Laser Light

Laser is an acronym for Light Amplification by Stimulated Emission of Radiation. The term light is used in a broad sense to include radiation at frequencies in the infra-red, visible or ultraviolet regions of the electromagnetic wave spectrum. In common parlance the term laser refers more to a device based on this principle than to the principle itself. The term laser action is often used when refering to the process. Lasers are devices that generate cohrerent light.

The physical principle ((stimulated emission) responsible for laser action was discovered by Albert Einstein in 1916. A device called MASER (microwave amplification by stimuleted emission of radiation) based on this principle was first operated in the microwave regime. The laser is an extension of this principle to the visible part of the electromagnetic spectrum. A summary of principal developments follows

1916: A. Einstein introduces stimulated emission as a fundamental process, in addition to already known absorption, and spontaneous emission, of light-matter interaction.

1960: Laser action was first observed in Ruby by T.H. Maiman in 1960 [Nature **187**, 493 (1960)]. It is now known to be one of the most difficult laser systems to operate.

1964: C.H. Townes, N.G. Basov and A.M. Prokhorov were awarded Nobel prize for their fundamental work in Quantum Electronics; Townes for demonstrating the ammonia (NH_3) maser and subsequent work in masers and lasers and Basov and Prokhorov for contributing to the development masers and lasers.

1971: Dennis Gabor was awarded Nobel prize in physics for his invention and development of the holographic method.

1981: N. Bloembergen and Arthur Schawlow were awarded Nobel prize in physics for their contributions to masers, nonlinear optics and spectroscopy.

1989 Norman F. Ramsey for the invention of the separated oscillatory fields method and its use in the hydrogen maser and other atomic clocks and Hans G. Dehmelt and Wolfgang Paul for the development of the ion trap technique.

1997: Steven Chu, Claude Cohen-Tannoudji and William D. Phillips for development of methods to cool and trap atoms with laser light.

2001: Eric A. Cornell, Wolfgang Ketterle and Carl E. Wieman for the achievement of Bose-Einstein condensation in dilute gases of alkali atoms, and for early fundamental studies of the properties of the condensates.

2006: J. Hall and T. W. Hänsch were awarded the Nobel prize for their contributions to the development of laser-based precision spectroscopy, including the optical frequency comb technique and to Roy Glauber for his contribution to the quantum theory of optical coherence.

There are three essential elements of a laser:

1. A gain or amplifying medium consisting of atoms, molecules, ions, or charged carriers in a semiconductor medium along with a pumping

mechanism to excite these species to their higher quantum mechanical states;

2. A suitable arrangement of optical elements to allow a single passage of light through the gain medium as in a laser amplifier or an optical resonator for repeated passage as in a laser oscillator; and

3. A loss mechanism to extract energy from the device. Absorption, diffraction, scattering, and transmission of light through mirrors contribute to this mechanism.

A coupling of these three elements via the electromagnetic field leads to laser action. These three elements come in a great variety of forms and shapes and often form the basis for classifying lasers.

1.1 Laser Light Characteristics

The light output from a laser is electromagnetic radiation and is not fundamentally different from the light emitted by other sources of electromagnetic radiation. There are, however, several important differences in detail between laser light and the light emitted by thermal sources. The output beam produced by lasers have much more in common with the output of conventional low-frequency electronic oscillators than they do with any kind of thermal light sources. We will briefly review laser beam characteristics that distinguish them from other light sources. The numbers stated below for lasers should not be taken as the final word. They change as progress is made in improving the performance of lasers.

Monochromaticity

The light emitted by a laser has high degree of spectral purity. This means it has relatively well-defined frequency or wavelength so that the output signal from an ideal laser is very nearly a highly stable, constant amplitude single-frequency sine wave. Two factors contribute to the narrowness of laser linewidth. First, only an EM wave with frequency within a certain range of the atomic transition frequency $\omega_0 = (E_2 - E_1)/\hbar$ can be amplified. This range of frequencies is called linewidth.

Second, the laser cavity forms a resonant system.This means oscillation can occur only at the resonance frequencies of this cavity. This leads to the further narrowing of the laser linewidth, which can be as large as 10 orders of magnitude!

The spectral width of a good quality single-frequency laser ranges from $\Delta\nu = 10^6$ Hz for moderatley stabilized lasers to $\Delta\nu = 10$ Hz or less in well-stabilized lasers. Since visible light has frequencies of order $\nu = 5 \times 10^{14}$ Hz, the spectral purity of laser light is $\Delta\nu/\nu = 2 \times 10^{-13} \equiv \frac{1}{Q}$. This ratio is a measure of the quality Q of the laser oscillator. Such large Q values are difficult to achieve in mechanical or electronic oscillators.

Table (1.1) shows a comparison of the spectral purity of light emitted by different sources of light [FWHM=Full width at half maximum].

Light Source	Center Wavelength	FWHM ($\Delta\lambda$)	FWHM ($\Delta\nu$)
He: Ne Laser	633 nm	10^{-8} nm	7.5×10^3 Hz
Cadmium low pressure lamp	644 nm	10^{-3} nm	9.4×10^8 Hz
Ordinary discharge (sodium discharge)	590 nm	0.1 nm	9×10^{10} Hz

Table 1.1: A comparison of spectral purity of different light sources

Thermal sources such as the sun and incandescent solids generally emit a broadband spectrum of light. There are, however, some thermal sources such as discharge lamps, that emit only a few spectral lines or narrow bands of wavelengths, but the spectral widths of the light emitted by even the best such sources are still limited by the linewidths of the atomic transitions in the discharge which range from $10^8 - 10^{11}$ Hz.

The ultimate limit on laser spectral purity is set by quantum noise fluctuations due to spontaneous emission from the atoms in the gain medium. This limit, however, can be reached with great difficulty only on the very best and highly stabilized lasers.

Source	$\Delta\nu$(Hz)	$\Delta\nu/\nu$
Broadband dye or semiconductor materials	$(10 - 100) \times 10^9$	10^{-2}
Atomic transitions in gases	10^9	10^{-6}

The primary reason for the spectral purity of laser light is the optical resonator (optical cavity) and not the atomic transition. The laser cavity permits continuous oscillations only at certain discrete frequencies (cavity resonance frequencies). The atoms serve primarily to provide gain at these resonance frequencies.

Coherence

The amplitude and the phase of the sine wave from a laser oscillator will in fact change slowly over space and time. Coherence refers to the property of a light wave that its phase and amplitude at one space-time point may be correlated to its phase and amplitude at some other space and time point. There are two types of coherences:

1. Temporal coherence refers to strong correlations between the amplitude and/or phase of the signal at different times. Laser light has high degree of temporal coherence. The amplitude and phase of the output sine wave at any one instance are strongly correlated with their values at some other instances over a time interval τ_c that spans millions of optical cycles. For stationary beams, the time interval τ_c, called coherence time, is of the order

$$\tau_c \approx \frac{1}{2\pi\Delta\nu} \tag{1.1}$$

The distance traversed by light in one coherence time τ_c is called the (longitudinal) coherence length

$$\ell_c = c\tau_c\,. \tag{1.2}$$

Coherence time τ_c may be thought of as the interval over which the amplitude and phase of the sine wave may be considered constant. Longitudinal coherence ℓ_c may then be considered as the average length of perfect sine waves emitted by the laser. For lasers ℓ_c can range from 300 m to 10^4 m or even larger.

2. Spatial coherence refers to correlation between laser field at different points in a plane transverse to the direction of wave propagation. At the output of a laser oscillator, laser light has almost perfect transverse spatial coherence. For thermal sources transeverse spatial coherence does not extend over distances much large than a few wavelengths. Transverse spatial coherence of laser light is also a consequence of the presence of a laser cavity.

Source	τ_c (s)	ℓ_c (m)
He:Ne laser	$10^{-4} - 10^{2}$	$3 \times 10^{4} - 3 \times 10^{10}$
Sodium lamp	10^{-10}	3×10^{-2}

Directionality/Collimation

Thermal sources emit light in random directions over a broad wavelength range. We can capture some fraction of this radiation and collimate it with a lens or mirror as in a searchlight or flashlight. The resulting degree of collimation (amount of radiation emitted per unit solid angle) is still much smaller than that for a laser. Consequently, thermal beams spread very rapidly with propagation.

A single-transverse mode laser oscillator, on the other hand, can produce a beam that can propagate for sizable distances with very little diffraction spread. The angular divegence (in radians) of a laser beam in the far zone is given by

$$\theta \approx \frac{\lambda}{\pi w_0} \tag{1.3}$$

where w_0 is the radius of laser beam spot at its waist (where the beam has narrowest transverse size). The distance over which this beam stays approximately collimated before diffraction spreading significantly increases is given by

$$b \approx \frac{2\pi w_0^2}{\lambda}\,. \tag{1.4}$$

This is a direct consequence of the fact that laser beam comes from a resonant cavity, and only waves propagating along the optical axis can be sustained in the cavity.

For laser light of wavelength $\lambda = 1.06 \times 10^{-3}$ mm, $w_0 = 3$ mm, then $\theta = \lambda/\pi D = \frac{1.06\times10^{-6}}{\pi 3\times10^{-3}} = 1.1 \times 10^{-4}$ rad=0.006^o . Compare this value with a normal flashlight, the divergence is about 25^o , a searchlight has a divergence angle of 10^o , the high directionality of laser light is obvious.

For a small Helium Neon (He:Ne) laser emiting at $\lambda = 633$ nm, the beam waist might be $w_0 \approx 0.5$ mm. This corresponds to an angular divergence of $\theta \approx \frac{0.633\times10^{-6}}{\pi 0.5\times10^{-3}} = 4 \times 10^{-4}$ radians=0.02^o and a collimation distance of $b = 2.3$ m.

For an Argon-ion (Ar-ion) laser at 514 nm, the waist might be 5 mm corresponding to angular divergence of $\theta = 3 \times 10^{-5}$ and a collimation distance of $b = 310$ m.

Laser Beam Focusing

A laser beam can also be focused by a lens to a small spot only a few laser wavelengths in diameter. The diameter of the focused spot is given by the formula

$$w \approx \frac{\lambda}{\pi w_o} \times f \tag{1.5}$$

where f is the lens focal length. If the laser beam fills the lens aperture, the ratio f/w_0 is simply the f-numebr of the lens. For best lenses this number is of order one. It follows that a laser beam can be focused to spots which are only a few wavelngth in diameter.

The directionality of laser beam is also a consequence of the presence of a resonator.

Brightness

Brightness of a source is defined as the power efflux (emitted) per unit surface area into unit solid angle. Its units are W/m^2·Sr. Spectral brightness is power emitted per unit area of the source per unit solid angle per unit bandwidth. Its units are W/m^2·Sr·Hz. The idea of brightness can be understood by considering a source that emits through a surface area $\mathcal{A}$. Each area element can emit light into a solid angle 2π steradian. Then, if a surface element $\delta\mathcal{A}$ emits power δP into a solid angle $\delta\varOmega$, the brightness of the source is given by

$$B = \frac{\delta^2 P}{\delta\mathcal{A}\delta\varOmega}, \qquad [B] = \mathrm{W/m^2{\cdot}Sr}$$

The power emitted by a black body at temperature T is given by Stefan-Boltzmann law to be

$$I = \sigma T^4, \qquad [I] = \mathrm{W/m^2}$$

For the sun with T=6000 K this gives an efflux of

$$I_{\mathrm{sun}} = 5.6705 \times 10^{-8}(6 \times 10^3)^4 = 7 \times 10^7\,\mathrm{W/m^2}$$

A small He:Ne laser of modest power $P = 1$ mW and beam waist of 0.5 mm will produce an efflux

$$I_{\mathrm{laser}} = \frac{P}{\pi w_o^2} = \frac{10^{-3}}{\pi(0.5 \times 10^{-3})^2} = 1.3 \times 10^3\,\mathrm{W/m^2}$$

This looks even more impressive when we take into accunt the directionality. The efflux from the sun is emitted, isotropically, into a solid angle of 2π steradian (Sr) whereas, because of the directionality of the laser beam, the laser emits it into a solid angle $\pi\theta^2 = \pi(\lambda/\pi w_o)^2 = \lambda^2/\pi w_o^2$. This leads to brightness

$$B = \frac{P}{\pi w_0^2 \pi\theta^2} = \frac{P}{\lambda^2} = \frac{10^{-3}}{(0.6328 \times 10^{-6})^2} = 2.5 \times 10^8 \, \mathrm{W/m^2 \cdot Sr} \, .$$

This exceeds the brightness of the sun. This comparison looks even more impressive if we examine the spectral brightness. The sun radiates like a blackbody over wide spectral bandwidth. Let us compare the spectral brightness of the sun near the peak (yellow) of visible spectrum. The spectral energy density [energy per unit volume per unit bandwidth ($\mathrm{J/m^3 \cdot Hz}$)] of a black-body is given by

$$\rho(\nu) = \frac{8\pi\nu^2}{c^3}\frac{h\nu}{e^{\beta h\nu} - 1} , \quad \beta = \frac{1}{k_{_B} T} \, . \tag{1.6}$$

Each surface area element of blackbody radiates isotropically into the solid angle 4π. Then its spectral brightness will be

$$B_\nu = \frac{1}{4}\frac{\rho(\nu)c}{\delta\Omega} = \frac{1}{4}\left[\frac{8\pi\nu^2}{c^3}\frac{h\nu}{e^{\beta h\nu} - 1}\right]\frac{c}{2\pi} = \frac{2\nu^2}{c^2}\left[\frac{h\nu}{e^{\beta h\nu} - 1}\right]$$

where we have used the fact that the radiation escaping the radiator is distributed over the solid angle 2π. The factor of $\frac{1}{4}$ accounts for two factors. First note that only half of the radiation is propagating outward from the black body. Furthermore, since each element radiates isotropically (into and out of the source) only half of the radiation incident on each area element actually escapes the black-body. For the sun at $T \approx 6000$ K and

$h\nu = 2.5$ eV (yellow light) we have

$$kT = k \cdot 300 \frac{T}{300} = \frac{1}{40} \times 20 \approx \frac{1}{2}\,\text{eV}$$
$$\frac{h\nu}{kT} \approx 5\,, \quad e^{\beta h\nu} = e^5 \approx 150$$
$$B_\nu = \frac{2}{(633 \times 10^{-9})^2} \frac{2.5 \times 1.6 \times 10^{-19}}{150 - 1} \approx 2 \times 10^{-8}\,\text{W/m}^2\text{-Sr-Hz}$$
$$= 2 \times 10^{-12}\,\text{W/cm}^2\text{.Hz}\,.$$

For a l mW He:Ne laser ($\lambda = 633$ nm) of spectral width $\Delta\nu = 10^4$ Hz and spot size 0.5 mm, the spectral brightness will be

$$B_\nu = \frac{P}{\lambda^2 \Delta\nu} = \frac{1 \times 10^{-3}}{(0.633 \times 10^{-6})^2 \times 10^4} \approx 2.5 \times 10^5\,\text{W/m}^2\cdot\text{Sr}\cdot\text{Hz}$$
$$= 25\,\text{W/cm}^2\cdot\text{Sr}\cdot\text{Hz}$$

For a Nyodymium glass (Nd-glass) laser with a power of $P = 10^4$ MW, $\lambda = 1.06\,\mu$, and bandwidth limited by pulse duration of 30 ps, we have

$$\Delta\nu = \frac{1}{2\pi\tau_p} = \frac{1}{2\pi \times 3 \times 10^{-11}} \approx 5 \times 10^9\,\text{Hz}$$
$$B_\nu = \frac{1 \times 10^4 \times 10^6}{(10^{-6})^2 \times 5 \times 10^9} \approx 2 \times 10^{12}\,\text{W/m}^2\text{.Sr.Hz}$$
$$= 2 \times 10^8\,\text{W/cm}^2\text{.Sr.Hz}$$

Laser Performance Records

1. Wide power range : Continuous wave (CW) powers of upto hundreds of kilowatts are available from certain IR chemical lasers. In pulsed mode peak powers in excess of 10^{13} watts, which exceeds the total electrical power generated in the U.S. for very short times (pico-seconds) are available.

2. Extreme frequency stability: The short term frequency stability of a highy stabilized laser can be as good as 1:10^{13}. A He:Ne laser operating at 3.39 μm stabilized against a methane absorption line has absolute reproducibility of 1 part in 10^{10}. Frequency stabilities of $\Delta\nu = 10^{-2}$ Hz have been achieved.

3. Wide tunability : Most common lasers are limited to sharply defined discrete frequencies. However, widely tunable sources of coherent radiation including dye lasers, titanium-doped sapphire (Ti-sapphire) lasers and optical parametric oscillators (OPOs) can provide tunable radiation over bandwidths of order $\Delta\lambda \approx 200$ nm. This corresponds to a frequency bandwidth of $\Delta\nu \approx 24 \times 10^{12}$ Hz.

4. Ultra-short pulses : Mode-locked laser pulses shorter than 1 ps (picosecond) are routine. Mode-locked compressed dye laser pulses of only a few femtosecond long (FWHM) (few optical cycles) have been produced.

5. Very efficient: Power conversion efficiency is defined to be the ratio of optical power radiated by the laser to the power supplied to operate the laser. Power conversion efficiencies of lasers range from 0.001 to 0.1% for gas lasers, 1 to 2 % for solid state lasers and 50-70% for carbon dioxide (CO_2) and semiconductor lasers.

In computing the values of various physical quantities we will find the following table of constants useful.

Constant Name	Usual Symbol	Current Value
Speed of light	c	$2.997\,924\,58 \times 10^{8}$ m/s
Elementary charge	e	$1.602\,177 \times^{-19}$ C
Planck's constant	$\hbar$	$1.054\,571\,628(53) \times 10^{-34}$
Gravitational constant	G	$6.67428(67)\,6 \times 10^{-11}$ N·
Fine structure constant	$\alpha = e^2/4\pi\epsilon_0\hbar c$	$1/137.036 \approx 1/137$
Stefan-Boltzmann constant	σ	5.6705×10^{-8} W/m^2·K^4
Boltzmann constant	k_B	$1.380\,7 \times 10^{-23}$ J/K
Avogadro's number	N_A	$6.022\,137 \times 10^{23}$ mol^{-1}
Solar mass	M_e	1.989×10^{30} kg
Solar radius	R_e	6.96×10^{8} m
Earth mass	M_E	5.976×10^{24} kg
Earth radius	R_E	6.374×10^{6} m
Earth-Sun distance (mean radius)	AU	1.496×10^{11} m
Earth-Moon distance (mean radius)		3.844×10^{8} m
Volume of a mole of an ideal gas at STP		22.414 liter
Air desnity at STP		1.293 kg/m^3
Triple point temperature of water		273.16 K

In practice, it is more convenient and useful to remember certain combinations of fundamental constants rather than the constants themselves.

Conversions

Fine structure constant $\alpha = \dfrac{e^2}{4\pi\epsilon_0\hbar c}$	$1/137.036 \approx 1/137$
$\hbar c$	197.3271 eV· nm
k_BT	1/40 eV at 293K and 1/39 eV at 300K
Electron rest energy (mass) m_ec^2	0.5110 MeV
Proton rest energy (mass) m_pc^2	938.28 MeV
Neutron rest energy m_nc^2	939.57 MeV
Proton-electron mass ratio m_p/m_e	1836.15
Bohr radius $a_0 = \hbar/m_ec\alpha$	0.5292×10^{-10} m
Planck time $\sqrt{\hbar G/c^5}$	5.4×10^{-44} s
Compton wavelength of electron $\hbar/m_ec$	3.8616×10^{-13} m
1 calorie	4.1860 J
1 atm	1.1013×10^5 Pa
1 degree	1.745×10^{-2} rad
1 parsec	3.09×10^{16} m

Energy of a photon can be calculated by using the formula

$$E = h\nu = \frac{2\pi\hbar c}{\lambda} = \frac{1238}{\lambda(\text{in nm})}\,\text{eV}$$

1.2 Survey of Laser Elements

Gain media and pumps come in great variety of forms and generate radiation ranging from far infra-red (far-IR) to soft X-rays. Important laser gain media and wavelengths include
HCN far-IR laser (311, 337, 545, 676, 744 μm)
H_2O far-IR laser (28,48, 120 μm)
CO_2 laser (9.6-10.6 μm)
CO laser (5.1-6.5 μm)
HF chemical laser (2.7-3.0 μm)

Nd:YAG laser (1.06 μm)
He:Ne laser (1.15 μm, 633 nm)
Ga-As semiconductor laser (870 nm)
Ruby laser (694 nm)
Rhodamine 6G dye laser (560-640 nm)
Argon-ion laser (488-514 nm)
Pulsed N_2 discharge laser (337 nm)
Pulsed H_2 discharge laser (160 nm)

Laser pumping methods include

Gas discharge including dc, rf, and pulsed electrical discharges involving both direct electron excitation and two-stage collision pumping.

Optical pumping using flash lamps, arc lamps, semiconductor LEDs, otgher lasers and even direct sunlight.

Chemical reactions including chemical mixing, photolysis, and combustion.

Direct electrical pumping includes high-voltage electron beams directed into high-pressure gas cells and direct current injection into semiconductor injection lasers.

Nuclear pumping of gases by nuclear fission fragments when a gas laser tube is placed in close proximity of a nuclear reactor.

Supersonic expansion of gases, usually pre-heated by chemical reaction or electrical discharge, through supersonic expansion nozzles, to create the so-called gas-dynamic lasers.

Plasma pumping in hot dense plasmas, created by plasma piches, focused high-power laser pulses, or electrical pulses. There are reports of X-ray laser action in some laser materials pumped by the explosion of a nuclear bomb.

Resonators store electromagnetic energy. At microwave frequencies these resonators are closed metallic boxes but at optical frequencies they can be open. The need for open resonators arises because we want only a few modes interacting with the atoms to grow[1]. The presence of boundaries in resonators (boundary conditions) allows only certain field configurations (frequency and spatial variation) to exist inside the resonators. These allowed field configurations are called modes of the resonator. For open resonators modes corresponding to propagation away from the resonator axis will have very large losses because any light emitted into such modes will be quickly lost. Only a group of paraxial modes where energy is localized near the axis will experience build up. Even these modes will experience losses due to diffraction, absorption, and transmission by the end mirrors. As a result mode definition in the sense of stationary field configurations can not be used. However, modes with quasi-stationary (long lived) patterns do exist. These modes experience very little losses in one cavity round trip so that any energy emitted into these modes will remain in the cavity for a long time.

These long lived modes then interact with the gain medium and extract energy from the gain medium. The gain medium can either emit light spontaneously into many of the modes - long or short lived - or it can be stimulated to emit into a particular mode by the light energy stored in it. Stimulated emission of light into a mode increases as the light energy of a mode increases. Clearly the long lived modes have the best chance for amplification of light energy if they can overcome losses. Thus it is the competition between gain due to stimulated emisssion and loss due to

[1]At optical frequencies the density (# per unit frequency band per unit volume of the resonator) of modes near a frequencvy ν is very large $8\pi\nu^2/c^3$. If the atoms interacts with modes in a band $\delta\nu$ (typically 10^9 Hz or more), their number $8\pi\nu^2\delta\nu/c^3$ is enormous. Such a situation is not conducive to the growth of any mode amplitude to a significant level.

diffraction and transmission in one or more of these long lived modes that determines whether light amplification by stimulated emission of radiation can occur or not. If gain exceeds loss laser action can occur. Although the loss is detrimental to achieving laser action, not all loss is undesirable. In fact the loss of stored light energy through end mirror transmission is what emerges as a beam of light which has made the laser into such a useful tool that hardly any aspect of our life has remain untouched by it.

An understanding of lasers will require us to study an interacting matter and light system inside a resonator. It is an open system as energy can be added to or extracted from the system. This may seem like a formidable problem. We will see, however, that this seemingly complex problem can be dealt with quite effectively.

We will first study atoms and light separately by ignoring their interaction and then couple the two. The atoms are described by Schrodinger equation. Light, being an electromagnetic wave phenomenon, is described by Maxwell's equations. Its propagation as rays or waves, diffraction, and inteference, all follow from Maxwell's equations. These equations admit wave-like solutions. Plane waves are the best known of these solutions. For lasers we require beam like wave solutions which, like plane waves, have a pre-dominant direction of propagation but they have finite extent in directions perpendicular to the direction of propagation. In view of the remarks in the preceding paragraph one might suspect that these solutions are the right type to fit the boundary conditions imposed by open optical resonators. We will see that this is indeed true.

1.3 Electromagnetic Waves in Homogeneous Media

We will use the complex analytic representation of the fields so that the real physical fields are given by

$$F(\vec{r},t) = \Re\mathcal{F}(\vec{r},t)\,, \tag{1.7}$$

where $F(\vec{r},t)$ represents any of the components of the fields. In a homogeneous transparent isotropic medium chracterized by dielectric permittivity ϵ and magnetic permeability μ the constitutive relations are simple proportionalities $\vec{\mathcal{D}}(\vec{r},t) = \epsilon\vec{\mathcal{E}}(\vec{r},t)$ and $\vec{\mathcal{B}}(\vec{r},t) = \mu\vec{\mathcal{H}}(\vec{r},t)$. These relations hold for arbitrarily rapid variation of the fields as long as the most significant part of the field spectrum lies in the transparency range of the medium. Maxwell's equations then read

$$\vec{\nabla}\cdot\vec{\mathcal{E}}(\vec{r},t) = 0 \tag{1.8}$$

$$\vec{\nabla}\cdot\vec{\mathcal{B}}(\vec{r},t) = 0 \tag{1.9}$$

$$\vec{\nabla}\times\vec{\mathcal{E}}(\vec{r},t) = -\frac{\partial\vec{\mathcal{B}}(\vec{r},t)}{\partial t} \tag{1.10}$$

$$\vec{\nabla}\times\vec{\mathcal{B}}(\vec{r},t) = \mu\epsilon\frac{\partial\vec{\mathcal{E}}(\vec{r},t)}{\partial t} \tag{1.11}$$

These are coupled first-order partial differentail equations. By eliminating the magnetic (electric) field we can obtain a closed equation for the electric (magnetic) field. For example, on taking the curl of Eq. (1.10) we obtain

$$\vec{\nabla}\times(\vec{\nabla}\times\vec{\mathcal{E}}) = -\frac{\partial}{\partial t}\vec{\nabla}\times\vec{\mathcal{B}}\,.$$

Using the identity $\vec{\nabla}\times(\vec{\nabla}\times\vec{\mathcal{E}}) = \vec{\nabla}(\vec{\nabla}\cdot\vec{\mathcal{E}}) - \nabla^2\vec{\mathcal{E}}$ together with $\vec{\nabla}\cdot\vec{\mathcal{E}} = 0$ and eliminating $\vec{\mathcal{B}}$ with the help of Eq. (1.11), we find that the electric field obeys the equation

$$\left[\nabla^2 - \mu\epsilon\frac{\partial^2}{\partial t^2}\right]\vec{\mathcal{E}}(\vec{r},t) = 0. \tag{1.12}$$

Similarly, by eliminating $\vec{\mathcal{E}}$ we find that the magnetic field $\vec{\mathcal{B}}$ also satisfies the same equation. The quantity $1/\mu\epsilon$ has the dimensions of square of a speed. This speed v is

$$v = \sqrt{\frac{1}{\mu\epsilon}} = \sqrt{\frac{1}{\mu_0\epsilon_0}}\sqrt{\frac{\mu_0\epsilon_0}{\mu\epsilon}} = \frac{c}{n} \tag{1.13}$$

where c and n are, respectively, the speed of light in free space and the refractive index of the medium,

$$c = 2.99792458 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s}. \tag{1.14}$$

$$n = \sqrt{\frac{\mu\epsilon}{\mu_0\epsilon_0}}\,. \tag{1.15}$$

Equation (1.12) is then readily identified as homogeneous (source free) wave equation with v as the wave speed.

1.4 Solutions of the Wave Equation

To gain some insight into these new wave-like solutions of Maxwell's equations, consider the scalar wave equation for fields that depend only on a single spatial variable

$$\left(\frac{\partial^2}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)\mathcal{F}(z,t) = 0\,. \tag{1.16}$$

where $\mathcal{F}$ stands for any of the field components. By means of the change of variables $\xi = t - z/v$ and $\eta = t + z/v$, the derivatives in the wave equation can be transformed to

$$\frac{\partial^2}{\partial z^2} = \left(-\frac{1}{v}\frac{\partial}{\partial\xi} + \frac{1}{v}\frac{\partial}{\partial\eta}\right)\left(-\frac{1}{v}\frac{\partial}{\partial\xi} + \frac{1}{v}\frac{\partial}{\partial\eta}\right), \tag{1.17}$$

$$\frac{1}{v^2}\frac{\partial^2}{\partial t^2} = \frac{1}{v^2}\left(\frac{\partial}{\partial\xi} + \frac{\partial}{\partial\eta}\right)\left(\frac{\partial}{\partial\xi} + \frac{\partial}{\partial\eta}\right). \tag{1.18}$$

In terms of ξ and η the wave equation becomes

$$\frac{1}{v^2}\frac{\partial^2}{\partial\xi\partial\eta}\mathcal{G}(\xi,\eta) = 0\,. \tag{1.19}$$

where $\mathcal{G}(\xi,\eta)$ is the function $\mathcal{F}(z,t)$ expressed in terms of ξ and η. The solution to this equation are of the form

$$\mathcal{G}(\xi,\eta) = K_1 f(\xi) + K_2 g(\eta), \tag{1.20}$$

where K_1 and K_2 are some constants and $f(\xi)$ and $g(\eta)$ are the two independent solutions of Eq. (6.10). Transforming back to z and t we find the solutions to the wave equation are of the form

$$\mathcal{F}(z,t) = C_1 f(t - z/c) + C_2 g(t + z/c)\,, \tag{1.21}$$

where C_1 and C_2 are constants. We see that $\mathcal{F}(z,t)$ is not an arbitrary function of z and t but in it the space and time variables occur in the combination $t - z/c$ or $t + z/c$. Examples of acceptable waveforms are periodic functions such as $e^{-i\omega(t \pm z/v)}$ or functions of finite duration such as $e^{-a(t \pm z/v)^2/2}$, where ω and a are constants having dimensions of frequency and frequency squared, respectively.

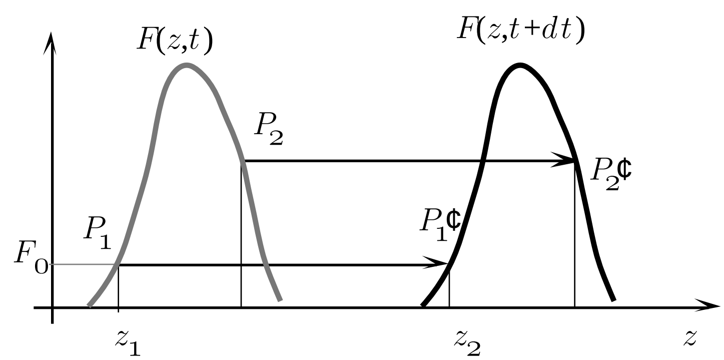


Figure 1.1: Pulse profile at times t and $t + dt$ for $\mathcal{F}(z,t) = f(t - z/v)$.

The function $f(t - z/v)$ represents a distutrbance propagating in the $+z$ direction with speed v and $g(t + z/v)$ represents a disturbance propagating in the $-z$ direction with speed v. To see this consider a real valued disturbance $F(z,t) = f(t - z/c)$ in the form of a pulse consisting only of the first term. A plot of $F(z,t)$ as function of z when t is held fixed represents pulse shape or profile. Figure (1.1) shows a profile of this pulse at times t. To see what happens to this pulse at a later time, consider a point, such as P_1, where the pulse amplitude has a value F_0. At a later time $t + dt$, the pulse will have this same value F_0 at point z_2 such that

$F(z_2, t+dt) = F_0 = f(z_1, t)$. For this to happen the argument of f at time t and $t + dt$ must be the same. This means $t - z_1/v = t + dt - z_2/v$ or $z_2 = z_1 + vdt$. Hence at time t, the disturbance has value F_0 at position z_1 and at time $t + dt$ it has the same value at point $z_2 = z_1 + vdt$. If dt is an infinitesimal interval, z_2 and z_1 will differ by an infinitesimal amount dz so that $dz = vdt$. Similar considerations for other points on the pulse profile at time t show that at time $t+dt$ the pulse has moved to the right, without change of shape, by an amount $dz = vdt$. It is clear that the disturbance propagates with speed $v = dz/dt$ in the positive z-direction. Similar considerations show that $F(z, t) = g(t + z/v)$ represents a wave propagating in the $-z$ direction with constant speed v. We have considered a simple case where wave propagates without change of shape. Propagation with change of pulse shape is possible when the speed of the wave depends on frequency.

In general, a wave propagating in the direction specified by the unit vector $\hat{k}$ is given by

$$F(\vec{r}, t) = C_1 f(t - \hat{k} \cdot \vec{r}/v) + C_2 g(t + \hat{k} \cdot \vec{r}/v) \tag{1.22}$$

The term plane wave is used for such solutions because the surfaces of constant wave amplitude, called wavefronts, are planes. For example the wavefronts for $\mathcal{F}(\vec{r}, t) = Cf(t-\hat{k}\cdot\vec{r}/v)$ are given by $t-\hat{k}\cdot\vec{r}/v = const$, which for different values of the constant defines a family of planes perpendicular to $\hat{k}$. A wavefront corresponding to a definite value $t - \hat{k} \cdot \vec{r}/v = C_0$ moves with velocity given by $d\vec{r}/dt = v\hat{k}$. If an energy density is associated with the modulus squared of $\mathcal{F}(\vec{r}, t) = Cf(t - \hat{k} \cdot \vec{r})$, transport of energy occurs along trajectories called rays, which for the waves considered here are straight lines parallel to $\hat{k}$.

It is possible to have other types of scalar solutions. For example solutions of the form $\mathcal{F}(\vec{r}, t) = \mathcal{F}(r, t)$ exist which satisfy

$$\left(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)\mathcal{F}(r, t) = 0. \tag{1.23}$$

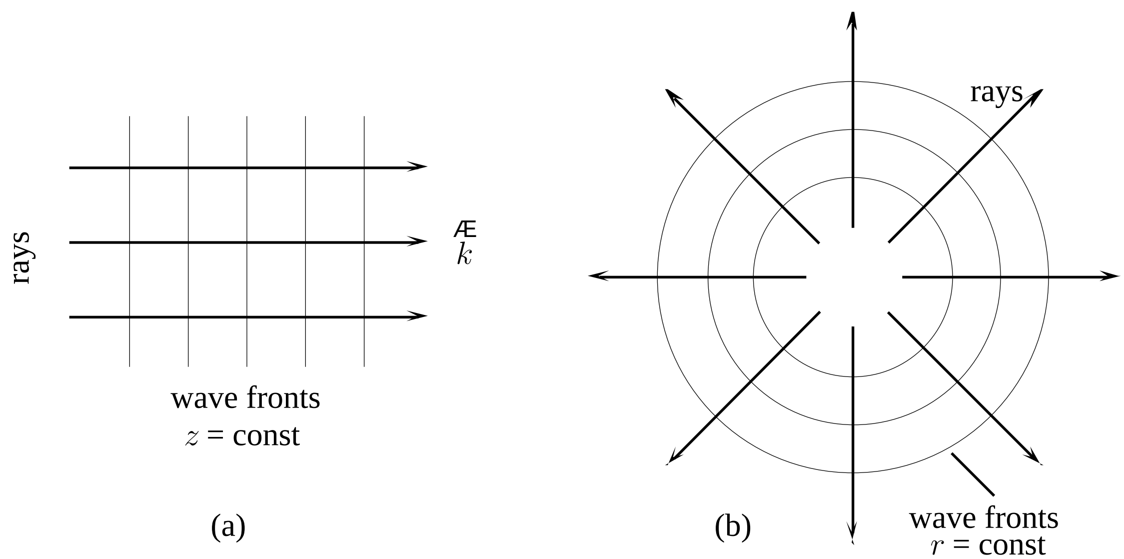


Figure 1.2: Wavefronts and rays for (a) plane and (b) spherical waves.

Since $\mathcal{F}$ is assumed to depend only on r and t, we can use the identity $\nabla^2\mathcal{F}(r) = (1/r)(\partial^2/\partial r^2)r\mathcal{F}$ in source free regions. Then the wave equation reduces to

$$\left(\frac{\partial^2}{\partial r^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) r\mathcal{F}(r,t) = 0 \tag{1.24}$$

Noting the similarity of this equation with the one-dimensional wave equation (1.16), its solution can be written down at once as $r\mathcal{F}(r,t) = C_1 f(t - r/v) + C_2 g(t + r/v)$. This can be rewritten in the form

$$\mathcal{F}(r,t) = C_1\frac{f(t - r/v)}{r} + C_2\frac{g(t + r/v)}{r}\,. \tag{1.25}$$

Here the first term represents a spherically diverging wave and the second term represents a spherically converging wave. Note that for $\mathcal{F}(r,t) = C_1 f(t - r/v)/r$ the surfaces of constant wave amplitude (wavefronts) are spheres centered at the origin and energy transport occurs along radial lines diverging from the origin. The wave amplitude falls off as $1/r$ so that the energy of the wave as it propagates remains constant. More complex scalar spherical wave solutions which behave like outgoing or incoming waves far from the origin are

$$\mathcal{F}(\vec{r},t) = \begin{cases} \text{const} \times \; h_\ell^{(1)}(kr) Y_\ell^m(\theta,\varphi)\, e^{-i\omega t}\,, & \text{outgoing wave} \\ \text{const} \times \; h_\ell^{(2)}(kr) Y_\ell^m(\theta,\varphi)\, e^{-i\omega t}\,, & \text{incoming wave} \end{cases} \tag{1.26}$$

where $h_\ell^{(1)}(kr)$ and $h_\ell^{(2)}(kr)$ are spherical Hankel functions of the first and second kind. In cylindrical coordinates we have two-dimensional cylindrical waves

$$\mathcal{F}(\rho,t) = C_1 \frac{f(t-\rho/v)}{\sqrt{\rho}} + C_2 \frac{g(t+\rho/v)}{\sqrt{\rho}}, \tag{1.27}$$

where the wavefronts are (surfaces of constant wave amplitude) are cylinders coaxial with the z-axis and rays are radial lines diverging from the axis. Other solutions that involve Hankel functions and behave like outgoing or incoming cylindrical waves far from the z-axis also exist. We have mentioned only some of the simplest traveling wave solutions of the scalar wave equation. Many other solutions with more complex wavefronts representing standing or traveling waves are possible.

Vector waves predicted by Maxwell's equations can be constructed from the solutions of the scalar wave equation. If the vector wave field has a fixed direction $\hat{e}$ in space, then plane wave solutions of the form $\vec{\mathcal{F}}(\vec{r},t) = \hat{e} f(t - \hat{k}\cdot\vec{r}/c)$ exist. Vector spherical or cylindrical wave solutions are more complex even in the simplest case. For example, the simplest vector spherical wave allowed by Maxwell's equations has the form

$$\vec{F}(\vec{r},t) = \text{const} \times \left[\frac{\cos\omega(t-r/v)}{kr} - \frac{\sin\omega(t-r/v)}{(kr)^2} \right] \sin\theta\, \hat{e}_\varphi\,, \quad k = \omega/v\,, \tag{1.28}$$

where ω is a frequency. We encounter this and other types of vector spherical waves in the context of scattering and radiation problems.

Thus Maxwell's equations admit solutions that satisfy the wave equation. It must be remembered, however, that not all solutions of the wave equation are admissible as solutions of Maxwell's equations; only those solutions of wave equation that also satisfy the constraints imposed by the Maxwell's equations describe electromagnetic waves.

1.5 Geometrical Optics

The branch of optics characterized by the neglect of wavelength λ in comparison to the characteristic length ℓ of the problem - for example, the length scale on which the refractive index changes significantly or the radius of curvature of interfaces - is known as geometrical optics. In this approximation the laws of optics may be formulated in the language of geometry and optical energy may be regarded as being transported along certain curves called the light rays. It must be mentioned that geometrical optics deals with transparent media.

We will see that for small wavelength the field has the same general character as that of a plane wave and that within the approximation of geometrical optics the laws of reflection and refraction established for plane waves incident upon a plane boundary remain valid under more general conditions. Hence if a light ray falls on a sharp boundary (for example the surface of a lens) it is split into a refelcted ray and trasnmitted ray obeying the laws of reflection and refraction. The preceding remarks imply that, when wavelength is small enough, optical phenomena may be deduced from genometrical considerations by determining the path of the light rays.

We will show that geometrical optics is the $k_0 \longrightarrow \infty$ or $\lambda_0 \longrightarrow 0$ limit of wave optics. In this limit the wavelike monochromatic solutions of Maxwell's equations are of the form

$$\vec{\mathcal{E}}(\vec{r},t) = \vec{e}(\vec{r})\, e^{-i(\omega t - k_0 \mathcal{S}(\vec{r}))}\,, \qquad \vec{\mathcal{B}}(\vec{r},t) = \vec{b}(\vec{r})\, e^{-i(\omega t - k_0 \mathcal{S}(\vec{r}))}\,. \tag{1.29}$$

Here the envelope functions $\vec{e}(\vec{r})$ and $\vec{b}(\vec{r})$ are some slowly varying[2] func-

[2]Fractional change in their values is negligible over distances of the order of a wavelength

tions of position and $\mathcal{S}(\vec{r})$ is a real scalar function which remains to be determined. Substituting these into Maxwell's equations for transparent media (μ and ϵ real),

$$\begin{array}{rclcrcl} \vec{\nabla}\cdot\epsilon\vec{\mathcal{E}} &=& 0 & \qquad & \vec{\nabla}\cdot\vec{\mathcal{B}} &=& 0 \\ \vec{\nabla}\times\vec{\mathcal{E}} &=& i\omega\vec{\mathcal{B}} & \qquad & \vec{\nabla}\times\frac{\vec{\mathcal{B}}}{\mu} &=& -i\omega\epsilon\vec{\mathcal{E}} \end{array} \tag{1.30}$$

we find, with $\Psi(\vec{r},t) = \omega t - k_0\mathcal{S}(\vec{r})$,

$$\begin{aligned} \epsilon\left[\vec{e}\cdot\vec{\nabla}\ln\epsilon + \vec{\nabla}\cdot\vec{e} + ik_0\vec{e}\cdot\vec{\nabla}\mathcal{S}(\vec{r})\right]e^{-i\Psi(\vec{r},t)} &= 0 \\ \left[\vec{\nabla}\cdot\vec{b} + ik_0\vec{b}\cdot\vec{\nabla}\mathcal{S}(\vec{r})\right]e^{-i\Psi(\vec{r},t)} &= 0 \\ \left[\vec{\nabla}\times\vec{e} + ik_0\vec{\nabla}\mathcal{S}(\vec{r})\times\vec{e}\right]e^{-i\Psi(\vec{r},t)} &= i\omega\vec{b}\, e^{-i\Psi(\vec{r},t)} \\ \frac{1}{\mu}\left[-\vec{\nabla}\ln\mu\times\vec{b} + \vec{\nabla}\times\vec{b} + ik_0\vec{\nabla}\mathcal{S}(\vec{r})\times\vec{b}\right]e^{-i\Psi(\vec{r},t)} &= -i\omega\epsilon\vec{e}\, e^{-i\Psi(\vec{r},t)} \end{aligned} \tag{1.31}$$

Rearranging these we find

$$\vec{e}\cdot\vec{\nabla}\mathcal{S}(\vec{r}) = \frac{1}{ik_0}\left[\vec{e}\cdot\vec{\nabla}\ln\epsilon + \vec{\nabla}\cdot\vec{e}\right] \tag{1.32a}$$

$$\vec{b}\cdot\vec{\nabla}\mathcal{S}(\vec{r}) = \frac{1}{ik_0}\left[\vec{\nabla}\cdot\vec{b}\right] \tag{1.32b}$$

$$\vec{\nabla}\mathcal{S}(\vec{r})\times\vec{e} - c\vec{b} = -\frac{1}{ik_0}\left[\vec{\nabla}\times\vec{e}\right] \tag{1.32c}$$

$$\vec{\nabla}\mathcal{S}(\vec{r})\times\vec{b} + c\mu\epsilon\vec{e} = -\frac{1}{ik_0}\left[-(\vec{\nabla}\ln\mu)\times\vec{b} + \vec{\nabla}\times\vec{b}\right] \tag{1.32d}$$

If the changes in μ, ϵ and the envelopes $\vec{e}$ and $\vec{b}$ over distances of the order of a few wavelengths are small we can ignore the terms on the right hand side of each of these equations,

$$\vec{e}(\vec{r}) \cdot \vec{\nabla}\mathcal{S}(\vec{r}) = 0 \tag{1.33a}$$

$$\vec{b}(\vec{r}) \cdot \vec{\nabla}\mathcal{S}(\vec{r}) = 0 \tag{1.33b}$$

$$\vec{\nabla}\mathcal{S}(\vec{r}) \times \vec{e}(\vec{r}) - c\vec{b}(\vec{r}) = 0 \tag{1.33c}$$

$$\vec{\nabla}\mathcal{S}(\vec{r}) \times \vec{b}(\vec{r}) + c\mu\epsilon\vec{e}(\vec{r}) = 0 \tag{1.33d}$$

Note that the first two equations follow from the last two by taking scalar product with $\vec{\nabla}\mathcal{S}(\vec{r})$. It follows from Eqs.(1.33) that $\vec{e}(\vec{r})$, $\vec{b}(\vec{r})$, and $\vec{\nabla}\mathcal{S}(\vec{r})$ form a right handed triad of mutually orthogonal vectors at each point $\vec{r}$. Furthermore, since the vector $\vec{\nabla}\mathcal{S}(\vec{r})$ is perpendicular to the surface $\mathcal{S}(\vec{r})$=const, vectors $\vec{e}(\vec{r})$ and $\vec{b}(\vec{r})$ are tangential to the surface $\mathcal{S}(\vec{r})$=const. This surface may be called the geometrical optics wave surface or the geometrical wavefront.

By eliminating $\vec{e}$ or $\vec{b}$ from Eqs. (1.28c) and (1.28d), we find that the condition for nontrivial solutions of Eqs.(1.33) is

$$[\vec{\nabla}\mathcal{S}(\vec{r})]^2 = \mu\epsilon c^2 = n^2(\vec{r}) \,, \tag{1.34}$$

where the refractive index of the medium

$$n(\vec{r}) = \sqrt{\frac{\mu(\vec{r})\epsilon(\vec{r})}{\mu_0\epsilon_0}} \tag{1.35}$$

is a function of position because ϵ and μ are functions of position. Equation (1.34), known as the eikonal equation, is the basic equation of geometrical optics. The function $\mathcal{S}$ is known as eikonal. It follows from the eikonal equation that the vector

$$\hat{s} = \frac{\vec{\nabla}\mathcal{S}(\vec{r})}{n(\vec{r})} \tag{1.36}$$

has unit magnitude and is perpendicular to the surface $\mathcal{S}(\vec{r})$ =const. The surfaces $\mathcal{S}(\vec{r})$ =const are called the geometrical wavefronts

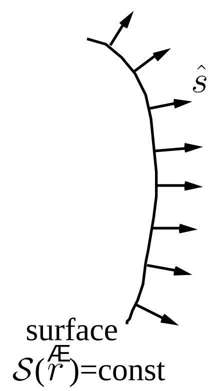


Figure 1.3: Geometrical wavefronts and the direction of the unit vector $\hat{s}$.

With the help of Eqs. (1.28c) and (1.28d) we can express the electric and magnetic field vectors as

$$\vec{b}(\vec{r}) = \frac{\vec{\nabla}\mathcal{S}(\vec{r}) \times \vec{e}(\vec{r})}{c} = \frac{\hat{s} \times \vec{e}(\vec{r})}{v}\,, \tag{1.37a}$$

$$\vec{e}(\vec{r}) = -v\hat{s} \times \vec{b}(\vec{r})\,, \tag{1.37b}$$

where $v = c/n$ is the wave speed in the medium. The time averaged electric and magnetic energy densities are then

$$\overline{u}_e = \frac{1}{4}\epsilon|\vec{e}|^2 \tag{1.38a}$$

$$\overline{u}_m = \frac{1}{4}\frac{|\vec{b}|^2}{\mu} = \frac{1}{4\mu}\frac{\vec{e}\cdot\vec{e}^*}{v^2} = \frac{1}{4}\epsilon\frac{|e|^2}{\mu\epsilon v^2} = \frac{1}{4}\epsilon|e|^2 = \overline{u}_e \tag{1.38b}$$

Thus in the limit of geometrical optics, the time averaged electric and magnetic energy densities associated with a monochromatic wave in a transparent medium are equal.

The time averaged Poynting vector (energy flux density) is given by

$$\overline{\vec{S}} = \frac{1}{2\mu}\Re\left[\vec{e}\times\vec{b}^*\right] = \frac{\vec{e}\times[\vec{\nabla}\mathcal{S}(\vec{r})\times\vec{e}^*]}{2\mu c} = \frac{1}{2}\left(\frac{c}{n}\right)\epsilon\vec{e}\cdot\vec{e}^*\frac{\vec{\nabla}\mathcal{S}(\vec{r})}{n} = v\overline{u}_{em}\hat{s}\,. \tag{1.39}$$

The average Poynting vector (energy flux vector) is in the direction of normal to the geometrical wavefront $\mathcal{S}(\vec{r})$ =const. Its magnitude is equal to the product of the average energy density and speed $v = c/n$ of the wave in the medium. It follows from Eqs. (1.28)-(1.33) that the fields in the geometrical optics limit have the same local character as a plane wave.

1.6 Ray Propagation

We can now define the geometrical light rays as the orthogonal trajectories to the geometrical wavefronts $\mathcal{S}(\vec{r})$ = const. We regard them as oriented curves whose direction coincides everywhere with the direction of the average Poynting vector. We may then say that in geometrical optics light energy is transported along the light rays. The differential equation obeyed by the ray is easily derived as follows. Let $\vec{r}(s)$ denote the position vector of a point P on a ray, considered as a function of the arc length s along the ray measured from some fixed point on it. Then the unit vector $d\vec{r}/ds = \hat{s}$ is tangential to the ray in the direction of energy flow. Using the relation of $\hat{s}$ to $\vec{\nabla}\mathcal{S}(\vec{r})$, the equation for the ray can be written as

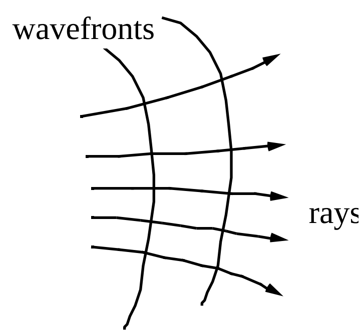


Figure 1.4: Relation betwen wavefronts and rays.

$$\frac{d\vec{r}}{ds} = \hat{s} \equiv \frac{\vec{\nabla}\mathcal{S}(\vec{r})}{n} \qquad \Longrightarrow \qquad n\frac{d\vec{r}}{ds} = \vec{\nabla}\mathcal{S}(\vec{r})\,. \tag{1.40}$$

This equation is purely formal as it specifies rays in terms of $\mathcal{S}(\vec{r})$ which must be determined from Eikonal equation. We can derive another differential equation for the rays directly in terms of the refractive index $n(\vec{r})$ which is much more useful. Differentiating the equation for the ray with respect to arc length we obtain

$$\frac{d}{ds}n\frac{d\vec{r}}{ds} = \frac{d}{ds}\vec{\nabla}\mathcal{S}\,. \tag{1.41}$$

We can express the right hand sides in terms of $n(\vec{r})$ as

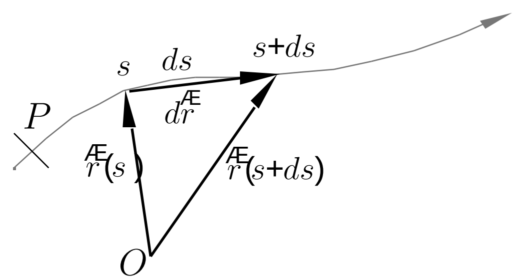


Figure 1.5: Rays are curves along which energy of a wave is transported.

$$\begin{aligned}\frac{d}{ds}\vec{\nabla}\mathcal{S} &= (\hat{s}\cdot\vec{\nabla})\vec{\nabla}\mathcal{S} \\ &= \left(\frac{\vec{\nabla}\mathcal{S}}{n}\cdot\vec{\nabla}\right)\vec{\nabla}\mathcal{S} = \frac{1}{2n}\vec{\nabla}(\vec{\nabla}\mathcal{S})^2 = \frac{1}{2n}\vec{\nabla}(n^2) = \vec{\nabla}n\,. \end{aligned} \tag{1.42}$$

Using these results we find the equation of a ray in terms of the variation of the refractive index

$$\frac{d}{ds}n\frac{d\vec{r}}{ds} = \vec{\nabla}n\,. \tag{1.43}$$

We can gain some insight into what this equation says by writing this in yet another form. We first note that $\hat{s} = \vec{\nabla}\mathcal{S}/n$ [Eq. (1.36)] is a unit vector so that $\hat{s}\cdot\hat{s} = 1$. By differentiating this with respect to s we find

$\hat{s} \cdot (d\hat{s}/ds) = 0$ which means that $d\hat{s}/ds$ is a vector perpendicular to $\hat{s}$. In fact from differential geometry

$$\frac{d\hat{s}}{ds} = \frac{\hat{\nu}}{R} = \frac{d^2\vec{r}}{ds^2}, \tag{1.44}$$

where R is the radius of curvature of the trajectory and $\hat{\nu}$ is a unit vector along the principal normal. Using these results we find the equation of a ray can also be written as

$$\frac{dn}{ds}\hat{s} + n\frac{\hat{\nu}}{R} = \vec{\nabla} n. \tag{1.45}$$

Rewriting this equation as

$$\frac{\hat{\nu}}{R} = \frac{1}{n}\left[\vec{\nabla} n - \hat{s}\frac{dn}{ds}\right], \tag{1.46}$$

and taking the scalar product with $\hat{\nu}$ we find

$$\frac{1}{R} = \frac{\hat{\nu}}{n} \cdot \vec{\nabla} n = \hat{\nu} \cdot \vec{\nabla} \ln n \tag{1.47}$$

This equation says that a ray is bent toward the region of higher refractive index.

Homogeneous Medium

As an application of this equation we consider a homogeneous medium $n = \text{const}$. In this case the equation for the ray becomes

$$\frac{d^2\vec{r}}{ds^2} = 0. \tag{1.48}$$

On integrating this equation we obtain

$$\begin{aligned} \vec{r}\,'(s) &= \vec{r}\,'_0 \\ \vec{r}(s) &= \vec{r}\,'_0 s + \vec{r}_0. \end{aligned} \tag{1.49}$$

where $\vec{r}\,'_0$ is the initial slope of the ray at point $\vec{r}_0$. This is the equation of a straight line.

Rays in a Duct

As another example we consider a medium with quadratic index variation in the transverse direction $n = n_0 - \frac{1}{2}n_2(x^2 + y^2)$ with $n_2 > 0$. The problem has cylindrical symmetry about the z-axis. Then in the paraxial approximation $ds = dz$ and the equation for paraxial rays becomes

$$\frac{d^2r}{dz^2} + \frac{n_2}{n_0}r = \frac{d^2r}{dz^2} + \alpha^2 r = 0\,, \quad \alpha = \sqrt{n_2/n_0} \tag{1.50}$$

with solution

$$r(z) = r_0\cos(\alpha z) + \left(\frac{r_0'}{\alpha}\right)\sin(\alpha z) \tag{1.51}$$

where r_0 is the ray displacement from the z axis and r_0' is the initial slope of the ray at $z = 0$. Thus the ray oscillates back and forth about the z-axis. We note also that a family of parallel rays (rays with the same slope but different displacement) periodically converge in planes $z = (4n+1)\pi/2\alpha$ at points at a height $r_f = (r_o'/\alpha)$ from the z-axis and emerge from these planes with their original slope. These planes are separated from each other by $z_f = 2\pi/\alpha$.

Exercise 1:

Show that a section of the quadratic medium of length ℓ surrounded by a medium of refractive index n_m acts as a lens. [Hint: Show that a family of parallel rays entering at $z = 0$ with different displacement converge after emerging at $z = \ell$ to a common focus at a distance $f = \frac{1}{n_m\alpha}\cot\alpha\ell$]

Exercise 2:

Find the equation for paraxial rays when $n = n_0 + \frac{1}{2}n_2(x^2 + y^2)$.
The answer is [$\alpha = \sqrt{n_2/n_0}$]

$$r(z) = r_0\cosh\alpha z + \left(\frac{r_0'}{\alpha}\right)\sinh\alpha z$$

Laws of Reflection and Refraction

These laws follow from the fact that $n\hat{s} = \vec{\nabla}\mathcal{S}$. This means $\vec{\nabla} \times n\hat{s} = 0$. Integrating the normal component of $\vec{\nabla} \times n\hat{s}$ over the area of a small rectangle straddling the interface between two media

$$\hat{n}_{12} \times (n_2\hat{s}_2 - n_1\hat{s}_1) = 0 \quad \Longrightarrow \quad \hat{n}_{12} \times n_2\hat{s}_2 = \hat{n}_{12} \times n_1\hat{s}_1 \tag{1.52}$$

where $\hat{n}_{12}$ is a unit normal to the interface directed from medium 1 to 2. This equation implies the normal to the boundary $\hat{n}_{12}$, and the ray vectors $\hat{s}_1$ and $\hat{s}_2$ are coplanar and

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \,. \tag{1.53}$$

Here θ_1 and θ_2 are the angles that the rays in the two media make with $\hat{n}_{12}$. A similar procedure for the reflected ray leads to the laws of reflection. These laws are usually derived for plane waves incident on a refracting plane surface. Here we find that they are valid for more general waves and refracting surface provided that the wavelength is sufficiently small. This means the radius of curvature of the incident wave and of the interface must be large compared to the wavelength of the incident light.

Thus in the limit of short wavelength Maxwell's equations lead to a description of light propgation in terms of rays which are geometric curves along which light energy is transported. We also see that in this limit wave nature of light is masked and phenomena such as diffraction are neglected. It may seem a drastic simplification but it is very useful in instrument optics and wqe will see that laws of ray optics are very useful even for understanding diffraction effects. In what follows we will limit our discussion to paraxial rays.

1.6.1 Paraxial Rays

In problems involving instrument optics or laser resonators we are interested in rays that stay close to the optical axis, usually taken to be the $z-$

axis, of the system. Such rays are called paraxial rays. They make angles with the optical axis that are small enough that the sine and tangent of the angle can be approximated by the angle (expressed in radians) itself,

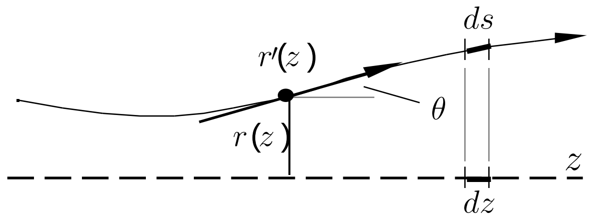


Figure 1.6: Paraxial rays make small angles with the optical axis (usually taken to be the $z-$axis) and stay close to it as it propagates.

$$\tan\theta \approx \theta \approx \sin\theta\,. \tag{1.54}$$

This approximation is good to within 3% for angles less than about 18^o ($0.31 \approx 1/\pi$ radian). For such rays the arc length $ds \approx dz$ and the equation for the ray becomes

$$\frac{d}{dz}n\frac{d\vec{r}}{dz} = \vec{\nabla}n \tag{1.55}$$

For media with cylindrical symmetry we can write $\vec{r} = \hat{e}_{\perp}r + \hat{e}_z z$, where r is the lateral displacement of the ray from the z axis, we find the equation of a ray becomes

$$\frac{d}{dz}n\frac{dr}{dz} = \hat{e}_{\perp}\cdot\vec{\nabla}n\,. \tag{1.56}$$

This equation determines a ray given the lateral displacement r_0 and slope $r_0' = [dr/dz]_{z=z_0}$ of a ray at some fixed point z_0. In what follows we will use the abbreviation $r' = dr/dz$ to denote the slope.

Example 1. In a homogeneous medium $n =$const so that the equation of a ray is

$$\frac{d}{dz}n\frac{dr}{dz} = 0\,. \tag{1.57}$$

with solution

$$\begin{aligned} r'(z) &= r'_0\,, \\ r(z) &= r_0 + r'_o(z - z_0)\,, \end{aligned} \tag{1.58}$$

where r_0 and r'_0 are the ray displacement and slope, respectively, at $z = z_0$.

If n changes, as for example when a ray is incident from one homogeneous medium (refractive index n_1) with slope r'_1 and displacement r_1 onto another homogeneous medium (refractive index n_2), an integration of the ray equation across the interface ($z = 0$) into the second medium gives

$$\begin{aligned} r'_2(z) &= \frac{n_1}{n_2}\,r'_1\,, \\ r_2(z) &= r_1 + \frac{n_1}{n_2}\,r'_1 z\,. \end{aligned} \tag{1.59}$$

The first of these equations is Snell's law in the paraxial approximation. The second equation gives the ray displacement in the second medium in terms of the ray displacement r_1 and slope $r'_2 = (n_1/n_2)r'_1$ at the boundary just inside the second medium.

Treating light as a collection of geometrical optics rays ignores diffraction. Although the study of ray propagation is important in its own right in instrument optics, laws of paraxial ray propagation turn out to be very useful in understanding the full diffractive propagation of light in optical resonators and laser beams.

In paraxial optics a ray is specified by its displacement r from the optical axis and its slope $r' = dr/dz$. Both vary as a function of z as the ray propagates through an optical system. If we introduce a column matrix with r and r' its elements by

$$\tilde{r} = \begin{pmatrix} r \\ r' \end{pmatrix}\,, \tag{1.60}$$

we can describe the effect of an optical element on ray parameters (r, r')

by a 2×2 matrix of the form

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} . \tag{1.61}$$

For most purposes we have to know the transformation properties of three basic elements:

(i) free propgation in a homogeneous medium of length L and refractive index n;

(ii) reflection from a curved surface of radius of curvature R;

(iii) refraction at a curved interface (radius of curvature R) between two media with refractive indices n_1 and n_2 when a ray is incident from medium n_1.

Matrices for most others elements can be derived from these.

Propagation in a homogeneous medium

Consider a ray propagating from plane z_1 to plane $z_2 = z_1 + L$ in a homogenoeus medium of refractive index n, the ray parameters at the input and output faces are related by [Eq. (1.58)]

$$\begin{aligned} r_2 &= r_1 + r_1' L \, , \\ r_2' &= r_1' \end{aligned} \quad \Rightarrow \quad \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \tag{1.62}$$

Reflection at a curved surface

From Fig. (1.7) it is clear that the ray displacement remains unchanged in reflection.

$$r_{\text{out}} = r_{\text{in}} \tag{1.63}$$

To find a relation between the input and output ray slopes we use the law of reflection, which leads to

$$\begin{aligned} &\theta_2 = \theta_1 && \text{(law of reflection)} \\ \text{or} \quad &\theta_2' - \phi = \phi - \theta_{\text{in}} && \\ \text{or} \quad &\theta_2' = 2\phi - \theta_{\text{in}} && (1.64) \end{aligned}$$

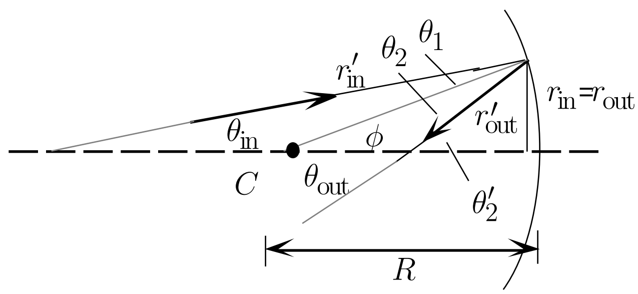


Figure 1.7: According to the laws of reflection the angles of incidence and reflection are equal $\theta_2 = \theta_1$. Note that $r_{\text{out}}' = \theta_{\text{out}}$ is negative so that $\theta_2' = -\theta_{\text{out}}$.

To relate these angles to ray slopes we note that r_{in}' is positive whereas r_{out}' is negative,

$$\begin{aligned} \theta_{\text{in}} &= r_{\text{in}}' , \\ \theta_2' &= -\theta_{\text{out}} = -r_{\text{out}}' , \\ \phi &= \frac{r_{\text{in}}}{R} , \end{aligned} \qquad (1.65)$$

where we consider R to be ﬀﬂsﬂﬁvﬂ ﬀﬂr ﬂ ﬀﬂﬂﬀwﬂ ﬂ ﬂrrﬂr ﬂrﬀﬂﬀﬂtﬀﬂﬂ surﬂ ﬀﬂﬀﬀﬂ Using these, we find that the exit ray slope is given by

$$r_{\text{out}}' = -\frac{2r_{\text{in}}}{R} + r_{\text{in}}' \qquad (1.66)$$

With the help of Eqs. (1.63) and (1.66) we obtain the ABCD matrix for a reflecting surface

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{2}{R} & 1 \end{pmatrix} \qquad (1.67)$$

Refraction at a curved interface.

From Fig. (1.8) we see that the input an output displacements are the same

$$r_{\text{out}} = r_{\text{in}} \,. \tag{1.68}$$

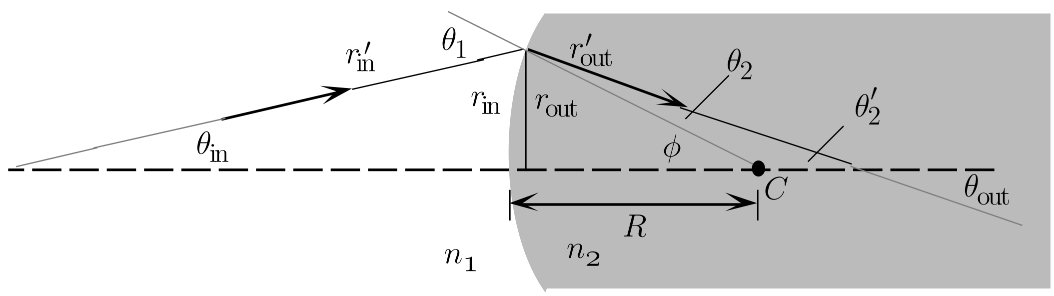


Figure 1.8: Refraction at a curved interface between two media. Note that in the figure $r'_{\text{out}} = \theta_{\text{out}}$ is negative so that $\theta'_2 = -\theta_{\text{out}} = -r'_{\text{out}}$.

To relate input and output ray slopes we consider the relation between angles,

$$
\begin{aligned}
& n_1\theta_1 = n_2\theta_2 \quad \text{Snell's law} \\
\text{or} \qquad & n_1(\theta_{\text{in}} + \phi) = n_2(\phi - \theta'_2) \\
\text{or} \qquad & \theta'_2 = \frac{(n_2 - n_1)}{n_2}\phi - \frac{n_1}{n_2}\theta_{\text{in}} \qquad (1.69)
\end{aligned}
$$

Now noting $r'_{\text{in}} = \theta_{\text{in}}$ and $r'_{\text{out}} = \theta_{\text{out}} = -\theta'_2$ and $\phi = r_{\text{in}}/R$ where we choose R ﬄsﬅﬁvﬂ ﬀr ﬂ ﬃﬂvﬁx rﬀﬁﬂﬅﬃﬂ surﬀﬂﬃ. This leads us to

$$r'_{\text{out}} = -\frac{(n_2 - n_1)}{n_2}\frac{r_{\text{in}}}{R} + \frac{n_1}{n_2}r'_{\text{in}} \tag{1.70}$$

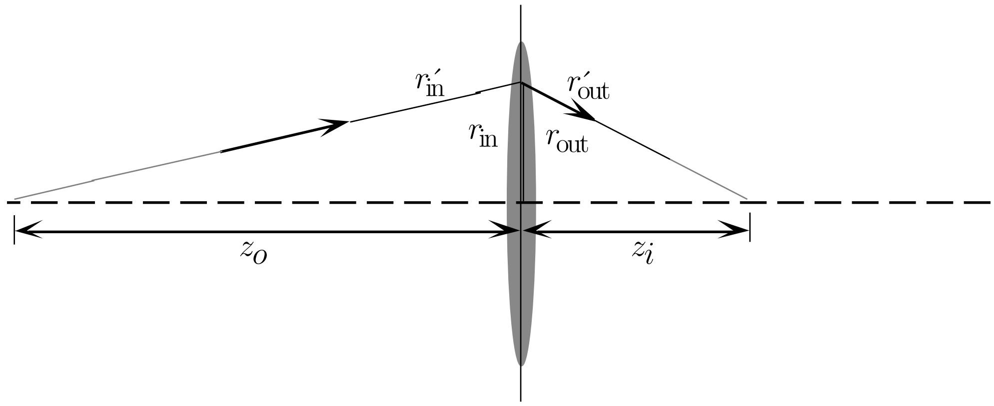


Figure 1.9: Relation between input and output ray parameters for a thin lens.

From Eqs. (1.68) and (1.70) we find the ABCD matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{n_2 - n_1}{n_2}\frac{1}{R} & \frac{n_1}{n_2} \end{pmatrix} \tag{1.71}$$

For a thin lens we can then find the ABCD matrix by multiplying the ABCD matrix for each of its surfaces,

$$\begin{pmatrix} 1 & 0 \\ -\frac{n_1 - n_2}{n_1}\frac{1}{R_2} & \frac{n_2}{n_1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{n_2 - n_1}{n_2}\frac{1}{R_1} & \frac{n_1}{n_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{n_2 - n_1}{n_1}\left(\frac{1}{R_1} - \frac{1}{R_2}\right) & 1 \end{pmatrix} \tag{1.72}$$

We can, of course, derive the matrix of a thin lens of focal length f directly with the help of Fig. (1.9). Let $r_{\rm in}$ and $r'_{\rm in}$ denote the displacement and slope of the incident ray just before the lens and $r_{\rm out}$ and $r'_{\rm out}$ their values just after the lens. Then it follows

$$r_{\rm out} = r_{\rm in} \,. \tag{1.73}$$

To determine the slope after the lens we note that the incident ray may be thought of as coming from the axial point z_0 and the emergent ray may thought of as proceeding towards the point z_i. These distances are related by the thin lens formula

$$\frac{1}{z_o} + \frac{1}{z_i} = \frac{1}{f} \,. \tag{1.74}$$

Multiplying both sides by r_{in} and noting that $r_{\text{in}}/d_o = r'_{\text{in}}$ and $r_{\text{in}}/z_i = -r'_{\text{out}}$ (the emergent ray has negative slope) and rearranging the terms we find the slope of the emergent ray

$$r'_{\text{out}} = r_{\text{in}} - \frac{1}{f} r_{\text{in}} \,. \tag{1.75}$$

From the preceding examples, it is clear that we can write the relation between the input and output ray parameters in matrix form as

$$\begin{pmatrix} r_{\text{out}} \\ r'_{\text{out}} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} r_{\text{in}} \\ r'_{\text{in}} \end{pmatrix} \tag{1.76}$$

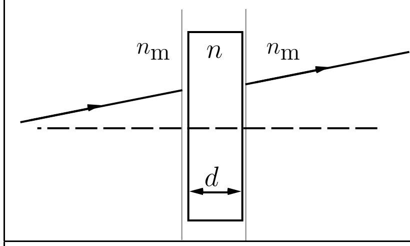
This relation represents a transformation of input ray parameters into putput ray parameters. The matrix of transformation, also called the *ABCD* matrix, depends on the nature of the optical element inside the black box. For example the ABCD matrix for propagation over a section of length L in a homogeneous medium is

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \tag{1.77}$$

ABCD matrices for a number of optical elements are given in Table (1.2).

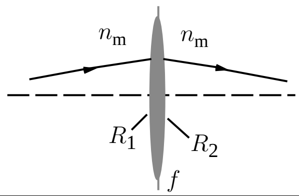
Once these basic matrices are known we can calculate the overall ABCD matrix for any system of optical elements. Consider, fro example, the passage of a ray through a sequence of optical elements shown in Fig. [a thin lens, followed by free space and a dilectric slab(refractive index n)]. By labeling various optical elelments as $1\,,2\,,3\cdots$ in the order in which

Plate of thickness d, refractive index n in a medium of refractive index n_m, normal incidence



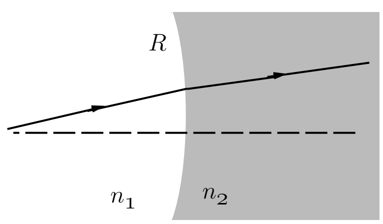
$$\begin{bmatrix} 1 & \frac{d}{n} \\ 0 & 1 \end{bmatrix}$$

Thin lens of focal length f in a medium of refractive index n_m, normal incidence



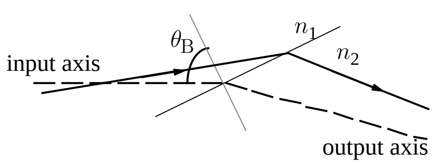
$$\begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix} \qquad \frac{1}{f} = \frac{n - n_m}{n_m}\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

Dielectric interface with radius of curvature R (+ for convex and − for concave interface).



$$\begin{bmatrix} 1 & 0 \\ -\frac{n_2 - n_1}{n_2}\frac{1}{R} & \frac{n_1}{n_2} \end{bmatrix}$$

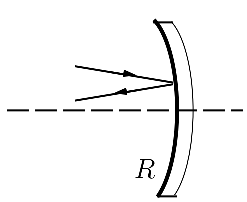
Dielectric interface at Brewster's angle.



$$\begin{bmatrix} \frac{n_2}{n_1} & 0 \\ 0 & \frac{n_1^2}{n_2^2} \end{bmatrix}$$ plane of incidence (tangential plane)

$$\begin{bmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{bmatrix}$$ perpendicular to the plane of incidence (sagittal plane)

Spherical mirror of radius of curvature R, normal incidence ($R > 0$ for concave mirror and $R < 0$ for convex mirror)



$$\begin{bmatrix} 1 & 0 \\ -\frac{2}{R} & 1 \end{bmatrix}$$

Table 1.2: ABCD matrices of some common optical elements.

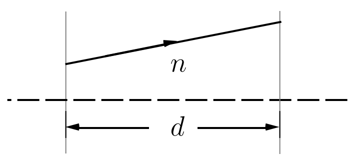
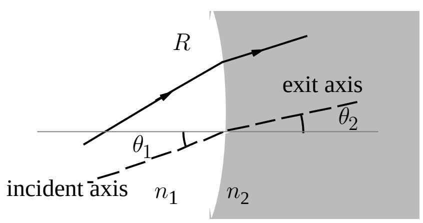
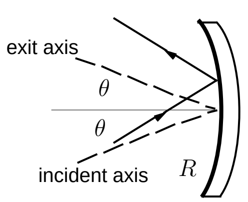
<table>
<tr><td colspan="3">Straight section of length d in a homogeneous medium of refractive index n</td></tr>
<tr><td>
</td><td>$$\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}$$</td><td></td></tr>
<tr><td colspan="3">Dielectric interface, radius of curvature R (+ for convex and − for concave refracting surface), arbitrary angle of incidence
$\Delta n_T = \dfrac{n_2}{\cos\theta_1} - \dfrac{n_1}{\cos\theta_2}$,
$\Delta n_S = n_2\cos\theta_2 - n_1\cos\theta_1$</td></tr>
<tr><td rowspan="2">
</td><td>$$\begin{bmatrix} \dfrac{\cos\theta_2}{\cos\theta_1} & 0 \\ -\dfrac{\Delta n_T}{n_2 R} & \dfrac{n_1\cos\theta_1}{n_2\cos\theta_2} \end{bmatrix}$$</td><td>plane of incidence (tangential plane)</td></tr>
<tr><td>$$\begin{bmatrix} 1 & 0 \\ -\dfrac{\Delta n_S}{n_2 R} & \dfrac{n_1}{n_2} \end{bmatrix}$$</td><td>perpendicular to the plane of incidence (sagittal plane)</td></tr>
<tr><td colspan="3">Spherical mirror of radius of curvature R (+ for concave mirror and − for convex mirror), arbitrary angle of incidence</td></tr>
<tr><td rowspan="2">
</td><td>$$\begin{bmatrix} 1 & 0 \\ -\dfrac{2}{R\cos\theta} & 1 \end{bmatrix}$$</td><td>plane of incidence (tangential plane)</td></tr>
<tr><td>$$\begin{bmatrix} 1 & 0 \\ -\dfrac{2\cos\theta}{R} & 1 \end{bmatrix}$$</td><td>perpendicular to the plane of incidence (sagittal plane)</td></tr>
</table>

Table 1.3: ABCD matrices for paraxial rays for three basic optical elements.

they are encountered, we can write the ABCD matrix as

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = M_7 \cdot M_6 \cdots M_2 \cdot M_1 \,. \tag{1.78}$$

Note that the matrices are written from right to left in the order in which they are encountered by the ray.

1.6.2 Periodic Focusing System

An interesting and important application of ray matrices comes in the analyses of periodic focusing systems in which the same sequence of elements is repeated many times down a cascaded chain. An optical resonator can be modeled by such an iterated periodic focusing system because propagation through repeated round trips in the resonator is physically equivalent to propagation through repeated sections of a periodic lens guide.

As an example consider an optical resonator formed by two spherical mirrors of radii of curvature R_1 and R_2 placed a distance L apart. Imagine a ray propagating to the right starting at the left end of the resonator. After a round trip, this ray will have been transformed by a straight section of length L, a spherical mirror of radius of curvature R_1 another section of length L, and finally a spherical mirror of radius of curvature R_2. In each roundtrip, the ray encounters the same transformation. This is equivalent to a lens waveguide where lenses of focal length $f_1 = R_1/2$ and $f_2 = R_2/2$ are placed alternately separated by L.

The ABCD matrix M describing the ray transformation in a roundtrip through the resonator is given by $M = M_1 \cdot M_L \cdot M_2 \cdot M_L$, where M_1 and M_2 are the ray matrices of mirrors R_1 and R_2, respectively, and M_L is the ray matrix of a section of length L. Note that the matrices are written from

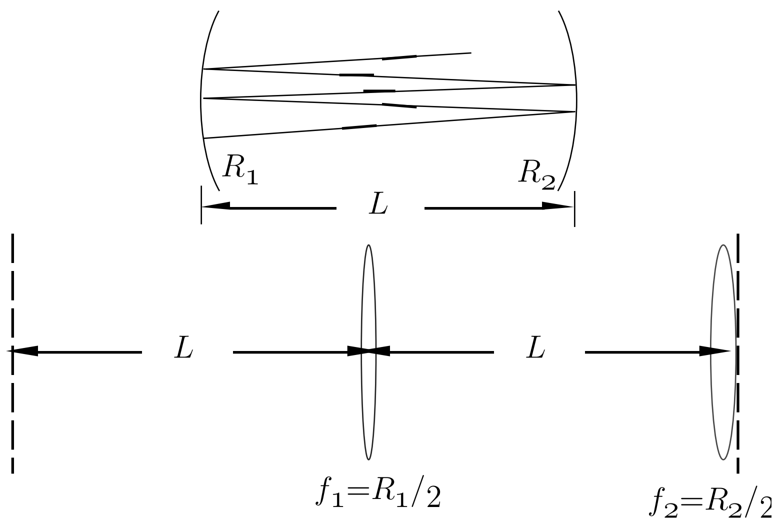


Figure 1.10: An optical cavity formed by two mirrors of radii R_1 and R_1 separated by distance L is equivalent to a lens waveguide.

right to left in the order in which the optical element was encountered.,

$$\begin{aligned}\begin{pmatrix} A & B \\ C & D \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ -\dfrac{2}{R_1} & 1 \end{pmatrix} \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\dfrac{2}{R_2} & 1 \end{pmatrix} \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 - \dfrac{2L}{R_2} & 2L\left(1 - \dfrac{L}{R_2}\right) \\ -\dfrac{2}{R_1} - \dfrac{2}{R_2} + \dfrac{4L}{R_1 R_2} & 1 - \dfrac{4L}{R_1} - \dfrac{2L}{R_2} + \dfrac{4L^2}{R_1 R_2} \end{pmatrix} \\ &\equiv \begin{pmatrix} 2g_2 - 1 & 2Lg_2 \\ \frac{2}{L}(2g_1 g_2 - g_1 - g_2) & 4g_1 g_2 - 2g_2 - 1 \end{pmatrix} \quad g_i \equiv 1 - \frac{L}{R_i} \end{aligned} \tag{1.79}$$

After $(n+1)$th roundtrip, the ray is given by

$$\widetilde{r}_{n+1} = M\widetilde{r}_n = M^2 \tilde{r}_{n-1} = M^n \tilde{r}_0 \tag{1.80}$$

A cascaded system such as this is best analyzed by finding the eigenvalues and eigen solutions of the matrix M. This means we look for a set of eigen rays and eigenvalues Λ satisfying

$$M\widetilde{r} = \Lambda\widetilde{r} \qquad \text{or} \qquad \begin{pmatrix} A-\Lambda & B \\ C & D-\Lambda \end{pmatrix} \begin{pmatrix} r \\ r' \end{pmatrix} = 0 \tag{1.81}$$

Nontrivial solutions require

$$det \begin{pmatrix} A-\Lambda & B \\ C & D-\Lambda \end{pmatrix} = \Lambda^2 - (A+D) + (AD-BC) = 0 \tag{1.82}$$

This equation can be simplified by noting that $AD - BC$ is the determinant of M and that this determinant is unity. The easiest way to check this is to recall that the determinant of the product of a set of matrices is the product of the determinants of individual matrices $det(M) = det(M_1)det(M_L)det(M_2)det(M_L)$ and that the determinant of each of these matrices is unity. We also note that the the coefficient of Λ is the trace of the matrix. By introducing a parameter m by

$$m = \frac{1}{2}\mathrm{tr}M = \frac{1}{2}(A+D)\,, \tag{1.83}$$

we can solve Eq. (1.82) for Λ and write the two eigenvalues as

$$\Lambda_a = m + \sqrt{m^2-1}\,, \qquad \Lambda_b = m - \sqrt{m^2-1}\,. \tag{1.84}$$

We also have

$$\Lambda_a + \Lambda_b = 2m\,, \quad \text{and} \quad \Lambda_a\Lambda_b = 1\,. \tag{1.85}$$

There are two matching eigenrays $\widetilde{r}_a$ and $\widetilde{r}_b$ such that

$$M\widetilde{r}_a = \Lambda_a\widetilde{r}_a, \qquad M\widetilde{r}_b = \Lambda_b\widetilde{r}_b\,. \tag{1.86}$$

This equation means that the propagation of each eigen ray is specified simply by multiplying it by the corresponding eigenvalue.

In terms of these eigen rays, any input ray $\widetilde{r}_0$ can be written as

$$\widetilde{r}_0 = C_a\widetilde{r}_a + C_b\widetilde{r}_b\,, \tag{1.87}$$

where C_a and C_b are some constants which are determined by the intial ray displacement r_0 and slope r'_0. After n roundtrips this ray becomes

$$\widetilde{r}_n = M^n \widetilde{r}_0 = C_a \Lambda_a^n \widetilde{r}_a + C_b \Lambda_b^n \widetilde{r}_b \,. \tag{1.88}$$

Here we have used the fact that the propagation of each eigen ray is specified simply by multiplying it by the corresponding eigenvalue raised to the number of round trips.

We can now see from Eq. (1.88) that if any of the eigenvalues is greater than unity, the ray displacement will continue to grow with n. Thus whether or not a periodic array has stable ray trajectories depends on the eigenvalues. We can divide periodic focusing systems into stable or unstable systems depending the nature of the eigenvalues which depend on the parameter m via Eq.(1.84). If the parameter m is less than unity, we have complex eigenvalues. If, on the other hand, $|m| > 1$, we have real eigenvalues and at least one of them is greater than unity. Let us consider these cases separately.

Case 1: $m^2 < 1$ **or** $- \leq m \leq 1$

In this case, we can write

$$m = \frac{1}{2}(A + D) \equiv \cos\theta \,, \tag{1.89}$$

$$\Lambda_{a,b} = \cos\theta \pm i\sqrt{1-\cos^2\theta} = \cos\theta \pm i\sin\theta = e^{\pm i\theta} \tag{1.90}$$

Then ray propagationin the periodic system takes the form

$$\widetilde{r}_n = C_a e^{in\theta} \widetilde{r}_a + C_b e^{-in\theta} \widetilde{r}_b = \widetilde{c}_0 \cos n\theta + \widetilde{s}_0 \sin n\theta \tag{1.91}$$

where

$$\widetilde{c}_0 = C_a \widetilde{r}_a + C_b \widetilde{r}_b \,, \qquad \widetilde{s}_0 = i(C_a \widetilde{r}_a - C_b \widetilde{r}_b) \,. \tag{1.92}$$

The displacment after n round-trips is then a periodic function

$$r_n = c_0 \cos n\theta + s_0 \sin n\theta \,. \tag{1.93}$$

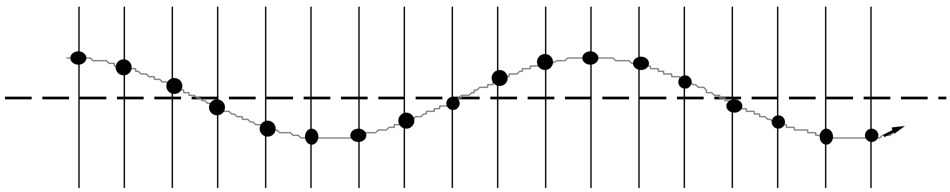


Figure 1.11: The dots indicate the ray position at successive reference planes.

Thus a periodic focusing system is stable if $|m| < 1$. Rays in this system will oscillate back and forth about the axis as shown in Fig. (1.11). Note that r_n only gives the displacement at the successive reference planes; it does not say anything about what happens to the ray inside the periodic section between the reference planes. Note also that it is n that is varying. The period of oscillation of the ray is $2\pi/\theta$ periods of oscillation of the periodic system.

Case 2 : $m^2 > 1\,(m \leq -1\,\textbf{or}\,m \geq 1)$**.**

In this case the eigenvalues of M are real given by

$$\begin{aligned} \Lambda_a &= m + \sqrt{m^2-1} \equiv \Lambda\,, \\ \Lambda_b &= m - \sqrt{m^2-1} \equiv \frac{1}{\Lambda}\,. \end{aligned} \tag{1.94}$$

It can be seen that one of the eigenvalues has a magnitude greater than unity. Then after n steps the ray coordinates are given by

$$\widetilde{r}_n = C_a\, e^{n\theta}\widetilde{r}_a + C_b\, e^{-n\theta}\widetilde{r}_b\,, \quad \theta = \ln\Lambda \tag{1.95}$$

The ray displacement

$$r_n = C_a\, e^{n\theta} r_a + C_b\, e^{-n\theta} r_b \tag{1.96}$$

grows with each step. For $\Lambda > 1$ the ray displacement grows monotonically whereas for $\Lambda < -1$ ($\ln\Lambda = \ln|\Lambda| + i\pi$ for $\Lambda < -1$) it grows in an

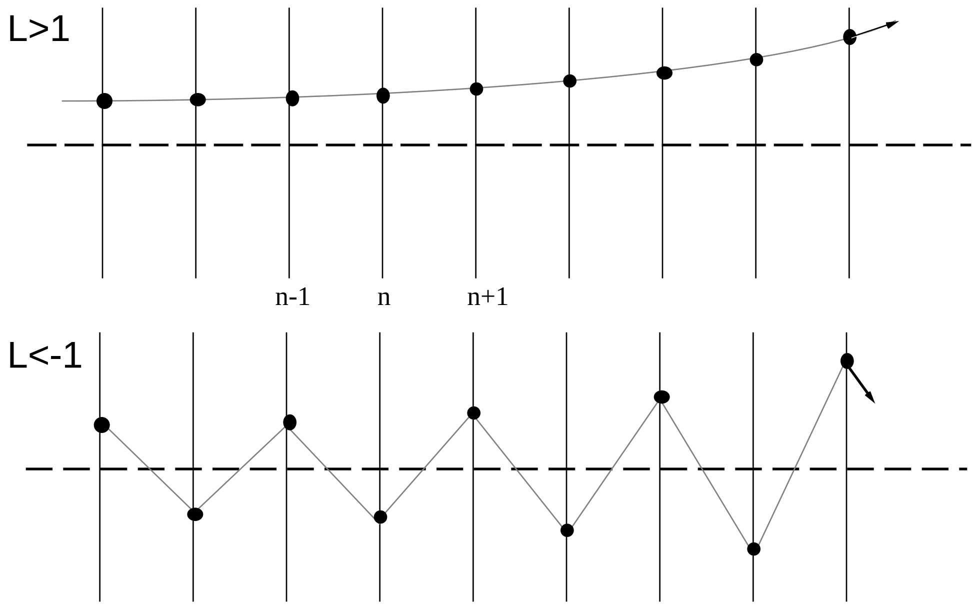


Figure 1.12: In an unstable periodic array the ray displacement increases without limit.

oscillatory fashion as seen in Fig. (1.12). Periodic systems with $m > 1$ are clearly unstable.

Two-Mirror Resonator Stability

For a two-mirror cavity, the round trip ABCD matrix is given by Eq. (1.79). The parameter $m = \frac{1}{2}(A + D)$ for a two mirror cavity is then

$$m = 1 - \frac{2L}{R_1} - \frac{2L}{R_2} + \frac{2L^2}{R_1 R_2} = 2\left(1 - \frac{L}{R_1}\right)\left(1 - \frac{L}{R_2}\right) - 1 \qquad (1.97)$$

The condition for stability, $-1 \leq m \leq 1$, then implies

$$\begin{aligned} -1 \leq& 2\left(1 - \frac{L}{R_1}\right)\left(1 - \frac{L}{R_2}\right) - 1 \leq 1 \\ 0 \leq& 2\left(1 - \frac{L}{R_1}\right)\left(1 - \frac{L}{R_2}\right) \leq 2 \\ 0 \leq& \left(1 - \frac{L}{R_1}\right)\left(1 - \frac{L}{R_2}\right) \leq 1 \end{aligned} \tag{1.98}$$

On introducing the parameters

$$g_1 = 1 - \frac{L}{R_1}, \qquad g_2 = 1 - \frac{L}{R_2}. \tag{1.99}$$

the stability condition can be written more compactly as

$$0 \leq g_1 g_2 \leq 1. \tag{1.100}$$

This important relation specifies the range of stability of two-mirror optical resonators. A plot of this inequality is shown in Fig. (1.13).

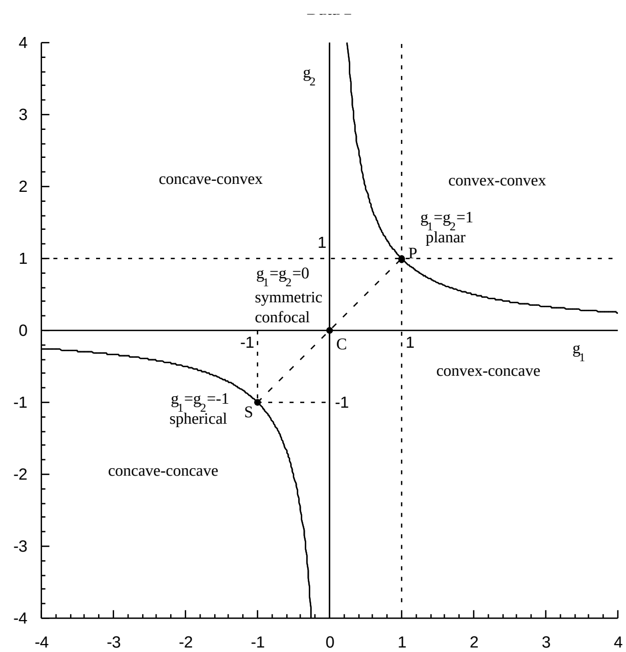


Figure 1.13: Stable resonators lie in the region bounded the coordinate axes and the two branches of hyperbola $g_1 g_2 = 1$. Symmetric resonators lie along the line $g_2 = g_1$.