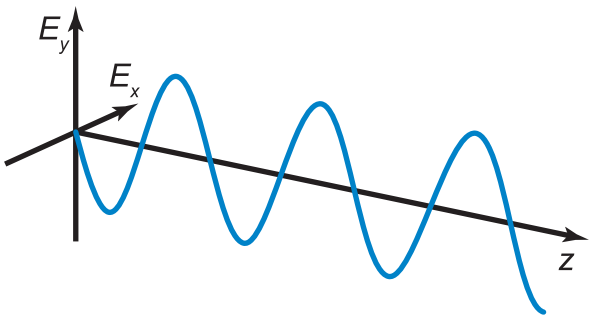
Decreasing phase convention : $U = Ae^{i(kz - \omega t)} \rightarrow N = n + i\kappa$

Increasing phase convention : $U = Ae^{-i(kz - \omega t)} \rightarrow N = n - i\kappa$

It may be parsimoniously argued that the **increasing phase convention** describes the behavior of a source whose phase is increasing with time, and so the phase should decrease with increasing distance from the source.

The **decreasing phase convention** used in this text implies that the phase decreases with time but increases with optical path. This convention is convenient in problems that are either monochromatic or can be treated as sums of monochromatic systems. In either case, the time part of the phase can be dropped providing a simplification in notation since the minus sign in the exponent does not need to be carried throughout.

Linear Polarization



The electric field of a **linearly polarized** harmonic plane traces a sinusoid in a plane parallel to the direction of propagation. This plane is known as the **plane of vibration** and is pictured as the vertical plane in the figure above.

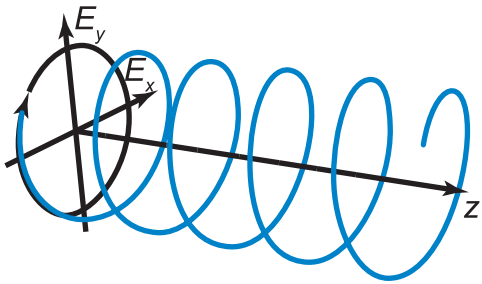
In a plane that is perpendicular to the direction of propagation, at some fixed location z, the electric field traces out a line segment over time that is also contained in the plane of polarization:

$$\begin{aligned}\mathbf{E}_\theta &= \mathrm{Re}[Ae^{i(kz-\omega t+\phi_0)}(\hat{x}\cos\theta+\hat{y}\sin\theta)]\\ &= A(\hat{x}\cos\theta+\hat{y}\sin\theta)\end{aligned}$$

Because the electric field adds in a linear medium, the **principle of linear superposition** means that any elliptical polarization state can be expressed as a combination of two orthogonal linear polarization states whose relative phase and amplitude determine the shape and orientation of the polarization ellipse.

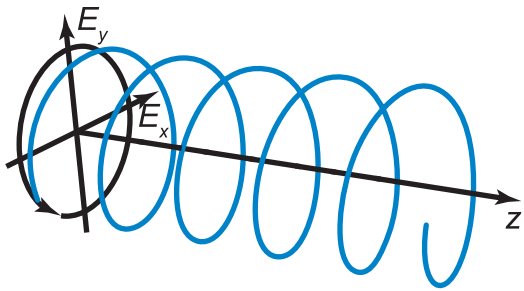
For this reason, many problems are broken into separate problems for two linear polarization states. For example, it is convenient to treat s and p polarization states separately in thin films.

Right- and Left-Hand Circular Polarization



$$\mathbf{E}_{RCP} = \mathrm{Re}\left[\frac{A}{\sqrt{2}} e^{i(kz-\omega t+\phi_0)}(\hat{x}+\hat{y}e^{-i\pi/2})\right]$$
$$= \frac{A}{\sqrt{2}}\left[\hat{x}\cos(kz-\omega t+\phi_0)+\hat{y}\sin(kz-\omega t+\phi_0)\right]$$

At a moment in time, the electric field of a **right-hand circular polarized** plane wave traces out a right-hand corkscrew along the direction of propagation. In a fixed position, the electric field sweeps out a clockwise circle in time as seen by an observer looking at the source.



$$\mathbf{E}_{LCP} = \mathrm{Re}\left[\frac{A}{\sqrt{2}} e^{i(kz-\omega t+\phi_0)}(\hat{x}+\hat{y}e^{i\pi/2})\right]$$
$$= \frac{A}{\sqrt{2}}\left[\hat{x}\cos(kz-\omega t+\phi_0)-\hat{y}\sin(kz-\omega t+\phi_0)\right]$$

In **left-hand circular polarization**, the corkscrew, frozen in time, is left-handed, and at a fixed position the electric field sweeps out a counter-clockwise circle as seen by an observer looking at the source.

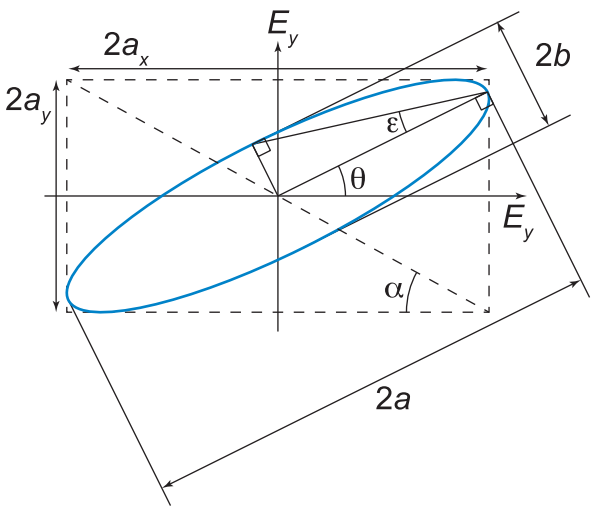
Elliptical Polarization

The electric field vector of a harmonic plane wave traveling along the z axis can be written somewhat generally:

$$\mathbf{E} = a_x\,\hat{x}\cos(kz - \omega t + \delta_x) + a_y\,\hat{y}\cos(kz - \omega t + \delta_y)$$

This general state is known as **elliptical polarization** because the field at a fixed z traces an ellipse as described by

$$\left(\frac{E_x}{a_x}\right)^2 + \left(\frac{E_y}{a_y}\right)^2 - 2\left(\frac{E_x}{a_x}\right)\left(\frac{E_y}{a_y}\right)\cos\phi = \sin^2\phi$$



$$\phi = \delta_y - \delta_x \qquad \tan\alpha = a_y/a_x \qquad (0 \le \alpha \le \pi/2)$$

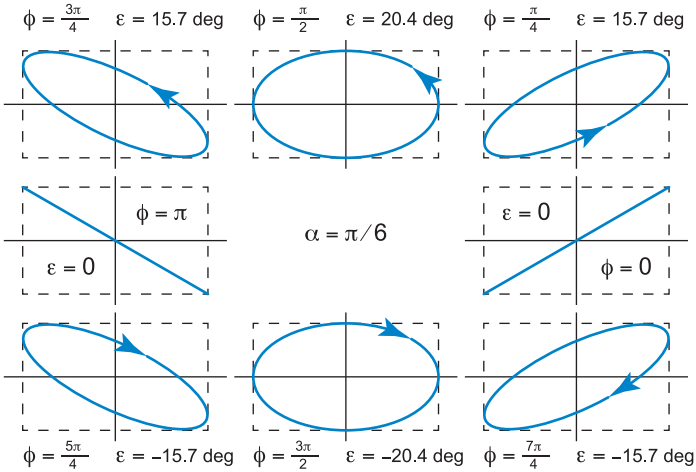
$$\tan\varepsilon = \mp\frac{b}{a} \qquad \tan 2\theta = \tan 2\alpha\cos\phi \qquad (0 \le \theta \le \pi)$$

$$\sin 2\varepsilon = \sin 2\alpha\sin\phi \qquad (-\pi/4 \le \varepsilon \le \pi/4)$$

Elliptical Polarization Handedness

Elliptical polarization handedness depends on the sign of ε. For right-handed elliptical polarization ($\varepsilon < 0$) and the electrical field traces a clockwise oval when looking toward the source, or equivalently, makes a right-handed corkscrew in space at a moment in time.

Right handed	*Clockwise*	$\varepsilon < 0$
Linear		$\varepsilon = 0$
Left handed	*Counter-clockwise*	$\varepsilon > 0$



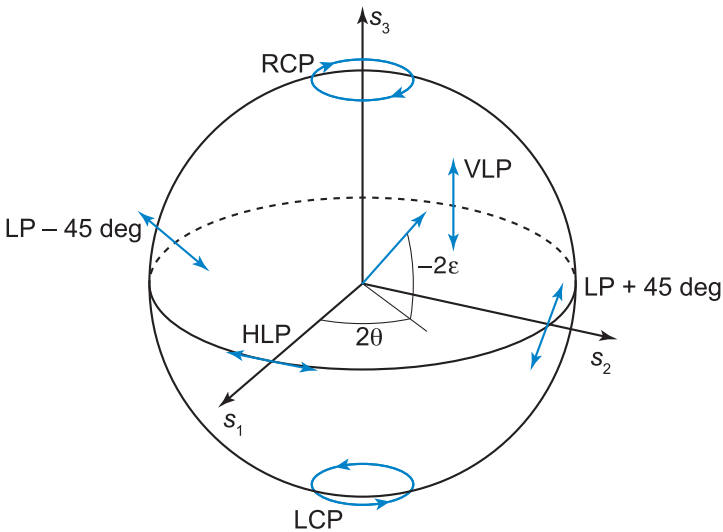
The **Jones vector** **E** and **Stokes vector** **S** describing elliptical polarization are

$$\mathbf{E} = \begin{pmatrix} a_x \\ a_y\, e^{i\phi} \end{pmatrix} = \sqrt{a_x^2 + a_y^2} \begin{pmatrix} \cos\alpha \\ e^{i\phi} \sin\alpha \end{pmatrix}$$

$$\mathbf{S} = \begin{pmatrix} a_x^2 + a_y^2 \\ a_x^2 - a_y^2 \\ 2a_x a_y \cos\phi \\ -2a_x a_y \sin\phi \end{pmatrix} = \begin{pmatrix} S_0 \\ S_0\, 2\cos 2\alpha \\ S_0 \sin 2\alpha \cos\phi \\ -S_0 \sin 2\alpha \sin\phi \end{pmatrix} = S_0 \begin{pmatrix} 1 \\ \cos 2\varepsilon \cos 2\theta \\ \cos 2\varepsilon \sin 2\theta \\ -\sin 2\varepsilon \end{pmatrix}$$

Poincaré Sphere

The **Poincaré sphere** visually represents polarization states expressed in terms of Stokes parameters or the angles ε and θ.



Normalized Stokes vectors are commonly used so that polarization states residing on the surface of the sphere have a maximum **degree of polarization (DOP)**: DOP = 1.

Linear polarization states reside along the equator and vary continuously from **horizontal linear polarization (HLP)**, to linear at 45 deg (LP + 45 deg), to **vertical linear polarization (VLP)**. Right-hand elliptical states reside above the equator with **right-hand circular polarization (RCP)** at the north pole. The southern hemisphere contains the left-hand states with **left-hand circular polarization (LCP)** at the south pole. Note that orthogonal polarizations are always on opposite sides of the origin.

Jones Vectors

Since the optical field of an elliptically polarized plane wave can be expressed as a sum of orthogonal linear-polarization-basis states having the same frequency and wavelength, the time and space dependency can be factored from a constant complex vector $\boldsymbol{E}$ known as the **Jones vector**:

$$\mathbf{E} = \hat{\mathbf{x}}\, a_x\, e^{i(kz - \omega t + \delta_x)} + \hat{\mathbf{y}}\, a_y\, e^{i(kz - \omega t + \delta_y)} = \mathbf{E}\, e^{i(kz - \omega t)}$$

$$\mathbf{E} = \begin{pmatrix} J_x \\ J_y \end{pmatrix} = A \begin{pmatrix} a_x \\ a_y e^{i\phi} \end{pmatrix}$$

The following table provides a sampling of unit-amplitude Jones vectors.

Horizontal linear	$\mathbf{E}_{HLP} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$
Vertical linear	$\mathbf{E}_{VLP} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
Linear at 45 deg	$\mathbf{E}_{L+45\ \mathrm{deg}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
Linear at −45 deg	$\mathbf{E}_{L-45\ \mathrm{deg}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$
Linear at angle θ	$\mathbf{E}_{L\theta} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$
Right-hand circular	$\mathbf{E}_{RCP} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}$
Left-hand circular	$\mathbf{E}_{LCP} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ +i \end{pmatrix}$

Jones Matrices and Eigenpolarizations

Many polarization problems can be treated with **Jones calculus**, where an input polarization state represented by the **Jones vector** $\mathbf{E}_1$ is modified by an optical element or group of elements described by a 2×2 **Jones matrix** $\mathbf{J}$ to produce the output polarization $\mathbf{E}_2$:

$$\mathbf{E}_2 = \mathbf{J}\,\mathbf{E}_1 = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} E_{1x} \\ E_{1y} \end{pmatrix} = \begin{pmatrix} m_{11}E_{1x} + m_{12}E_{1y} \\ m_{21}E_{1x} + m_{22}E_{1y} \end{pmatrix}$$

If there are N polarization elements, the system Jones matrix is the product of individual matrices.

$$\mathbf{E}' = \mathbf{J}_N \cdots \mathbf{J}_2 \mathbf{J}_1\, \mathbf{E} = \mathbf{J}_{system}\, \mathbf{E}$$

A Jones matrix can have two orthogonal **eigenpolarizations**, represented by the Jones eigenvectors ψ_1 and ψ_2 with the amplitude transmissions represented by eigenvalues λ_1 and λ_2. In terms of the trace $T = m_{11} + m_{22}$ and determinant $D = m_{11}m_{22} - m_{12}m_{21}$ of $\mathbf{J}$:

$$\lambda_1 = T/2 + \sqrt{T^2/2 - D}, \quad \lambda_2 = T/2 - \sqrt{T^2/2 - D}$$

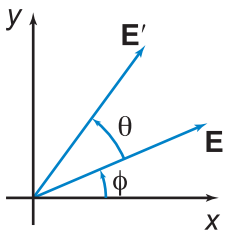
$$\begin{aligned} &\textit{if } m_{12} \neq 0, \quad \psi_1 = (m_{12}, \lambda_1 - m_{11})^T, \quad \psi_2 = (m_{12}, \lambda_2 - m_{11})^T \\ &\textit{if } m_{21} \neq 0, \quad \psi_1 = (\lambda_1 - m_{22}, m_{21})^T, \quad \psi_2 = (\lambda_2 - m_{22}, m_{21})^T \\ &\textit{if } m_{12} = m_{21} = 0, \qquad \psi_1 = (1,0)^T, \quad \psi_2 = (0,1)^T \end{aligned}$$

The eigenpolarizations $\psi_1 = (a,\, b)^T$, $\psi_2 = (-b^*,\, a^*)^T$ and eigenvalues λ_1 and λ_2 (where a^T denotes a matrix transpose) can be used to derive the Jones matrix:

$$\mathbf{J} = \frac{1}{a\,a^* + b\,b^*} \begin{pmatrix} a & -b^* \\ b & a^* \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} a^* & b^* \\ -b & a \end{pmatrix}$$

Jones Rotation and Reflection Matrices

Counter-clockwise rotation of a vector about the z axis can be expressed as a matrix operator $\mathbf{J}_{Rot}(\theta)$.



$$\mathbf{E}' = \begin{pmatrix} E\cos(\phi+\theta) \\ E\sin(\phi+\theta) \end{pmatrix}$$

$$= E\begin{pmatrix} \cos\phi\cos\theta - \sin\phi\sin\theta \\ \cos\phi\sin\theta + \sin\phi\cos\theta \end{pmatrix}$$

$$\mathbf{E}' = \mathbf{J}_{Rot}(\theta)\mathbf{E}$$

$$\mathbf{J}_{Rot}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

The **Jones rotation matrix** $\mathbf{J}_{Rot}(\theta)$ can be used to rotate a Jones vector as shown above, but it may also be used to rotate any Jones matrix such as a retarder or polarizer.

$$\mathbf{J}' = \mathbf{J}_{Rot}(\Delta\theta)\,\mathbf{J}\,\mathbf{J}_{Rot}^{T}(\Delta\theta) = \mathbf{J}_{Rot}(\Delta\theta)\,\mathbf{J}\,\mathbf{J}_{Rot}(-\Delta\theta)$$

Properties of the Jones rotation matrix:

$$\mathrm{Det}[\mathbf{J}_{Rot}(\theta)] = 1$$
$$\mathbf{J}_{Rot}^{T}(\theta) = \mathbf{J}_{Rot}(-\theta)$$
$$\mathbf{J}_{Rot}(\theta_2)\mathbf{J}_{Rot}(\theta_1) = \mathbf{J}_{Rot}(\theta_1+\theta_2)$$

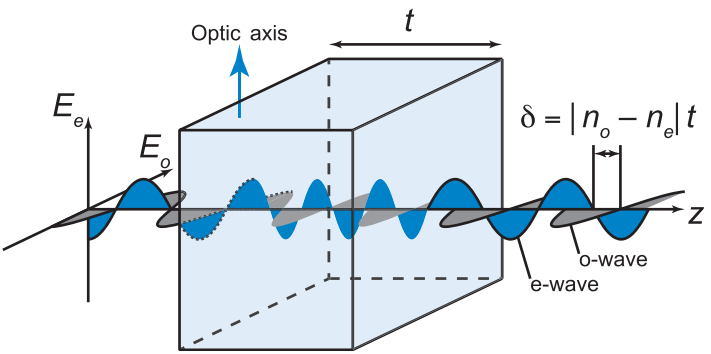
Reflection reverses the direction of propagation. By convention, this reverses the x component of the field so that the **Jones reflection matrix** is

$$\mathbf{J}_{Mirror} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

Note that the reflection matrix is not invariant under the above form of rotation, so care must be taken when using the rotation matrix in a system with a mirror.

Retardance

A linearly polarized plane wave propagating within a **uniaxial crystal**—such as calcite ($CaCO_3$) or quartz (SiO_2)—with its electric field aligned to the optic axis (crystal-lattice axis of symmetry) is known as an extraordinary wave, or **e-wave**, and experiences a refractive index n_e. When the electric field is perpendicular to the optic axis, it is known as an ordinary wave, or **o-wave**, and experiences a refractive index n_o. The difference in phase between the e-wave and o-wave is the **retardance** and may be expressed as a phase angle, such as degrees or radians, or in terms of optical path in waves or nanometers.



In a **positive crystal**, such as quartz, $n_e > n_o$, the e-wave travels slower by virtue of its higher refractive index so that the optic axis is known as the **slow axis**, while the **fast axis** is perpendicular to the optic axis. In a **negative crystal**, $n_e < n_o$, and the fast axis is parallel to the optic axis.

Biaxial birefringent crystals, such as mica and borax, exhibit three characteristic refractive indices. Cubic crystals, such as NaCl or CaFl, are not usually considered birefringent except in the ultraviolet where they present a retardance with a higher-order angular dependence known as cubic birefringence.

Retardance (cont.)

As indicated by the phase difference between *s* and *p* polarization in the Fresnel equations, reflection can also produce **retardance** that depends on the refractive indices, wavelength, and angle of incidence. Thin-film theory shows that this effect also depends on the materials and thicknesses that make up the thin-film stack.

Retardance of bulk material can also be attributed to mechanical stress from an applied force, thermal effects, or nonuniformities intrinsic to the transmitting materials, such as glass or plastic.

An optical component that is designed to introduce a phase difference between orthogonal polarization states is known as a retarder. If the polarization states are horizontal and vertical, and the retardance is given in radians as ϕ, then the Jones matrix for a retarder can be derived as follows:

$$\left.\begin{aligned}\psi_1 &= \begin{pmatrix}1\\0\end{pmatrix} \lambda_1 = e^{i\phi/2}\\ \psi_2 &= \begin{pmatrix}0\\1\end{pmatrix} \lambda_2 = e^{-i\phi/2}\end{aligned}\right\} \rightarrow \mathbf{J}_R(\phi) = \begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}e^{i\phi/2}&0\\0&e^{-i\phi/2}\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}$$

$$\mathbf{J}_R(\phi) = \begin{pmatrix}e^{i\phi/2}&0\\0&e^{-i\phi/2}\end{pmatrix}$$

If $\phi > 0$, then this Jones matrix describes a retarder with a horizontal slow axis, since the *x* polarization is retarded more than the *y* polarization.

This text uses the **decreasing phase convention**, and it should be noted that under the **increasing phase convention**, this Jones matrix would indicate a vertical slow axis for $\phi > 0$.

Dichroism and Diattenuation

A dichroic material has an **extinction coefficient** that depends on the orientation of the optical field. A dichroic uniaxial crystal, such as tourmaline, has anisotropic conductivity that produces different extinction coefficients parallel and perpendicular to the optic axis. When the **dichroism** is large, the material may be used as a polarizer, but uniaxial crystals are seldom used in this way. It is more common to use polymer-based sheet polarizers, which produce a more controllable dichroism.

The terms dichroic and dichroism also refer to the separation of colors—the safer term for referring only to the difference in amplitude transmission between eigenpolarizations is **diattenuation**. The Jones matrix for linear diattenuation with horizontal amplitude transmission a_x and vertical amplitude transmission a_y can be derived as follows:

$$\left.\begin{aligned}\psi_1 &= \begin{pmatrix}1\\0\end{pmatrix} \quad \lambda_1 = t_x\\ \psi_2 &= \begin{pmatrix}0\\1\end{pmatrix} \quad \lambda_2 = t_y\end{aligned}\right\} \rightarrow \mathbf{J}_D(t_x,t_y) = \frac{1}{1^2+0^2}\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}a_x&0\\0&a_y\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}$$

$$\mathbf{J}_D(a_x,a_y) = \begin{pmatrix}a_x&0\\0&a_y\end{pmatrix}$$

The diattenuation D of an element is the contrast between the maximum $T_{\max}$ and minimum $T_{\min}$ possible transmittance:

$$D = \frac{T_{\max} - T_{\min}}{T_{\max} + T_{\min}}$$

In the case of diattenuation of linear polarization,

$$D = |a_x^2 - a_y^2|/(a_x^2 + a_y^2)$$

Polarizers and Malus' Law

A **polarizer** is a device designed to remove all of the polarization components but one: the preferred polarization. The degree to which the polarizer accomplishes this is characterized by the **polarization extinction ratio (PER)**:

$$PER = \frac{T_{\max}}{T_{\min}} = \frac{1+D}{1-D}$$

A **linear polarizer** that perfectly transmits linear polarization oriented at an angle θ with the x axis can be characterized by its eigenpolarizations and Jones matrix.

$$\left.\begin{aligned} \psi_1 &= \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} & \lambda_1 &= 1 \\ \psi_2 &= \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} & \lambda_2 &= 0 \end{aligned}\right\} \rightarrow \mathbf{J}_{LP}(\theta) = \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}$$

The transmittance of linearly polarized light at an angle $\theta + \Delta\theta$ and directed into such a device is governed by **Malus' law**.

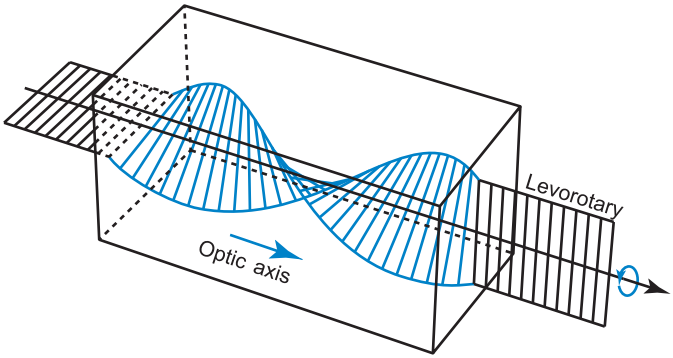
$$I_t/I_i = \cos^2\Delta\theta$$

A circular polarizer passes only one circular-polarization state. The Jones matrices for **right- and left-hand circular polarizers** can be constructed as

$$\left.\begin{aligned} \mathbf{E}_{RCP} &= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix} & \lambda_{RCP} &= 1 \\ \mathbf{E}_{LCP} &= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} & \lambda_{LCP} &= 0 \end{aligned}\right\} \rightarrow \mathbf{J}_{RCP} = \frac{1}{2}\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$$

$$\left.\begin{aligned} \mathbf{E}_{RCP} &= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix} & \lambda_{RCP} &= 0 \\ \mathbf{E}_{LCP} &= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} & \lambda_{LCP} &= 1 \end{aligned}\right\} \rightarrow \mathbf{J}_{LCP} = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}$$

Optical Activity



A material that rotates the polarization state is **optically active**. The Jones matrix for optical activity is a phase-modified **rotation matrix** $\mathbf{J}_{ROT}$ whose eigenpolarizations are left- and right-hand circular polarization:

$$\mathbf{E}_{RCP}=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix},\ \lambda_{RCP}=e^{i\frac{2\pi}{\lambda}n_R t},\ \mathbf{E}_{LCP}=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix},\ \lambda_{LCP}=e^{i\frac{2\pi}{\lambda}n_L t}$$

$$\mathbf{J}_\beta=\frac{1}{2}\begin{pmatrix}1&-i\\-i&1\end{pmatrix}\begin{pmatrix}e^{i\frac{2\pi}{\lambda}n_R t}&0\\0&e^{i\frac{2\pi}{\lambda}n_L t}\end{pmatrix}\begin{pmatrix}1&i\\i&1\end{pmatrix}=e^{i\phi}\begin{pmatrix}\cos\beta&-\sin\beta\\\sin\beta&\cos\beta\end{pmatrix}$$

Optical activity generates a phase ϕ and a counter-clockwise rotation $\beta>0$ of the polarization state about the direction of propagation:

$$\phi=\pi t(n_R+n_L)/\lambda,\beta=\pi t(n_R-n_L)/\lambda$$

Traditionally, materials that rotate the polarization clockwise when looking toward the source are **dextrorotatory** or d-rotary ($\beta<0$), while counter-clockwise rotation when looking toward the source indicates a **levorotatory** or l-rotary material ($\beta>0$). The amount of optical activity varies with material and wavelength, and is given by the **specific rotary power** in units of angle per length:

$$Specific\ rotary\ power=\frac{\beta}{t}$$

Stokes Vectors and Degree of Polarization

The **Stokes parameters** can be understood in terms of the measured irradiances $\{I_0, I_H, I_{45\text{ deg}}, I_R\}$ transmitted through four filters, each transmitting 50% of natural, unpolarized light. Conventionally, the first filter passes all polarizations equally, the second is a linear polarizer with a horizontal transmission axis, the third is a linear polarizer with a 45 deg transmission axis, and the fourth is a right-hand circular polarizer. The Stokes parameters conventionally expressed as a columnar four vector are

$$\mathbf{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} 2I_0 \\ 2(I_H - I_0) \\ 2(I_{45\text{ deg}} - I_0) \\ 2(I_R - I_0) \end{pmatrix}$$

Unlike Jones vectors, **Stokes vectors** can describe partial polarization, and when a beam is a result of combined incoherent and partially polarized sources, the resulting Stokes vector is a sum of the component Stokes vectors.

$\hat{\mathrm{S}}$ is the normalized Stokes vector, and $\mathbf{s}_p$ is the Stokes-three vector.

$$\hat{\mathrm{S}} = \mathbf{S}/S_0 = \begin{pmatrix} 1 \\ s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{s}_p \end{pmatrix}$$

which is useful in defining the **degree of polarization**:

$$DOP = |\mathbf{s}_p| = \sqrt{S_1^2 + S_2^2 + S_3^2}/S_0, \quad 0 \le DOP \le 1$$

For a perfectly polarized beam, $DOP = 1$, and the Stokes vector lies on the surface of the Poincaré sphere. When $DOP < 1$, the beam is partially polarized, and the Stokes vector is somewhere inside the Poincaré sphere. When $DOP = 0$, the Stokes vector is at the origin, corresponding to totally unpolarized light.

Mueller Matrices

When using Stokes vectors, polarization elements can be described with 4×4 **Mueller matrices**.

$$\mathbf{S}' = \mathbf{M}\,\mathbf{S}$$

The Mueller matrix for **diattenuation** aligned with the x and y axes:

$$\mathbf{M}_D(T_x, T_y) = \frac{1}{2}\begin{pmatrix} T_x + T_y & T_x - T_y & 0 & 0 \\ T_x - T_y & T_x + T_y & 0 & 0 \\ 0 & 0 & 2\sqrt{T_x T_y} & 0 \\ 0 & 0 & 0 & 2\sqrt{T_x T_y} \end{pmatrix}$$

Horizontal and **vertical linear polarizers:**

$$\mathbf{M}_{HLP} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \mathbf{M}_{VLP} = \frac{1}{2}\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Right- and **left-hand circular polarizers:**

$$\mathbf{M}_{RCP} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{M}_{LCP} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

The Mueller matrix for a retarder with **retardance** ϕ and slow axis aligned to the y direction:

$$\mathbf{M}_R(\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\phi & -\sin\phi \\ 0 & 0 & \sin\phi & \cos\phi \end{pmatrix}, \quad \begin{aligned} \mathbf{M}_{QWP} &= \mathbf{M}_R(\pi/4) \\ \mathbf{M}_{HWP} &= \mathbf{M}_R(\pi/2) \end{aligned}$$

Mueller Matrices and Rotation

The **Mueller matrix** that performs a rotation of the laboratory coordinate axes is

$$\mathbf{M}_{Rot}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta & 0 \\ 0 & -\sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The **rotation matrix** can be used to rotate a state of polarization and therefore behave like optical activity with $\theta = -\beta$.

$$\mathbf{S}' = \mathbf{M}_{Rot}(-\beta)S = \begin{pmatrix} S_0 \\ S_1 \cos 2\theta + S_2 \sin 2\theta \\ S_1 \cos 2\theta - S_2 \sin 2\theta \\ S_3 \end{pmatrix}$$

Or it can be used to model arbitrarily aligned components such as a rotated diattenuator/polarizer or a rotated retarder:

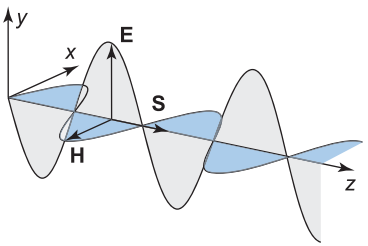
$$\mathbf{M}(\theta) = M_{Rot}(-\theta)\mathbf{M}\mathbf{M}_{Rot}(\theta)$$

For example, a polarizer with a transmission axis at an angle θ from horizontal is

$$\mathbf{M}_{LP}(\theta) = \frac{1}{2}\begin{pmatrix} 1 & \cos 2\theta & \sin 2\theta & 0 \\ \cos 2\theta & \cos^2 2\theta & \cos 2\theta \sin 2\theta & 0 \\ \sin 2\theta & \cos 2\theta \sin 2\theta & \sin^2 2\theta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Poynting Vector, Irradiance, and Optical Admittance

In unbounded, isotropic, nonabsorbing media without charges or currents, plane waves travel as transverse waves, and the electric and magnetic fields form an orthogonal triad with the **Poynting vector** **S**.



The Poynting vector, with units of W/m^2, is interpreted as the instantaneous radiant-energy-flux density. The time-averaged Poynting vector is then the **irradiance** perpendicular to the direction of travel $I_\perp$. In complex notation with constant amplitudes $\mathbf{E}_0$ and $\mathbf{H}_0$, the irradiance is

$$\langle \mathbf{S} \rangle_t = \frac{1}{2} \mathrm{Re}\, \{\mathbf{E} \times \mathbf{H}^*\}$$

$$I_\perp = \frac{1}{2} y |E_0|^2, \quad \text{where} \quad y = \frac{N}{c\mu}$$

The **optical admittance** (or **characteristic optical admittance**) y has SI units of siemens and is the ratio of magnetic and electric field strength of the electromagnetic wave.

If the medium is nonmagnetic and nonconductive, the optical admittance becomes

$$y = \frac{n}{c\mu_0} = n\, c\, \varepsilon_0$$

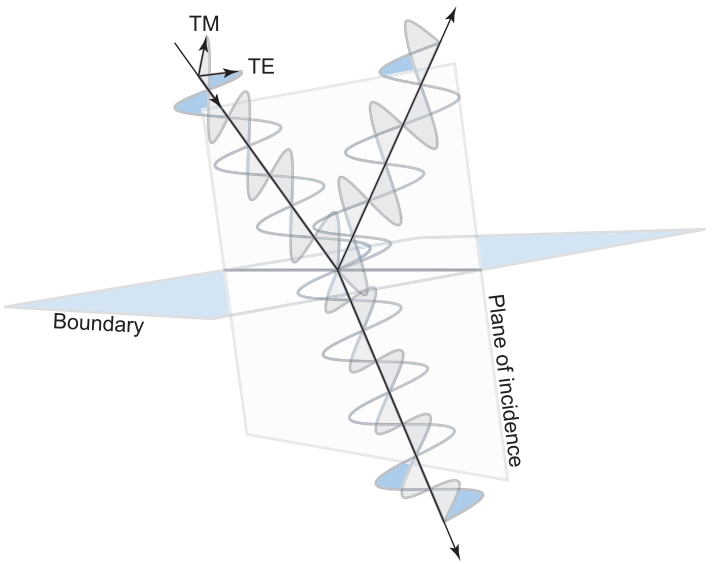
and in vacuum

$$y_0 = \sqrt{\frac{\varepsilon_0}{\mu_0}} = 2.6544 \times 10^{-3} \mathbf{S}$$

Some authors use the optical inductance $1/y$ instead.

Plane of Incidence

The diagram below illustrates the interaction of a plane wave incident on a planar boundary. The direction of propagation of the incident, transmitted, and reflected waves are contained within the **plane of incidence**.



Polarization is defined in terms of the orientation of the electric field, and it is often convenient to define the polarization state as either perpendicular or parallel to the plane of incidence.

When the electric field is perpendicular to the plane of incidence, the wave is **transverse electric** or TE polarized, and is known as ***s* polarization**.

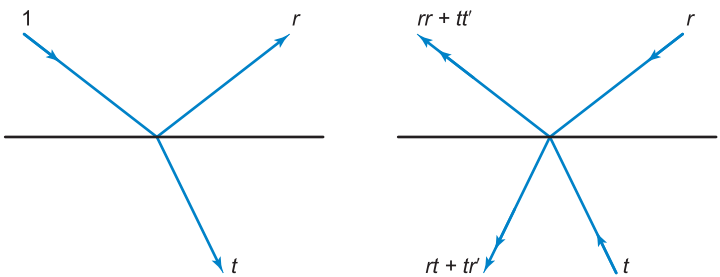
When the electric field is parallel to the plane of incidence, the wave is **transverse magnetic**, or TM polarized, and is known as ***p* polarization**.

The "*s*" in *s* polarization stands for *senkrecht*—German for perpendicular. The "*p*" in *p* polarization stands for *parallel*, which is also German... for parallel.

Stokes Relations

The **principle of reversibility** states that if there are no losses, then wave propagation is reversible.

The left diagram illustrates a unit-amplitude plane wave incident on a planar boundary between two homogeneous dielectric materials with the reflected and transmitted amplitudes r and t. Reversing the directions of the two waves, as shown in the right figure, results in one wave returning toward the source and another component directed back into the second material.



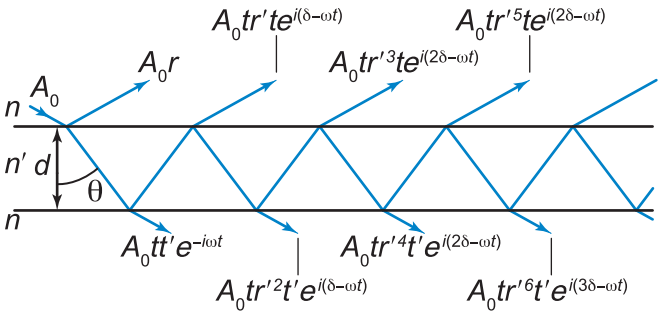
The principle of reversibility requires that the amplitude of the wave returning to the source is equal to the incident amplitude, while the amplitude directed into the medium must be zero.

The two equations below are known as **Stokes relations**. The second equation implies that there is always a 180-deg phase difference between waves reflected from either side of the boundary.

$$\begin{aligned} r\,r + t\,t' &= 1 \\ r\,t + t\,r' &= 0 \end{aligned} \quad \rightarrow \quad \begin{aligned} t\,t' &= 1 - r\,r' \\ r &= -r' \end{aligned}$$

In addition to reversibility, the Stokes relations also assume that the waves add in amplitude, meaning that they obey the principle of linear superposition.

Airy Formulae and Airy Function



A plane wave of amplitude A_0 is incident on a plane-parallel plate of index n'. By the principle of linear superposition, the total reflected or transmitted amplitude is the sum of all of the components directed toward the respective space. The coefficients of reflection and transmission (r, r', t, and t') are those used in the Stokes relations, and the phase difference between adjacent rays is δ. Assuming a lossless boundary, applying the Stokes relations, and simplifying the infinite sums produces

$$\frac{A_r}{A_0} = A_0(r + r'tt'e^{i\delta} + r'^3tt'e^{i2\delta} + \cdots) = r\frac{1 - e^{i\delta}}{1 - r^2e^{i\delta}}$$

$$\frac{A_t}{A_0} = tt'e^{i\delta} + tr'^2t'e^{i2\delta} + tr'^3t'e^{i3\delta} + \cdots = \frac{1 - r^2}{1 - r^2e^{i\delta}}$$

The **Airy formulae** describe the ratio of reflected and transmitted irradiance relative to the incident irradiance, while the **Airy function** specifically refers to I_t/I_0.

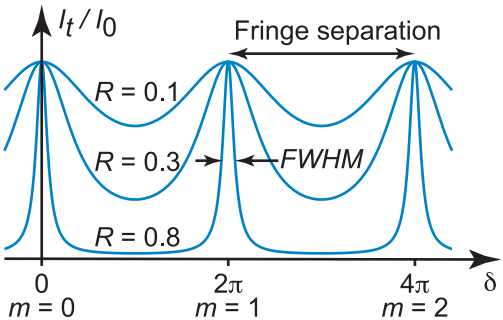
$$\frac{I_r}{I_0} = \frac{\left(\frac{2r}{1-r^2}\right)^2 \sin^2 \delta/2}{1 + \left(\frac{2r}{1-r^2}\right)^2 \sin^2 \delta/2}, \quad \frac{I_t}{I_0} = \frac{1}{1 + \left(\frac{2r}{1-r^2}\right)^2 \sin^2 \delta/2}$$

where the phase difference between adjacent rays is

$$\delta = \frac{4\pi}{\lambda} nd \cos \theta$$

Airy Function and Finesse

A plot of the relative transmitted intensity (**Airy function**) illustrates fringes of the order $m = \delta/2\pi$.



Analogous to conservation of energy, the reflected intensity distribution is the complement of the transmitted intensity, with dark fringes in place of bright fringes.

The **finesse** of the fringes can be defined as the ratio of fringe separation to the fringe full-width at half-max (FWHM):

$$finesse = \mathfrak{F} = \frac{fringe\ separation}{FWHM} \approx \frac{\pi}{2}\sqrt{F}$$

Where the approximation is valid for small δ, and F is defined as the **coefficient of finesse**,

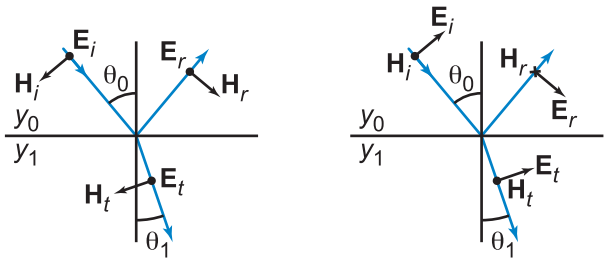
$$F = \left(\frac{2r}{1-r^2}\right)^2$$

The **Airy formulae** can then be conveniently expressed in terms of the coefficient of finesse:

$$\frac{I_r}{I_0} = \frac{F\sin^2\delta/2}{1+F\sin^2\delta/2}\ , \quad \frac{I_t}{I_0} = \frac{1}{1+F\sin^2\delta/2}$$

Fresnel Equations

The **amplitude coefficients of transmission and reflection** for s and p polarization (TE and TM) are given by the **Fresnel equations** and may be derived with the condition that the total electric and magnetic fields parallel to the boundary are continuous across the boundary.



TE polarization
s polarization

$$t_s = \frac{2y_0 \cos\theta_0}{y_0 \cos\theta_0 + y_1 \cos\theta_1}$$

$$r_s = \frac{y_0 \cos\theta_0 - y_1 \cos\theta_1}{y_0 \cos\theta_0 + y_1 \cos\theta_1}$$

TM polarization
p polarization

$$t_p = \frac{2y_0 \cos\theta_0}{y_0 \cos\theta_1 + y_1 \cos\theta_0}$$

$$r_p = \frac{y_0 \cos\theta_1 - y_1 \cos\theta_0}{y_0 \cos\theta_1 + y_1 \cos\theta_0}$$

In nonmagnetic media, the optical admittances y_0 and y_1 can be replaced by the complex refractive indices N_0 and N_1, since μ_0 cancels from the ratios:

$$t_s = \frac{2N_0 \cos\theta_0}{N_0 \cos\theta_0 + N_1 \cos\theta_1} \qquad t_p = \frac{2N_0 \cos\theta_0}{N_0 \cos\theta_1 + N_1 \cos\theta_0}$$

$$r_s = \frac{N_0 \cos\theta_0 - N_1 \cos\theta_1}{N_0 \cos\theta_0 + N_1 \cos\theta_1} \qquad r_p = \frac{N_0 \cos\theta_1 - N_1 \cos\theta_0}{N_0 \cos\theta_1 + N_1 \cos\theta_0}$$

In nonabsorbing and nonmagnetic media, the complex refractive index can be replaced by the real index. Snell's law then supplies the form originally derived by Fresnel:

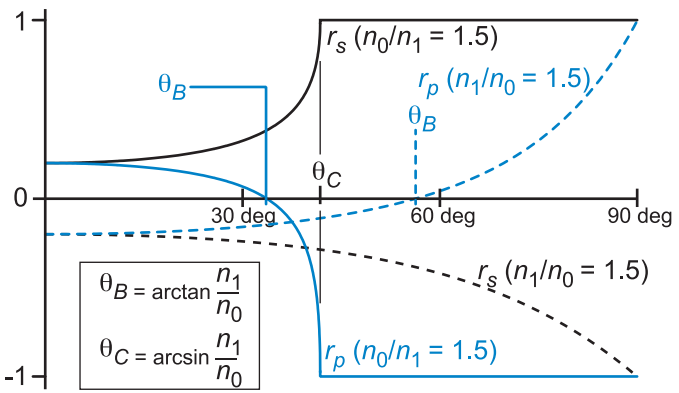
$$t_s = \frac{2\cos\theta_0 \sin\theta_1}{\sin(\theta_0 + \theta_1)} \qquad t_p = \frac{2\cos\theta_0 \sin\theta_1}{\sin(\theta_0 + \theta_1)\sin(\theta_0 - \theta_1)}$$

$$r_s = -\frac{\sin(\theta_0 - \theta_1)}{\sin(\theta_0 + \theta_1)} \qquad r_p = -\frac{\tan(\theta_0 - \theta_1)}{\tan(\theta_0 + \theta_1)}$$

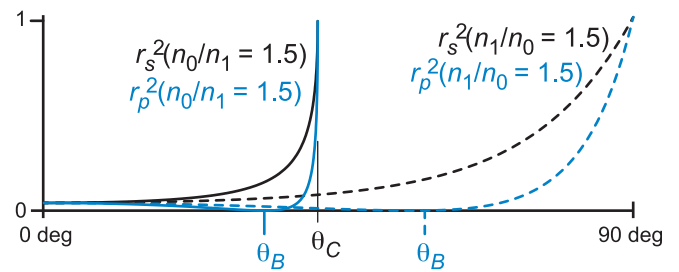
Coefficient of Reflection

In this field guide, the coordinate system is defined so that $r_s = r_p$ at normal incidence. However, many authors choose coordinate systems such that $r_s = -r_p$ at normal incidence.

Typical amplitude **coefficients of reflection** in dielectric media are shown in the following diagram:

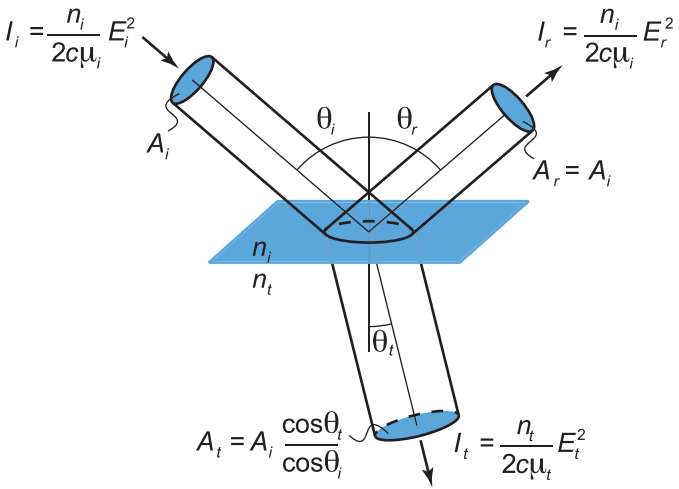


The coefficient of reflection for p polarization goes to zero and changes phase by π at the **Brewster angle** θ_B. This fact is often used in **Brewster windows**, which allow very high transmission for TM or p polarization.



Reflectance r^2 becomes maximum at the **critical angle** θ_C, where both polarizations experience **total internal reflection** (TIR).

Reflectance and Transmittance



Refraction causes a beam with a perpendicular area A to be transmitted as a beam that has been foreshortened by the ratio of cosines as shown above.

The **reflectance** and **transmittance** are defined, respectively, as the ratio of reflected and transmitted flux relative to the incident flux. In the case of reflectance R, the area of the beam is unchanged, so the cosines and refractive indices cancel but not so for transmittance T:

$$R = \frac{I_r A_r}{I_i A_i} = \frac{E_r^2}{E_i^2} = |r|^2$$

$$T = \frac{I_t A_t}{I_i A_i} = \frac{\frac{n_t}{2c\mu_t} E_t^2}{\frac{n_i}{2c\mu_i} E_i^2} \frac{\cos\theta_t}{\cos\theta_i} = \frac{n_t/\mu_t}{n_i/\mu_i} \frac{\cos\theta_t}{\cos\theta_i} |t|^2$$

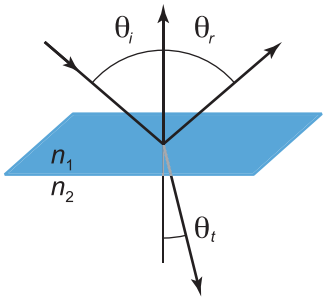
where r and t are the polarization-dependent coefficients of reflection and transmission.

In nonmagnetic media, $\mu_i = \mu_t = \mu_0$, resulting in

$$R = |r|^2 \quad T = \frac{n_t}{n_i} |t|^2$$

Laws of Refraction and Reflection

Snell's law is a consequence of Maxwell's equations and represents the relationship between the refractive indices across a boundary and the angles of incidence and refraction.



$$n_1 \sin\theta_i = n_2 \sin\theta_t$$

When the first medium has a lower refractive index than the second ($n_1 > n_2$), it is possible for the angle of refraction to reach 90 deg. The angle at which this occurs is known as the **critical angle** θ_C. As may be verified with the **Fresnel equations**, angles beyond the critical angle experience **total internal reflection**.

$$\theta_C = \sin^{-1}\left(\frac{n_2}{n_1}\right)$$

In vector form, the **law of refraction** can take these forms:

$$n_1(\hat{\mathbf{k}}_i \times \hat{\mathbf{n}}) = n_2(\hat{\mathbf{k}}_t \times \hat{\mathbf{n}})$$

$$n_2\hat{\mathbf{k}}_t = n_1\hat{\mathbf{k}}_i + \hat{\mathbf{n}}(n_2\cos\theta_t - n_1\cos\theta_i)$$

The **law of reflection** can be stated as $\theta_r = \theta_i$, as a special case of the law of refraction where $n_2 = -n_1$, or as the simplified and convenient vector forms:

$$\hat{\mathbf{k}}_r = \hat{\mathbf{k}}_i - 2(\hat{\mathbf{k}}_i \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$$

$$\hat{\mathbf{k}}_r \cdot \hat{\mathbf{n}} = -\hat{\mathbf{k}}_i \cdot \hat{\mathbf{n}}$$

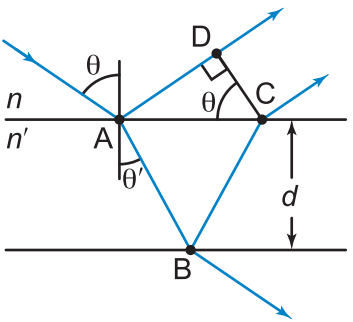
Phase Difference between Parallel Reflections

The phase difference δ between two partial reflections made by **parallel boundaries**, illustrated in the figure as ray segments ABC and AD, is determined by the distance between the boundaries d, the refractive index n', and the ray angle within the medium θ'.

$$\delta = \frac{2\pi}{\lambda}\Delta\tau = \frac{2\pi}{\lambda} n\,\Delta t$$

$$= \frac{2\pi}{\lambda}[n'(\overline{AB} + \overline{BC}) - n\,\overline{AD}]$$

$$\delta = \frac{4\pi}{\lambda} n' d\cos\theta'$$



Note that many authors define δ as half this quantity, which corresponds to the differential phase acquired in traversing the space between the boundaries.

Provided that the coherence length of the source is larger than δ, the reflected beams will interfere, and the condition for constructive interference between two such reflected plane waves is

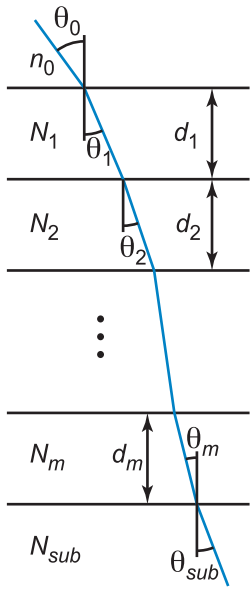
$$2\pi\, m = \delta + \phi = \frac{4\pi}{\lambda} n' d\cos\theta' + \phi$$

$$m\lambda = 2n' d\cos\theta' + \frac{\phi}{2\pi}\lambda$$

Where there is a difference in phase due to the optical path, there is generally also a phase difference due to reflection, denoted here by ϕ.

Characteristic Matrix of Thin Films

To simplify expressions, the description of **thin films** often involves the **tilted optical admittance** η, which is the ratio of the magnetic and electric field strengths parallel to the planar boundary and depends on the **complex refractive index** $N_j = n_j + ik_j$, the admittance of free space y_{fs}, and the angle of propagation within the layer θ_j.



The tilted optical admittance is different for TE and TM polarization.

$$TE \quad \rightarrow \quad \eta_j = y_{fs} N_j \cos\theta_j$$
$$TM \quad \rightarrow \quad \eta_j = y_{fs} N_j / \cos\theta_j$$

To avoid inadvertantly producing a layer with gain, the product $N_j \cos\theta_j$ should be in the same quadrant as N_j, which, in this text, is the first quadrant.

The behavior of the j^{th} layer can then be described in terms of its **characteristic matrix** $\mathbf{M}_j$ and the phase $\delta/2$ acquired by a ray traversing a layer:

$$M_j = \begin{pmatrix} \cos\delta_j/2 & \dfrac{-i\sin\delta_j/2}{\eta_j} \\ -i\eta_j \sin\delta_j/2 & \cos\delta_j/2 \end{pmatrix}, \quad \delta = \frac{4\pi N_j}{\lambda} d_j \cos\theta_j$$

The characteristic matrix of the full stack of thin-film layers is

$$\mathbf{M} = \prod_{j=1}^{m} \mathbf{M}_j = \mathbf{M}_1 \mathbf{M}_2 \cdots \mathbf{M}_m$$

There is a separate characteristic matrix for TE and TM polarization found by inserting the appropriate tilted optical admittance.

Reflectance and Transmittance of Thin Films

The behavior of a thin-film stack can be treated as a single surface with an equivalent optical admittance $\eta_{eq} = C/B$, which depends on the characteristic matrix and the tilted optical admittance of the substrate η_{sub}.

$$\begin{pmatrix} B \\ C \end{pmatrix} = \left(\prod_{j=1}^{m} \mathbf{M}_j\right)\begin{pmatrix} 1 \\ \eta_{sub} \end{pmatrix} = \begin{pmatrix} m_{11} + m_{12}\eta_{sub} \\ m_{21} + m_{22}\eta_{sub} \end{pmatrix} \rightarrow \eta_{eq} = \frac{C}{B}$$

The generally complex **amplitude coefficients of reflection and transmission** are, respectively,

$$\rho = \frac{\textit{reflected amplitude}}{\textit{incident amplitude}} = \frac{\eta_0 - \eta_{eq}}{\eta_0 + \eta_{eq}}$$
$$\tau = \frac{\textit{transmitted amplitude}}{\textit{incident amplitude}} = \frac{2\eta_0}{\eta_0 + \eta_{eq}}$$

The amplitude coefficients of reflection and transmission depend on the polarization, and so TE or TM coefficients are computed using the associated tilted admittance.

The **reflectance and transmittance** (also polarization dependent) are

$$R = \rho\rho^* = \left(\frac{\eta_0 - \eta_{eq}}{\eta_0 + \eta_{eq}}\right)\left(\frac{\eta_0 - \eta_{eq}}{\eta_0 + \eta_{eq}}\right)^*$$

$$T = \frac{\eta_{sub}\cos\theta_{sub}}{\eta_0\cos\theta_0}\tau\tau^* = \frac{\eta_{sub}\cos\theta_{sub}}{\eta_0\cos\theta_0}\frac{4\mathrm{Re}(\eta_0)\mathrm{Re}(\eta_{eq})}{(\eta_0 + \eta_{sub})(\eta_0 + \eta_{sub})^*}$$

These expressions assume that the incident medium is nonabsorbing, or that the refractive index is real, but these results are valid even if the other layers and substrate are absorbing.

Superposed Plane Waves

An expression of the complex monochromatic plane wave is written in terms of its wavevector $\mathbf{k}$, frequency ω, and vector amplitude $\mathbf{A}$:

$$\mathbf{U}(\mathbf{r},t) = \mathbf{A}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$$

The physical field $\mathbf{E}$ is the real part of $\mathbf{U}$.

$$\mathbf{E}(\mathbf{r},t) = \mathrm{Re}\{\mathbf{U}(\mathbf{r},t)\} = \frac{1}{2}\mathbf{U}(\mathbf{r},t) + \frac{1}{2}\mathbf{U}^*(\mathbf{r},t)$$

By the **principle of linear superposition**, N **superposed plane waves** add in amplitude in a linear medium.

$$\mathbf{U}(\mathbf{r},t) = \sum_{j=1}^{N}\mathbf{U}_j(\mathbf{r},t) = \sum_{j=1}^{N}\mathbf{A}_j e^{i(\mathbf{k}_j\cdot\mathbf{r}-\omega_j t)}$$

is generally a complex vector so that it includes the orientation of the optical field, polarization ellipticity, and the phase at $\mathbf{r} = 0$, $t = 0$. Furthermore, it may be written as the product of the complex unit vector $\hat{\mathbf{a}}$ and the scalar amplitude A:

$$\mathbf{A} = A\hat{\mathbf{a}} = A(a_x\hat{\mathbf{x}} + a_y\hat{\mathbf{y}} + a_z\hat{\mathbf{z}}), \quad \sqrt{|a_x|^2 + |a_y|^2 + |a_z|^2} = 1$$

Total irradiance is the time average of the square of the electric field. In a nonmagnetic, nonabsorbing medium:

$$I(\mathbf{r}) = \frac{1}{T}\int_{-T/2}^{T/2}\frac{cn\varepsilon_0}{2}\left|E(\mathbf{r},t)\right|^2 dt = \frac{cn\varepsilon_0}{2}\left\langle\left|E(\mathbf{r},t)\right|^2\right\rangle_T$$

$$= \frac{cn\varepsilon_0}{2}\left\langle E_x^2 + E_y^2 + E_z^2\right\rangle_T = I_x(\mathbf{r}) + I_y(\mathbf{r}) + I_z(\mathbf{r})$$

The quantities I_x, I_y, and I_z are independent contributions and may be computed separately. The factor of $cn\varepsilon_{0/2}$ is admittance, converting squared field to power per unit area.

Interference of Two Plane Waves (Different Frequency)

Consider one component contribution to the total irradiance for a sum of two **plane waves**. Using the simplifying notation $A_j\hat{\mathbf{a}}_{jx} = A_{jx}e^{i\phi_{jx}}\hat{\mathbf{x}}$ and:

$$A_{\Sigma x} = A_{1x} + A_{2x} \quad \mathbf{k}_\Sigma = \mathbf{k}_1 + \mathbf{k}_2 \quad \omega_\Sigma = \omega_1 + \omega_2 \quad \phi_{\Sigma x} = \phi_{1x} + \phi_{2x}$$
$$A_{\Delta x} = A_{1x} - A_{2x} \quad \mathbf{k}_\Delta = \mathbf{k}_1 - \mathbf{k}_2 \quad \omega_\Delta = \omega_1 - \omega_2 \quad \phi_{\Delta x} = \phi_{1x} - \phi_{2x}$$

The x-component contribution to the total field becomes

$$\begin{aligned} E_x &= \mathrm{Re}\left\{A_{1x}e^{i(\mathbf{k}_1\cdot\mathbf{r}-\omega_1 t+\phi_1)} + A_{2x}e^{i(\mathbf{k}_2\cdot\mathbf{r}-\omega_2 t+\phi_2)}\right\} \\ &= A_{\Sigma x}\cos\left(\frac{\mathbf{k}_\Sigma\cdot\mathbf{r}-\omega_\Sigma t+\phi_\Sigma}{2}\right)\cdot\cos\left(\frac{\mathbf{k}_\Delta\cdot\mathbf{r}-\omega_\Delta t+\phi_\Delta}{2}\right) \\ &\quad - A_{\Delta x}\sin\left(\frac{\mathbf{k}_\Sigma\cdot\mathbf{r}-\omega_\Sigma t+\phi_\Sigma}{2}\right)\cdot\sin\left(\frac{\mathbf{k}_\Delta\cdot\mathbf{r}-\omega_\Delta t+\phi_\Delta}{2}\right) \end{aligned}$$

The resulting optical field is made up of a sum of amplitude components and a difference of amplitude components that are $\pi/2$ out of phase. Each component includes a high-frequency factor (10^{14} to 10^{15} Hz—usually undetectable) and a low-frequency factor.

The x-component contribution to the irradiance is proportional to the time average (indicated by angle brackets) of the square of the x component of the electric field. So for integration time, $\Delta t \gg 1/\omega_\Sigma$, the x-component contribution to the total irradiance in a nonmagnetic, nonabsorbing medium becomes

$$\begin{aligned} I_x &= \frac{cn\varepsilon_0}{2}\left\langle E_x^2\right\rangle = \frac{cn\varepsilon_0}{2}\left\langle A_{\Sigma x}^2\cos^2\tfrac{\mathbf{k}_\Delta\cdot\mathbf{r}-\omega_\Delta t+\phi_\Delta}{2} + A_{\Delta x}^2\sin^2\tfrac{\mathbf{k}_\Delta\cdot\mathbf{r}-\omega_\Delta t+\phi_\Delta}{2}\right\rangle_{\Delta t} \\ &= \frac{cn\varepsilon_0}{4}\left\{A_{\Sigma x}^2 + A_{\Delta x}^2 + \left(A_{\Sigma x}^2 - A_{\Delta x}^2\right)\mathrm{sinc}\tfrac{\omega_\Delta\Delta t}{2}\cos\left(\mathbf{k}_\Delta\cdot\mathbf{r}-\omega_\Delta t+\phi_\Delta\right)\right\} \\ &= \frac{cn\varepsilon_0}{2}\left\{A_{1x}^2 + A_{2x}^2 + 2\sqrt{A_{1x}^2A_{2x}^2}\,\mathrm{sinc}\tfrac{\omega_\Delta\Delta t}{2}\cos\left(\mathbf{k}_\Delta\cdot\mathbf{r}-\omega_\Delta t+\phi_\Delta\right)\right\} \\ &= I_{1x} + I_{2x} + 2\sqrt{I_{1x}I_{2x}}\,\mathrm{sinc}\tfrac{\omega_\Delta\Delta t}{2}\cos\left(\mathbf{k}_\Delta\cdot\mathbf{r}-\omega_\Delta t+\phi_\Delta\right) \end{aligned}$$

The same applies to the y- and z-component contributions.

Interference of Two Plane Waves (Same Frequency)

The superposition of two plane waves having the same polarization and direction of propagation is a highly instructive example of **interference** but also forms a model of several important interferometric systems. The field from two superposed **plane waves** of the same frequency but different amplitudes and phase, traveling along the z axis, can be expressed in complex notation as

$$\begin{aligned} U &= |A_1|e^{i(kz-\omega t+\phi_1)} + |A_2|e^{i(kz-\omega t+\phi_2)} \\ &= \left(|A_1|e^{i\phi_2} + |A_2|e^{i\phi_2}\right)e^{i(kz-\omega t)} \end{aligned}$$

The resulting irradiance is independent of position and time but varies with the cosine of the phase difference.

$$\begin{aligned} I &= \frac{cn\varepsilon_0}{2}U(z,t)U^*(z,t) \\ &= \frac{cn\varepsilon_0}{2}\left[|A_1|^2 + |A_2|^2 + 2|A_1||A_2|\cos(\phi_1-\phi_2)\right] \\ &= I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\phi_\Delta \end{aligned}$$

When the phase difference is in the first or fourth quadrant (between $-\pi/2$ and $\pi/2$) the irradiance is greater than the sum of the irradiances produced by each wave separately and is known as **constructive interference**. At integer multiples of 2π, the constructive interference is maximum.

When the phase difference is in the second or third quadrants, the irradiance is less than the sum of the irradiances produced by each wave separately and is known as **destructive interference**. At odd integer multiples of π, the destructive interference produces a minimum:

$\phi_\Delta = m\pi$, $m \in$ Even $\rightarrow$ Maximum constructive interference

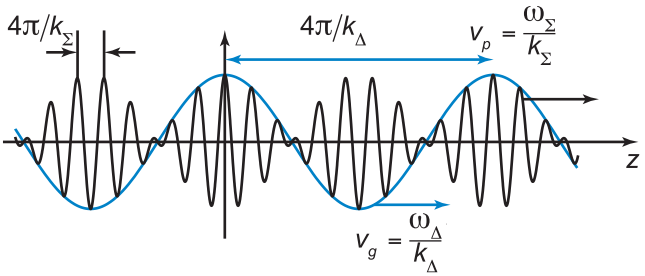
$\phi_\Delta = m\pi$, $m \in$ Odd $\rightarrow$ Maximum destructive interference

Phase Velocity and Group Velocity

When two interfering waves have different frequencies, the electric field appears as the product of two harmonic functions of time and position.

$$\begin{aligned} U &= Ae^{i(k_1 z - \omega_1 t + \phi_1)} + Ae^{i(k_2 z - \omega_2 t + \phi_2)} \\ &= 2A\cos\left(\frac{k_\Delta z - \omega_\Delta t + \phi_\Delta}{2}\right) e^{i(k_\Sigma z - \omega_\Sigma t + \phi_\Sigma)/2} \end{aligned}$$

The field behaves like a single-frequency plane wave traveling at the speed v_p, known as the **phase velocity**, and a lower-frequency modulating wave form, traveling at the speed v_g, known as the **group velocity**.



Phase velocity is the speed of the constant phase surfaces. Group velocity is the speed of the envelope and, in a mildly dispersive medium, the speed with which energy is transported. In other cases, this quantity can exceed the vacuum speed of light or even go in reverse directions.

In the limit of small ω_Δ, the finite difference can be replaced by a derivative:

$$v_g = \frac{d\omega}{dk} = v_p + k\frac{dv_p}{dk} = v_p\lambda\frac{dv_p}{d\lambda}$$

The **group index of refraction** n_g can then be defined as

$$n_g = \frac{c}{v_g} = c\frac{dk}{d\omega} = n + w\frac{dn}{d\omega} = n - \lambda\frac{dn}{d\lambda}$$

Interference of Two Plane Waves (3D)

The field and irradiance in nonabsorbing, nonmagentic media from two **superposed plane waves** with the same frequency but different wavevectors and amplitudes is

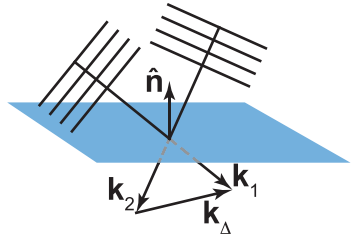
$$\mathbf{U} = \mathbf{A}_1 e^{i(\mathbf{k}_1\cdot\mathbf{r}-\omega t+\phi_1)} + \mathbf{A}_2 e^{i(\mathbf{k}_2\cdot\mathbf{r}-\omega t+\phi_2)}$$

$$I = \frac{cn\varepsilon_0}{2}\mathbf{U}\cdot\mathbf{U}^* = \frac{cn\varepsilon_0}{2}\left[A_1^2 + A_2^2 + 2\mathbf{A}_1\cdot\mathbf{A}_2\cos\left(\mathbf{k}_\Delta\cdot\mathbf{r}+\phi_\Delta\right)\right]$$

$$= I_1 + I_2 + 2\sqrt{I_1 I_2}\,p_{12}\cos\left(\mathbf{k}_\Delta\cdot\mathbf{r}+\phi_\Delta\right)$$

The factor of $p_{12} = \mathbf{A}_1\cdot\mathbf{A}_2/\sqrt{A_1^2A_2^2}$ depends on the relative orientations of the respective optical fields (their polarizations).

The **interference** fringes are arranged in parallel planes that are all perpendicular to the difference in wavevectors $\mathbf{k}_\Delta$ and separated by the distance $2\pi/|\mathbf{k}_\Delta|$.



In the case of an arbitrarily oriented observation plane, the fringes can be described in terms of a **fringe vector** $\mathbf{k}_f$ that is simply a projection of $\mathbf{k}_\Delta$ into the observation plane whose normal is $\hat{\mathbf{n}}$

$$\mathbf{k}_f = \mathbf{k}_\Delta - (\mathbf{k}_\Delta\cdot\hat{\mathbf{n}})\hat{\mathbf{n}} = \frac{2\pi}{\Lambda}\hat{\mathbf{v}},\quad \Lambda = \frac{2\pi}{|\mathbf{k}_f|}$$

$$I_{\hat{n}} = I_1 + I_2 + 2\sqrt{I_1 I_2}\,p_{12}\cos(\mathbf{k}_f\cdot\mathbf{r}+\phi_\Delta)$$

The **fringe vector** describes both the direction of the fringes $\hat{\mathbf{v}}$ and the fringe period Λ. The fringe vector and the related concept of the **grating vector** are useful when considering diffraction from 1D gratings.

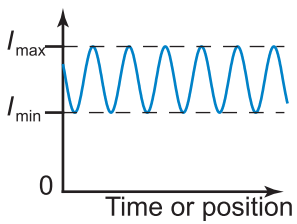
Fringe Visibility / Modulation / Contrast

Interference of two or more wave fields results in fringes that have maximum and minimum irradiance either in space or time as a function of a variable phase difference. **Complete destructive interference** refers to the case in which the resulting irradiance is zero. **Constructive interference** refers to cases where the resulting irradiance is greater than the sum of the individual irradiances.

The level of interference is characterized by **fringe contrast** V, which is also known as **fringe visibility** or **fringe modulation**, and is the ratio of the difference over the sum of maximum and minimum irradiance:

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2} p_{12}$$

Loss of fringe visibility can result from a number of factors. In the case of just two plane waves, nonparallel polarization expressed by the factor p_{12} will cause a loss of visibility that goes with the cosine of the angle between polarizations.



In addition to a loss of visibility in the irradiance itself, there can be a loss of contrast in the detected signal due to the size of the detector. When a is the width of a rectangular detector, and Λ is the fringe pitch, the contrast of a 1D-fringe pattern that would otherwise be 100% becomes

$$V = \frac{\sin(\pi a/\Lambda)}{\pi a/\Lambda}$$

This demonstrates that the visibility is maximum for $a = 0$ and goes through a **contrast reversal** when the detector size a drops below the fringe pitch Λ.

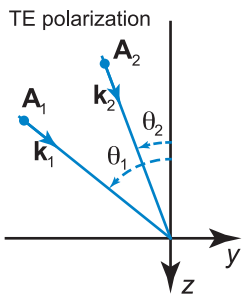
Interference of Two Polarized Plane Waves

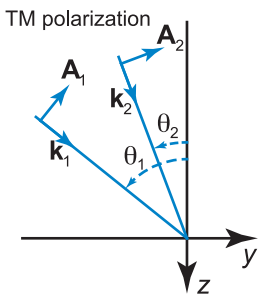
It is useful to consider TE and TM polarization separately when two wavevectors are in the same plane of incidence.

$$\mathbf{k}_1 = \frac{2\pi}{\lambda}\,(\hat{\mathbf{y}}\sin\theta_1 + \hat{\mathbf{z}}\cos\theta_1),\ \mathbf{k}_2 = \frac{2\pi}{\lambda}(\hat{\mathbf{y}}\sin\theta_2 + \hat{\mathbf{z}}\cos\theta_2)$$

$$\mathbf{k}_f = \frac{2\pi}{\lambda}\,(\sin\theta_1 - \sin\theta_2)\hat{\mathbf{y}}$$

Assuming that the frequencies, amplitudes, and phases of the two **plane waves** are the same:

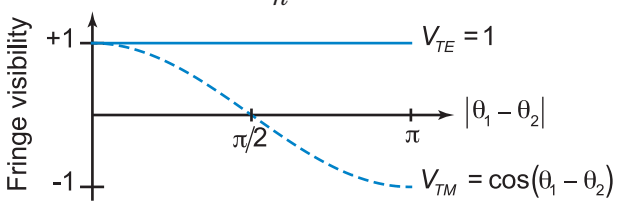




$\mathbf{A}_1 = A\hat{\mathbf{x}}$ $\quad$ $\mathbf{A}_1 = A(\hat{\mathbf{y}}\cos\theta_1 + \hat{\mathbf{z}}\sin\theta_1)$

$\mathbf{A}_2 = A\hat{\mathbf{x}}$ $\quad$ $\mathbf{A}_2 = A(\hat{\mathbf{y}}\cos\theta_2 + \hat{\mathbf{z}}\sin\theta_2)$

$$I_{TE} = \frac{c\varepsilon}{n}A^2(1+\cos k_f y) \quad I_{TM} = \frac{c\varepsilon}{n}A^2[1+\cos(\theta_1-\theta_2)\cos k_f y]$$



TE fringe visibility is independent of the difference in angles $|\theta_1 - \theta_2|$, but TM fringe visibility varies with the relative angle and can become negative for differences larger than $\pi/2$ resulting in **contrast reversal**.

This interesting result has serious implications for imaging systems with $NA > 1/\sqrt{2}$, where TE polarization can produce contrast-reversed images for some spatial frequencies.

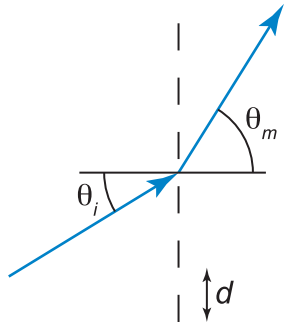
Grating Equation

In a planar 1D grating of period Λ, constructive interference of scattered plane waves occurs at angles θ_m, where m is known as the order of diffraction. The diffracted angles obey the **grating equation**, and at normal incidence

$$\Lambda \sin\theta_m = m\lambda$$

When the direction of the beam is perpendicular to the grating lines but at an angle θ_i to the plane of the grating, this equation applies:

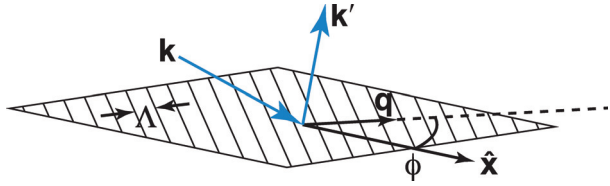
$$\sin\theta_m - \sin\theta_i = m\lambda/\Lambda$$



When the direction of the incident beam is oblique to both the grating lines and the plane of the grating, then it is convenient to express the diffracted wavevector $\mathbf{k}'$ in terms of the incident wavevector $\mathbf{k}$ and a grating vector $\mathbf{q}$. The grating vector magnitude is $\frac{2\pi}{\Lambda}$ with a direction assumed to be in the x-y plane so that ϕ is the angle that the grating vector makes with the x axis.

$$k'_x = k_x + m\frac{2\pi}{\Lambda}\cos\phi, \quad k'_y = k_y + m\frac{2\pi}{\Lambda}\sin\phi$$

$$k'_z = \sqrt{|k|^2 - k'^2_x - k'^2_y}$$

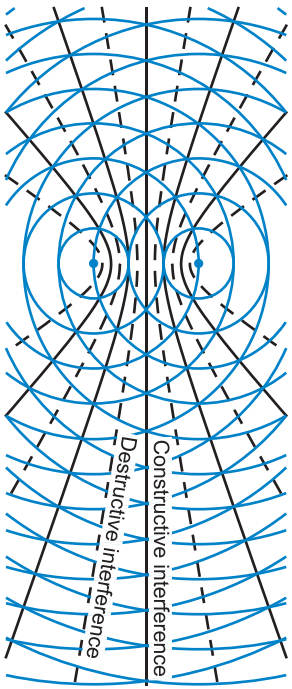


The choice of the sign in the last equation depends on whether $\mathbf{k}'$ refers to a transmitted or reflected order.

Interference of Two Spherical Waves

Many systems can be modeled based on the idealized scalar spherical optical wave, even though such waves do not occur in nature.

The crests of two **spherical waves** with initial phases ϕ_1 and ϕ_2 and wavelength λ is schematically shown at the right. **Constructive interference** occurs where crests meet crests. **Destructive interference** occurs where crests meet troughs.

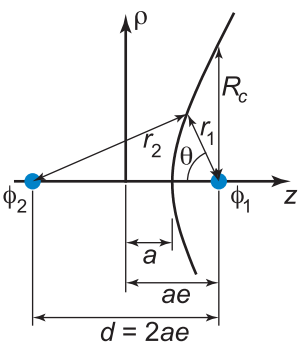


Surfaces of equal phase difference are hyperboloids revolved about the axis passing through the two point sources. Three equivalent expressions of these hyperbolae are

$$\frac{z^2}{a^2} - \frac{\rho^2}{b^2} = 1, \quad \rho = \frac{a(e^2-1)}{1 + e\cos\theta}$$

$$z = \frac{c\rho^2}{1\sqrt{1 - \frac{4\rho^2}{(m+\Delta m)^2\lambda^2 - d^2}}}$$

where Δm is the fractional fringe order, e is the **eccentricity**, R_c is the radius of curvature, and

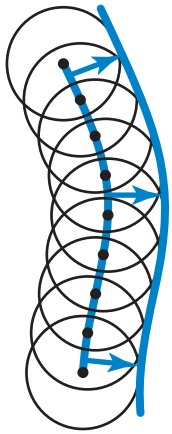


$$R_c = \frac{b^2}{a} = \frac{d^2 - (m+\Delta m)^2\lambda^2}{2(m+\Delta m)\lambda} = \frac{1}{c}$$

$$r_2 - r_1 = (m+\Delta m)\lambda = 2a$$

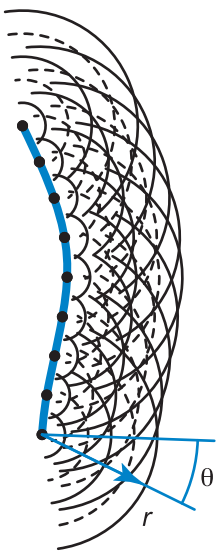
$$\Delta m = \frac{\phi_2 - \phi_1}{2\pi}, \quad e = \frac{d}{(m+\Delta m)\lambda}$$

Huygens' and Huygens–Fresnel Principles



The **Huygens' principle** states that each point on a wavefront is considered the source of a new spherical wave. The wavefront propagated downstream is then constructed from the envelope of the set of wavelets.

This is usually presented as the seed idea of a diffraction theory. The problem is that it does not actually account for diffraction because it does not include the concept of interference. Also, it implies that in addition to the forward-moving wave, there should be a backward-moving wave (which is not observed experimentally).



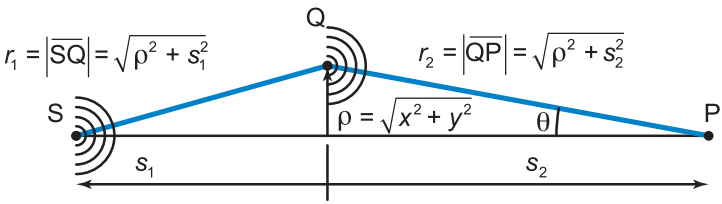
The **Huygens–Fresnel principle** is a modification of Huygens' principle where the back-propagating portion of each secondary wave is dropped and furthermore has some angular dependence known as an obliquity factor.

The **scalar theory** development shows that the so-called **Huygens' wavelet** takes the form of something close to a spherical wave.

$$h_z^H(r) = -\frac{e^{ikr}}{2\pi r}\left(ik - \frac{1}{r}\right)\cos\theta$$
$$= \frac{e^{ikr\,-\,\tan^{-1}kr}}{2\pi r}\sqrt{k^2 + 1/r^2}\,\cos\theta$$

Fresnel Diffraction

The Huygens–Fresnel principle is applicable when a spherical wave is emitted from point S and secondary waves are emitted from an intermediate plane at point Q to construct the field U at point P.



$$U = \frac{A}{i\lambda}\iint \frac{e^{ikr_1}}{r_1}\frac{e^{ikr_2}}{r_2}K(\theta)\rho\, d\phi\, d\rho = -iAe^{ik(s_2 - s_1)}\int \frac{e^{ik\tau}}{r_1 r_2}K(\theta)\rho d\rho$$

τ is the optical path difference from the on-axis path:

$$\tau = (r_1 + r_2) - (s_2 - s_1) \approx \frac{\rho^2}{2(s_2 - s_1)}$$

Fresnel zones are concentric annular regions where the outer edge of each is defined by an optical path difference of $\lambda/2$ from the inner edge. Using the above quadratic approximation in phase, and a linear approximation in the denominator ($r_1 r_2 \approx -s_1 s_2$), and assuming that the obliquity factor $K(\theta)$ is constant within a zone, the field due to the n^{th} zone of outer radius ρ_n is

$$U_n \approx -i\frac{Ae^{ik(s_2 - s_1)}}{s_2 - s_1}K_n \int_{(n-1)\lambda/2}^{n\lambda/2} e^{ik\tau} k d\tau = 2(-1)^{n+1} U\, K_n$$

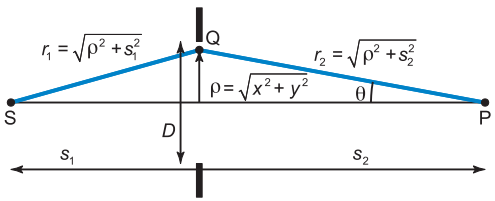
So the contributions of odd zones are 180 deg from the even zones.

In the approximation $K_n \approx 1$ (especially good on axis), the contribution of a single zone is twice the field produced by all of the zones together.

On-Axis Irradiance behind a Circular Aperture

The **Fresnel number** N_f is the number of Fresnel zones "seen" from the observation point P.

$$N_f = \frac{D^2}{4\lambda}\left(\frac{1}{s_2} - \frac{1}{s_1}\right)$$

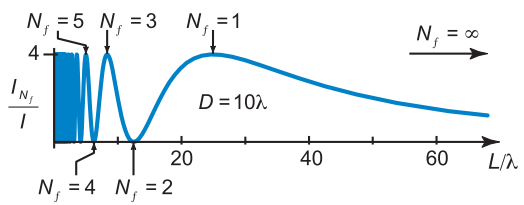


A point source at S behind a **circular aperture** will produce an on-axis field and irradiance at P that depends on the Fresnel number.

$$U_{N_f} = -iU \int_0^{N_f\lambda/2} e^{ik\tau} k \, d\tau = U(1 - e^{i\pi N_f})$$

$$I_{N_f} = \frac{c\varepsilon}{2n}|U_P|^2 = 4I\left[1 - \cos(\pi N_f)\right]$$

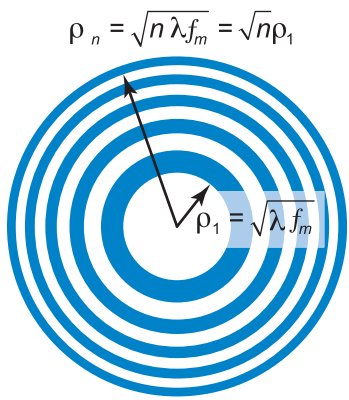
where U and I are the field and irradiance without an aperture, and τ is the optical path difference for point Q. The **on-axis irradiance** oscillates with distance from the aperture and with the aperture size.



In the above diagram, L is defined as

$$\frac{1}{L} = \frac{1}{s_2} - \frac{1}{s_1}$$

Fresnel Zone Plate



The **Fresnel zone plate**, or **zone plate**, is an aperture that blocks every other Fresnel zone, as shown at the right.

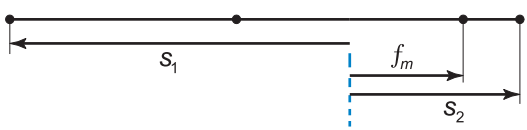
The result is that only odd (or even) zones contribute to the on-axis irradiance.

Under the parabolic approximation, each zone contributes twice the optical field that would exist in the absence of the device, so that the relative on-axis field is proportional to the number of open Fresnel zones (approximately half of the Fresnel number), and the irradiance is proportional to the square of the Fresnel number.

$$U_P = 2N_{OFZ}\ U \approx N_f U, \quad I_P = 4N_{OFZ}^2 I \approx N_f^2 I$$

At positions where an odd number of Fresnel zones is contained within each zone of the zone plate, another bright spot appears on-axis. Each of the axial bright spots is an image of the source, so zone plates behave like thin lenses with multiple focal lengths (of order m), both positive and negative. The thin-lens equation can be used to find the conjugate locations for each order.

$$-\frac{1}{s_1} + \frac{1}{s_2} = \frac{1}{f_m} \qquad f_m = \frac{\rho_1^2}{m\lambda},\ p \in \text{odd integers}$$



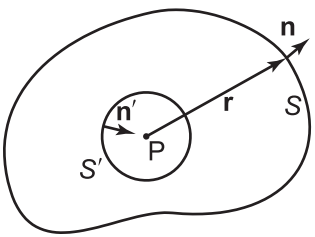
Each image order is dimmer than the first-order image, and the background irradiance tends to reduce the contrast of extended images.

Integral Theorem of Helmholtz and Kirchhoff

Green's theorem relates the scalar field U and a **Green's function** G in a volume V and on the enclosing surface S.

$$\iiint_V (U\nabla^2 G - G\nabla^2 U)dV = \iint_S \left(U\frac{\partial G}{\partial n} - G\frac{\partial U}{\partial n}\right)dS$$

The spherical surface S' is an additional portion of the surface of integration that separates P from the volume.



If U and G both satisfy the Helmholtz equation and have continuous first and second partial derivatives on the surface of integration, then the volume integral vanishes.

$$\iint_{S'} \left(G\frac{\partial U}{\partial n'} - U\frac{\partial G}{\partial n'}\right)dS = -\iint_S \left(G\frac{\partial U}{\partial n} - U\frac{\partial G}{\partial n}\right)dS$$

The Kirchhoff Green's function G_K has the form of a spherical wave centered on the observation point so that

$$G_K = \frac{e^{ikr}}{r} \rightarrow \frac{\partial G_K}{\partial n} = \left(ik - \frac{1}{r}\right)\frac{e^{ikr}}{r}\cos(\mathbf{n},\mathbf{r}) = \left(ik - \frac{1}{r}\right)\frac{e^{ikr}}{r}\frac{\mathbf{n}\cdot\mathbf{r}}{|\mathbf{n}\cdot\mathbf{r}|}$$

In the limit as S' collapses on P, the **integral theorem of Helmholtz and Kirchhoff** gives the field at P.

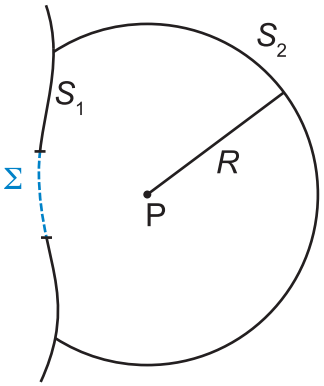
$$U_p = \frac{1}{4\pi}\iint_S \left(\frac{e^{ikr}}{r}\frac{\partial U_S(\mathbf{r})}{\partial n} - U_S(\mathbf{r})\frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r}\right)\right)dS$$

Given:

1. Scalar theory holds.
2. U satisfies the Helmholtz equation.
3. U has continuous first and second partial derivatives on and within S.

Sommerfeld Radiation and Kirchhoff Boundary Conditions

In the case of an observation point P and an aperture Σ within a generally curved surface, the contribution to the diffraction integral of the field U on surface S_2 (a portion of a sphere with radius R) can be neglected if



$$\lim_{R\to\infty} R\left(\frac{\partial U}{\partial n} - ikU\right) = 0$$

This condition is known as the **Sommerfeld radiation condition** and is satisfied when the scalar field U radiating away from the aperture vanishes at least as quickly as a spherical wave.

The diffraction problem is further simplified when the **Kirchhoff boundary conditions**—also known as the **Cauchy boundary conditions**—are applied.

1. Across the transparent portion of the screen Σ, both the field and its derivative are exactly the same as they would be in the absence of the screen.
2. In the surface region outside the screen S_1, both the field and its derivative are zero.

Strictly speaking, both conditions cannot be true for both the field and its derivative at the same time, but results of these assumptions are so useful and sufficiently accurate that we can generally accept the inconsistency and use them in a wide variety of problems.

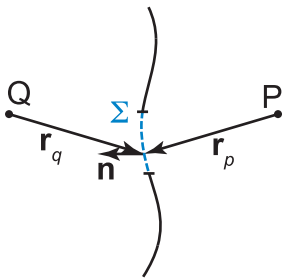
Fresnel–Kirchhoff Diffraction Integral

The diffraction integral can be evaluated as a sum of integrals over three portions of a closed surface that includes an aperture.

$$U_P = \frac{1}{4\pi}\left[\underbrace{\iint}_{\text{Inside Aperture}} + \overbrace{\iint}^{\approx 0}_{\text{Outside Aperture}} + \overbrace{\iint}^{\approx 0}_{\text{Spherical Cap}}\right]\left\{G\frac{\partial U_S}{\partial n} - U_S\frac{\partial G}{\partial n}\right\}dS$$

Kirchhoff Sommerfeld

The Kirchhoff boundary conditions and Sommerfeld radiation condition leave only the integral over the open aperture. If it is also assumed that the aperture is illuminated by a spherical wave of amplitude A centered on Q, then we arrive at the **Fresnel–Kirchhoff diffraction integral**:



$$U_p = \frac{A}{i2\lambda}\iint_{\Sigma}\frac{e^{ik(r_q + r_p)}}{r_q r_p}\left\{\frac{\mathbf{n}\cdot\mathbf{r}_p}{|\mathbf{n}\cdot\mathbf{r}_p|} - \frac{\mathbf{n}\cdot\mathbf{r}_q}{|\mathbf{n}\cdot\mathbf{r}_q|}\right\}dS$$

Given:

1. Scalar theory holds,
2. Sommerfeld radiation condition,
3. Kirchhoff boundary conditions,
4. Illumination by single point source,
5. P is several wavelengths from Σ.

The Fresnel–Kirchhoff diffraction integral is interesting in that it does not require a planar aperture, but its use of the Kirchhoff boundary conditions means that it requires an inconsistency in the field after the aperture. Nonetheless, this formula is generally considered very accurate for distances of several wavelengths or more from the aperture.

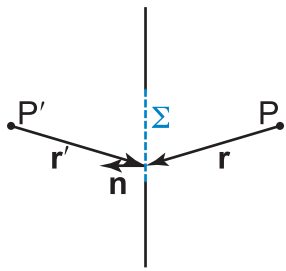
Rayleigh–Sommerfeld Diffraction Integral

With a planar boundary, a Green's function consisting of a linear combination of two Kirchhoff-type Green's functions can remove the inconsistency in the Kirchhoff boundary conditions. In this scheme, one is placed at the observation point, and the other is placed directly opposite the boundary, outside the volume of integration.

Two convenient cases are used: in the first case, the Green's function vanishes on the boundary, and in the second case, its derivative vanishes on the boundary.

$$G_{\mathrm{I}} = \frac{e^{ikr}}{r} - \frac{e^{ikr'}}{r'} \rightarrow G_{\mathrm{I}}|_{\Sigma} = 0$$

$$G_{\mathrm{II}} = \frac{e^{ikr}}{r} + \frac{e^{ikr'}}{r'} \rightarrow \frac{\partial G_{\mathrm{I}}}{\partial n}|_{\Sigma} = 0$$



These produce the first and second **Rayleigh–Sommerfeld diffraction integrals**, U_{I} and U_{II}, respectively.

$$U_{\mathrm{I}} = \frac{1}{i\lambda}\iint U\frac{e^{ikr}}{r}\cos(\mathbf{n},\mathbf{r})dS$$

$$U_{\mathrm{II}} = \frac{1}{2\pi}\iint \frac{\partial U}{\partial n}\frac{e^{ikr}}{r}\cos(\mathbf{n},\mathbf{r})dS$$

Given:

1. Scalar theory holds.
2. Sommerfeld radiation condition holds.
3. P is several wavelengths from Σ.
4. I. U is unchanged within the planar aperture but vanishes outside the aperture (also known as the **Dirichlet boundary conditions**).
   II. $\partial U/\partial n$ is unchanged within the planar aperture but vanishes outside the aperture (also known as the **Neumann boundary conditions**).

Boundary Conditions and Obliquity Factors

The Fresnel–Kirchhoff and Rayleigh–Sommerfeld diffraction integrals contain **obliquity factors** K that depend on the choice of both Green's function and boundary conditions. Take the case of a planar boundary illuminated by a single spherical wave originating at q. Then, all three diffraction integrals express the field at p with the form:

$$U_p = \frac{A}{i\lambda}\iint \frac{e^{ik(r_p + r_q)}}{r_p r_q} K \, dS$$

When the **Kirchhoff boundary conditions** are satisfied (outside of the aperture $U = \partial U/\partial n = 0$), then

$$G = G_K = \frac{e^{ikr_p}}{r_p} \rightarrow K = \frac{\cos(\mathbf{n},\mathbf{r}_p) - \cos(\mathbf{n},\mathbf{r}_q)}{2}$$

The first Rayleigh–Sommerfeld solution uses the **Dirichlet boundary conditions** (outside of the aperture $U = 0$), and

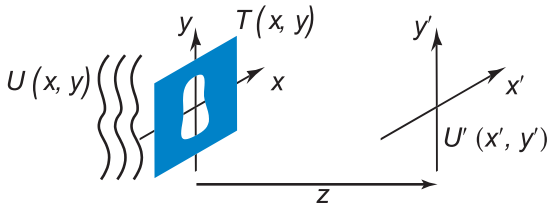
$$G = G_{\mathrm{I}} = \frac{e^{ikr_p}}{r_p} - \frac{e^{ikr_q}}{r_q} \rightarrow K = \cos(\mathbf{n},\mathbf{r}_p)$$

The first Rayleigh–Sommerfeld solution uses the **Neumann boundary conditions** (outside the aperture $\partial U/\partial n = 0$), and

$$G = G_{\mathrm{II}} = \frac{e^{ikr_p}}{r_p} + \frac{e^{ikr_q}}{r_q} \rightarrow K = -\cos(\mathbf{n},\mathbf{r}_q)$$

Interestingly, the Kirchhoff boundary conditions yield an obliquity factor that is the arithmetic mean of the obliquity factors obtained with the Dirichlet and Neumann boundary conditions.

Fresnel Diffraction Formula



The first Rayleigh–Sommerfeld diffraction integral depends on r and a cosine factor (obliquity factor). Expanding these into Taylor series, and truncating at the second order in phase and first order in amplitude, composes the **Fresnel approximation**.

$$\frac{\cos(\mathbf{n},\mathbf{r})}{r} \approx \frac{1}{z}, \quad r \approx z\left[1 + \frac{1}{2}\left(\frac{x'-x}{z}\right)^2 + \frac{1}{2}\left(\frac{y'-y}{z}\right)^2\right]$$

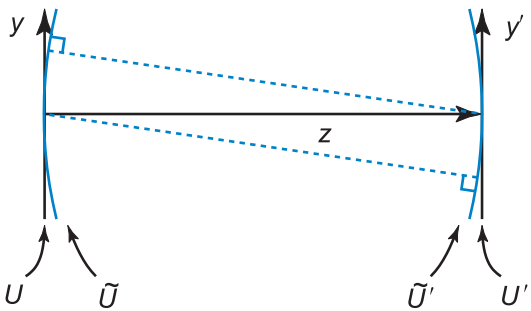
The validity of these approximations is said to depend on the relative size of the next term in the Taylor series. Inserting the Fresnel approximation into the Rayleigh–Sommerfield diffraction integral produces the **Fresnel diffraction integral**, which is valid within the so-called Fresnel region or near-field.

$$U'(x',y') = \frac{e^{ikz}}{i\lambda z} e^{\frac{ik}{2z}(x'^2+y'^2)} \iint_{\Sigma} U(x,y)T(x,y) e^{\frac{ik}{2z}(x^2+y^2)} e^{-\frac{i2\pi}{\lambda z}(xx'+yy')} dx\, dy$$

where $$z^3 \gg \tfrac{\pi}{4\lambda} \max\left[(x'-x)^2 + (y'-y)^2\right]^2$$

This condition on z^3 is sufficient because it allows us to neglect the next term in the Taylor series. However, in practice, the validity of the Fresnel approximation is somewhat more relaxed since the only requirement of the higher-order contributions is that they remain constant. This explains why the Fresnel diffraction integral is often useful at small distances from the aperture.

Fresnel Diffraction between Confocal Surfaces



Two surfaces are **confocal surfaces** if their curvatures are centered on each other, forming the rotationally symmetric system shown above. A nonrigorous but instructive way of relating the fields on confocal surfaces (indicated with a tilde), is to estimate the phase on the curved surfaces in terms of the tangent planes and substitute this into the **Fresnel diffraction formula**.

$$\tilde{U} = U e^{\frac{i\pi}{\lambda z}(x^2+y^2)} \qquad \tilde{U}' = U' e^{-\frac{i\pi}{\lambda z}(x'^2+y'^2)}$$

$$U' = \frac{e^{ikz}}{i\lambda z} e^{\frac{i\pi}{\lambda z}(x'^2+y'^2)} \iint U e^{\frac{i\pi}{\lambda z}(x'^2+y'^2)} e^{-i\frac{k}{z}(xx'+yy')} dxdy$$

$$\tilde{U}' = \frac{e^{ikz}}{i\lambda z} \iint \tilde{U} e^{-i\frac{k}{z}(xx'+yy')} dxdy$$

This interesting result shows that propagation between confocal surfaces causes both quadratic phases (inside and outside the integral) to vanish.

A more rigorous derivation, starting from the integral theorem of Helmholtz and Kirchhoff—which can be applied directly to the spherical surfaces or by using the principle of stationary phase—produces the same result.

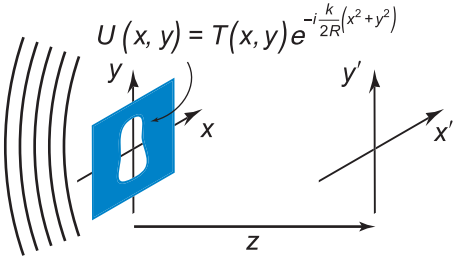
Fraunhofer Diffraction Formula

The **Fraunhofer approximation** assumes that the distance from the aperture is great enough that the quadratic phase in the Fresnel diffraction integral becomes negligible, resulting in the **Fraunhofer diffraction integral**:

$$U'(x',y') = \frac{e^{ikz}}{i\lambda z} e^{\frac{ik}{2z}(x'^2+y'^2)} \iint_{\Sigma} U(x,y)T(x,y)e^{\frac{i2\pi}{\lambda z}(xx'+yy')}dxdy$$

where $$z \gg \frac{\pi}{2\lambda}\max[x^2+y^2]$$

As illustrated, a transparency illuminated by a uniform spherical wave converges on the observation plane. The spherical wave is approximated in the Fresnel diffraction integral by a multiplicative quadratic phase with a radius of curvature R.

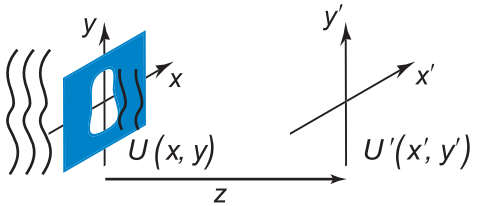


$$U'(x',y') = \frac{e^{ikz}}{i\lambda z} e^{\frac{i\pi}{\lambda z}(x'^2+y'^2)} \iint T\, e^{\frac{i\pi}{\lambda}\left(\frac{1}{z}-\frac{1}{R}\right)(x^2+y^2)} e^{-i\frac{2\pi}{\lambda z}(xx'+yy')}dx\,dy$$

When the curvature of the illuminating wavefront is centered on the observation plane $R = z$, the quadratic phase inside the integral vanishes, and the field becomes a scaled Fourier transform of the amplitude transmission—again arriving at the Fraunhofer result:

$$U'(x',y') = \frac{e^{ikz}}{i\lambda z} e^{\frac{i\pi}{\lambda z}(x'^2+y'^2)} \iint T(x,y)\, e^{-i\frac{2\pi}{\lambda z}(xx'+yy')}dxdy$$

Huygens' Wavelet



The Rayleigh–Sommerfeld diffraction integral assumes that the observation point is several wavelengths from the aperture ($z \gg \lambda$). Without this assumption, the field transmitted by the aperture U becomes U' at a distance z through this convolution:

$$U'(x',y') = -\frac{1}{2\pi}\iint_{\Sigma} U(x,y)\frac{e^{ikr}}{r}\left(ik - \frac{1}{r}\right)\cos\theta\, dS$$

where $r = \sqrt{(x'-x)^2 + (y'-y)^2 + z^2}$, and $\cos\theta = z/r$.

This represents a mathematical expression of **Huygens' principle**, which states that each point in a wavefront is a source of a spherical wave. However, this integral indicates that the convolution kernel h_z^H is slightly different from a simple spherical wave and is sometimes referred to as the **Huygens' wavelet**:

$$h_z^H(r) = -\frac{e^{ikr}}{2\pi r}\left(ik - \frac{1}{r}\right)\cos\theta = \frac{e^{ikr\,-\,\tan^{-1}kr}}{2\pi r}\sqrt{k^2 + 1/r^2}\,\cos\theta$$

Under the **Fresnel approximation**, Huygens' wavelet becomes the **Fresnel wavelet** (a parabolic wave) and provides an alternative way of deriving the Fresnel diffraction integral in terms of convolution:

$$h_z^F(r) \approx \frac{1}{i\lambda}e^{ikz}e^{\frac{ik}{2z}(x^2+y^2)}$$

Angular Spectrum of Plane Waves

The angular-spectrum approach to diffraction is the idea that a field can be decomposed into a set of plane waves, each of which are then easily propagated individually. Each plane wave in the decomposition has the form

$$Ae^{i(\mathbf{k}\cdot\mathbf{r}\,-\,\omega t)} = Ae^{i\frac{2\pi}{\lambda}(\alpha x+\beta y+\gamma z)}e^{-\omega t} = Ae^{i2\pi(\xi x\,+\,\eta y\,+\,\gamma z)}e^{i(k\gamma z\,-\,\omega t)}$$

$\xi = \alpha/\lambda$ and $\eta = \beta/\lambda$ are spatial frequencies in a Fourier transform, while $\{\alpha,\beta,\gamma\}$ are generally interpreted as direction cosines in the **angular spectrum of plane waves** $A(\xi,\eta;z)$. At $t = 0$ and $z = z_i$,

$$A(\xi,\eta;z_i) = \iint_{\infty} U(x,y;z_i)e^{i2\pi(\xi x+\eta y)}dxdy$$

$$U(x,y;z_i) = \iint_{\infty} A(\xi,\eta;z_i)e^{-i2\pi(\xi x+\eta y)}d\xi d\eta$$

This decomposition is permitted if the medium is linear and the field satisfies the **Dirichlet conditions** (not to be confused with the Dirichlet boundary conditions used in the first Rayleigh–Sommerfeld diffraction integral):

- Absolutely integrable;
- A finite number of maxima and minima;
- A finite number of finite discontinuities;
- No infinite discontinuities (bounded).

When the angular spectrum is propagated using the **transfer function of free space** $H_z(\xi,\eta;\Delta z)$,

$$A(\xi,\eta;z_f) = A(\xi,\eta;z_i)H(\xi,\eta;z_f - z_i)$$

An inverse Fourier transform recovers the field at $z = z_f$:

$$U(x,y;z_f) = \iint_{\infty} A(\xi,\eta;z_f)e^{-i2\pi(\xi x\,+\,\eta y)}d\xi d\eta$$

Transfer Function of Free Space

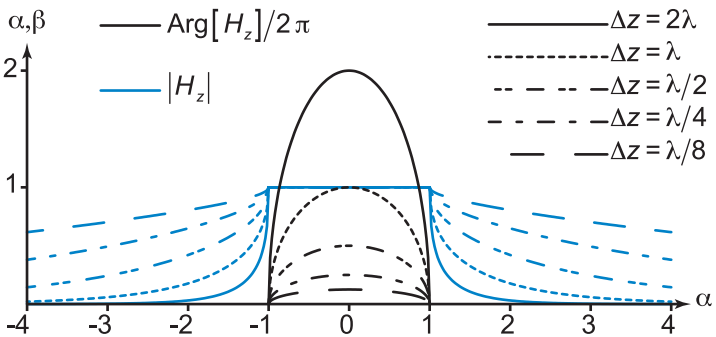
The exponential phase applied to the plane-wave spectrum upon propagation is known as the **transfer function of free space** $H_z(\xi,\eta;\ \Delta z)$ and is variously written as

$$H_z(\xi,\eta;\Delta z) = e^{i\frac{2\pi}{\lambda}\Delta z\sqrt{1-\lambda^2\xi^2-\lambda^2\eta^2}} = e^{i\frac{2\pi}{\lambda}\Delta z\sqrt{1-\alpha^2-\beta^2}} = e^{i\frac{2\pi}{\lambda}\gamma\,\Delta z}$$

where

$$\Delta z = z_f - z_i$$
$$\{\alpha,\beta\} = \{\lambda\xi,\lambda\eta\}$$
$$\gamma = \sqrt{1-\alpha^2-\beta^2}$$

The plot below illustrates the phase (black) and amplitude (blue) of the transfer function of free space versus α with $\beta = 0$ for several small propagation distances.



Within $\alpha^2+\beta^2 < 1$ (the spatial frequencies less than $1/\lambda$), H_z has a unit amplitude and a curved phase that increases in curvature with propagation distance.

In the range $\alpha^2+\beta^2 > 1$ (spatial frequencies greater than $1/\lambda$), the transfer function is purely real. These plane-wave components are **evanescent**: they form standing waves directed orthogonally to the z axis that decay exponentially with propagation distance.

Method of Stationary Phase

Integrals of the form below occur frequently in diffraction theory, and in many cases w is large enough to cause the phase to oscillate rapidly with changing $g(r)$:

$$f = \int u(\mathbf{r}) e^{iwg(r)} d\mathbf{r}$$

In those regions of rapid oscillation, the integrand tends to average to zero so that the overall integral is heavily weighted to regions where $g(r)$ is slowly varying around **stationary points** (x_0,y_0) where the first derivatives are zero so that a Taylor expansion becomes

$$g(x,y) \approx g(x_0,y_0) + (x-x_0)(y-y_0)g_{xy}(x_0,y_0) + \frac{1}{2}\left[(x-x_0)^2 g_{xx}(x_0,y_0) + (y-y_0)^2 g_{yy}(x_0,y_0)\right]$$

where $$g_{xx} = \frac{\partial^2 g}{\partial x^2},\ g_{xy} = \frac{\partial^2 g}{\partial x\,\partial y},\ g_{yy} = \frac{\partial^2 g}{\partial y^2}$$

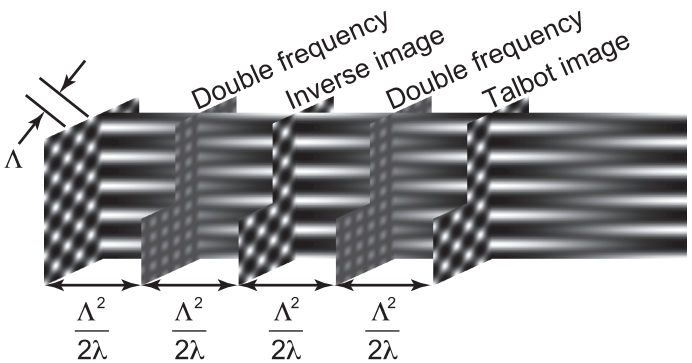
Under the **stationary phase approximation**, the Taylor series is substituted back into the integral, and the amplitude is evaluated at the stationary point outside the integral. When there is more than one stationary point, the integral becomes a summation over the stationary points:

$$f = \frac{i2\pi}{w} \sum_{x_0,y_0} u(x_0,y_0) \frac{a e^{iwg(x_0,y_0)}}{\sqrt{g_{xx}(x_0,y_0) g_{yy}(x_0,y_0) - g_{xy}^2(x_0,y_0)}}$$

The stationary phase approximation can be used to derive the Fresnel and Fraunhofer diffraction integrals from the plane-wave-spectrum approach to diffraction. The stationary phase approximation is also useful in a wide variety of other problems, optical and otherwise.

Talbot Images

If an infinite periodic transparency with period Λ is illuminated by a normally incident plane wave of wavelength λ, then the irradiance distribution will repeat itself as the field propagates, forming what are known as **Talbot images**, or **self images**, every Talbot distance $z_T = 2\Lambda^2/\lambda$.



In addition to producing an image of the initial periodic transmittance, an inverse Talbot image is formed at Λ^2/λ, and in between each Talbot image and inverse Talbot image there is a double-frequency image.

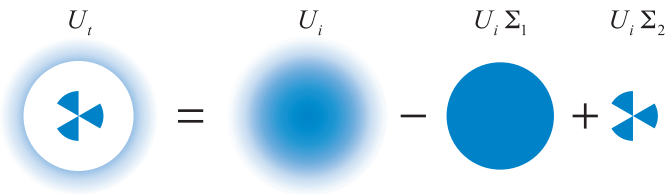
Although the above diagram illustrates a simple 2D grating having a single pitch in each orthogonal direction, the Talbot effect is more general and applies to any periodic structure.

Periodic structures of finite extent also display this effect. However, the contrast begins to decay after a few cycles.

A common application of the Talbot effect is the so-called **Talbot interferometer**. When the illuminating wavefront contains some aberration, the self images are deformed. The deformation is roughly linear with the slope of the abberrated wavefront, and the self image can be processed to retrieve the wavefront information.

Babinet's Principle

A consequence of the principle of linear superposition is that the field leaving a complex aperture can be decomposed into contributions from complimentary subapertures. The diagram illustrates the total field U_t in terms of the incident field U_i and the two aperture transmission functions Σ_1 and Σ_2.



Babinet's principle states that at some distance from the complex aperture, the field is the sum of the propagated fields from each subaperture. This is a consequence of the linearity of beam propagation, as indicated below with complimentary Σ_1 and Σ_2 and the linear operator $\mathbf{L}$:

$$\mathbf{L}(U_t) = \mathbf{L}(U_i) - \mathbf{L}(U_i\Sigma_1) + \mathbf{L}(U_i\Sigma_2)$$

Although this schematic diagram employs irradiance plots computed with the Fresnel diffraction integral, remember that Babinet's principle applies to the field (not the irradiance).

Babinet's principle is useful in the problem of an infinite half-plane aperture illuminated by a normally incident plane wave. Any point on the edge of the geometric shadow has an equal contribution from either half-plane of the unobstructed beam, so the field amplitude on the edge of the geometric shadow is 50% of the unobstructed field, and the irradiance is 25% of the unobstructed irradiance.

Fresnel Diffraction by a Rectangular Aperture

Assuming a normally incident plane wave of amplitude A, the **Fresnel diffraction integral** gives the amplitude U in a parallel plane at a distance z.

$$U = -ie^{ikz}$$

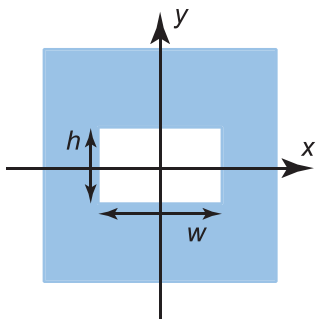
$$\times \left\{ C\left(\frac{w - 2x'}{\sqrt{\lambda z}}\right) + C\left(\frac{-w - 2x'}{\sqrt{\lambda z}}\right) + i\,\mathrm{sign}\left[S\left(\frac{w - 2x'}{\sqrt{\lambda z}}\right) + S\left(\frac{-w - 2x'}{\sqrt{\lambda z}}\right)\right]\right\}$$

$$\times \left\{ C\left(\frac{h - 2y'}{\sqrt{\lambda z}}\right) + C\left(\frac{-h - 2y'}{\sqrt{\lambda z}}\right) + i\,\mathrm{sign}\left[S\left(\frac{h - 2y'}{\sqrt{\lambda z}}\right) + S\left(\frac{-h - 2y'}{\sqrt{\lambda z}}\right)\right]\right\}$$

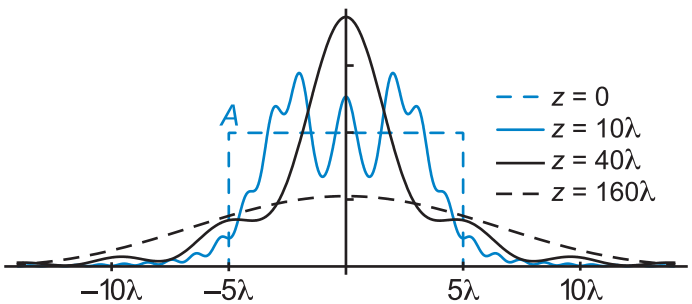
where $\mathrm{sign}(z)$ is $+1$ or -1 depending on the size of z, and the **Fresnel integrals** (not to be confused with the Fresnel diffraction integral) are

$$C(x) = \int_0^x \cos\frac{\pi u^2}{2}\,du$$

$$S(x) = \int_0^x \sin\frac{\pi u^2}{2}\,du$$

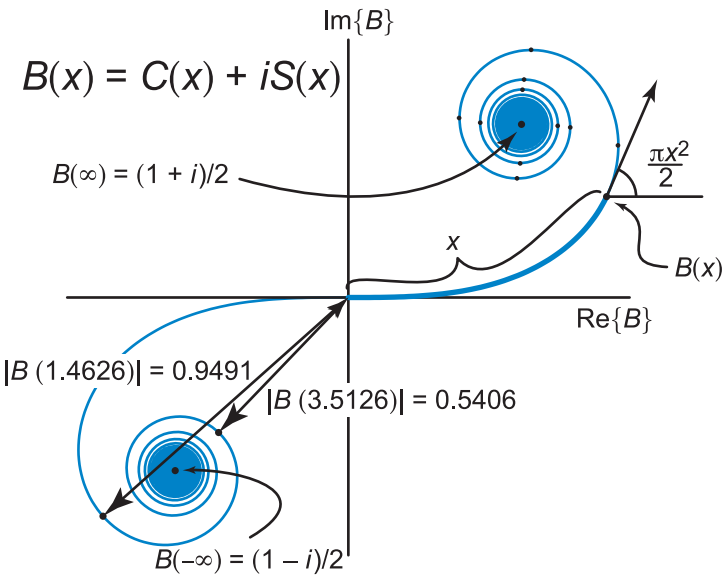


Many mathematical software packages include Fresnel integral calculators, making evaluation of expressions in terms of $C(x)$ and $S(x)$ very easy. The example below shows the case of a slit that is 10 × 100 wavelengths in size and where the irradiance is evaluated at 10, 40, and 160 wavelengths from the slit.



Cornu Spiral

The **Cornu spiral** is a plot of the phasor $B = C + iS$, where C and S are Fresnel integrals. The Cornu spiral can be used to make quantitative diffraction calculations in certain situations, but today its main use is in providing a deeper understanding of diffraction problems.



For example, it is easy to see from the Cornu spiral that at the border of the geometric shadow of a knife edge, illuminated by a unit-amplitude plane wave, the field amplitude is halved, while the irradiance is one-fourth that of the unobstructed field because the arc starts at the origin and wraps around one-half of the full spiral.

As the observation point moves into the shadow, the arc wraps around one end of the spiral, and the length of the phasor (and thus the irradiance) decreases monotonically. As the observation point moves away from the shadow, the arc wraps around the other end, and the phasor length goes through a series of maxima and minima.

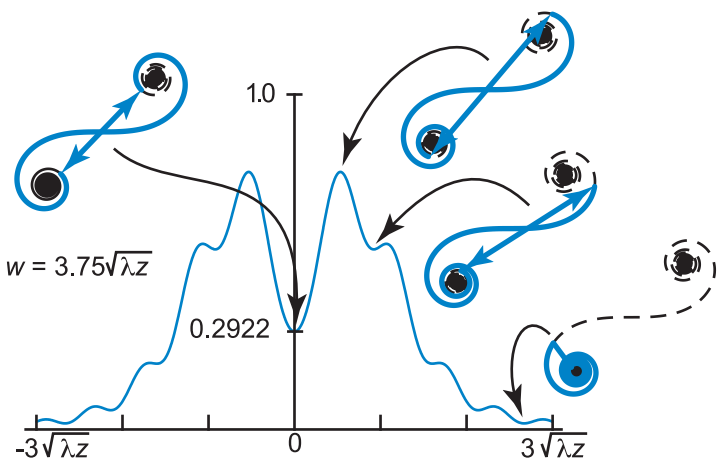
Cornu Spiral (cont.)

It is also instructive to consider the case of a long, narrow slit of width w. In this case, the diffraction integral reduces to

$$U = -ie^{ikz} \times \left[B\left(\frac{w - 2x'}{\sqrt{\lambda z}}\right) + B\left(\frac{-w - 2x'}{\sqrt{\lambda z}}\right) \right]$$

When the observation point is behind the center of the slit, the path along the **Cornu spiral** is symmetric with respect to its origin. In the example plot below, the width of the slit and the distance to the observation point are chosen so that the center of the diffraction pattern is relatively dark.

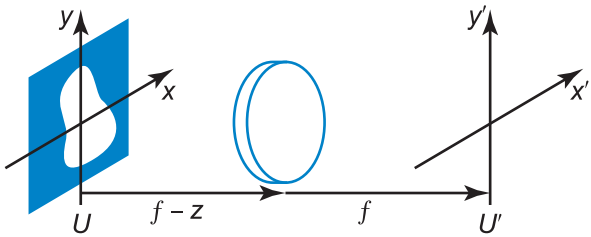
As the observation point shifts, the length of the arc along the Cornu spiral does not change, but its position does, so the irradiance oscillates between relative maxima and minima on either side of the center.



As the observation point moves well outside of the geometric shadow, the path wraps tightly around the spiral so that the phasor length becomes smaller and the local maxima and minima approach zero.

Propagation through a Lens

A transparency illuminated by a normally incident unit-amplitude plane wave, a lens placed at a distance $f - z$, and the field is observed in the rear focal plane.



$z > 0$ indicates that the transparency is to the right of the front focal plane.

1. Propagation of the field transmitted by the transparency is performed using the spectrum of plane waves and the paraxial approximation to the transfer function of free space.
2. The action of the lens is represented by a quadratic phase $e^{-ik\,(x^2\,+\,y^2)/2f}$.
3. The **Fresnel diffraction integral** propagates the field to the observation plane.

Because the observation plane is at the back focal plane, the quadratic phase inside the Fresnel integral is canceled by the action of the lens. If the lens is assumed to be infinite in extent, then the simplification produces

$$U'(x',y') = \frac{e^{ik(2f-z)}}{i\lambda f} e^{\frac{ikz}{2f^2}(x'^2+y'^2)} \iint t_A(x,y)\,U(x,y)\,e^{-i\frac{2\pi}{\lambda f}(xx'+yy')}\,dx\,dy$$

If the transparency, as defined by the transparency function t_A, is placed in the front focal plane, the leading quadratic phase disappears so that without the **Fraunhofer approximation**, the lens produces a final field that is a scaled Fourier transform of the initial field.

Airy Disk

Consider scalar distribution U' within the pupil at a distance z from the front focal plane of a lens:

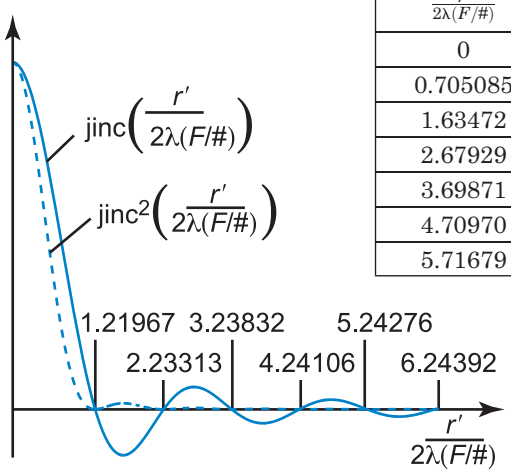
$$U' = \frac{e^{ik(2f+z)}}{i\lambda f} e^{\frac{ikz}{2f^2}r'^2} \int_{-\pi}^{\pi}\int_{0}^{\infty} U e^{-i\frac{2\pi}{\lambda f}rr'\cos(\theta-\theta')} r\,dr\,d\theta$$

$$= 2\pi \frac{e^{ik(2f+z)}}{i\lambda f} e^{\frac{ikz}{2f^2}r'^2} \int_{0}^{\infty} U J_0\left(\frac{2\pi}{\lambda f} rr'\right) r\,dr$$

J_0 is the zero-order Bessel function of the first kind. The field in the focal plane of the lens is the **Fourier–Bessel transform** of U. When $U = \mathrm{circ}(2r/\rho)$, the field is proportional to the **jinc function** and becomes the familiar **Airy disk**, which is proportional to jinc^2.

$$U' = \frac{\pi D e^{i\phi}}{i2\lambda(F/\#)} \frac{J_1\left(\frac{\pi r'}{\lambda(F/\#)}\right)}{\frac{\pi r'}{\lambda(F/\#)}} = \frac{\pi D e^{i\phi}}{i4\lambda(F/\#)} \mathrm{jinc}\left(\frac{r'}{2\lambda(F/\#)}\right)$$

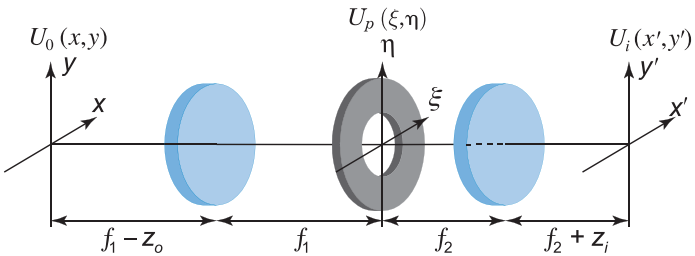
$$|U'|^2 = \frac{\pi^2 D^2 e^{i\phi}}{16\lambda^2 (F/\#)^2} \mathrm{jinc}^2\left(\frac{r'}{2\lambda(F/\#)}\right) = \textit{Airy disk}$$



$\frac{r'}{2\lambda(F/\#)}$	$\mathrm{jinc}\left(\frac{r'}{2\lambda F/\#}\right)$
0	1
0.705085	0.5
1.63472	−0.132279
2.67929	0.0644825
3.69871	−0.040008
4.70970	0.0279185
5.71679	−0.0209052

Double-Telecentric Imaging System

Two positive elements, separated by the sum of their focal lengths, with the stop in the common focal plane, form a **double-telecentric** system, because the entrance and exit pupils are at infinity.



$U_0(x,y)$ is the field in a plane z_o from the front focal plane, and the field incident on the stop $U_p(\xi,\eta)$ is

$$U_p = \frac{e^{ik(2f_1 - z_o)}}{i\lambda f_1} e^{\frac{ikz_o}{2f_1^2}(\xi^2 + \eta^2)} \iint U_o e^{-i\frac{2\pi}{\lambda f_1}(x\xi + y\eta)} dx\, dy$$

The field at a plane z_i from the final back focal plane $U_i(x', y')$ in terms of the pupil function $P(\xi,\eta)$ is

$$U_i = \frac{e^{ik(2f_2 + z_i)}}{i\lambda f_2} \iint P U_p e^{-\frac{ikz_i}{2f_2^2}(\xi^2 + \eta^2)} e^{-i\frac{2\pi}{\lambda f_2}(\xi x' + \eta y')} d\xi\, d\eta$$

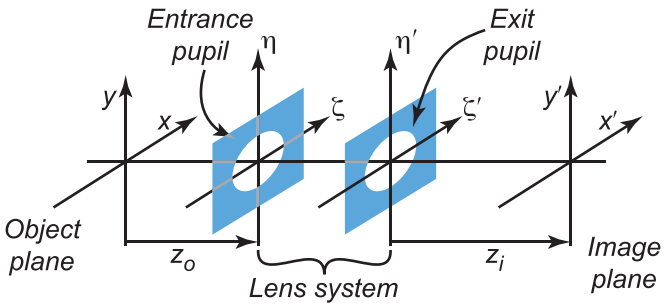
When substituting U_p, the quadratic phases cancel if the first and last planes are conjugate:

$$z_i / z_o = f_2^2 / f_1^2$$

These two integrals show that the image is estimated by first Fourier transforming the field leaving the object plane, then multiplying by the pupil function, and finally Fourier transforming again to arrive at the image plane. Defocus is included by means of a phase factor.

Linear and Shift-Invariant Imaging System

Most imaging problems can be cast in a simplified optical model where an object plane is at a distance z_o from an entrance pupil, which is relayed to an exit pupil with some magnification m_p, followed by an image plane with magnification m_i at a distance z_i beyond the exit pupil.



$$\{x',y'\} = m_i\{x,y\}, \quad \{\xi',\eta'\} = m_p\{\xi,\eta\}$$

It is convenient to assume that the imaging system is **linear and shift invariant (LSI)**. If the operator F describes the imaging system where $F\{f_1(x)\} = g_1(x)$ and $F\{f_2(x)\} = g_2(x)$ while a and b are constants, then F is linear and shift invariant when

$$F\{a\, f_1(x) + b\, f_2(x)\} = a\, g_1(x) + b\, g_2(x)$$
$$F\{f_1(x - x_0)\} = g_1(x - x_0)$$

Complex exponentials such as $f(x) = e^{-i2\pi\xi x}$ are eigenfunctions of LSI systems. Because they also form the kernel of the Fourier transform and the Fraunhofer diffraction integral (which assumes LSI), a cornucopia of mathematical tools are available to describe LSI imaging systems.

Shift invariance is not satisfied if there is an image flip, so it is also convenient to carry out calculations in either the object or image space using the geometric object or image as the input.

Coherent and Incoherent Point Spread Function

The **point spread function (PSF)** of an optical system $h(x,y)$ is the image of a point object. The image produced by an optical system can be thought of as a convolution of geometric image and the PSF:

$$f(x,y) = \iint g(x,y)h(x-\alpha, y-\beta)d\alpha\, d\beta$$

Although negative magnification can violate shift invariance, it can always be recovered by using image space coordinates.

The **coherent point spread function** of an LSI coherent imaging system is the Fraunhofer diffraction integral of the pupil function $P(\xi,\eta)$:

$$h_c(x,y) = \frac{A}{\lambda z_i}\iint P(\xi,\eta)e^{-i\frac{2\pi}{\lambda z_i}(\xi x,\eta y)}d\xi\, d\eta$$

The **central ordinate theorem** states that the central value of the Fourier transform of a function is the integral over the function itself and shows that the **on-axis irradiance** is proportional to the square of transmission-weighted area of the aperture.

$$I(0,0) = \propto h_c^2(0,0) = \left|\frac{A}{\lambda z_i}\iint P(\xi,\eta)d\xi\, d\eta\right|^2$$

Incoherent imaging occurs when the fields at any two points in the object are completely uncorrelated, causing each imaged point to add in intensity rather than amplitude. So the **incoherent point spread function** is the squared modulus of the coherent point spread function:

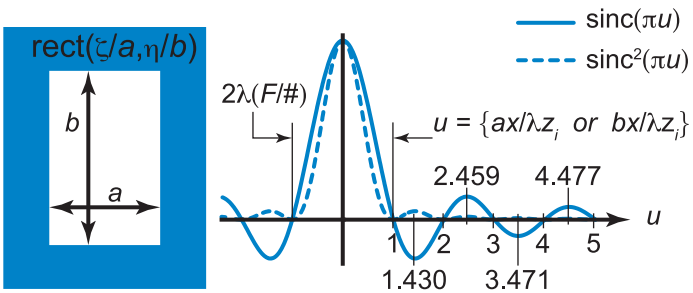
$$h_i(x,y) = |h_c(x,y)|^2 = \left|\frac{A}{\lambda z_i}\iint P(x,\eta)e^{-i\frac{2\pi}{\lambda z_i}(\xi x+\eta y)}d\xi\, d\eta\right|^2$$

PSF for Rectangular and Circular Apertures

The coherent **PSF** of a **rectangular aperture** is the product of sinc functions, and the incoherent PSF is the product of sinc-squared functions:

$$h_c(x,y) = ab\frac{A}{\lambda z_i}\operatorname{sinc}(\pi a x/\lambda z_i)\operatorname{sinc}(\pi b y/\lambda z_i)$$

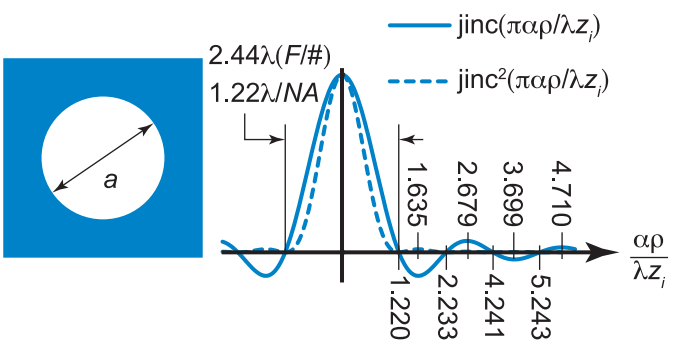
$$h_i(x,y) = |h_c(x,y)|^2 = \left(ab\frac{A}{\lambda z_i}\right)^2 \operatorname{sinc}^2(\pi a x/\lambda z_i)\operatorname{sinc}^2(\pi b y/\lambda z_i)$$



The PSF of a **circular aperture** is a jinc function. The incoherent PSF is a jinc-squared function.

$$h_c(x,y) = \frac{\pi a^2}{4}\frac{A}{\lambda z_i}\frac{2J_1(\pi a r/\lambda z_i)}{\pi a r/\lambda z_i} = \frac{\pi a^2}{4}\frac{A}{\lambda z_i}\operatorname{jinc}(\pi a r/\lambda z_i)$$

$$h_i(x,y) = |h_c(x,y)|^2 = \left(\frac{\pi a^2}{4}\frac{A}{\lambda z_i}\right)^2 \operatorname{jinc}^2(\pi a r/\lambda z_i)$$



Transfer Function

Although the image can be computed as a convolution of the geometric image with the PSF, it is often more convenient to use the **transfer function** $H(\xi,\eta)$, which is simply the Fourier transform of the PSF.

If the image $f(x,y)$ is a convolution of the geometric image $g_i(x,y)$ with the PSF $h(x,y)$

$$f(x,y) = \iint g_i(x',y')h(x-x',y-y')dx'dy'$$

then the Fourier transform of the image is the product of the geometric image spectrum $G(\xi,\eta)$ and the transfer function of the imaging system $H(\xi,\eta)$.

$$F(\xi,\eta) = \iint f(x,y)e^{-i2\pi(\xi x+\eta y)}dx\,dy$$

$$= G_i(\xi,\eta)H(\xi,\eta)$$

where

$$G_i(\xi,\eta) = \iint g_i(x,y)e^{-i2\pi(\xi x+\eta y)}dx\,dy$$

$$H(\xi,\eta) = \iint h(x,y)e^{-i2\pi(\xi x+\eta y)}dx\,dy$$

The transfer function itself depends on the type of imaging (coherent, incoherent, or some form of partial coherence) and the pupil function, which could include aberrations and apodization.

Coherent Transfer Function (CTF)

In the case of a coherent imaging system, the PSF of a linear-shift-invariant system is related to the Fourier transform of the pupil function $P(\xi,\eta)$:

$$h_c(x,y) = \frac{A}{\lambda z_i} \iint P(\xi,\eta) e^{i\frac{2\pi}{\lambda z_i}(\xi x + \eta y)} d\xi\, d\eta$$

The transfer function is the Fourier transform of the PSF:

$$\begin{aligned} H_c(f_x,f_y) &= \iint h_c(x,y) e^{-i2\pi(xf_x+yf_y)} dx\, dy \\ &= \iint \left(\frac{A}{\lambda z_i} \iint P(\xi,\eta) e^{-i\frac{2\pi}{\lambda z_i}(\xi x+\eta y)} \right) e^{-i2\pi(xf_x+yf_y)} dx\, dy \end{aligned}$$

Carrying out this integral shows that the **coherent transfer function (CTF)**, also known as the **amplitude transfer function**, is a scaled version of the pupil function.

$$H_c(f_x,f_y) = A\lambda z_i P(-\lambda z_i f_x, -\ \lambda z_i f_y)$$

The leading factor of $A\lambda z_i$ is often dropped because only the general shape of the image is of interest. Also, whenever the pupil is symmetric to inversion (which is usually the case), the negative signs in the arguments to P can be dropped, and the CTF becomes

$$CTF(f_x,f_y) = H_c(f_x,f_y) = P(\lambda z_i f_x, \lambda z_i f_y)$$

Note that the pupil function includes the aperture and also any wavefront aberrations and apodization.

Incoherent Transfer Function and the Optical Transfer Function

In an incoherent imaging system, each point in the object (and therefore in the geometric image) does not interfere with the others, so the image points add in irradiance rather than in amplitude. As a result, the point spread function of an incoherent imaging system is the squared modulus of the coherent point spread function:

$$h_i(x,y) = |h_c(x,y)|^2 = \left| \frac{A}{\lambda z_i} \iint P(\xi,\eta) e^{-i\frac{2\pi}{\lambda z_i}(\xi x + \eta y)} d\xi\, d\eta \right|^2$$

The **incoherent transfer function** H_i is the Fourier transform of the PSF, which is the autocorrelation of the CTF:

$$\begin{aligned} H_i(f_x,f_y) &= \iint h_i(x,y) e^{-i2\pi(xf_x+yf_y)} dx\, dy \\ &= \iint \left(\left| \frac{A}{\lambda z_i} \iint P(\xi,\eta) e^{-i\frac{2\pi}{\lambda z_i}(\xi x+\eta y)} d\xi\, d\eta \right|^2 \right) e^{-i2\pi(xf_x+yf_y)} dx\, dy \\ &= |A\lambda z_i|^2 \iint P(\lambda z_i f_x, \lambda z_i f_y) P(\lambda z_i f_x + \xi', \lambda z_i f_y + \eta') d\xi\, d\eta \\ &= \iint H_c(f_x', f_y') H_c^*(f_x' - f_x, f_y' - f_y) df_x'\, df_y' \end{aligned}$$

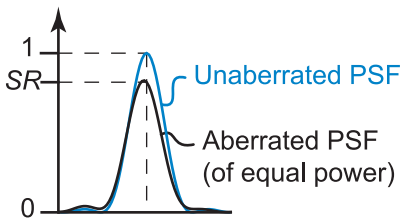
Normalizing by the central value yields the **optical transfer function (OTF)**: the normalized autocorrelation of the CTF (a scaled and flipped version of the pupil function).

After a change of variables, the OTF in a symmetric form is

$$\begin{aligned} OTF(f_x f_y) &= \frac{H_i(f_x,f_y)}{H_i(0,0)} \\ &= \frac{\iint H_c\left(f_x' + \frac{f_x}{2}, f_y' + \frac{f_y}{2}\right) H_c^*\left(f_x' - \frac{f_x}{2}, f_y' - \frac{f_y}{2}\right) df_x'\, df_y'}{\iint H_c(f_x', f_y') H_c^*(f_x', f_y') df_x'\, df_y'} \end{aligned}$$

Strehl Ratio

The **Strehl ratio** is the ratio of the central value of an aberrated and/or apodized PSF to that of an unaberrated and unapodized PSF. At a distance z_i from the exit pupil, close to the Gaussian focus distance z_g, the Fresnel diffraction formula gives



$$SR = \frac{\left| z_i^{-1} \iint A(\mathbf{r}) P(\mathbf{r}) e^{i\phi(\mathbf{r})} d\mathbf{r} \right|^2}{\left| z_g^{-1} \iint P(\mathbf{r}) d\mathbf{r} \right|^2} = \frac{\iint OTF_a(\mathbf{v}) d\mathbf{v}}{\iint OTF_u(\mathbf{v}) d\mathbf{v}}$$

where A is real and represents apodization, while ϕ represents phase aberration, and P is unity everywhere inside the exit pupil. The right-most expression is derived from the Fourier transform relationship between the PSF and OTF, where a and u indicate aberrated/apodized and unaberrated/unapodized OTFs.

The limit of small aberrations in the Strehl ratio can be estimated through the variance of the phase aberrations:

$$\sigma_\phi^2 = \langle \phi^2 \rangle - \langle \phi \rangle^2, \quad \text{where } \langle f \rangle = \frac{\iint P(\mathbf{r}) f(\mathbf{r}) d\mathbf{r}}{\iint P(\mathbf{r}) d\mathbf{r}}$$

$$SR \cong (1 - \sigma_\phi^2/2)^2 \quad \text{Maréchal formula}$$

$$SR \cong 1 - \sigma_\phi^2$$

$$SR \cong e^{1-\sigma_\phi^2}$$

The first expression is the well-known **Maréchal formula**, and although the second expression is less expensive computationally, the third expression usually yields more accurate results.

Properties of the OTF and MTF

If the pupil function has inversion symmetry, then we can also state that the **OTF** is the normalized autocorrelation of the pupil function:

$$OTF = \frac{\iint P(x,y)P^*(x - \lambda z_i f_x, y - \lambda z_i f_y)\,dx\,dy}{\iint |P(x,y)|^2\,dx\,dy}$$

The **modulation transfer function (MTF)** is the absolute value of the OTF, and the pair have several important properties.

$$MTF(f_x,f_y) = |OTF(f_x,f_y)|$$

1. The MTF and OTF are normalized so that

$$OTF(0,0) = MTF(0,0) = 1.$$

2. The OTF has Hermitian symmetry and the MTF has inversion symmetry:

$$OTF(f_x f_y) = OTF^*(-f_x, -f_y)$$
$$MTF(f_x f_y) = MTF(-f_x, -f_y)$$

3. The central value is always larger than, or equal to, any other value:

$$|OTF(0,0)| \geq |OTF(f_x,f_y)|$$
$$MTF(0,0) \geq OTF(f_x,f_y)$$

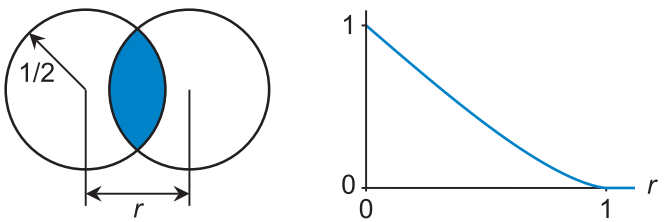
4. The OTF of an aberration-free system is always real and nonnegative so that OTF = MTF.
5. Aberrations cannot increase the value of the MTF over an unaberrated MTF at any particular frequency.

CTF and OTF of a Circular Aperture

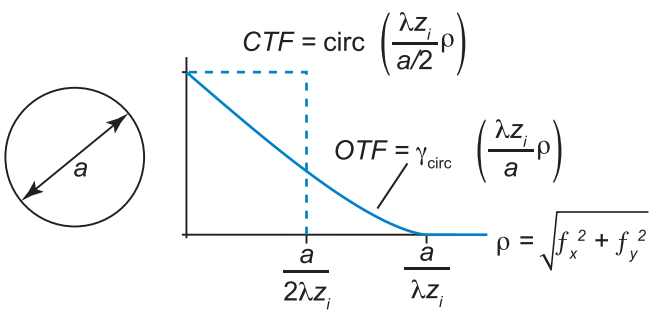
The **OTF** for an unaberrated imaging system with a **circular aperture** is the autocorrelation of a circ function, which corresponds to the overlapping area between two relatively displaced circ functions:

$$\text{circ}(r) = \begin{cases} r < 1 & \rightarrow & 1 \\ r = 1 & \rightarrow & 1/2 \\ r > 1 & \rightarrow & 0 \end{cases}$$

$$\gamma_{\text{circ}}(r) = \frac{2}{\pi}\left(\cos^{-1}(r) - r\sqrt{1 - r^2}\right)\text{circ}(r)$$



For an aperture diameter of a, the **CTF** and OTF become



Note that the **cutoff frequency** of the OTF is twice that of the CTF and that we can write these in terms of the f-number (F/#) or numerical aperture (NA) of the beam.

$$f_{cutoff}^{incoherent} = \frac{a}{\lambda z_i} = \frac{1}{\lambda(F/\#)} = 2\frac{NA}{\lambda} = 2f_{cutoff}^{coherent}$$

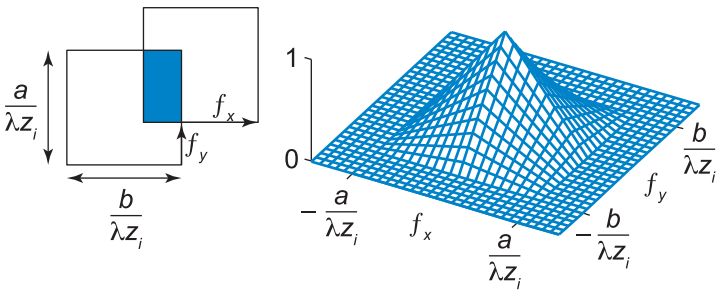
CTF and OTF of a Rectangular Aperture

With a rectangular aperture of dimensions $a \times b$, the **CTF** is just the product of rect functions.

$$CTF = \text{rect}\left(\frac{\lambda z_i f_x}{a}\right)\text{rect}\left(\frac{\lambda z_i f_y}{b}\right)$$

The **OTF** is then the product of tri functions:

$$OTF = \text{tri}\left(\frac{\lambda z_i f_x}{2a}\right)\text{tri}\left(\frac{\lambda z_i f_y}{2b}\right)$$



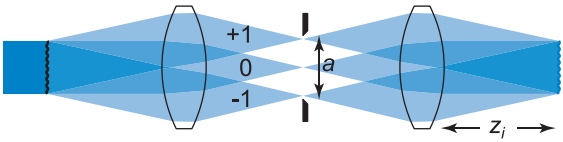
$$\text{tri}(x) = \begin{cases} |x| \le 1 & \rightarrow \quad 1 - |x| \\ |x| > 1 & \rightarrow \quad 0 \end{cases}$$

The incoherent cutoff frequency is twice the coherent cutoff but depends on the direction in the f_x-f_y plane. In a direction along one of the sides, the cutoff is $a/\lambda z_i$. But the maximum cutoff will be along one of the diagonals.

$$f_{cutoff}^{incoherent} = \frac{\sqrt{a^2 + b^2}}{\lambda z_i} = 2f_{cutoff}^{coherent}$$

Coherent and Incoherent Cutoff Frequency

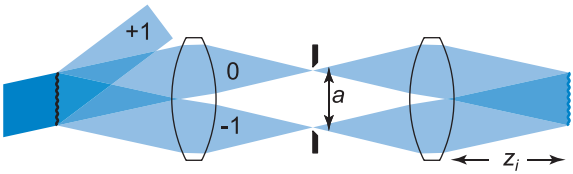
The classic **coherent imaging** system uses a normally incident monochromatic plane wave for illumination.



When the object is a grating with a single frequency at the coherent cutoff, the $+1$ and -1 orders are just barely accepted by the aperture of the imaging system. These three beams then produce a fringe pattern at the **cutoff frequency**. If the grating frequency is any greater than the cutoff frequency, then only the zero-order diffracted beam is passed by the imaging system, and the modulation is reduced to zero.

$$f_{cutoff}^{coherent} = \frac{a/2}{\lambda z_i} = \frac{1}{2\lambda(F/\#)} = \frac{NA}{\lambda}$$

In an **incoherent imaging** system, the grating is illuminated from every direction. Although frequencies just beyond the coherent cutoff are still blocked for a normally incident wavefront, the zero order and either the $+1^{st}$ or -1^{st} orders can be passed by other angles of illumination and allowed to interfere at the image plane. The off-axis light continues to produce modulation until only one of the three beams can be passed; this occurs at twice the coherent cutoff.

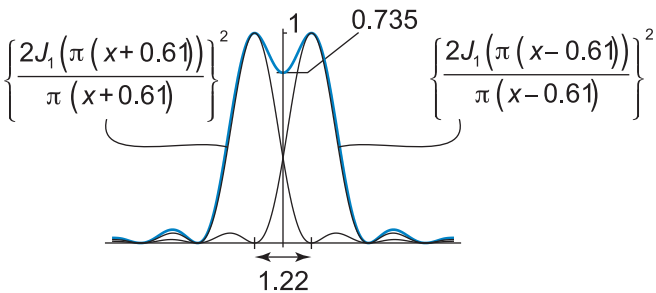


$$f_{cutoff}^{incoherent} = \frac{a}{\lambda z_i} = \frac{1}{\lambda(F/\#)} = 2\frac{NA}{\lambda}$$

Rayleigh Criterion

The **Rayleigh criterion** is a common metric for determining whether two points or lines are resolved. It states that the first zeros of two neighboring point or line spread functions overlap with the maximum of one over the first zero of the other.

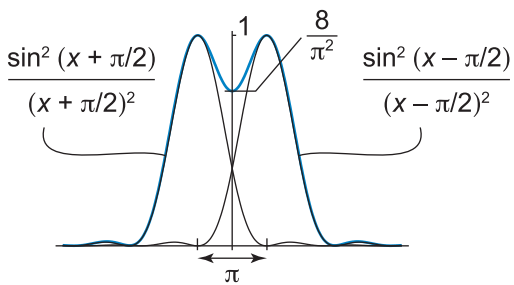
In the case of point objects and a circular aperture, the dip in the middle of the two images is about 74% of the maximum.



By the Rayleigh criterion, the angular resolution of a diffraction-limited telescope with a circular-entrance-pupil diameter D and imaging wavelength λ is

$$\textit{Angular resolution} = 1.22\,\lambda/D$$

In spectroscopy, neighboring slit images must be resolved, so it is also common to apply the Rayleigh criterion to sinc functions as shown below.



Rotationally Symmetric Gaussian Beams

The classical Gaussian beam closely approximates many real beams—primarily coherent laser beams—and is therefore of great importance in wave optics.

The basic form of the **Gaussian beam**, which is a solution to the paraxial Helmholtz equation, is a parabolic approximation to a spherical wave with complex radius curvature q:

$$U = ue^{ikz} = -\frac{iA}{q}e^{i\frac{kr^2}{2q}}e^{ikz}$$

The paraxial Helmholtz equation requires that $\partial q/\partial z = 1$ or that $q = z + a + iz_0$, where a and z_0 are real constants:

$$u = -\frac{iA}{q}e^{i\frac{kr^2}{2|q|^2}(z+a+iz_0)} = -\frac{iA}{q}e^{i\frac{kr^2}{2|q|^2}(z+a)}e^{-\frac{kz_0r^2}{2|q|^2}}$$

It is customary to set $a = 0$ so that the wavefront is flat at $z = 0$. The **Rayleigh range** z_0—which must be positive so that u has the form of a paraxial spherical wave with a Gaussian irradiance profile—is the distance between the maximum and minimum absolute curvature.

Imposing a form with a real radius of curvature R, standard deviation, σ, total power Φ, and **Gouy shift** θ,

$$u = -\frac{iA}{q}e^{i\frac{kr^2}{2|q|^2/z}}e^{\frac{kz_0r^2}{2|q|^2}} = \sqrt{\frac{\Phi}{\pi\sigma^2}}e^{i\frac{kr^2}{2R}}e^{-\frac{r^2}{2\sigma^2}}e^{-i\theta}$$

where

$$q = z - iz_0 \qquad R = z + z_0^2/z \qquad A = k\sigma_0\sqrt{\Phi/\pi}$$

$$\sigma^2 = \sigma_0^2(1 + z^2/z_0^2) \qquad z_0 = k\sigma_0^2 \qquad \tan\theta = z/z_0$$

Rotationally Symmetric Gaussian Beams (cont.)

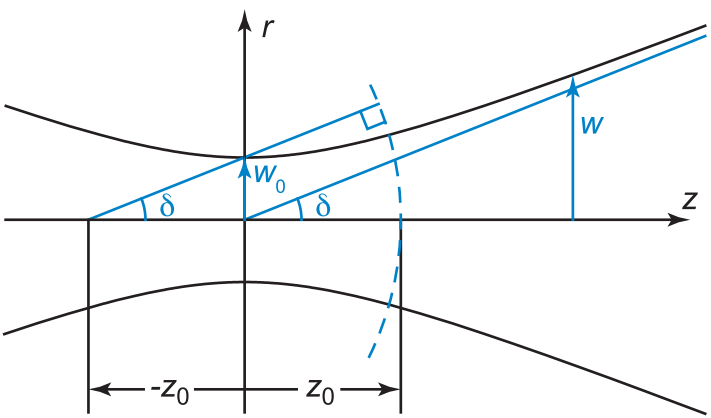
Another common form of the **Gaussian beam** uses the e^{-2} beam radius w.

$$u = -\frac{iA}{q} e^{-i\frac{kr^2}{2q}} = \sqrt{\frac{2\Phi}{\pi w^2}} e^{-i\frac{kr^2}{2R}} e^{-\frac{r^2}{w^2}} e^{-i\theta}$$

$$w = w_0\sqrt{1 + z^2/z_0^2} \quad \tan\theta = z/z_0 \quad \tan\delta = \frac{w_0}{z_0}$$

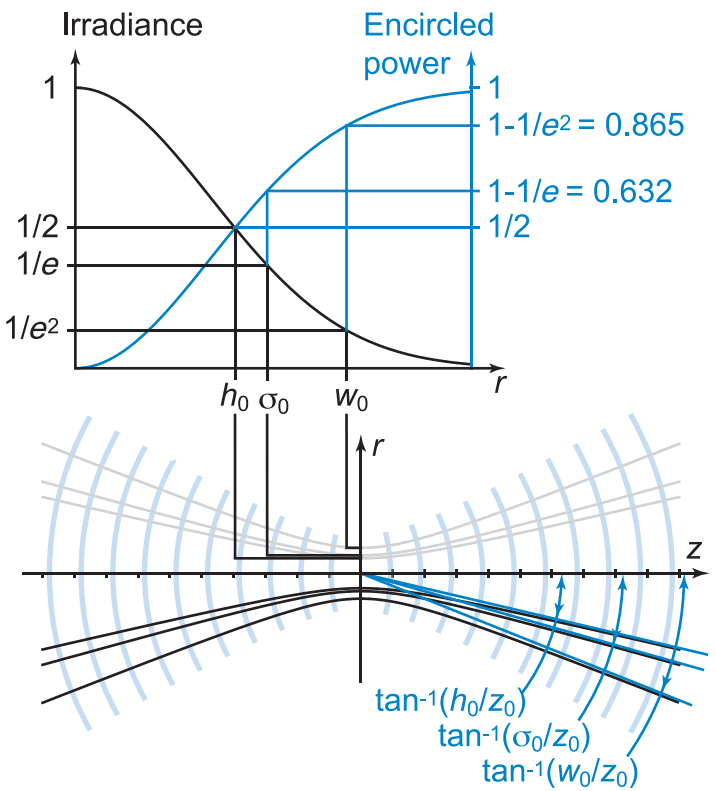
$$R = -z - z_0^2/z \qquad q = z + i\, z_0 \quad z_0 = \frac{1}{2} k\, w_0^2$$

Independent of how the radius is defined (with σ, w, or something else), the radius of the beam traces out a hyperbola with propagation. The divergence of a Gaussian beam δ, or sometimes $\tan\delta$, is the angle defined by the asymptote of the beam radius and happens to be the same angle defined by the beam-waist radius as viewed from the **Rayleigh range**.



Gaussian Beam Size

The size of rotationally symmetric (or two-fold symmetric) **Gaussian beams** can be described in several ways. The beam size is usually given in terms of the irradiance, and the most common size parameters are the half-width at half-max h, the $1/e$ radius σ, and the $1/e^2$ radius w. These plots illustrate the relationship between these size parameters relative to the irradiance profile of the beam (top) and the **beam divergence**, which may also be defined using the same size parameters.



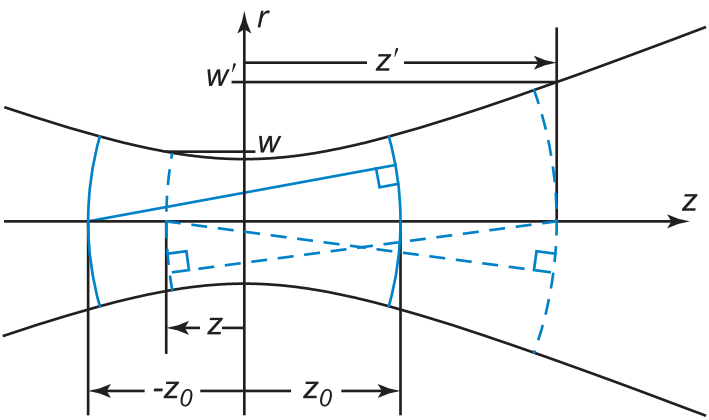
Although each of these parameters is a half-width, full-widths are sometimes encountered as well.

Rayleigh Range and Sister Surfaces

The **Rayleigh range** z_0 is perhaps best defined as the distance from the beam waist where the **Gouy shift** θ is $\pi/4$ because this is true for general Gaussian beams, which are not necessarily rotationally symmetric and do not exhibit separability in x and y.

In the case of a rotationally symmetric Gaussian beam, the Rayleigh range is simultaneously the distance from the waist where the wavefront curvature is maximum, where the beam area doubles, and where the axial irradiance halves. It also corresponds to a pair of **sister surfaces** positioned at equal distances on either side of the waist.

Sister surfaces are pairs of Gaussian beam wavefronts that are located at each other's center of curvature, as illustrated here.



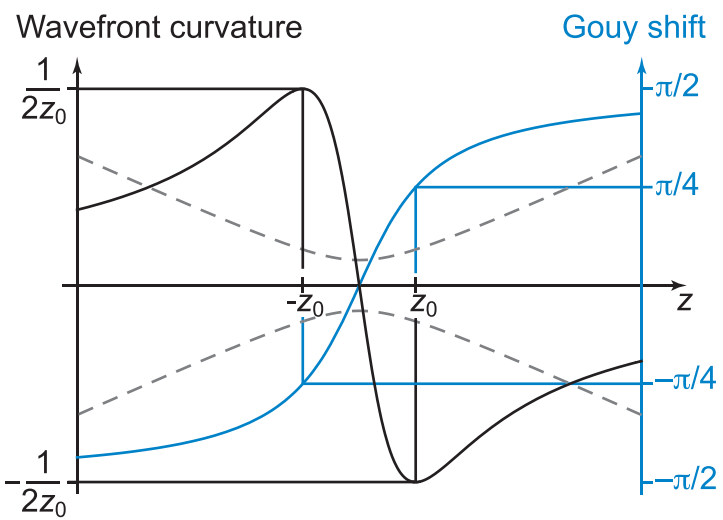
The positions, radii, and Gouy shifts of sister surfaces obey these relationships:

$$zz' = -z_0^2, \quad \frac{1}{w^2} + \frac{1}{w'^2} = \frac{1}{w_0^2}, \quad |\theta - \theta'| = \pi/2$$

Gouy Shift and Wavefront Curvature

The phase of a Gaussian beam relative to a parallel-propagating plane wave will lose 180 deg as it goes from $-\infty$ to $+\infty$. This effect, which also exists in any focused beam, is known as the **Gouy shift** (often misspelled as "Guoy") and is captured in the Gaussian beam parameter θ:

$$\tan\theta = \frac{z}{z_0}$$



In this plot, the rotationally symmetric Gaussian beam profile is shown schematically with a dashed line superimposed on the plots of **wavefront curvature** and Gouy shift. It is clear that the **Rayleigh range** corresponds to the location of Gouy shift of $\pm\pi/4$, as well as the positions of maximum absolute wavefront curvature (for the case of a rotationally symmetric Gaussian beam).

ABCD Method for Gaussian Beams

The complex curvature q completely specifies the unit-power rotationally symmetric **Gaussian beam**, and the **ABCD method** relates input and output complex curvatures q and q' through the elements of a paraxial system matrix:

$$q' = \frac{Aq + B}{Cq + D}$$

The system matrix (which can be formed from the submatrices shown in the table below) determines the propagation of a ray of initial height y and the optical angle $\omega = nu$ (paraxial angle u and refractive index n):

$$\begin{pmatrix} y' \\ \omega' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} y \\ \omega \end{pmatrix} = \mathbf{M}_n \ldots \mathbf{M}_2 \mathbf{M}_1 \begin{pmatrix} y \\ \omega \end{pmatrix} = \mathbf{M} \begin{pmatrix} y \\ \omega \end{pmatrix}$$

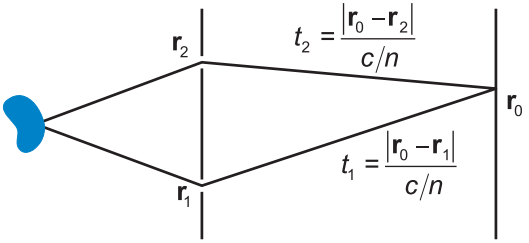
Transfer by reduced thickness $\tau = t/n$ (thickness t and index n)	$\mathbf{M}_T = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix}$
Thin lens of power ϕ	$\mathbf{M}_R = \begin{pmatrix} 1 & 0 \\ -\phi & 1 \end{pmatrix}$
Conjugates at magnification m	$\mathbf{M}_C = \begin{pmatrix} m & 0 \\ -\phi & 1/m \end{pmatrix}$
Focal planes for lens of power ϕ	$\mathbf{M}_F = \begin{pmatrix} 0 & 1/\phi \\ -\phi & 0 \end{pmatrix}$
Nodal planes with initial and final refractive indices n and n'	$\mathbf{M}_N = \begin{pmatrix} n/n' & 1/\phi \\ -\phi & n'/n \end{pmatrix}$

Any system matrix can also be derived in reverse, given initial and final heights and optical directions of any two linearly independent rays:

$$\mathbf{M} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} y_1' & y_2' \\ \omega_1' & \omega_2' \end{pmatrix} \begin{pmatrix} \omega_1 & -y_2 \\ -\omega_1 & y_2 \end{pmatrix} \Big/ \begin{vmatrix} y_1 & y_2 \\ \omega_1 & \omega_2 \end{vmatrix}$$

Young's Double Pinhole

In the classic **Young's double pinhole** experiment, a source illuminates a pair of separated pinholes located at $\mathbf{r}_1$ and $\mathbf{r}_2$, and the irradiance is observed at $\mathbf{r}_0$ as shown.



The irradiance at the observation point is the time-averaged square of the combined fields arriving at the observation point. The fields generally include transmission coefficients and obliquity factors $\sqrt{K_1}$ and $\sqrt{K_2}$, and the fields themselves further depend on the time delays t_1 and t_2 in reaching the observation point, which in turn depends on the relative position of the three points and the source.

$$I_{obs}(\mathbf{r}_0;\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{2}cn\varepsilon_0\left\langle\left|\sqrt{K_1}U(\mathbf{r}_1,t+t_1)+\sqrt{K_2}U(\mathbf{r}_2,t+t_2)\right|^2\right\rangle_t$$

$$= \frac{1}{2}cn\varepsilon_0\left[K_1\left\langle|U(\mathbf{r}_1,t+t_1)|^2\right\rangle_t + K_2\left\langle|U(\mathbf{r}_2,t+t_2)|^2\right\rangle_t\right.$$

$$\left. +2\sqrt{K_1K_2}\left\langle|U(\mathbf{r}_1,t+t_1)U^*(\mathbf{r}_2,t+t_2)|^2\right\rangle_t\right]$$

$$I_{obs}(\mathbf{r}_0;\mathbf{r}_1,\mathbf{r}_2) = K_1I_1 + K_2I_2 + cn\varepsilon_0\sqrt{K_1K_2}\Gamma(r_1,r_2,t_1,t_2)$$

The first two terms are the irradiance due to each pinhole (as if the other were not present), while the third term represents the presence of any interference and depends on the **mutual coherence function**:

$$\Gamma(r_1,r_2,t_1,t_2) = \left\langle|U(\mathbf{r}_1,t+t_1)U^*(\mathbf{r}_2,t+t_2)|^2\right\rangle_t$$

Mutual Coherence Function

In most cases, the source is considered stationary in time so that the **mutual coherence function** is only a function of the difference in time delay τ.

$$\Gamma(r_1,r_2,\tau) = \langle U(\mathbf{r}_1,t+\tau)U^*(\mathbf{r}_2,t)\rangle_t$$

Either form of the mutual coherence function is generally complex and can be normalized to form the **complex degree of coherence**: $\gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)$.

$$\gamma(\mathbf{r}_1,\mathbf{r}_2,\tau) = |\gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)|e^{i\,\phi_\gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)} = \frac{\Gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)}{\sqrt{\left\langle |U(\mathbf{r}_1,t)|^2\right\rangle_t \left\langle |U(\mathbf{r}_2,t)|^2\right\rangle_t}}$$

Referring to **Young's double pinhole**, assuming temporal stationarity and that the source has a narrow bandwidth with average angular frequency $\overline{\omega}$, and absorbing the obliquity factors into the irradiances, the field at the observation point is

$$I_{obs}(\mathbf{r}_1,\mathbf{r}_2,\tau) = I_1 + I_2 + 2|\gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)|\sqrt{I_1 I_2}\cos[\overline{\omega}\tau - \phi(\mathbf{r}_1,\mathbf{r}_2,\tau)]$$

This represents a set of fringes whose visibility at each point is determined by the modulus of the complex degree of coherence:

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = |\gamma(\mathbf{r}_1,\mathbf{r}_2,\tau)|$$

Because the position of the observation point is included in the time delay τ, the phase of the complex degree of coherence $\phi_\gamma(\mathbf{r}_1,\mathbf{r}_1,\tau)$ produces a shift of the fringes. This could be caused, for example, by a source that is not equidistant from the pinholes, or by a phase delay introduced as a block of nondispersive glass in front of one of the pinholes.

Spatial Coherence: Mutual Intensity

Spatial coherence, characterized by the mutual intensity, refers to the ability of two source points to interfere when there is no time delay, or when the time delay is short compared to the coherence time: the quasi-monochromatic condition.

The **mutual intensity** is the mutual coherence function under the special case of zero time delay ($\tau = 0$) and temporal stationarity (the mutual coherence function only depends on the difference in time delay).

$$\Gamma(\mathbf{r}_1,\mathbf{r}_2) = \Gamma(\mathbf{r}_1,\mathbf{r}_2,0)$$

Or, in normalized form

$$\gamma(\mathbf{r}_1,\mathbf{r}_2) = \frac{\Gamma(\mathbf{r}_1,\mathbf{r}_2)}{\sqrt{\left\langle |U(\mathbf{r}_1,t)|^2 \right\rangle_t \left\langle |U(\mathbf{r}_2,t)|^2 \right\rangle_t}}$$

$\Gamma(\mathbf{r}_1,\mathbf{r}_2)$ is sometimes called the **spatial coherence function** or **mutual optical intensity** (mutual intensity is used here because it is somewhat more common). At $\mathbf{r}_1 = \mathbf{r}_2$, there is perfect correlation, and

$$\Gamma(\mathbf{r}_1,\mathbf{r}_1) = I_1 \text{ and } \gamma(\mathbf{r}_1,\mathbf{r}_1) = 1$$

If $\mathbf{r}_2 = \mathbf{r}_1 + \Delta\mathbf{r}$, then the width of $\Gamma(\mathbf{r}_1,\mathbf{r}_1 + \Delta\mathbf{r})$ as a function of $\Delta\mathbf{r}$ is known as the **correlation length**. The correlation length can be specified in terms full-width at half-max, radius of the first zero, or any other convenient size parameter; the coherence area is then the area within the same specification.

Note that the correlation length sometimes goes by the name of "coherence length," but one should be careful with this term, which is reserved here to refer to the path length associated with the coherence time.

Van Cittert–Zernike Theorem

The **Van Cittert–Zernike theorem** is a key result of coherence theory that explains how propagation produces spatial coherence from an incoherent quasi-monochromatic source.

The mutual intensity of a **spatially incoherent source** drops off rapidly with the separation of the two points, so the width of the mutual intensity is generally on the order of one wavelength and can be approximated as

$$\Gamma(\mathbf{r}_1,\mathbf{r}_2) = A_c\delta(\mathbf{r}_1 - \mathbf{r}_2)I(\mathbf{r}_1)$$

where $I(\mathbf{r}_1)$ is the radiant exitance of the source, δ is a Dirac delta function, and A_c is the correlation area.

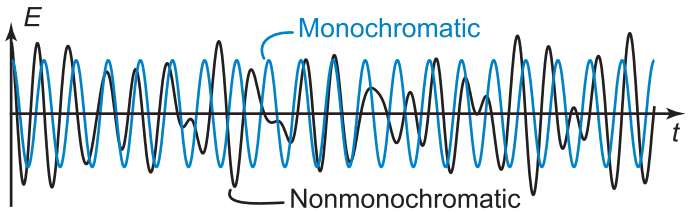
Assuming shift invariance, the downstream mutual intensity can be expressed in terms of a point spread function, and if that point spread function is the Fresnel wavelet $h = h_z^F$, then

$$\begin{aligned}
\Gamma(\mathbf{r}_2',\mathbf{r}_2') &= \langle U(\mathbf{r}_1')U^*(\mathbf{r}_2')\rangle \\
&= \left\langle \left(\iint U(\mathbf{r}_1)h(\mathbf{r}_1'-\mathbf{r}_1)dx_1dy_1\right)\left(\iint U(\mathbf{r}_2)h(\mathbf{r}_2'-\mathbf{r}_2)dx_2dy_2\right)^*\right\rangle \\
&= \frac{A_c}{\lambda^2 z^2}\iint I(\mathbf{r}_1)e^{-\frac{ik}{2z}(\mathbf{r}_1'-\mathbf{r}_1)^2}e^{\frac{ik}{2z}(\mathbf{r}_2'-\mathbf{r}_1)^2}dx_1dy_1 \\
&= \frac{A_c}{\lambda^2 z^2}e^{\frac{ik}{2z}(\mathbf{r}_2'-\mathbf{r}_1')}\iint I(\mathbf{r}_1)e^{-\frac{ik}{z}\{x_1(x_2'-x_1')+y_1(y_2'-y_1')\}}dx_1dy_1
\end{aligned}$$

This is the essence of the Van Cittert–Zernike theorem, which states that if the source is spatially incoherent, and the Fresnel approximation (not the more stringent Fraunhofer approximation) is satisfied, then the mutual intensity at a distance is the Fourier transform of the source. So a collection of light waves that starts out as perfectly incoherent (spatially) becomes coherent with propagation.

Temporal Coherence

The perfectly monochromatic harmonic wave is an idealization of **temporal coherence** that implies a sinusoidal wave that has existed and will continue to exist *forever*.



The length of time that a wave continues to resemble a sinusoid of a single frequency is the **coherence time**. Although a source with an infinite coherence time is unlikely to be encountered, many lasers can, in certain circumstances, have practically infinite coherence times.

Although sources of different frequencies may be made to interfere, it is generally assumed that the interference between different frequencies occurs on a time scale too short for normal detection. This allows us to model sources with a finite bandwidth as a sum of single-frequency systems. Often, one source is enough.

When a source can be treated as if it is monochromatic, it is referred to as **quasi-monochromatic**. A source behaves as if it is monochromatic if the fringe modulation (as a function of time delay) remains constant within the range defined by the application, or as long as the maximum optical path difference is somewhat less than the coherence length:

$$OPD_{\max} \ll \Delta l$$

Coherence Length

Several types of interferometers make two copies of a wavefront so that one is directed down a path that is Δz longer than the other and then recombined to interfere at an observation plane. If the recombined amplitudes are equal, then the irradiance for a monochromatic source in vacuum is

$$I(\Delta z;\nu) = \frac{c\varepsilon_0}{2} 2A_\nu^2[1+\cos(2\pi\ \Delta z\ \nu/c)]$$

and the **fringe contrast** is 1 and is independent of path difference.

If this source cannot be considered monochromatic but rather has a finite power spectrum A_ν^2, then the irradiance is the weighted sum over all of the frequencies:

$$I(\Delta z) = \int_0^\infty I(\nu;\Delta z)d\nu = c\varepsilon_0 \int_0^\infty A_\nu^2[1+\cos(2\pi\ \Delta z\ \nu/c)]d\nu$$

When the power spectrum is symmetric about ν_0,

$$I(\Delta z) = I_\infty + c\varepsilon_0 \cos(2\pi\ \Delta z\ \nu_0/c) \int_{-\nu_0}^\infty A_{\nu'+\nu_0}^2 \cos(2\pi\ \Delta z\ \nu'/c)d\nu'$$

and the **fringe visibility** is

$$V = \frac{1}{\Phi}\int_{-\nu_0}^\infty A_{\nu'+\nu_0}^2 \cos(2\pi n\nu' z/c)d\nu'$$

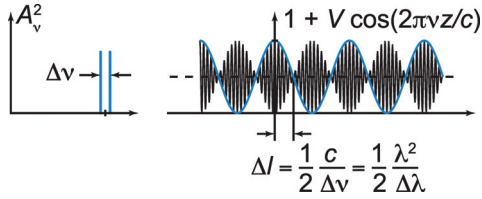
Depending on the width $\Delta\nu$ of the power spectrum, the fringe visibility as a function of the path difference will also have some finite width Δl, which is known as the **coherence length** of the source. In addition, the **coherence time** $\Delta t = \Delta l/c$ is the length of time over which the source emission remains correlated with a delayed copy of itself. The definition of the spectrum width and coherence length/time varies; it might refer to half-width at half-max or full-width between zeros, etc.

Coherence Length for Simple Spectra

Two very narrow frequency components:

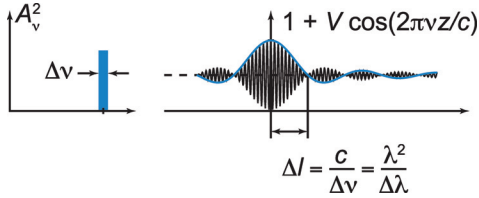
$$A_\nu^2 = \frac{1}{2}\delta(\nu - \nu_0 + \Delta\nu/2) + \frac{1}{2}\delta(\nu - \nu_0 - \Delta\nu/2)$$

$$V = \cos\left(\frac{\pi z}{c}\Delta\nu\right)$$



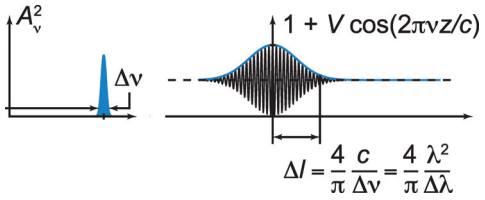
Top-hat frequency distribution:

$$A_\nu^2 = \frac{1}{\Delta\nu} rect\left(\frac{\nu - \nu_0}{\Delta\nu}\right), \quad V = \frac{\sin(n\pi z\Delta\nu/c)}{n\pi z\Delta\nu/c}$$



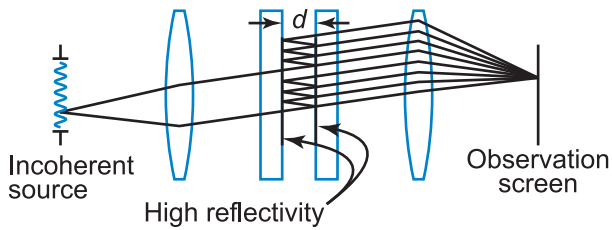
Gaussian frequency distribution:

$$A_\nu^2 = \frac{1}{\sqrt{\pi/8}\Delta\nu} e^{-2(\nu-\nu_0)^2/(\Delta\nu/2)^2}, \quad V = e^{-\frac{1}{8}\left(\frac{\pi n\Delta\nu}{c}z\right)^2}$$



Fabry–Pérot Interferometer

The **Fabry–Pérot interferometer** consists of parallel reflective surfaces through which a source is imaged onto a screen, as shown.



The resulting circular fringes are then described by the Airy formulae. With the ray angle θ inside the gap where the refractive index is n, and an additional phase factor ϕ, the order of interference is

$$m = \frac{\delta}{2\pi} = \frac{2nd}{\lambda}\cos\theta + \frac{\phi}{\pi}$$

On axis, the interference order m_0 is maximum. Generally not an integer, it can be defined in terms of the fractional part Δm and the order of the smallest circular fringe m_1.

$$m_0 = m_1 + \Delta m = \frac{2nd}{\lambda} + \frac{\phi}{\pi}$$

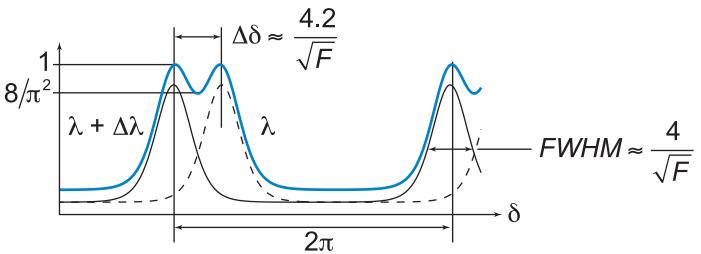
The angular coordinate of the j^{th} bright fringe for small θ_j is

$$\theta_j = \sqrt{\frac{\lambda}{nd}(j - 1 + \Delta m)}$$

When used in spectrometry, overlapping rings must be discerned for different wavelengths. This leads to the concepts of **resolving power** and **free spectral range**, both of which are essential to the practical use of Fabry–Pérot spectrometers.

Fabry–Pérot Spectrometer

Closely separated fringes (at different wavelengths) are considered resolved when the intensity dip between equal peaks is $8/\pi^2 \approx 81\%$ of the peak intensity and $\Delta\delta$ is the separation of these peaks in a plot of the fringe irradiance versus the path difference δ.



Under this 81% condition (derived from the **Rayleigh criteria** for imaging point sources) the **resolving power**—which is the ratio of the primary wavelength to the least resolvable-wavelength difference $\Delta\delta_{\text{min}}$—depends on the order of the fringe m:

$$\Re = \frac{\lambda}{\Delta\lambda_{\text{min}}} \approx m\mathfrak{F} = \frac{m\pi\sqrt{F}}{2}$$

where F is the **coefficient of finesse** that appears in Airy function, and $\mathfrak{F}$ is the **finesse** (the ratio of the fringe separation to the width of a monochromatic fringe).

When the gap between the mirrors is large enough, fringes from different wavelengths may overlap at different orders. The range of wavelengths (or frequencies) over which this does not occur is known as the **free spectral range**.

$$\Delta\lambda_{fsr} \approx \frac{\lambda^2}{2nd}, \quad \Delta\nu_{fsr} = \frac{c}{\lambda}\frac{\Delta\lambda_{fsr}}{\lambda} = \frac{c}{2nd}$$

Special Functions and Fourier Transforms

$f(x) = \int_{-\infty}^{\infty} F(\xi) e^{i2\pi x\xi}\, d\xi$	$F(\xi) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi x\xi} dx$
$f(x)$	$F(\xi)$
$F(x)$	$f(-\xi)$
$a\, f_1(x) + b\, f_2(x)$	$F_1(\xi) + F_2(\xi)$
$f_1(x) * f_2(x)$	$F_1(\xi)\, F_2(\xi)$
$f((x - x_0)/a)$	$\lvert a\rvert F(\xi) e^{-i2\pi x_0 \xi}$
$f(a\, x) e^{i2\pi x \xi_0}$	$\lvert 1/a\rvert F(\xi - \xi_0)$
$2\cos(2\pi x\, \xi_0)$	$\delta(\xi - \xi_0) + \delta(\xi + \xi_0)$
$2i \sin(2\pi x\, \xi_0)$	$\delta(\xi - \xi_0) - \delta(\xi + \xi_0)$
$\delta(x - x_0)$	$e^{-i2\pi x_0 \xi}$
$\mathrm{rect}(x)$	$\mathrm{sinc}(\pi\, x)$
e^{-x^2}	$\sqrt{\pi} e^{-\pi^2 \xi^2}$
$e^{i\pi x^2}$	$e^{-i\pi(\xi^2 - 1/4)}$
$\mathrm{comb}(x/a)$	$\lvert a\rvert \mathrm{comb}(a\, x)$
$\frac{\partial^n}{\partial x^n} f(x)$	$(i2\pi\xi)^n F(\xi)$

Equation Summary

Vector wave equations:

$$\nabla^2\mathbf{E} = \varepsilon\mu\frac{\partial^2\mathbf{E}}{\partial t^2} + \mu\sigma\frac{\partial\mathbf{E}}{\partial t}, \quad \nabla^2\mathbf{H} = \varepsilon\mu\frac{\partial^2\mathbf{H}}{\partial t^2} + \mu\sigma\frac{\partial\mathbf{H}}{\partial t}$$

Scalar wave equation:

$$\nabla^2 V = \varepsilon\mu\frac{\partial^2 V}{\partial t^2} + \mu\sigma\frac{\partial V}{\partial t}$$

Helmholtz equation:

$$\nabla^2 F(\mathbf{r}) + k^2 F(\mathbf{r}) = 0$$

Paraxial Helmholtz equation:

$$\nabla_{xy}^2 u + 2ik\frac{\partial u}{\partial z} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + 2ik\frac{\partial u}{\partial z} = 0$$

Plane wave, spherical wave, cylindrical wave:

$$\mathbf{U} = Ae^{i(\mathbf{k}\cdot\mathbf{r} - \omega t + \phi)}$$

$$U(\mathbf{r},t) = \frac{A}{|\mathbf{r} - \mathbf{r}_0|}e^{i(k|\mathbf{r}-\mathbf{r}_0| - \omega t + \phi_0)} \approx \frac{A}{z - z_0}e^{\frac{ik}{2(z-z_0)}[(x-x_0)^2 + (y-y_0)^2]}e^{i(k(z-z_0) - \omega t + \phi_0)}$$

$$U(r,t) = \frac{A}{\sqrt{r}}e^{i(kr - \omega t + \phi_0)}$$

Beer–Lambert law:

$$T = \|U/U_0\|^2 = e^{-\frac{2\omega}{c}\kappa z} = e^{-\frac{4\pi}{\lambda}\kappa z} = e^{-\alpha z}, \quad \alpha = \frac{2\omega}{c}\kappa = \frac{4\pi}{\lambda}\kappa$$

Degree of polarization:

$$DOP = |\boldsymbol{s}_p| = \sqrt{S_1^2 + S_2^2 + S_3^2}/S_0, \quad 0 \le DOP \le 1$$

Polarization extinction ratio:

$$PER = \frac{T_{\max}}{T_{\min}} = \frac{1 + D}{1 - D}$$

Equation Summary

Malus' law:

$$I_t/I_i = \cos^2\Delta\theta$$

Jones vectors for linear polarization states:

$$\mathbf{E}_{L\theta} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$

$$\mathbf{E}_{HLP} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{E}_{VLP} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{E}_{L\,+\,45\ \mathrm{deg}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$$\mathbf{E}_{L\,-\,45\ \mathrm{deg}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Jones vectors for circular polarization states:

$$\mathbf{E}_{RCP} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}, \quad \mathbf{E}_{LCP} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ +i \end{pmatrix}$$

Jones vectors from eigenpolarizations and eigenvalues:

Eigenpolarizations: $\psi_1 = (a,b)^T, \quad \psi_2 = (-b^*,a^*)^T$

Eigenvalues: λ_1, λ_2

Jones matrix:

$$\mathbf{J} = \frac{1}{a\,a^*+b\,b^*}\begin{pmatrix} a & -b^* \\ b & a^* \end{pmatrix}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} a^* & b^* \\ -b & a \end{pmatrix}$$

Jones matrix for retardance:

$$\mathbf{J}_R(\phi) = \begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{pmatrix}$$

Jones matrix for diattenuation:

$$\mathbf{J}_D(t_x,t_y) = \begin{pmatrix} \alpha_x & 0 \\ 0 & \alpha_y \end{pmatrix}$$

Equation Summary

Jones matrix for a linear polarizer at angle θ:

$$\mathbf{J}_{LP}(\theta) = \begin{pmatrix} \cos^2\theta & \cos\theta\,\sin\theta \\ \cos\theta\,\sin\theta & \sin^2\theta \end{pmatrix}$$

Jones matrix for horizontal and vertical linear polarizers:

$$\mathbf{J}_{HLP} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{J}_{VLP} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Jones matrix for right- and left-hand circular polarizers:

$$\mathbf{J}_{RCP} = \frac{1}{2}\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}, \quad \mathbf{J}_{LCP} = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}$$

Jones matrix for optical activity:

$$\mathbf{J}_{\beta} = e^{i\phi}\begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix}$$

Jones rotation matrix:

$$\mathbf{J}_{Rot}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Jones reflection matrix:

$$\mathbf{J}_{Mirror}(\theta) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

Equation Summary

Degenerate normalized Stokes vectors:

$$\hat{\mathbf{S}}_{HLP} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{S}}_{VLP} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{S}}_{L+45\ \text{deg}} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

$$\hat{\mathbf{S}}_{L-45\ \text{deg}} = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{S}}_{RCP} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \hat{\mathbf{S}}_{LCP} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

Elliptical polarization—Jones and Stokes vectors:

$$\mathbf{E} = \begin{pmatrix} a_x \\ a_y e^{i\phi} \end{pmatrix} = \sqrt{a_x^2 + a_y^2} \begin{pmatrix} \cos\alpha \\ e^{i\phi} \sin\alpha \end{pmatrix}$$

$$\mathbf{S} = \begin{pmatrix} a_x^2 + a_y^2 \\ a_x^2 - a_y^2 \\ 2a_x a_y \cos\phi \\ -2a_x a_y \sin\phi \end{pmatrix} = \begin{pmatrix} S_0 \\ S_0\, 2\cos 2\alpha \\ S_0 \sin 2\alpha \cos\phi \\ -S_0 \sin 2\alpha \sin\phi \end{pmatrix} = S_0 \begin{pmatrix} 1 \\ \cos 2\varepsilon \cos 2\theta \\ \cos 2\varepsilon \sin 2\theta \\ -\sin 2\varepsilon \end{pmatrix}$$

Mueller matrix for diattenuation:

$$\mathbf{M}_D(T_x, T_y) = \frac{1}{2} \begin{pmatrix} T_x + T_y & T_x - T_y & 0 & 0 \\ T_x - T_y & T_x + T_y & 0 & 0 \\ 0 & 0 & 2\sqrt{T_x T_y} & 0 \\ 0 & 0 & 0 & 2\sqrt{T_x T_y} \end{pmatrix}$$

Equation Summary

Mueller matrix for horizontal and vertical linear polarizers:

$$\mathbf{M}_{HLP} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \mathbf{M}_{VLP} = \frac{1}{2}\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Mueller matrix for retardance:

$$\mathbf{M}_R(\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\phi & -\sin\phi \\ 0 & 0 & \sin\phi & \cos\phi \end{pmatrix}, \quad \mathbf{M}_{QWP} = \mathbf{M}_R(\pi/4) \quad \mathbf{M}_{HWP} = \mathbf{M}_R(\pi/2)$$

Mueller rotation matrix:

$$\mathbf{M}_{Rot}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta & 0 \\ 0 & -\sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Fresnel equations:

$$t_s = \frac{2N_0 \cos\theta_0}{N_0 \cos\theta_0 + N_1 \cos\theta_1} \qquad t_p = \frac{2N_0 \cos\theta_0}{N_0 \cos\theta_1 + N_1 \cos\theta_0}$$

$$r_s = \frac{N_0 \cos\theta_0 - N_1 \cos\theta_1}{N_0 \cos\theta_0 + N_1 \cos\theta_1} \qquad r_p = \frac{N_0 \cos\theta_1 - N_1 \cos\theta_0}{N_0 \cos\theta_1 + N_1 \cos\theta_0}$$

Brewster angle and critical angle:

$$\theta_B = \arctan\frac{n_1}{n_0} \quad \theta_C = \arcsin\frac{n_1}{n_0}$$

Equation Summary

Phase difference for reflection between parallel planes:

$$\delta = \frac{4\pi}{\lambda} n' d \cos\theta'$$

Thin films:

$$\eta_j^{TE} = y_{fs} N_j \cos\theta_j$$

$$\eta_j^{TM} = y_{fs} N_j / \cos\theta_j$$

$$\mathbf{M}_j = \begin{pmatrix} \cos\delta_j/2 & \dfrac{-i\sin\delta_j/2}{\eta_j} \\ -i\,\eta_j \sin\delta_j/2 & \cos\delta_j/2 \end{pmatrix}, \quad \delta = \frac{4\pi N_j}{\lambda} d_j \cos\theta_j$$

$$\begin{pmatrix} B \\ C \end{pmatrix} = \left(\prod_{j=1}^{m} \mathbf{M}_j\right) \begin{pmatrix} 1 \\ \eta_{sub} \end{pmatrix} = \begin{pmatrix} m_{11} + m_{12}\eta_{sub} \\ m_{21} + m_{22}\eta_{sub} \end{pmatrix}$$

$$\eta_{eq} = \frac{C}{B} \quad \rho = \frac{\eta_0 - \eta_{eq}}{\eta_0 + \eta_{eq}} \quad \tau = \frac{2\eta_0}{\eta_0 + \eta_{eq}}$$

Reflectance and transmittance of thin films:

$$R = \rho\rho^* = \left(\frac{\eta_0 - \eta_{eq}}{\eta_0 + \eta_{eq}}\right)\left(\frac{\eta_0 - \eta_{eq}}{\eta_0 + \eta_{eq}}\right)^*$$

$$T = \frac{\eta_{sub}\cos\theta_{sub}}{\eta_0\cos\theta_0}\tau\tau^* = \frac{\eta_{sub}\cos\theta_{sub}}{\eta_0\cos\theta_0}\frac{4\mathrm{Re}(\eta_0)\mathrm{Re}(\eta_{eq})}{(\eta_0 + \eta_{sub})(\eta_0 + \eta_{sub})^*}$$

Group velocity:

$$v_g = \frac{d\omega}{dk} = v_p + k\frac{dv_p}{dk} = v^p\lambda\frac{dv_p}{d\lambda}$$

Group index:

$$n_g = \frac{c}{v_g} = c\frac{dk}{d\omega} = n + w\frac{dn}{d\omega} = n - \lambda\frac{dn}{d\lambda}$$

Equation Summary

Fringe visibility or contrast:

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2} p_{12}$$

Grating equation—normal incidence:

$$\Lambda \sin\theta_m = m\lambda$$

Grating equation—standard mount:

$$\sin\theta_m - \sin\theta_i = m\lambda/\Lambda$$

Grating equation—general:

$$k'_x = k_x + m\frac{2\pi}{\Lambda}\cos\phi,\ k'_y = k_y + m\frac{2\pi}{\Lambda}\sin\phi$$

$$k'_z = \pm\sqrt{|k|^2 - k'^2_x} - k'^2_y$$

Fresnel number:

$$N_f = \frac{D^2}{4\lambda L} = \frac{D^2}{4\lambda}\left(\frac{1}{s_2} - \frac{1}{s_1}\right)$$

Fresnel diffraction formula:

$$U'(x',y') = \frac{e^{ikz}}{i\lambda z} e^{\frac{ik}{2z}(x'^2+y'^2)} \iint_{\Sigma} U(x,y) T(x,y) e^{\frac{ik}{2z}(x^2+y^2)} e^{-\frac{i2\pi}{\lambda z}(xx'+yy')} dx\,dy$$

Fraunhoffer diffraction formula:

$$U'(x',y') = \frac{e^{ikz}}{i\lambda z} e^{\frac{ik}{2z}(x'^2+y'^2)} \iint_{\Sigma} U(x,y) T(x,y) e^{\frac{i2\pi}{\lambda z}(xx'+yy')} dx\,dy$$

Equation Summary

Huygens' wavelet:

$$h_z^H(r) = -\frac{e^{ikr}}{2\pi r}\left(ik - \frac{1}{r}\right)\cos\theta = \frac{e^{ikr - \tan^{-1}kr}}{2\pi r}\sqrt{k^2 + 1/r^2}\cos\theta$$

Fresnel wavelet:

$$h_z^F(r) \approx \frac{1}{i\lambda}e^{ikz}e^{\frac{ik}{2z}(x^2 + y^2)}$$

Plane wave spectrum:

$$A(\xi,\eta;\, z_i) = \iint_{\infty} U(x,y;z_i)e^{i2\pi(\xi x + \eta y)}dx\,dy$$

$$U(x,y;\, z_i) = \iint_{\infty} A(\xi,\eta;z_i)e^{-i2\pi(\xi x + \eta y)}d\xi\,d\eta$$

$$A(\xi,\eta;\, z_f) = A(\xi,\eta;z_i)H(\xi,\eta;\; z_f - z_i)$$

$$U(x,y;\, z_f) = \iint_{\infty} A(\xi,\eta;z_f)e^{-i2\pi(\xi x + \eta y)}d\xi\,d\eta$$

$$H_z(\xi,\eta;\Delta z) = e^{i\frac{2\pi}{\lambda}\Delta z\sqrt{1-\lambda^2\xi^2-\lambda^2\eta^2}} = e^{i\frac{2\pi}{\lambda}\Delta z\sqrt{1-\alpha^2-\beta^2}} = e^{i\frac{2\pi}{\lambda}\gamma\,\Delta z}$$

Talbot distance:

$$z_T = \frac{2\Lambda^2}{\lambda}$$

Propagation through a lens (distance z from front focal plane to back focal plane):

$$U'(x',y') = \frac{e^{ik(2f-z)}}{i\lambda f}e^{\frac{ikz}{2f^2}(x'^2+y'^2)}\iint t_A(x,y)U(x,y)e^{-i\frac{2\pi}{\lambda f}(xx'+yy')}dx\,dy$$

Equation Summary

Propagation between conjugates of a double-telecentric system:

$$U_i = \frac{e^{ik(2f_2+z_i)}}{i\lambda f_2}\iint P\ U_p e^{-\frac{ikz_i}{2f_2^2}(\xi^2+\eta^2)} e^{-i\frac{2\pi}{\lambda f_2}(\xi x'+\eta y')} d\xi\, d\eta$$

Strehl ratio approximations:

$$SR \cong (1-\sigma_\phi^2/2)^2 \cong 1-\sigma_\phi^2 \cong e^{1-\sigma_\phi^2}, \quad \text{where} \quad \sigma_\phi^2 = \langle\phi^2\rangle - \langle\phi\rangle^2$$

Coherent cutoff frequency:

$$f_{cutoff}^{coherent} = \frac{a/2}{\lambda z_i} = \frac{1}{2\lambda(F/\#)} = \frac{NA}{\lambda}$$

Incoherent cutoff frequency:

$$f_{cutoff}^{incoherent} = \frac{a}{\lambda z_i} = \frac{1}{\lambda(F/\#)} = 2\frac{NA}{\lambda}$$

Gaussian beam:

$$u = -\frac{iA}{q} e^{i\frac{kr^2}{2|q|^2/z}} e^{\frac{kz_0 r^2}{2|q|^2}} = \sqrt{\frac{\Phi}{\pi\sigma^2}} e^{i\frac{kr^2}{2R}} e^{-\frac{r^2}{2\sigma^2}} e^{-i\theta}$$

$$q = z - iz_0 \qquad R = z + z_0^2/z \qquad A = k\sigma_0\sqrt{\Phi/\pi}$$
$$\sigma^2 = \sigma_0^2(1+z^2/z_0^2) \qquad z_0 = k\sigma_0^2 \qquad \tan\theta = z/z_0$$

ABCD method for Gaussian beams:

$$\begin{pmatrix} y' \\ \omega' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} y \\ \omega \end{pmatrix} = \mathbf{M}_n \ldots \mathbf{M}_2\mathbf{M}_1\begin{pmatrix} y \\ \omega \end{pmatrix} = \mathbf{M}\begin{pmatrix} y \\ \omega \end{pmatrix}$$

$$q' = \frac{Aq+B}{Cq+D}$$

Paraxial system matrix from two independent rays:

$$\mathbf{M} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} y_1' & y_2' \\ \omega_1' & \omega_2' \end{pmatrix}\begin{pmatrix} \omega_1 & -y_2 \\ -\omega_1 & y_2 \end{pmatrix} \Bigg/ \begin{vmatrix} y_1 & y_2 \\ \omega_1 & \omega_2 \end{vmatrix}$$

Equation Summary

Basic paraxial system matrices:

$$\mathbf{M}_T = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix} \quad \mathbf{M}_R = \begin{pmatrix} 1 & 0 \\ -\phi & 1 \end{pmatrix} \quad \mathbf{M}_F = \begin{pmatrix} 0 & 1/\phi \\ -\phi & 0 \end{pmatrix}$$

$$\mathbf{M}_C = \begin{pmatrix} m & 0 \\ -\phi & 1/m \end{pmatrix} \quad \mathbf{M}_N = \begin{pmatrix} n/n' & 1/\phi \\ -\phi & n'/n \end{pmatrix}$$

Coherence length:

$$\text{Two narrowband sources: } \Delta l = \frac{1}{2}\frac{c}{\Delta\nu} = \frac{1}{2}\frac{\lambda^2}{\Delta\lambda}$$

$$\text{Top-hat spectrum: } \Delta l = \frac{c}{\Delta\nu} = \frac{\lambda^2}{\Delta\lambda}$$

$$\text{Gaussian spectrum: } \Delta l = \frac{4}{\pi}\frac{c}{\Delta\nu} = \frac{4}{\pi}\frac{\lambda^2}{\Delta\lambda}$$

Coherence time:

$$\Delta t = \frac{\Delta l}{c}$$

Mutual coherence function:

$$\Gamma(r_1, r_2, t_1, t_2) = \left\langle |U(\mathbf{r}_1, t + t_1) U^*(\mathbf{r}_2, t + t_2)|^2 \right\rangle_t$$

Finesse:

$$\mathfrak{F} = \frac{\textit{fringe separation}}{\textit{FWHM}} \approx \frac{\pi}{2}\sqrt{F}$$

Coefficient of finesse:

$$F = \left(\frac{2r}{1 - r^2}\right)^2$$

Free spectral range:

$$\Delta\lambda_{fsr} \approx \frac{\lambda^2}{2nd}, \quad \Delta\nu_{fsr} = \frac{c}{\lambda}\frac{\Delta\lambda_{fsr}}{\lambda} = \frac{c}{2nd}$$

Equation Summary

Fourier transform and inverse Fourier transform:

$$F(\xi,\eta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-i2\pi(x\xi + y\eta)}dx\,dy$$

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\xi,\eta)e^{i2\pi(x\xi + y\eta)}d\xi\,d\eta$$

Central ordinate theorem:

$$F(0,0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dx\;dy \quad f(0,0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\xi,\eta)d\xi\,d\eta$$

Central moment theorem:

$$\int_{-\infty}^{\infty} x^n f(x)dx = \frac{1}{(-i2\pi)^n}\frac{\partial^n}{\partial\xi^n}F(\xi)\bigg|_{\xi=0}$$

Parseval's theorem:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(x,y)\,f_2^*(x,y)dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_1(\xi,\eta)\;F_2^*(\xi,\eta)d\xi\,d\eta$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f_1(x,y)|^2\,dx\;dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |F_1(\xi,\eta)|^2\;d\xi\,d\eta$$

Convolution:

$$f_1(x,y) * f_2(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(x',y')\,f_2(x-x',y-y')dx'\;dy'$$

Equation Summary

Special functions:

$$\text{rect}(x) = \begin{cases} 1 & |x| < 1/2 \\ 1/2 & |x| = 1/2 \\ 0 & |x| > 1/2 \end{cases} \quad \text{circ}(r) = \begin{cases} 1 & r < 1 \\ 0 & r \geq 1 \end{cases}$$

$$\text{sinc}(x) = \frac{\sin(x)}{x} \qquad \text{jinc}(x) = \frac{2J_1(x)}{x}$$

Delta function (a practical definition):

$$\delta(x - x_0) = 0, \quad x \neq x_0$$

$$\int_a^b \delta(x - x_0) f(x) dx = f(x_0), \quad a < x_0 < b$$

Delta function properties:

$$\delta(a\,x) = \frac{1}{|a|}\delta(x)$$

$$\delta(x - x_0) f(x) = \delta(x - x_0) f(x_0)$$

$$\delta(x - x_0)\delta(x) = 0, \quad x_0 \neq 0$$

Comb function:

$$\text{comb}(x) = \sum_{m=-\infty}^{\infty} \delta(x - m)$$

$$\text{comb}(x,y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(x - m, y - n) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(x - m)\delta(y - n)$$

Bibliography

Barrett, H. H. and K. J. Myers, *Foundations of Image Science*, John Wiley & Sons, Hoboken, New Jersey (2004).

Born, M. and E. Wolf, *Principles of Optics*, 7th ed. Cambridge University Press, London (1999).

Brosseau, C., *Fundamentals of Polarized Light : A Statistical Optics Approach*, John Wiley & Sons, New York (1998).

Collett, E., *Polarized Light: Fundamentals and Applications*, Marcel Dekker, New York (1993).

Gaskill, J. D., *Linear Systems, Fourier Transforms, and Optics*, John Wiley & Sons, New York (1978).

Goodman, J. W., *Introduction to Fourier Optics*, 3rd ed. McGraw-Hill, New York (2005).

Goodman, J. W., *Statistical Optics*, John Wiley & Sons, New York (1985).

Hecht, E., *Optics*, 4th ed. Addison-Wesley Longman, Reading, MA (2001).

Jenkins, F. A. and H. E. White, *Fundamentals of Optics*, 4th ed. McGraw-Hill, New York (2001).

Macleod, H. A., *Thin-Film Optical Filters*, 4th ed. Institute of Physics Publishing, Bristol (2010).

Mahajan, V. N., *Optical Imaging and Aberrations: Wave Diffraction Optics* (Part II), SPIE Press, Bellingham, WA, (2011).

Mansuripur, M., *Classical Optics and its Applications*, Cambridge University Press, New York (2002).

Pedrotti, F. L. and L. S. Pedrotti, *Introduction to Optics*, 3rd ed. Prentice Hall, Englewood Cliffs, NJ (2006).

Saleh, B. E. A. and M. C. Teich, *Fundamentals of Photonics*, 2nd ed. John Wiley & Sons, New York (2007).

Index

Index

Index

Index

Index

Daniel G. Smith has been with Nikon Research Corporation of America since 2004 where he works on a wide variety of optical engineering problems that usually require a good deal of physical optics to solve. He received a BS in Mathematics and Physics from the University of Nebraska at Omaha. He received a PhD in Optical Sciences from the University of Arizona College of Optical Sciences where he enjoyed the privilege of being a Kodak Fellow and recipient of the Arthur G. DeBell Memorial Scholarship. He also serves as co-chair of the Physical Optics conference within the SPIE Optical Systems Design symposium.

Physical Optics

Daniel G. Smith

This Field Guide is a practical overview of physical optics, with specific emphasis on information most useful in the field of optical engineering. Within this book, the reader will find formulae and descriptions of basic electromagnetic wave phenomena that are fundamental to a wave theory of light. Tools are provided for describing polarization, and emphasis is placed on scalar diffraction and imaging theory, which are essential in solving most optical engineering problems.

SPIE Field Guides

The aim of each *SPIE Field Guide* is to distill a major field of optical science or technology into a handy desk or briefcase reference that provides basic, essential information about optical principles, techniques, or phenomena.

Written for you—the practicing engineer or scientist—each field guide includes the key definitions, equations, illustrations, application examples, design considerations, methods, and tips that you need in the lab and in the field.

John E. Greivenkamp
Series Editor

P.O. Box 10
Bellingham, WA 98227-0010
ISBN: 978081948548-9
SPIE Vol. No.: FG17

www.spie.org/press/fieldguides

SPIE PRESS