Introduction to Laser Physics

BELA A. LENGYEL
Professor, San Fernando Valley State College
Northridge, California

JOHN WILEY AND SONS, INC.
New York · London · Sydney
Background Material on Radiation and Atomic Physics

Laser art uses material that originates in different branches of physics and engineering. Only an exceptional person is familiar with all concepts and relations that are bound to play a role in the discussion of the various phases of this art. Therefore, it is probably worthwhile to present in summary form the ideas and results needed from physics and other disciplines. Most of this material is discussed in undergraduate courses in optics, modern physics, and atomic theory; therefore the summary may be a brief reminder. A short discussion of units, conversion tables, and measurement techniques serves only to orient the reader about practical matters and point the way to literature where more adequate information may be found.

1. LIGHT AND THE GENERAL LAWS OF RADIATION

The subject of this book is the generation and amplification of light by stimulated emission of radiation. It therefore seems appropriate to commence by gathering material from classical electromagnetic theory and ordinary optics that is needed as a background for the subject to be discussed. At the risk of boring the experienced reader, we begin with some elementary definitions and statements.

Our interest is in electromagnetic radiation in or near the visible region. The wavelength in this region varies from 0.3 to 3 μ, the frequency, from $10^{14}$ to $10^{15}$ cps. The emphasis is on the fact that we are dealing with electromagnetic radiation and not that it is visible. We shall avoid all terms, so common in ordinary optics, that assess light in terms of its effects on the human eye. Consequently, we shall not speak of luminous but of radiative quantities, which are determined by using a detector capable of registering the transport of energy by means of electromagnetic radiation. We recapitulate the basic terms used in connection with such transport of energy.
The counterpart of luminous flux in ordinary optics is \textit{radiative flux}. This is the rate at which radiant energy passes through a surface; it is measured in units of power; that is, in watts (joules per second) or in ergs per second. The intensity of radiation incident on a surface is the \textit{radiative flux density}, whose MKS unit is watts per square meter. To indicate the directional distribution of radiation of a radiating surface, we need the concept of the \textit{radiance in a given direction}. This is the radiant flux in a given direction per unit solid angle per unit projected area of the radiator. It is usually denoted by the symbol $N$, and its meaning can be clarified as follows: given a radiating surface of area $A$ and a direction at an angle $\vartheta$ from the surface normal, the radiative flux in a small cone of $d\Omega$ steradians around the given direction is $NA \cos \vartheta d\Omega$. When $N$ is independent of the direction, we say that the surface radiates or scatters according to \textit{Lambert's law}. In this case the total radiation from the surface is $\pi NA$. Related to $N$ is the energy density of the radiation $u$, which is simply the radiative energy contained in the unit volume.

We now make use of the existence of filters and monochromators, which enable us to classify radiation according to its frequency or wavelength. All quantities pertaining to radiation may be regarded as functions of the frequency $\nu$ or the wavelength $\lambda$; their symbols are then provided with appropriate subscripts. The symbol $u_\nu$ is defined as follows: the energy density of radiation between the frequencies $\nu$ and $\nu + d\nu$ is $u_\nu d\nu$. The symbol $u_\lambda$ refers to energy density in the wavelength interval $\lambda$ to $\lambda + d\lambda$; consequently, $u_\nu$ and $u_\lambda$ are related but different functions of the variables. The frequency interval $\nu, \nu + d\nu$ and the wavelength interval $\lambda, \lambda - d\lambda$ are equivalent descriptions of the same spectral region when $d\nu/\nu = d\lambda/\lambda$. It is easily shown that $u_\nu \nu = u_\lambda \lambda$. Here $\nu$ and $\lambda$ are in arbitrary units, but their product is the velocity of light.

It is usually convenient to characterize radiation by its wavelength whenever experiments or applications are concerned, but in theoretical calculations, particularly in those involving energy, frequency is a more suitable variable. When electromagnetic radiation in a cavity is in thermal equilibrium at the absolute temperature $T$, the distribution of radiation density according to frequency follows \textit{Planck's law}:

$$u_\nu d\nu = \frac{8\pi h \nu^3}{c^3} \frac{d\nu}{e^{h\nu/kT} - 1}.$$ \hspace{1cm} (1.1)

Here $h$ is Planck's constant, $k$ is Boltzmann's constant, and $c$ is the velocity of light. Their numerical values are given in Section 2.

Radiation will escape through a hole cut into the walls of such a cavity at the rate of $W = uc/4$ per unit area of the hole. This is the radiative flux density at the exit of the cavity; it is called black-body radiation, and many
solids radiate like this idealized black body. Therefore the frequency
distribution of the flux radiated from the surface of a solid may be
approximated by means of Planck's formula.

In experimental work distribution according to wavelength is preferred,
and the radiation formula takes the form

\[ W(\lambda, T) \, d\lambda = \frac{C_1 \lambda^{-5} \, d\lambda}{e^{C_2 \lambda / kT} - 1}, \]  
(1.2)

where \( C_1 = 2\pi \hbar c^2 \) and \( C_2 = \hbar c / k \). However, rather than work with pure
CGS or MKS units, it is often convenient to express \( W \) in watts per square
centimeter and to measure \( \lambda \) and \( d\lambda \) in Ångström units. With these changes
the constants become

\[ C_1 = 3.74 \times 10^{20} \text{ watts/cm}^2 (\text{Å})^4, \quad C_2 = 1.438 \times 10^8 \text{ Å}^6 \text{K}. \]

According to the Stefan-Boltzmann law, the total black-body radiation is

\[ W = \int_0^\infty W(\lambda, T) \, d\lambda = \sigma T^4, \quad \sigma = 5.679 \times 10^{-12} \text{ watt/cm}^2 \text{ deg}^4. \]  
(1.3)

It follows from what has been said that an incandescent solid is the
source of radiation whose energy is not concentrated in any frequency
region. For each temperature there is naturally a wavelength at which the
emitted radiation is maximum. This wavelength \( \lambda_M \) is calculable from
Wien's displacement law:

\[ \lambda_M T = 2.898 \times 10^7 \text{ Å}^6 \text{K}, \]  
(1.4)

and the peak of \( W(\lambda, T) \) for a given \( T \) is

\[ W_M(T) = W(\lambda_M, T) = 1.290 \times 10^{-19} T^8 \text{ watt/cm}^2 \text{ Å}. \]  
(1.5)

Numerical calculation of the black-body radiation in a given spectral
region is facilitated by the introduction of the variable \( x = \lambda T \) because the
functions \( W(\lambda, T) / W_M(T) \) and \( \int_0^\lambda W \, d\lambda / W_T \) are functions of the variable
\( x \) alone. These two functions are tabulated [1, pp. 6–64]. The calculation is
carried out by finding the total radiation or the peak value first by means
of (1.3) or (1.5). The relative radiation density, or the integrated relative
radiation density, is then found by using the tables.

A black body at the temperature of 5200°K has its radiation peak at
5575 Å, which is about the center of the visible spectrum, the part to which
the human eye is most sensitive. Yet only about 40 per cent of the radiation
of this body falls within the visible part of the spectrum, about 6 per cent
is in the ultraviolet, and the rest is in the infrared.

Gaseous sources of light when operated at low pressures emit radiation
consisting of groups of more or less sharp lines and possibly a continuous
spectrum of lesser intensity. The frequencies of the spectral lines depend on the composition of the gas; their intensities and linewidths depend on a number of factors, such as the pressure and the temperature of the gas and the method of excitation. At low pressures the lines will be sharp, but the brightness of the gas as a lamp will be low. As the pressure increases, the brightness will increase and so will the linewidths, some extending over tens of Ångströms, until at last the lines overlap and the discrete character of the spectrum disappears.

Sources of greatest brightness—greatest radiative flux in the visible—are the high-pressure arcs and flashtubes. In order to obtain maximum brightness, flashtubes are operated at an extremely high power level which they can sustain for only short periods of time. This requires intermittent operation with a low duty cycle. They are energized by discharging large capacitors ranging from 100 to a few thousand microfarads charged between 1000 and 3000 volts. Xenon tubes so activated provide a flash of the order of 1 msec with a spectral distribution approximating that of a black body between 6500 and 10,000°K temperature. According to the unpublished measurements of V. Evtuhov, between 15 and 20 per cent of the electrical input is converted into radiation in the spectral range 3500 to 6500 Å.

Light emanating from the sources discussed will be radiated in all available directions. From the flat surface of an incandescent solid it will fill a solid angle of $2\pi$ steradians (not with uniform intensity, but according to Lambert’s law!). To produce a parallel beam of radiation from the sources discussed so far, it is necessary to place the radiator in the focal plane of an optical system. Since the source is of finite size, the resulting beam will not be a parallel one, but will have an angular divergence equal to the angular size of the source viewed from one of the principal planes of the optical system. In order to get a sharp beam, only a small portion of an extended source may be utilized. In addition, not all energy radiated from this quasi point source will be utilized because the aperture of the optical system will act as an effective stop, eliminating a large part of the radiation. Therefore, it appears that only a minute fraction of the energy of an ordinary light source may be converted to a nearly parallel beam. The higher our requirements for parallelism, the smaller this fraction becomes.

A system of mirrors and lenses can be used to direct the radiation from a source onto an object. In this manner it is possible to concentrate light at a target, and we might be tempted to try to devise an optical system that would create on a surface an image brighter than the extended source from which the light originates. In the present terminology this would mean that an image of the source is formed so that the radiance at the image is higher than at the source. A famous theorem of classical optics states that
this cannot be done. More precisely, it cannot be done with Lambert law radiators if the refractive indices of the object and image space are the same. [2, p. 188.]

We may sum up the principal limitations of classical sources of light as follows:

Energy radiated from an intense source is distributed over a relatively broad spectral region. Powerful monochromatic sources do not exist.

The radiated energy is generally poorly collimated and the collimation cannot be improved without sacrificing the intensity available.

Radiation from an extended source cannot be imaged with an increase in brightness.

We shall see how these limitations are overcome in the case of coherent sources.

REFERENCES


2. UNITS AND PHYSICAL CONSTANTS

It is, in principle, most desirable to adopt one system of units and to use it consistently and exclusively. This procedure is cumbersome to follow in all details in the laser field, which encompasses atomic quantities as well as those of power engineering. It is doubtful that anyone’s preference for the rationalized MKS system extends to the point of measuring the volumes of ruby crystals in cubic meters. On the laboratory and atomic scale we shall give preference to the CGS system, thus ensuring compatibility with our sources of reference. Some deviations from the CGS system are customary in spectroscopy. The wavelength of visible radiation is most conveniently expressed in Ångström units (10^{-8} cm), whereas in the infrared the micron (10^{-6} meter) is a more convenient unit. No one measures atomic energy levels in ergs or joules. They are expressed and tabulated either in electron volts (1.602 \times 10^{-12} \text{ erg}) or in reciprocal centimeters. The use of the reciprocal centimeter as a unit of energy originates from the relationship \( \hbar \nu = E_2 - E_1 \), which may be written

\[
\frac{1}{\lambda} = \frac{E_2 - E_1}{\hbar c},
\]

(2.1)

The quantity \( E/\hbar c \) has the dimension of reciprocal length. It is frequently referred to as the energy, although correctly it should be called the wave number. The tabulation of energy levels in reciprocal centimeters enables us to obtain by direct subtraction of two tabulated entries the reciprocal of the wavelength (*in vacuo*) corresponding to a transition between the levels.
Wavelengths in the visible and the near infrared are measured in air and usually the value in air is quoted. For this reason the reciprocals of the wave numbers obtained by subtracting tabulated entries must be corrected to secure agreement with the measured values. The correction is determined by the deviation of $\eta$, the refractive index of air, from 1. Its magnitude can be gauged by the fact that $\eta - 1$ varies from $277 \times 10^{-6}$ to $274 \times 10^{-6}$ as the wavelength varies from 0.6 to 1.0 $\mu$. Conversion to wavelength in air is facilitated by the use of tables of wave numbers [1] in which wavelengths in air are listed against wave numbers. A short table of wavelength corrections appears in the Appendix (Table A.1).

As is often true, the representation convenient for the theorist is less convenient for the experimentalist. Most experimentalists working in atomic spectroscopy prefer to express their results in wavelengths measured in air. Those conditioned in the field of infrared molecular spectroscopy use wave numbers more frequently, because combinations of molecular vibrations lead to simple relations among the wave numbers and the organization of the experimental material is easier with the data displayed on a scale proportional to the frequency. Authors of articles about lasers regrettably often do not declare whether their wavelengths are measured in air or in vacuum. When the published data are not too accurate, this omission is of no consequence. Otherwise it is best to assume that a wavelength stated without qualification is a wavelength in air and that results published on the wave number scale were calculated after reduction to vacuum.

One of the useful atomic units of energy is the electron volt. One electron volt is equivalent to 8066 reciprocal centimeters. When the energy $E$ of a level is expressed in cm$^{-1}$ and it is necessary to find the energy in ergs required to raise a gram atom of the material to this state, starting from the ground level, then $E$ must be multiplied by $Nhc$, where $N = 6.02 \times 10^{23}$ is Avogadro’s number, and $hc = 1.986 \times 10^{-16}$ erg cm.

The following constants are frequently used in this book:

$$h = 6.625 \times 10^{-27} \text{ erg sec}, \quad c = 2.9979 \times 10^{10} \text{ cm/sec},$$
$$k = 1.380 \times 10^{-16} \text{ erg/deg}, \quad h = 1.054 \times 10^{-27} \text{ erg sec} [3].$$

It is sometimes desirable to relate the radiative flux density and the stored energy density of radiation to the peak values of the electric and magnetic fields. These fields are most conveniently measured in practical (MKS) units. The instantaneous power flow across a unit surface when expressed in rationalized MKS units is $W = E \times H$. The radiative flux density $\bar{W}$ is the time average of $W$. In a plane wave propagating in a medium with a dielectric constant $\varepsilon$ and permeability $\mu$ we have

$$W = \frac{1}{2} \sqrt{\varepsilon/\mu} E^2. \quad (2.2)$$
Therefore the magnitudes of the vectors concerned are related by the equation
\[ E = \sqrt{2W\sqrt{\mu/\varepsilon}}. \] (2.3)
Here \( E \) is the peak value of the electric field.

The values of \( \varepsilon \) and \( \mu \) that relate to vacuum (or air) are denoted by the subscript 0. It is known that \( \sqrt{\mu_0/\varepsilon_0} = 376.7 \text{ ohms} \). For a plane wave that creates a flux density of \( 1 \text{ MW/cm}^2 \) we obtain in vacuum a peak electric field of \( 2.74 \times 10^6 \text{ volts/meter} \). The stored-energy density is \( u = \frac{1}{2}\varepsilon E^2 \); therefore
\[ E = \sqrt{2u/\varepsilon_0} \sqrt{\varepsilon_0/\varepsilon}. \] (2.4)

For nonmagnetic materials, \( \sqrt{\varepsilon/\varepsilon_0} = \eta \), the index of refraction; therefore
\[ E = \frac{1}{\eta} \sqrt{2u/\varepsilon_0}, \] (2.5)
where \( \varepsilon_0 = 8.85 \times 10^{-12} \text{ farad/meter} \). For example, for ruby, whose refractive index is 1.76, an energy density of 0.01 joule/cm\(^3\) corresponds to a peak electric field of \( 2.70 \times 10^7 \text{ volts/meter} \). Note that these calculations are independent of the frequency. They should not be applied without further thought to situations involving other than a single plane wave. In the general case, the calculation of the energy density may involve both the electric and the magnetic vectors.

REFERENCES

3. OBSERVATION AND MEASUREMENT OF RADIATION

Anticipating the encounter with radiation originating from lasers, we survey some of the methods available for its observation and for quantitative measurement of its characteristics.

We expect to observe a highly collimated beam of radiation confined to a narrow spectral region. The intensity of the radiation is very high; frequently the radiation is available only in pulses of less than 1 msec duration. The frequency region of greatest interest includes the red portion of the visible and extends to about 3.5 \( \mu \) in the infrared.

Direct visual observation with a screen may be adequate to indicate the presence and the general shape of the beam in the case of visible radiation,
Observation and Measurement of Radiation

such as may be obtained from ruby (6943 Å) and neon (6328 Å). One must absolutely avoid looking into the laser beam directly, for severe eye damage may result from one pulse before the eyelid has time to close. Similarly, care must be taken that the laser beam does not strike a specularly reflecting object from which it may be reflected into someone's eyes. In the near infrared (up to about 1.5 μ) an image converter tube may be used to examine the laser beam visually. Farther in the infrared one may use a Liquid Crystal Viewer (Westinghouse) [1]. This device consists of a thin radiation-absorbing layer in contact with a liquid crystal layer. The color of the light scattered from the latter is a sensitive function of the temperature. The crystal layer is illuminated by white light and the color pattern is observed as the laser beam strikes the absorber.

More precise observations may be made by means of red-sensitive film or photographic plate. The laser beam may be permitted to strike the plate directly without an intervening lens. In this manner the spatial distribution of energy is recorded. One may also photograph interference patterns by interposing suitable apertures between the laser and the film.

Photographic records are convenient for recording the approximate spatial distribution of energy, but are not suitable for absolute energy or intensity measurements because the blackening of the plate has no simple relation to the radiant energy received. Energy-measuring devices in the visible and the infrared region are of either the thermal type or the photoelectric or quantum detector type.

A thermal detector absorbs the incident radiation. The energy is distributed over the receiving element as heat, and the measurement of the intensity of the incident radiation is reduced to the measurement of the temperature rise of the receiving element. Any physical property that is a known function of the temperature may serve as an indicator, and the detector may be calibrated by transferring to it a known amount of electrical energy. Thermoelectric detectors, bolometers, and calorimeters are constructed in this manner; the first two serve as indicators of intensity, the last one as an indicator of energy received during a pulse. We note that thermal detectors generally respond to incident radiation without discrimination to frequency, except that their ability to absorb may vary slowly with frequency. With moderate care, such variations may be made negligible over a wide band. Because of the finite mass of the receiving element that is heated, thermal detectors are not as sensitive as quantum detectors; moreover, the response time of thermal detectors is long compared with that of quantum detectors. For these reasons the application of thermal detectors in laser technology is limited to the measurement of the total energy of a short pulse and to the calibration of quantum detectors, which are generally useful only as indicators of relative intensity.
Calorimeters specifically constructed for the measurement of laser energy consist of an absorber of small heat capacity provided with a temperature-measuring element. The absorber may be a light carbon cone, a rat’s nest of copper wire, or razor blades assembled in a parallel stack [2, 3, 4]. The temperature rise may be measured by a thermistor, or a platinum resistance thermometer. These electrical elements are built in matched pairs and operated as parts of a balanced Wheatstone bridge. The instrument is calibrated by dissipating a known amount of electrical energy in a resistance incorporated in the absorber. The lighter absorber elements operate satisfactorily in the 0.01- to 1-joule range. Heavier instruments, such as the razor blade stack, are useful for higher energies. It is essential for accurate measurement that the laser pulse should produce only small temperature changes (a few degrees). Damon and Flynn [5] designed a liquid calorimeter for the energy range 1 to 100 joules.

As an alternative to direct energy measurement the momentum of a light pulse may be measured with a torsion pendulum, but this type of instrument is rather difficult to build [6].

Quantum detectors or photodetectors depend on the action of a light quantum on a single electron, not on the absorption and distribution of energy over an entire macroscopic body. A photodetector counts the number of quanta of radiation absorbed, in contrast to the thermal detector which responds to the total energy.

The operation of a photodetector is based on the photoelectric effect, that is, the liberation of an electron from its bound state. Free (mobile) electrons may be created in the interior of a solid by the absorption of appropriate quanta of radiation; this is the internal photoelectric effect. Alternatively, electrons may be ejected from the surface of a solid; this is the external photoelectric effect. Both effects may be used for the detection of radiation.

In the visible region the photomultiplier is the most convenient detector. It is based on the external photoelectric effect. Since it depends on the ejection of an electron from a surface by incident quanta, these quanta must have sufficient energy to overcome the work function of the photosurface. This requirement sets a long wavelength limit for the use of photodetectors. The work function of clean, pure metals has a value of several electron volts; therefore quanta of ultraviolet light are required to eject electrons from such materials. Special composite surfaces have been developed with effective work functions somewhat below 1 eV. With such surfaces the long wavelength limit to which photodetectors based on external photo effect may be used is extended to about 1.2 μ. The ratio of the number of electrons emitted to the number of photons incident is the quantum efficiency of the photosurface. The variation of this parameter
Observation and Measurement of Radiation

Fig. 1. Spectral response of various photocathodes in the visible and the near infrared region. (Source E.M.I. Electronics Ltd. catalog.)

with wavelength is shown in Fig. 1 for several commercially available photocathodes. The photoemissive surfaces form the cathodes of the photomultiplier tubes which conveniently amplify the current arising from electron emission. These tubes vary in size, in sensitivity, and in intrinsic noise which limits the detection of very weak signals. Technical data required for the selection and operation of photomultipliers are easily available in the literature [7, 8].

Internal photoeffect produces an increase in the conductivity of the material. The change in conductivity serves as an indicator of the incident radiation. The common materials usable as photoconductive sensors are
tellurides, selenides, and sulfides of thallium, bismuth, and lead, and quite a few doped semiconductors. Photoconductors are excellent detectors of infrared radiation in the 1- to 10-μ range. Their characteristics depend strongly on the temperature.

Internal photoeffect in a semiconductor produces other effects in addition to increasing the conductivity. When an electron-hole pair is created near a p-n junction the charges separate and a photovoltaic effect takes place. This means that a potential difference appears across the junction; this potential difference may be used as a measure of the photon absorption that caused the charge separation. Typical photovoltaic materials suitable as radiation detectors are InSb and InAs.

The choice of a detector in laser experiments depends not only on the frequency of the radiation employed, but also on what is to be detected. When, for example, the spatial distribution of radiation is to be detected, we choose a device suitable for image formation. A variety of television pickup cameras based on either the photoemissive or the photoconductive effect are sensitive in the near infrared. These cameras as well as numerous image converters operative between 0.8 and 1.3 μ are described by Kruse, McGlauchlin, and McQuistan [9].

When the application requires the monitoring of radiation at only one or a few points in space, and quantitative data concerning intensity are required, detectors other than image conversion devices will be used. The choice then depends on the intensity and the time variability of the signal to be detected or measured, the output device to which the detector is to be coupled, and the control the experimenter can afford to exercise over the environment of the detector. In addition to differing in intrinsic noise level, detectors of different types differ greatly in their response time, that is, in their ability to respond to rapid intensity fluctuations.

The detection of weak infrared signals developed into an advanced art during and after World War II. The theory as well as the necessary equipment are now available for the design of receivers for laser-based communications systems. The excellent and exhaustive treatises recently published about infrared detection [9, 10] make a detailed discussion unnecessary here. However, it is well to note that there are a number of criteria current in infrared technology which are intended to provide answers to the following questions:

1. What is the minimum level of power incident on the detector which gives rise to a signal voltage output equal to the noise voltage of the detector?
2. What output signal will be obtained for incident radiant power of unit intensity?
3. How does the output signal vary with the frequency of the incident radiation?

4. What is the minimum response time of the detector?

The parameters that provide at least partial answers to these questions are tabulated [7, 8, 9, 10]. We note briefly that devices based on external photoeffect (photomultipliers) are highly sensitive in the proper frequency range and provide a high output signal. Moreover, their response is fast—$10^{-8}$ sec. They are greatly limited in their spectral response, the long wavelength cutoff being around 1.2 $\mu$ at most, and their excellent characteristics deteriorate long before this limit is reached. Photoconductive detectors have a detectivity sensitivity about two orders of magnitude lower than the photomultipliers at their optimal frequency. The lead selenide, telluride, and sulfide types have response times of the order of $10^{-4}$ sec; some of the doped semiconductors, however, have shorter response times, down to $10^{-7}$ sec. The virtue of the photoconductive detectors is their response for wavelengths longer than 1 $\mu$. Their cutoff wavelengths depend on the material and on the temperature.

The 1- to 3.5-$\mu$ range can be covered adequately with detectors at room temperature, for example, photoconductive PbS or photovoltaic InAs detectors. From 2 to 6 $\mu$, detectors at liquid nitrogen temperature (77°K) are considerably superior. Beyond 6 $\mu$, special semiconductors are employed which must be refrigerated to a still lower temperature [9].

Photoelectric devices are eminently suitable for relative measurements of laser output. When absolute intensity measurements are required one must follow a tedious procedure of calibration. The integrated intensity may be measured by a calorimeter; then, if the waveform of a pulse is known, an intensity scale may be established. We shall see that the waveform of the laser is frequently very irregular. When that is so, the calibration may be accomplished by using a steady monochromatic beam of light derived from an ordinary light source as an intermediary reference. The mean frequency of this beam should approximate that of the laser. A complicating factor of this procedure is that we cannot get a monochromatic pulse powerful enough to register on the calorimeter from a classical source. It is therefore necessary to determine the absolute intensity of the auxiliary monochromatic beam by comparing its effect on a thermopile with that of the radiation derived from a standard lamp certified by the National Bureau of Standards and operated observing all precautions required for the use of such secondary radiation standards.

Wavelength is conveniently measured by means of a grating spectroscope. Difficulties arise if one attempts to determine the linewidth of a laser output with a spectroscope, especially if one attempts to resolve two or
more components separated by minute differences in wavelength, because the resolution of the spectroscope proves to be insufficient. The necessary high resolution may be obtained by means of a Fabry-Perot interferometer. This instrument is important in the laser art for reasons other than its great resolving power in spectroscopy. It deserves detailed attention.

A Fabry-Perot interferometer consists of an airspace or transparent medium* bounded by two parallel, partially transmitting mirrors. It is shown schematically in Fig. 2. Let a plane wave be incident on Mirror 1 at a small angle $\theta$ from the normal $ON$ of the mirrors. This plane wave is represented in Fig. 2 by the ray $SO$; the splitting of this wave by the first mirror is represented by the rays $OA$ and $OB$. The ray $OA$ is split again on the second mirror, and this procedure continues. The waves leaving the interferometer after a varying number of reflections finally combine into a resultant reflected and a resultant transmitted wave. The amplitude of the resultant transmitted wave is calculated by adding the amplitudes of all waves transmitted an odd number of times through the interferometer. These are represented by the rays $AC$, $A_1C_1$, and so on. Addition is performed with due regard to the phase differences resulting from passages between the mirrors. The phase difference resulting from one additional round-trip passage between mirrors separated by distance $d$ is

$$\delta = \frac{4\pi}{\lambda} d \cos \theta.$$  (3.1)

Let the (intensity) reflectivity and transmissivity of the mirrors be denoted by $R$ and $T$ respectively. Then for lossless mirrors, which we shall assume, we have

$$R + T = 1.$$  (3.2)

* For the sake of simplicity we confine ourselves to the case when the index of refraction is the same between the mirrors and outside.
The intensity $I_t$ of the light transmitted through the interferometer is then given by the equation [11, p. 324]

$$I_t = \frac{(1 - R)^2 I_i}{(1 - R)^2 + 4R \sin^2 \frac{\delta}{2}}$$

(3.3)

where $I_i$ is the incident intensity.

When light incident on the Fabry-Perot interferometer originates from a point source, the transmitted intensity will be the same in directions at equal inclination to the normal. Therefore, when the radiation emerging from the interferometer is focused on a screen, a set of concentric rings centered around the image of the normal is obtained. Such a pattern is shown in Fig. 3.

![Fig. 3. Fabry-Perot pattern obtained with red cadmium light.](image-url)
In practice $4\pi d/\lambda$ is a large number (tens of thousands); therefore, \(\sin(\delta/2)\) goes through a full cycle for a small variation of \(\theta\). Then, as \(\sin(\delta/2)\) varies from 0 to 1, the normalized transmitted intensity varies from 1 to \((1 - R)^2/(1 + R)^2\). The variation of the intensity with \(\delta\) is shown in Fig. 4 for two values of \(R\). Clearly the pattern associated with \(R = 0.87\) consists of much finer rings than the pattern with \(R = 0.64\). It is customary to define the fineness parameter as follows:

\[ F = 4R/(1 - R)^2. \]

The curves in Fig. 4 pertain to \(F = 20\) and \(F = 200\) respectively. When light containing several monochromatic components is incident on a Fabry-Perot interferometer, sets of rings pertaining to each wavelength will appear displaced with respect to each other. As the reflectivity of the mirrors, and thus the fineness \(F\), is increased, the individual rings become sharper, and the fringes produced by different monochromatic components are more easily separated. If the wavelength separation of two components is sufficiently large, the displacement of the two patterns may become greater than the distance between adjacent maxima of either pattern and the different orders may overlap. This overlapping limits the spectral range over which the Fabry-Perot interferometer may be used without auxiliary apparatus for the elimination of the overlap.

The virtue of the Fabry-Perot interferometer is its extremely high resolving power in a narrow spectral range. In the visible region lines differing by about 0.02 cm\(^{-1}\) in wave number may be resolved. For the complete theory of this instrument we refer to the literature [11].
REFERENCES


4. COHERENCE OF LIGHT

The classical theory of light describes optical phenomena in terms of electromagnetic oscillations. One of the basic tools of this theory is harmonic analysis. The variation of the electromagnetic field at a point is represented as the superposition of harmonic oscillations of the form

$$E = E_i \cos (2\pi v_i t - \varphi_i).$$  \hspace{1cm} (4.1)

Each oscillation has a definite amplitude $E_i$, frequency $v_i$, and phase $\varphi_i$. The phase varies in space from point to point in a linear manner. It is useful to think in terms of a monochromatic radiation, which is an electromagnetic oscillation of a single frequency.

In a physical experiment one always deals with superposition of harmonic oscillations of different frequencies, but it is possible to filter radiation in such a way that for most purposes it will behave as an ideal monochromatic radiation. When this is the case we call the radiation quasi-monochromatic, or briefly monochromatic. Whether a radiation is quasi-monochromatic or not depends on the experiment for which it is used.

An ideally monochromatic wave is necessarily of infinite duration; an oscillation that has the shape described by (4.1) for the finite time interval $0 < t < T$, and is zero outside of that interval, may be represented as a superposition of harmonic oscillations whose frequencies are confined to a narrow region of width approximately $1/T$ around the center frequency $v_i$.

Practically monochromatic radiation is characterized by a center frequency $v_0$ and a bandwidth $\Delta$ so defined that the frequency interval from
\[ I = I_1 + I_2 + 2\sqrt{I_1I_2} \cos \Phi, \]

where \( I_1 = \frac{1}{2}E_1^2 \) and \( I_2 = \frac{1}{2}E_2^2 \) are the (time average) intensities and \( \Phi = \varphi_2 - \varphi_1 \) is the phase difference. The latter is related to the path difference \( s_2 - s_1 \) as follows:

\[ \Phi = 2\pi v(s_2 - s_1)/c. \]

In those regions of the space where the phase difference is 0 or an even multiple of \( \pi \), the intensity is large, namely \( \frac{1}{2}(E_1 + E_2)^2 \), while at those points where the phase difference is an odd multiple of \( \pi \), the intensity is small, namely \( \frac{1}{2}(E_1 - E_2)^2 \).

A typical interference experiment is so arranged that the path difference \( s_2 - s_1 \) varies over a screen or over the field viewed by a telescope and the amplitudes \( E_1 \) and \( E_2 \) are adjusted to be nearly equal. Consequently, when a monochromatic point source is used, a series of alternating light and dark bands is observed. These are the interference fringes. The light band
Coherence of Light

which corresponds to zero path difference is particularly important; it is common to all frequencies. The positions of the other bands are frequency dependent; therefore, light fringes of one frequency overlap with dark ones of another. This overlap depends on the frequency difference and on the order of the fringe reckoned from the central light fringe which corresponds to zero path difference.

It is clear from (4.2) and (4.3) that reinforcement occurs when \( \frac{\nu(s_2 - s_1)}{c} \) is an integer, and cancellation when it is a half integer. Therefore the \( n \)th light fringe for wavelength \( \lambda \) will coincide with the \( n \)th dark fringe for wavelength \( \lambda - \Delta \lambda \) when

\[
s_2 - s_1 = n\lambda = (n + \frac{1}{2})(\lambda - \Delta \lambda).
\]

Hence when

\[
\frac{\Delta \lambda}{\lambda} = \left| \frac{\Delta \nu}{\nu} \right| = \frac{1}{2n + 1}
\]  

(4.4)

the interference pattern around the \( n \)th fringe will be seriously impaired when radiation of frequency \( \nu \) and \( \nu + \Delta \nu \) is present in approximately equal quantities.* Since actual interference experiments are performed with a quasi-monochromatic source and not an ideal monochromatic source, the number of fringes that may be clearly observed is limited by the spread of the spectrum of the source. Sometimes this fact is expressed in a different form: It is known that in a Michelson interferometer using an ordinary spectral line as a light source interference fringes are observed only when the path difference in the two branches is less than a few centimeters. When the path difference exceeds, say, 30 cm, interference is not observed. Apparently the phase of the radiation is not preserved over the length of time it takes the light to travel this distance. In an ideal monochromatic wave field the amplitude of vibrations at any fixed point is constant, whereas the phase varies linearly with time. This is not the case in a wave field produced by a real source; the amplitude and phase undergo irregular fluctuations, the rapidity of which is related to the width \( \Delta \nu \) of the spectrum. The time interval \( \Delta t = 1/\Delta \nu \) is the coherence time. During a time interval much shorter than \( \Delta t \), the radiation behaves like a truly monochromatic wave; this is not true for a longer time interval.

So far we have considered only the properties of light emanating from a point source. Such an ideal source may be approximately realized by a real source located so far from the observer that its physical dimensions are negligible compared to its distance. Now we turn to light sources of finite extent.

* The distinctness of the fringes is also affected by the finite extent of the illuminating source.
A common characteristic of all classical sources of light is the lack of coherence between light emanating from different points of the radiator. By the term coherence or spatial coherence we mean a correlation between the phases of monochromatic radiation emanating at two different points. To be precise, we ought to speak not of coherent and incoherent light but of different degrees of correlation. From a practical point of view, however, we may regard radiation emanating from two distinct sources as incoherent if we observe that the intensities of the radiation are additive. We take it as an established experimental fact that light emanating from two points of an ordinary source located well over a wavelength apart cannot be brought into interference even with extreme filtering to segregate a "monochromatic" component. We attribute this fact to a lack of correlation between the phases of distant radiators.

Related to the question of phase correlation on a light source is the question of phase correlation in a radiation field away from the source and its relation to the properties of the source as well as the geometry of the situation.

Consider more closely the electromagnetic vectors at two points $P_1$ and $P_2$ in a wave field produced by an extended monochromatic source many wavelengths removed from both $P_1$ and $P_2$. (See Fig. 5.) If $P_1$ and $P_2$ are so close to each other that the difference $SP_1 - SP_2$ between the paths from each source point $S$ is small compared to the wavelength $\lambda$, then it may be expected that the fluctuations at $P_1$ and $P_2$ will be effectively the same. Furthermore, it may be expected that some correlation will exist even for greater separations of $P_1$ and $P_2$, provided that the path difference

![Fig. 5. An interference experiment with light from an extended source.](image-url)
Coherence of Light does not exceed the coherence length $c \Delta t = c/\Delta \nu$. By using the correlation of the electromagnetic disturbance, we are led to define a region of coherence around any point in a wave field generated by an essentially monochromatic source.

The value of correlation can be tested experimentally by observing the illumination on the screen $B$ as a function of position. The observation is carried out generally so that the illumination is measured in a region which is approximately equally distant from $P_1$ and $P_2$. The illumination is determined with small apertures around $P_1$ and $P_2$. In addition to measuring the intensities $I_1$ and $I_2$ at $Q$ with the first and second apertures, respectively, opened, we may also measure interference effects arising from the superposition of radiation passing through these two apertures. Complete incoherence of the radiation field at $P_1$ and $P_2$ means that the intensity $I(Q)$ is

$$I(Q) = I_1 + I_2,$$

while in the case of complete coherence any value between $|I_1 - I_2|$ and $I_1 + I_2$ may be obtained depending on the path difference $s_1 - s_2$. In the general case an expression of the form

$$I(Q) = I_1 + I_2 + 2\sqrt{I_1 I_2} \Re \gamma$$

is obtained where $\gamma$ is a complex number of constant modulus $|\gamma| \leq 1$, whose phase varies linearly with the path difference. In fact, $\arg \gamma = \varphi_0 + 2\pi(s_1 - s_2)/\lambda$. The case $\gamma = 0$ corresponds to complete incoherence and $|\gamma| = 1$ to complete coherence, while an intermediate value of $|\gamma|$ characterizes a partially coherent field. As the point of observation $Q$ is moved parallel to the line $P_1P_2$, the intensity varies between a maximum of $I_M = I_1 + I_2 + 2\sqrt{I_1 I_2} |\gamma|$ and a minimum $I_m = I_1 + I_2 - 2\sqrt{I_1 I_2} |\gamma|$. The visibility of the interference fringes on screen $B$ is defined as

$$\nu = \frac{I_M - I_m}{I_M + I_m}.$$ 

When the intensities $I_1$ and $I_2$ are equal, $\nu$ reduces to $|\gamma|$.

The extension of the correlation concept to a polychromatic (non-monochromatic) field is quite straightforward but mathematically more demanding. The radiation must be represented in terms of Fourier integrals and cross correlation must be defined as it is in the theory of stationary random processes. This analysis is carried out in the literature [1, 2]. The result is that, given two points and a time interval $\tau$, a degree of coherence $\gamma_{12}(\tau)$ whose absolute value varies from 0 to 1 can be calculated. With this concept at hand, the variation of the degree of coherence of a wave field generated by an extended source can be discussed. The degree of coherence
of radiation between the points $P_1$ and $P_2$ can be related to the diffraction pattern of the source regarded as an aperture of specified amplitude and phase distribution. This is the substance of the van Cittert-Zernike theorem which permits the calculation of the variation of $\gamma(0)$ in a plane illuminated by an extended incoherent source.

The ideal case $\gamma(0) = 1$ represents a fully coherent plane wave with the phase front coincident with the plane of observation. This, of course, cannot be achieved with a finite incoherent source. With the point $P_1$ held fixed and $P_2$ moving away from $P_1$, the degree of coherence $|\gamma(0)|$ decreases. Arbitrarily, the tolerance limit $|\gamma| \geq 0.88$ is set to specify the region within which the radiation is called "almost coherent." With the aid of the van Cittert-Zernike theorem it can be shown that radiation derived from a uniform, quasi-monochromatic, noncoherent source of circular shape is almost coherent over a distance $d = 0.16\lambda/\alpha$, where $\alpha = \rho/r$ is the angular radius of the source as viewed from the point of observation, $\rho$ being the radius of the circle and $r$ its distance from the observer. (See Fig. 6.)

Therefore, an almost coherent beam of finite cross section can be obtained from a noncoherent source but only a minute fraction of the energy radiated can be utilized in the process. In order to obtain an almost coherent beam of 1 cm diameter at 5000 Å, the source must be so far removed optically that its angular radius $\alpha$ is $8 \times 10^{-6}$ rad. If a source of flux density $\omega_s$ and surface area $A$ radiates according to Lambert’s law, the flux density at a distance $R$ from the radiator near the normal to its surface is

$$w_p = AN = \frac{A\omega_s}{\pi R^2}.$$  

Consequently, for a circular radiator of radius $\rho$,

$$w_p = \frac{\rho^2\omega_s}{R^2}.$$  

In order to obtain coherence over 1 cm, we must have $\alpha = \rho/R = 0.16\lambda$; therefore $w_p = (0.16\lambda)^2\omega_s = 0.64 \times 10^{-16}\omega_s$. In the present case the flux
density in the almost coherent part of the beam is less than $10^{-10}$ times the flux density emitted by the source. An almost coherent beam can be focused into a region whose dimensions are of the order of the wavelength. Once an almost coherent beam is obtained, it is possible to concentrate this energy, and the degree to which this concentration is successful depends on the degree of coherence of the beam.

We can now readily appreciate some of the advantages gained by having a coherent, or almost coherent, source whose radiation is already in the form of a spherical wave or a plane wave of limited cross section. Such radiation can be concentrated by lenses and mirrors to images much brighter than the original source. Moreover, radiation emitted from a source already in the approximate form of a plane wave can be directed at a distant object with only negligible losses from diffraction effects, whereas only a small part of the radiation from a noncoherent source can be converted into an approximately plane wave.

REFERENCES


5. EMISSION AND ABSORPTION OF RADIATION BY ATOMS

It is well known that atomic systems such as atoms, ions, and molecules can exist in certain stationary states, each of which corresponds to a definite value of energy. The states are characterized by quantum numbers. The energy values are called the *levels* of the atomic system. When two or more states have the same energy, the level is called degenerate, and the number of states with the same energy is the *multiplicity* of the level. Frequently the word state is used to mean level; all states with the same energy are regarded as identical. Transitions between stationary states may occur with attendant emission or absorption of energy as radiation, or with the transfer of energy to or from another system. If the transition is radiative, the frequency of the radiation emitted or absorbed by the system is given by Bohr's frequency relation:

$$h\nu = E_2 - E_1, \quad \text{(5.1)}$$

where $E_1$ and $E_2$ are the energies of the states among which transition takes place and $h$ is Planck's constant.

The level of the system with the lowest energy is the *ground level*; every other level is an *excited level*. The terms ground state and excited state are also used. An atom in the ground level can only absorb radiation. Starting
with the ground level, we number the levels in increasing order of energy. When the atomic system is not in the ground level, it may change to a lower level with the emission of radiation without any external causation. This is the phenomenon of spontaneous emission. The probability that an atom in level \( n \) will spontaneously change to the lower level \( m \) within a unit of time is called the spontaneous transition probability. It is denoted by \( A_{nm} \). This quantity is a characteristic of the pair of energy levels in question. In multiple levels \( A_{nm} \) is obtained by summation over all pairs of states involved. If there is a large collection of atomic systems on hand and \( N_n \) is the number of systems in the \( n \)th level, the total number of transitions from level \( n \) to level \( m \) will be approximately \( N_n A_{nm} \) per second, and the power radiated at the frequency \( v_{nm} = (E_n - E_m)/h \) will be \( N_n (E_n - E_m) A_{nm} \). Spontaneous radiation will emerge from the atoms of the assembly in a random phase; therefore the assembly of independent atoms (gas) will emit this radiation as an incoherent source.

Transitions between different atomic or molecular energy levels take place not only spontaneously but also under stimulation by electromagnetic radiation of appropriate frequency. The total probability that an atomic system will change during a unit of time from a level of index \( n \) to a level of lower energy of index \( m \) is

\[
P_{nm} = A_{nm} + u_v B_{nm},
\]

where \( u_v \) is the radiation density at the frequency \( v = v_{nm} \); \( A_{nm} \) and \( B_{nm} \) are constants determined by the atomic system. In the presence of radiation of the proper frequency the atomic system may also pass from a lower to a higher energy level. The probability of such an event (absorption) is

\[
P_{mn} = u_v B_{mn}.
\]

Radiation emitted from an atomic system in the presence of external radiation consists of two parts. The part whose intensity is proportional to \( A_{nm} \) is the spontaneous radiation; its phase is independent of that of the external radiation. The part whose intensity is proportional to \( u_v B_{nm} \) is the stimulated radiation; its phase is the same as that of the stimulating external radiation. The main concern in this book is this stimulated radiation.

For the sake of simplicity, we assume at this point that the spectral extent of each atomic line is so narrow that the distribution of energy with frequency within the line is not resolved and what we observe is only the total energy emitted or absorbed. We imply here also that the radiation density \( u \) does not vary significantly over the frequency range of the spectral line.

* Strictly speaking, we are dealing with quantities that should be called rates, not probabilities. Their dimension is reciprocal time. This matter is discussed further on p. 31.
The relations between the $A$'s and $B$'s are known as *Einstein's relations*. They are usually stated in the form

$$B_{nm} = B_{mn}, \quad A_{nm} = \frac{8\pi \hbar \nu^3}{c^3} B_{nm}.$$  

These equations are valid in vacuum for particles having only non-degenerate energy levels. When the energy levels are degenerate, Einstein's first relation takes the form [1, p. 450]

$$g_n B_{nm} = g_m B_{mn},$$

where $g_n$ and $g_m$ are the multiplicities of levels $n$ and $m$, respectively. The second relation is not affected by the multiplicities. In solids in which the index of refraction $\eta$ differs appreciably from unity, the second relation must be replaced by

$$A_{nm} = \frac{8\pi \hbar \nu^3 \eta^3}{c^3} B_{nm}.$$  

The reason for the appearance of the $\eta$ is that the factor in front of $B_{nm}$ arises from the counting of the radiation modes in a volume element. The wave numbers that enter into this calculation are defined in terms of the frequency and the velocity of the radiation in the material. In fact, $k^2 = (2\pi \nu)^2 / \nu^2$, which leads to the replacement of $c$ in (5.4) by $v = c / \eta$.

We turn now to the concept of *lifetime*, which is frequently used in describing transitions between different states of an atom. The lifetime of a state is simply related to the probability of transition from that state. Let $p \, dt$ be the probability that an atom originally in state $s$ will leave that state during a short time interval $dt$. (This interval must be so short that $p \, dt \ll 1$.) Then, for a constant $p$, the number of atoms in state $s$ will decrease exponentially according to the formula $N(t) = N_0 e^{-pt}$. Hence the number of atoms leaving the state $s$ in the time interval from $t$ to $t + dt$ is $pN_0 e^{-pt} \, dt$. Therefore the average lifetime of the atom in state $s$ is

$$T = \frac{1}{N_0} \int_0^\infty tpN_0 e^{-pt} \, dt = \frac{1}{p}.$$  

In view of (5.7), the reciprocal of the transition probability of a process is called its lifetime. If an atomic state can be altered by several processes with lifetimes $\tau_1, \tau_2, \ldots, \tau_n$ and these processes are statistically independent, then the lifetime of the state is related to the lifetimes of the processes by means of which the state can be altered by the equation

$$\frac{1}{T} = \frac{1}{\tau_1} + \frac{1}{\tau_2} + \cdots + \frac{1}{\tau_n}.$$  

We shall now make a few statements about the nomenclature of the atomic energy levels and some of the rules governing the transitions.
between them. These are intended not as a brief presentation of the quantum theory of the atom, but as a recapitulation of some of the working rules of that discipline.

The electrons in an atom are characterized by three orbital quantum numbers \( n, l, \) and \( m, \) and the spin quantum number \( s. \) The orbital quantum numbers are integers; \( s \) is \( \frac{1}{2}. \) The first, \( n, \) governs the radial distribution of the wave function; it corresponds to the principal quantum number of Bohr's theory. In the case of atoms with relatively simple structure the value of \( n \) is the primary determinant of the energy of the electron in question. The quantum number \( l \) varies from \( 0 \) to \( n - 1; \) it determines the orbital angular momentum, whose largest component is \( \hbar. \) For brevity we shall say that the orbital angular momentum is \( l\hbar \) with apologies to the expert in quantum mechanics who knows that the magnitude of this vector is \( \sqrt{l(l + 1)}\hbar. \) The quantum number \( m \) describes the orientation of the angular momentum vector with respect to an external field; it may assume the values \( -l \) to \( +l; \) that is, for a fixed value of \( l \) a total of \( 2l + 1 \) values of \( m \) are possible. With different quantum numbers there are associated different wave functions, i.e., different states of the electron. When several states have the same energy, the level is said to be degenerate, as we have already noted. The degeneracy of a level may be removed by the application of an external field, or by the field of the other electrons of the same atom, or by that of neighboring atoms. This removal of degeneracy comes about most often by the appearance of a small energy difference associated with the reversal of the spin of an electron.

When an atom contains many electrons the electrons that form a closed shell may be disregarded and the energy differences associated with transitions in the atom may be calculated by considering only the electrons outside of the closed shell. Thus in the case of the alkali metals only one electron, and in the case of the alkali earths only two, need to be considered.

An electron is called an \( s, p, d, \) or \( f \) electron if its azimuthal quantum number is \( l = 0, 1, 2, \) or \( 3, \) respectively. For larger values of \( l \) the letters of the alphabet are used in their natural order starting with \( f \) for \( l = 3. \) The notation \( 3p \) indicates an electron with \( n = 3, l = 1. \)

In describing the state of a multielectron atom it is well to remember that the electrons within the atom are interchangeable and that no two electrons can have the same quantum numbers (including \( s). \) The ground state of \( Li, \) for example, is described by the symbol \( 1s^22s, \) which means that there are two electrons (with opposing spins) in the \( 1s \) state, and one in the \( 2s \) state. The sum of the orbital angular momenta of these electrons is 0, and the sum of the spin angular momenta is \( \frac{1}{2}, \) giving the total angular momentum \( \frac{1}{2}. \) All these momenta are in the units of \( \hbar = \hbar/2\pi. \)

The total angular momenta of most atoms and ions of interest to us may
field parameters; their calculation is among the most complicated tasks of quantum theory. [2, 3]

The transition probabilities between atomic states are related to the wave functions associated with these states. Our interest is in radiative transitions, i.e., transitions in which an energy exchange takes place between the atom and the electromagnetic field. The principal factor in this exchange is the electric dipole radiation, which is already familiar from classical electrodynamics. Given a distribution of charges $e_i$ with position vectors $r_i$ such that the sum of the charges is 0, the vector

$$\mathbf{\mu} = \sum_{i=1}^{n} e_i \mathbf{r}_i \quad (5.9)$$

is independent of the origin of the coordinate system. It is the electric dipole moment of the charge system. If the position coordinates undergo harmonic oscillations at the frequency $\nu$, the rate at which energy is radiated according to classical electrodynamics is proportional to $\nu^4 |\mathbf{\mu}|^2$. Here

$$|\mathbf{\mu}|^2 = |\mu_x|^2 + |\mu_y|^2 + |\mu_z|^2, \quad (5.10)$$

where

$$\mu_x = \sum_{i=1}^{n} e_i x_i, \quad (5.11)$$

In quantum theory the intensity of electric dipole radiation is calculated essentially by means of the same formula with the dipole moment replaced by the corresponding quantum mechanical matrix element. For a transition from state $q$ to state $q'$ this matrix element is defined by the equations*

$$\mu_x(q, q') = e \int \psi_{q'}^* x \psi_q \, dv, \quad (5.12)$$

$$\mu_y(q, q') = e \int \psi_{q'}^* y \psi_q \, dv, \quad (5.13)$$

$$\mu_z(q, q') = e \int \psi_{q'}^* z \psi_q \, dv, \quad (5.14)$$

$$|\mathbf{\mu}(q, q')|^2 = |\mu_x(q, q')|^2 + |\mu_y(q, q')|^2 + |\mu_z(q, q')|^2. \quad (5.15)$$

The rate (probability) of transition from $q$ to $q'$ is given by [1, p. 451]:

$$A_{q,q'} = \frac{64\pi^4\nu^3}{3\hbar c^3} |\mu(q, q')|^2. \quad (5.16)$$

When multiple levels are involved and these are designated with the letters $n$ and $m$, as at the beginning of this section, the probabilities are calculated

* The symbol $q$ represents all quantum numbers required to specify the initial state; $q'$ has similar meaning for the final state.
by summing over all terminal states and averaging over all initial states. This process leads to the expression

\[ A_{nm} = \frac{64\pi^3 v_0^3}{3h\epsilon^2} |\mu(n, m)|^2, \]

where

\[ |\mu(n, m)|^2 = \sum_{i=1}^{g_n} \sum_{j=1}^{g_m} |\mu(q_i, q_j)|^2. \]

As a consequence of the symmetry properties of the integrals (5.12) to (5.14) most combinations of states lead to a zero matrix element. Such transitions are called forbidden transitions. The term forbidden means that a transition among the states concerned does not take place as a result of the interaction of the electric dipole moment of the atom with the radiation field. The allowed transitions are specified by the selection rules, which help to sort out the pairs of states capable of yielding a non-zero matrix element for electric dipole radiation.

The parity of a state is even or odd according to the parity of the (scalar) sum \( \sum l_i \), where summation is extended over all electrons of the atom. Since the parity of a closed shell is always even, it is sufficient to sum over the orbital quantum numbers of the valence electrons.

The general selection rules are the following:

1. Transitions must change the parity.
2. \( \Delta J = 0 \) or \( \pm 1 \), but transition from \( J = 0 \) to \( J = 0 \) is excluded.

In atoms for which Russell-Saunders coupling is applicable the following additional selection rules hold:

3. \( \Delta L = 0 \), or \( \pm 1 \);
4. \( \Delta S = 0 \).

When the state of only one electron changes, the parity rule requires a change in \( L \). In this case the third rule becomes \( \Delta L = \pm 1 \). Transitions in which the states of two electrons change at the same time are considerably less probable than transitions involving a single electron.

In order to facilitate the application of the selection rules terms of odd parity are usually provided with an upper index \( o \). For the same reason it is practical to list or plot the energy levels of singlets, triplets, etc. in separate groups. Such groupings are illustrated on the example of He in Fig. 7. An abbreviated notation is employed in this figure; the symbols \( n^1S \) and \( n^1P^o \) stand for \( ns^1S \) and \( np^1P^o \) respectively.

The selection rules were derived by considering electric dipole radiation only. Other radiative mechanisms are less effective, they lead to less frequent transitions in general than electric dipole radiation. However,
when the electric dipole transition is forbidden, secondary mechanisms lead to transitions rates that are generally several orders of magnitude slower than the rates of permitted dipole transitions. Thus transitions forbidden by the selection rules will occur, but they will occur relatively rarely. The selection rules are most rigidly in force for elements at the beginning of the periodic table. They lose their effectiveness in complex atoms and in a strong interaction of an atom with another as it takes place in a collision or in a crystal lattice.

As a result of the selection rules, an atom may get into an excited state from which it will have difficulty returning to the ground state. A state

Fig. 7. Energy-level diagram for helium. Typical permitted transitions are shown by slant lines. The scale on the left is in electron volts measured from the ground state; that on the right is in wave numbers measured from the ionization level.
from which all transitions to lower energy states are forbidden is metastable; an atom entering such a state will generally remain in that state much longer than it would in an ordinary excited state from which an easy escape is possible. A metastable state has a long lifetime. An example of such a state is the lowest triplet state of He.

The purpose of this section was to introduce the concepts used in the discussion of the stimulated emission of radiation and to establish the basic relationships between the Einstein coefficients and the quantum mechanical matrix elements.

We reviewed some of the basic properties of atomic structure, introducing the spectroscopic nomenclature for the most common energy levels, and indicated some of the complications that arise under less usual circumstances, for example, when an atomic system is embedded in a crystal field. We stated the simplest selection rules applicable to electric dipole transitions. These concepts are constantly used in laser literature. An interested reader not already familiar with these subjects will find an elementary discussion of atomic energy levels and selection rules in Herzberg’s book [4] on atomic spectra. The quantum mechanical background is available in any of the numerous books on modern physics and quantum theory.

REFERENCES

6. INTERACTION OF RADIATION WITH ATOMIC SYSTEMS

In a practical situation observations are made not on a single atom but on a collection containing billions of atoms not necessarily in the same state.

Given a large number \( N_0 \) of atoms, it is known that in thermal equilibrium at absolute temperature \( T \) the distribution of these atoms among the different states will follow Boltzmann’s law; that is, the number of atoms in state \( j \) will be

\[
N_j' = \frac{N_0 e^{-E_j/kT}}{\sum_i e^{-E_i/kT}},
\]

where \( E_j \) is the energy in state \( j \). All states of the same level will be equally
populated; therefore the number of atoms in level $n$ is $N_n = g_n N'_n$, where $N'_n$ refers to the population of any of the states in level $n$. It follows, then, from (6.1), that the populations of the energy levels $n$ and $m$ are related by the formula

$$
\frac{N_n}{g_n} = \frac{N_m}{g_m} e^{-(E_n - E_m)/kT}.
$$

(6.2)

At absolute zero all atoms will be in the ground state. Thermal equilibrium at any temperature requires that a state with a lower energy be more densely populated than a state with a higher energy.

Consider now an ensemble of atoms initially at absolute zero. This ensemble will absorb only radiation whose frequency is contained in the sequence $(E_i - E_j)/\hbar$, where $i = 2, 3, \ldots$. If the ensemble is at equilibrium at a finite temperature $T$, then not only the ground state will be populated; consequently, radiation whose frequency corresponds to a transition between excited states may also be absorbed. As a matter of practical fact it is well to remember that the first excited levels of most atoms and ions are at least $2 \times 10^{-12}$ erg above the ground level and that for $T = 500^\circ$K the product $kT$ is approximately $0.07 \times 10^{-12}$ erg. Therefore, at moderate temperatures, generally, few atoms will occupy even the first excited level compared to the number present in the ground state, because the exponential factor in (6.2) is so small. The absorption of radiation requiring the transition from an excited level will be weak, for the number of transitions from the $n$th to the $m$th level is proportional to $N_n$.

As a consequence of the absorption of radiation, the equilibrium of the ensemble will be disturbed. Let us assume that monochromatic radiation is absorbed. Atoms that become excited above the first excited level by the absorption of radiation may return directly to the ground state by spontaneous or stimulated radiation or they may follow another path and change to a lower level other than the ground level. In this manner they may cascade down on the energy scale, emitting at each step radiation different in frequency from that which originally lifted them out of the ground state. Because of the relationship (5.1) connecting energy and frequency, the radiations emitted in the cascade process, which is called fluorescence, have lower frequencies than the exciting radiation.

Consider now an ensemble that is not necessarily in thermal equilibrium and again designate the number of atoms per unit volume in state $n$ by $N_n$. Assuming $n > m$, what is the response of the ensemble to collimated radiation of frequency $\nu_{nm}$ and density $\nu$? The number of downward

$*$ Actually the density $\nu$ should have the subscript $\nu_{nm}$ to indicate that we are discussing the radiation density within a spectral region surrounding the frequency $\nu_{nm}$. To simplify notation we drop subscripts in every instance where there is no ambiguity concerning the transition to which it refers.
Consider a parallel beam of light of frequency between $v$ and $v + dv$ and intensity $I_v$, traveling in the positive x-direction through a layer of atoms bounded by the planes $x$ and $x + dx$. Let the velocity of light in this medium be $v = c/\gamma$, where $\gamma$ is the index of refraction. The phase front will travel through the slab of thickness $dx$ in the time $dt = dx/v$. Suppose there are $N_1$ atoms/cm$^3$ in level 1, of which $dN_1$ are capable of absorbing in the frequency range $v$ to $v + dv$, and $N_2$ atoms/cm$^3$ in level 2, of which $dN_2$ are capable of emitting in the same range. Then, as the phase front advances from $x$ to $x + dx$, the decrease of energy in the beam is

$$-d(I_v dv) = h\nu (B_{12} dN_{1v} - B_{21} dN_{2v}) I_v \frac{dx}{v}. \quad (6.4)$$
Here we have made use of the fact that radiation emitted by stimulation is coherent with the stimulating radiation; therefore it will reinforce the beam. Radiation emitted spontaneously does not contribute significantly to the beam because it is not collimated. From (6.4) it follows that

\[
\frac{-1}{T_v} \frac{dI_v}{dx} dv = \frac{hv\eta}{c} (B_{12} dN_{1v} - B_{21} dN_{2v}).
\]  

(6.5)

We now recognize the left-hand member as \( k_v dv \), and by integration over the entire line centered around \( v_0 \) we obtain

\[
\int k_v dv = \frac{h\nu_0\eta}{c} (B_{12}N_1 - B_{21}N_2).
\]  

(6.6)

Here \( B_{12} \) may be eliminated by means of (5.5) and \( B_{21} \) may be expressed in terms of \( A_{21} \) by making use of (5.6). In this manner we get the Füchtbauer-Ladenburg formula:

\[
\int k_v dv = \frac{c^2A_{21}g_2}{8\pi\nu_0^2\gamma g_1} \left( N_1 - \frac{g_1}{g_2}N_2 \right).
\]  

(6.7)

This is a basic formula, which we write as

\[
\int k_v dv = \kappa \left( N_1 - \frac{g_1}{g_2}N_2 \right).
\]  

(6.8)

The constant \( \kappa \) may be written in various forms, one of which is

\[
\kappa = \frac{\lambda_0^2A_{21}g_2}{8\pi\eta^2g_1},
\]  

(6.9)

where \( \lambda_0 \) designates wavelength in vacuum.

Consider now the case in which level 1 is the ground level. In electrically excited gases, or in a laser material excited by means of intense irradiation, the number of excited atoms \( N_2 \) may become an appreciable fraction of the total number of atoms. In this case the last term in (6.8) cannot be neglected. However, when the only agency responsible for the formation of the excited atoms is the absorption of a beam of moderate intensity, the ratio \( N_2/N_1 \) is exceedingly small. Consequently, (6.8) may be written

\[
\int k_v dv = \kappa N_0.
\]  

(6.10)

where \( N_1 \) is identified with the total number \( N_0 \) of atoms present. It is interesting to note that under the conditions described the integral of the absorption coefficient is simply proportional to the number of atoms present and that its value is completely independent of the line shape (the
curve in Fig. 9). The constant $\kappa$ is the integrated absorption cross section per atom for the line in question; the quantity $\sigma_v = k_v/N_0$ is the absorption cross section per atom.

When the material is in thermal equilibrium, the distribution of atoms among the levels is described by (6.2). For any positive value of the absolute temperature we get

$$\frac{N_i}{g_i} > \frac{N_j}{g_j}$$

whenever the inequality $E_i > E_j$ holds.

The nonequilibrium situation in which the inequality (6.11) is reversed is frequently referred to in the literature as a state of negative temperature. A negative value of $T$ is calculated from the distribution of atoms among the energy levels by means of Boltzmann's formula (6.2). The idea is applicable only to a pair of levels, and it arises from the use of this formula in connection with a pair of levels in a system not in thermal equilibrium.

Temperature in this connection does not have its customary meanings: $kT/2$ is not the average energy of the system per degree of freedom, and nothing can be inferred from the value of $T$ about the distribution of the population in states other than the pair from which this negative value of $T$ was calculated. I feel that the use of the term "negative temperature" does not facilitate the understanding of nonequilibrium phenomena, and therefore will avoid the use of this term and speak of population inversion instead. The term negative temperature is introduced here merely to provide a connection with the language of the pertinent literature. The expression that "the negative temperature $T$ is established for levels $n$ and $m$" means nothing other than that $E_n > E_m$, $N_n > N_m$, and the value of $T$ is defined by (6.2).

When population inversion takes place for levels $n$ and $m$, formula (6.8) gives a negative value for the integrated absorption coefficient. We have a condition of negative absorption; that is, we have amplification. Negative absorption, or amplification, is the consequence of the excess of stimulated radiation over absorbed radiation. In a material that is in the condition of negative absorption for a frequency region, an incident light wave will grow according to the law (6.3), which in this case represents an exponential growth at the rate of $\alpha = -k_v$.*

* In complex materials, typically in semiconductors, the intensity of a light beam may be diminished by processes other than the one contemplated here. Light might be scattered, for example. When such additional loss mechanisms are present, population inversion may not always lead to amplification because the gain due to the excess of stimulated emission over absorption may be cancelled by losses of other kinds. In this case $N_n/g_n > N_i/g_1$ is necessary but not sufficient for negative absorption.
It is desirable now to take a closer look at the shape of the absorption and emission lines observed from an assembly of atomic systems such as a gas. Figure 9 gives a general indication of what is observed; the exact shape will vary from one situation to another depending on the principal causes of line broadening.

The natural or intrinsic linewidth of an atomic line is extremely small. This is the linewidth that would be observed from atoms at rest without interaction with one another. There is a theoretical limit for linewidth under such circumstances, but this may be disregarded in most instances because it is small compared with the broadening effects of other causes which are invariably present. The two major factors of line broadening are the frequency variations resulting from the thermal motion of the atoms and those resulting from the interruption of absorption or emission of radiation by atomic collisions.

The thermal motion of the atoms is the cause of the Doppler (or homogeneous) broadening whose frequency dependence is calculated as follows: The probability that a fixed (say $x$) component of the velocity of an atom in a gas at absolute temperature $T$ is between $v_x$ and $v_x + \Delta v_x$ is proportional to $[\exp - \left(\frac{m v_x^2}{2kT}\right)] \Delta v_x$. The Doppler shift in frequency is related to the relative velocity $v_x$ toward the observer according to the equation:

$$\frac{v - v_0}{v_0} = \frac{v_x}{c},$$

(6.12)

where $c$ is the velocity of light. Therefore the Doppler effect gives rise to the following Gaussian frequency distribution:

$$P(v) \, dv = P_0 e^{-\beta (v - v_0)^2/v_0^2} \, dv, \quad \text{normal distribution}$$

(6.13)

where $\beta = mc^2/2kT$.

The constant $P_0$ is determined from the requirement that the integral of the probability distribution $P(v)$ over all frequencies must be 1; therefore

$$P_0 = \frac{c}{v_0} \sqrt{\frac{m}{2\pi kT}}.$$

(6.14)

The width of the distribution (6.13) at half power is

$$\Delta v = 2 \frac{v_0}{c} \sqrt{\frac{2kT \log 2}{m}}.$$

(6.15)

Here $m$ is the mass of the molecule. One may introduce the molecular weight $M = N_0 m$ and the gas constant $R = kN_0$ by multiplying the atomic quantities by Avogadro's number. Upon substitution of the
proper numerical values, the following formula is obtained for the Doppler broadening of spectral lines:

\[ \Delta \nu = 7.162 \times 10^{-7} \sqrt{\frac{T}{M}} \nu_0. \]  

(6.16)

It is to be noted that this linewidth, for a given line, depends only on the temperature of the gas.

The second major cause of line broadening is the collision of radiating particles (atoms or molecules) with one another and the consequent interruption of the radiative process. A finite wavetrain is never purely monochromatic; the spectrum of a wavetrain is spread in inverse proportion to the length of the train in the time domain. As an atomic collision interrupts either the emission or the absorption of radiation, the long wavetrain which otherwise would be present becomes truncated. After the collision the process is restarted without memory of the phase of the radiation prior to the collision. The result of frequent collisions is the presence of many truncated radiative or absorptive processes. The linewidth of the radiation of this aggregate is, of course, greater than that of an individual uninterupted process. The lineshape, i.e., the distribution of frequencies, must be computed statistically.

The original classical computation of this kind was carried out around the turn of the century by H. A. Lorentz, who showed that when the frequency of collisions is small compared to the undisturbed frequency \( \nu_0 \), the following expression describes the frequency distribution of the collision-broadened line:

\[ g(\nu)_c = \frac{\Delta \nu}{2\pi (\nu - \nu_0)^2 + (\Delta \nu/2)^2}. \]  

(6.17)

Here \( \nu_0 \) is the center frequency, and \( \Delta \nu \) is the width between the half-power points of the curve. The factor \( \Delta \nu/2\pi \) assures normalization according to area under the curve:

\[ \int_{-\infty}^{+\infty} g(\nu) \, d\nu = 1. \]  

(6.18)

Actually, formula (6.17) is an approximation valid in the vicinity of \( \nu = \nu_0 \). The complete Lorentz formula is of the form

\[ g(\nu) = \frac{\Delta \nu}{2\pi} \left[ \frac{1}{(\nu - \nu_0)^2 + (\Delta \nu/2)^2} + \frac{1}{(\nu + \nu_0)^2 + (\Delta \nu/2)^2} \right], \]  

(6.19)

but the second term in the brackets is negligible in the optical range, where the linewidth \( \Delta \nu \) is much smaller than the central frequency. The linewidth
$\Delta \nu$ is related to the average time $\tau$ which elapses between consecutive interrupting collisions:

$$\Delta \nu = \frac{1}{\pi \tau}.$$  \hfill (6.20)

Since the frequency of collisions is proportional to the density of the gas, the Lorentz linewidth is proportional to the density.

Although the derivation of Lorentz's formula was based on the simple classical model of a radiating dipole, Van Vleck and Weisskopf [2] have shown that the final result, as embodied in (6.17), is still valid in the optical region in the vicinity of $\nu_0$, even after the requirements of quantum theory and other refinements are taken into consideration.

The Lorentz-type broadening is frequently referred to as inhomogeneous broadening. Although both homogeneous and inhomogeneous broadening result in bell-shaped curves for the distribution of frequencies, these curves are quite different. The difference is illustrated in Fig. 10, which shows

![Fig. 10. Gaussian and Lorentz lines of common linewidth. ($G_p$ and $L_p$ denote the peak intensities.)](image-url)
Interaction of Radiation with Atomic Systems

Gaussian and Lorentz-type curves of the same linewidth plotted on the same scale. The peak values of these curves are related to the linewidths as follows:

for the Gaussian curve

\[ G_p = \frac{2}{\Delta \nu} \sqrt{\frac{\log 2}{\pi}} = \frac{0.939}{\Delta \nu}, \]  

(6.21)

for the Lorentz curve

\[ L_p = \frac{2}{\pi \Delta \nu} = \frac{0.637}{\Delta \nu}. \]  

(6.22)

The peak of the Gaussian curve exceeds that of the Lorentz curve by almost 50 per cent. In an actual situation factors producing both homogeneous and inhomogeneous broadening may be present at the same time. The combination of these factors leads to more complex line shapes for which we refer to the literature [1, 3]. Frequently one of the factors predominates, and when that is the case, calculations based on that factor alone will lead to approximately correct results.

We return now to the process of amplification in a material with population inversion. When a beam propagates in the x-direction, its intensity varies according to the formula

\[ I = I_0 e^{\alpha x}, \]  

(6.23)

where \( \alpha = -k_v \) is a function of the frequency and the integral of \( k_v \) is determined by the population inversion according to equations (6.7), or (6.8) and (6.9). The integrated value of \( \alpha \) is related to the population distribution in the following manner:

\[ \int \alpha_v \, dv = \kappa N_0 \left( \frac{g_1}{g_2} N_2 - \frac{N_1}{N_0} \right), \]  

(6.24)

where \( \kappa N_0 \) is the integrated absorption of the unexcited material which can be determined readily. The quantity

\[ N = \frac{g_1}{g_2} N_2 - N_1 \]

is called the population inversion. Most useful is the quantity \( n = N/N_0 \), the relative population inversion, which is \(-1\) for the totally unexcited material and \(0\) for the material which neither absorbs nor amplifies. With the introduction of \( n \), equation (6.24) takes the form

\[ \int \alpha_v \, dv = \kappa N_0 n. \]  

(6.25)
The actual value of $\alpha_v$ for a given frequency $\nu$ depends on the lineshape, but as long as the lineshape does not change, we have

$$\alpha_v = k_{\nu_0}n,$$  \hfill (6.26)

where $k_{\nu_0}$ is the absorption of the unexcited material at the frequency $\nu$. The peak values are $0.939kN_0n/\Delta\nu$ for the Gaussian line and $0.637kN_0n/\Delta\nu$ for the Lorentz line.

REFERENCES