Introduction to Laser Physics

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

JOHN WILEY AND SONS, INC.
New York · London · Sydney
General Description and Theory of Lasers

1. THE LASER

The laser, by definition, is a device that amplifies light by means of stimulated emission of radiation. In practice, a laser is generally used as a source or generator of radiation. The generator is constructed by adding a feedback mechanism in the form of mirrors to the light amplifier. The basic physical problem is the creation of a material with a sufficient degree of negative absorption for some frequency, so that adequate amplification is available to overcome incidental losses of a device and to deliver useful power. We shall now describe the original ruby laser more or less as it was invented by Maiman [1, 2], and we shall introduce lasers of several other types.

The description will be brief and preliminary. Its purpose is to prepare the reader for the analytical discussions that follow. These in turn are necessary to provide a foundation for an adequate presentation of the current state of the art. Therefore we shall not attempt completeness or precision in this section, and we shall avoid digressions pertaining to improvements of the laser. We shall return to this subject in later chapters to provide an accurate and up-to-date picture.

The working element of the ruby laser is a cylinder of pink ruby containing 0.05 per cent chromium. The cylinder is usually between $\frac{1}{2}$ and 1 cm in diameter and 2 to 10 cm long; the end faces are plane and parallel to a high degree of accuracy. One of the end faces is provided with a completely reflecting surface; the other is partially reflecting. The ruby is irradiated on its side by light from a flashlamp operated usually for a few milliseconds at a time with an input energy of 1000 to 2000 joules. The schematic diagram of a typical configuration is shown in Fig. 11, a photograph of the essential parts in Fig. 12. Most of the input energy is dissipated as heat; a fraction of it, however, is emitted by the flashlamp as blue and green radiation, which is absorbed by the ruby. This energy provides the excitation. The ruby funnels the energy, which it absorbs over a broad spectral...
region, into a narrow emission line of the trivalent chromium ion around 6943 Å. The radiation emerges coherently through the partially reflecting end of the ruby.

Actually, this coherent radiation does not appear immediately at the beginning of the excitation. At first, fluorescent radiation of about the same wavelength appears. This is broader in spectral content; it radiates in all directions, and there is no coherence between radiation arising from different points of the ruby. This radiation arises by spontaneous transitions in the chromium ions of ruby. Unless the exciting radiation is sufficiently intense, this fluorescent radiation will be the only one emitted from the ruby. When, however, the exciting radiation exceeds a certain threshold, coherent radiation appears through the end with the partially reflecting surface approximately 0.5 msec after the start of the irradiation.
In the simplest case the partially reflecting surface is a phase front of the radiation that emerges perpendicular to the end face. The intensity of this radiation exceeds that of the spontaneous radiation by several orders of magnitude, and the spectral range of the coherent radiation is considerably narrower than that of the fluorescence. The narrowing of the linewidth is due to the effect of the resonant cavity formed by the mirrors. The exact laser linewidth depends on numerous circumstances, which are discussed in Sections II.3 and II.4.

A simple laser of the kind described may produce a radiant flux density of several kilowatts per square centimeter, and in a more elaborate laser the flux density may reach several megawatts per square centimeter. All this radiation is in the spectral interval about 0.1 \( \text{Å} \) wide.

Let us contrast this with black-body radiation, making use of the radiation laws introduced in Chapter I. By means of (1.4), we find that the black body whose peak emission is at 6943 \( \text{Å} \) has a temperature of 4174\(^\circ\)K. The total radiation of this body in the entire spectrum is approximately 1700 watts/cm\(^2\), as calculated from (1.3). However, we find from (1.2) that only 0.016 watt/cm\(^2\) falls into an 0.1-Å interval around the peak of the emission.

The variation of the intensity of the laser beam described is irregular. In experiments with the early lasers it was observed that even the time delay between the start of excitation and the onset of coherent oscillations was variable, in spite of efforts to keep all experimental variables under control. Once coherent light appeared, its intensity varied greatly and irregularly, with spikes or pulsations of duration of about 1 \( \mu\text{sec} \). The intensity and frequency of these pulsations depend on the temperature of the ruby. Their existence may be explained by the observation that, once stimulated emission sets in, the downward transitions proceed at a rate greater than that at which atoms are excited to the state from which the stimulated emission originates. When this occurs, the stimulated emission may drive the population of this upper state below the level at which the process first sets in, and thus the stimulated emission stops.

The coherence of the light radiated through the partially transmitted mirror can be inferred from the angular distribution of the radiation. The fluorescence of ruby with excitation below the threshold is essentially nondirectional, but with excitation above the threshold the beam is confined to a cone of the order of one degree or less in diameter.

Diffraction experiments with an aperture placed at the mirror result in the anticipated Fraunhofer diffraction pattern for an aperture illuminated by wavefronts approximately plane and coherent.

The processes of fluorescence and stimulated emission in ruby are readily comprehended with the aid of Fig. 13, which is the energy-level diagram of
Fig. 13. Simplified energy-level diagram of chromium ions in ruby.

a three-level fluorescent solid. The ground state is denoted by index 1. Excitation is supplied to the solid by radiation of frequencies which produce absorption into the broad band 3. Most of the absorbed energy is transferred by fast radiationless transitions into the intermediate sharp level 2. The emission of radiation associated with the spontaneous return from level 2 to ground level is ordinary fluorescence. Such fluorescence will take place even at a low level of excitation. When the exciting radiation is sufficiently intense, it is possible to obtain more atoms at level 2 than are left at ground level. The spontaneously emitted photons traveling through the crystal will stimulate additional radiation, and thus induced emission is superposed on the spontaneous emission. Stimulated emission will also take place when the population of the ground level is larger than that of level 2, but the stimulated absorption will be larger than the emission and the net result will be a loss in the number of photons.

Ordinary fluorescence acts as a drain on the population of level 2. After the exciting radiation is extinguished, level 2 is emptied by fluorescence at a rate that varies from material to material. In ruby, at room temperature, the lifetime of level 2 is 3 msec. When the ruby is excited by a light flash, no laser output is obtained for the energy invested in removing, by excitation, one-half of the atoms from the ground level. This inefficiency is an intrinsic property of every three-level solid laser. To overcome it, it is necessary to utilize a material possessing four levels capable of participating in laser action. The schematic energy-level diagram of such a material is shown in Fig. 14. In this case there is an additional normally unoccupied
level above ground level at which the relevant transitions terminate; therefore laser action can begin as soon as there is significant occupation of the initial level, which in this diagram bears index 3. A pulsed four-level laser need not operate with the energy waste of a three-level laser. Sorokin and Stevenson [3, 4, 5] constructed the first four-level laser, utilizing uranium or samarium ions embedded in a calcium fluoride crystal. These and others are described in Chapter III.

Laser action is possible only if the material can be placed in a condition of amplification (negative absorption) for some wavelength region and if, in addition, a minimum feedback is established by means of reflectors. Negative absorption takes place in a stationary, nonequilibrium state that depends on the rate at which excitation is provided and also on the rates of relaxation and transition, stimulated and spontaneous, that govern the passage of atoms through the cycles illustrated in Figs. 13 and 14.

It is necessary, in general, that the rate of radiationless transfer from the uppermost level to the level at which the laser action begins be fast compared with the other spontaneous transition rates in a three-level laser. In a four-level laser there is the additional requirement of a rapid rate of spontaneous transition from the laser terminal level to ground level; without this, a phenomenon analogous to the stopping up of a drain will occur. In a four-level laser we must also guard against overpopulation of the terminal state by electrons raised from the ground level by thermal fluctuations. Hence the separation of the terminal level from the ground level must exceed $kT$. If this is not the case at room temperature, the laser must be refrigerated.
The top levels in Figs. 13 and 14 are shown as broad bands in contrast with the other levels. This breadth at the top level is a practical necessity because there is not enough energy available from ordinary sources of radiation in a narrow band. If a laser were to be used to excite a second laser, a material with a narrow top level would be acceptable. Under ordinary circumstances powerful flashlamps are pushed to the limit of their capabilities to provide sufficient excitation for materials, such as ruby, that are capable of utilizing incident radiation from 3800 to about 6100 Å.

Solid-state lasers generally operate intermittently. The reasons for this are mostly technical. First, it is difficult to provide a sufficiently powerful source of exciting light capable of continuous operation; second, a great deal of heat evolves within the laser which must be dissipated. Ordinary ruby lasers are excited for periods of a few milliseconds, the length of the period being determined by the duration of the exciting flash.

In addition to chromium ions, which are the active ingredients in ruby, laser action has been observed in uranium, and in most rare-earth ions embedded in solids. These are four-level laser materials; the first continuously operating solid laser was of this type.

The lasers described so far operate by optical excitation.

Stimulated emission may also be obtained from semiconductors. Highly doped GaAs p-n junctions have been found to exhibit laser action when a large current passes through a junction refrigerated to 77 °K. The radiation is due to transitions of injected electrons from low-lying levels of the conduction band to the uppermost levels of the valence band; the frequency corresponds to the band gap energy. In this case the primary source of energy is not irradiation but an electric generator which causes the injection of the electrons into the semiconductor.

Gases offer interesting possibilities as laser materials because their atoms are more accessible for excitation by a variety of means. In addition to optical excitation, which is not very effective for gases, there are at least three other processes capable of producing population inversion in a gas. Collision by electron impact in a suitable discharge will in certain gases such as pure neon, argon, krypton, and xenon produce population inversion for some energy levels. Laser operation has been obtained in this manner in the infrared region in all these gases. By far the most effective method of achieving laser action involves the use of a gaseous mixture, such as helium and neon, in which there occurs, in addition to excitation by electron impact, a transfer of excitation between colliding atoms of different kinds. The upper state of neon is populated by the transfer of excitation from helium atoms in a metastable state. Finally, collisions between atoms and molecules which lead to dissociation may leave one of the dissociated fragments in an excited state from which a radiative
transition to a lower unpopulated level is possible. Stimulated emission has been produced in several elements in this manner.

We defer the detailed discussion of the considerable variety of gas lasers until Chapter V and confine ourselves here to the brief introduction of the helium-neon laser. The first such laser was built in 1960 by Javan, Bennett, and Herriott [6] at Bell Telephone Laboratories. It consists of a discharge tube 100 cm long with an inside diameter of 1.5 cm filled with helium at 1 torr pressure and with neon at 0.1 torr. Flat reflector plates, which must be adjusted parallel within a few seconds of arc, are included in the gas-filled section of the tube. A simplified diagram of this laser is given in Fig. 15, while Fig. 16 is the photograph of Javan’s original apparatus. More complete drawings and pictures showing the adjusting mechanism have been published by Herriott and Schawlow [7, 8]. The technical complications in the fabrication of this laser arise mostly from the inclusion of the bellows for the adjustment of the mirrors. These complications may be avoided by the simple expedient of using external reflectors, but when external reflectors are used care must be taken to minimize unwanted reflections from the glass or quartz windows at the ends of the tube. This is accomplished by terminating the tube with optical flats oriented at the Brewster angle.

The radio-frequency (rf) generator shown in Fig. 15 is usually operated in the 25- to 30-Mc region. It serves to establish an electric discharge in the gas, although a d-c discharge may serve the purpose as well. In the discharge electrons acquire energy from the driving electrical source; a part of their energy is communicated to the helium and neon atoms. The relevant stimulated transitions take place in neon; the presence of helium serves to aid in the establishment of population inversion in neon.
Javan and associates originally observed laser action at five different frequencies which correspond to wavelengths between 1.12 and 1.21 \( \mu \). The strongest stimulated emission occurs at \( \lambda = 1.1523 \mu \). Laser action was later observed in neon at many other frequencies. The most notable neon laser line has a wavelength of 6328 Å; it provides the most convenient, steady source of coherent radiation in the visible region of the spectrum. The conditions for the excitation of different neon laser lines are similar; one may change from the 1.15-\( \mu \) laser to the 6328-Å laser by replacing the reflectors which have a high reflectivity in the near infrared by reflectors with high reflectivity between 6300 and 6400 Å.

The great value of helium-neon lasers is their remarkable monochromaticity and stability under carefully controlled experimental conditions. The power output of helium-neon lasers is of the order of 1 mW, but this power level remains fairly constant and may be maintained essentially indefinitely.

We began this book with the statement that the laser is a light amplifier, yet so far our discussion has pertained only to light generation. Amplification was described as a part of generation, but not independently. Those readers who recall the exceptionally fine properties of the maser as a
microwave amplifier might wonder why the laser is not employed as an amplifier of an incident light of low intensity.

The reason for the unsuitability of a laser as a low-level amplifier is its extremely high noise level. The second of Einstein's relations (5.4) in Chapter I shows that the rate of spontaneous emission is proportional to \( \nu^3 \) times the rate of stimulated emission. In an amplifier the spontaneous emission is the noise; the stimulated emission is the amplified signal. In the microwave region the noise is extremely low, but changing to the optical region changes the frequency by a factor of about \( 10^4 \); therefore the signal-to-noise ratio deteriorates by a factor of about \( 10^{12} \). Under these circumstances the amplifying properties of a laser are useful only in situations in which the high noise is not objectionable. This is the case when the input signal arises from another laser and is already very intense. Further amplification can then be accomplished as demonstrated by Kisliuk and Boyle [9], who amplified the output of a ruby laser oscillator by means of a second ruby which was naturally not provided with reflecting mirrors.

REFERENCES

2. THRESHOLD CONDITION AND RATE EQUATIONS

The basic ideas concerning absorption and emission of radiation in an atomic ensemble were introduced in Section I.6. We defined an absorption coefficient \( k_\nu \), a quantity representable in the form

\[
k_\nu = K g(\nu; \nu_0),
\]

where \( K \) is the integrated absorption for the entire line

\[
K = \int k_\nu \, dv,
\]
and \( g(v; v_0) \) is the normalized lineshape. The quantity \( K \) is related to the distribution of atoms among the energy levels and to the lifetime of the relevant spontaneous transition. The relation is equation (6.7) of Chapter I. The peak value of the absorption coefficient is \( k_m = Kg(v_0; v_0) \); it depends on the integrated absorption and the lineshape. For Lorentz lineshape we have as a consequence of equation (6.22) of Chapter I

\[
k_m = \frac{2K}{\pi \Delta \nu}.
\]

(2.3)

For a given integrated absorption the peak value of the absorption coefficient is inversely proportional to the linewidth. If the peak absorption \( k_m \) of the material having a certain distribution of population among the levels is compared with the peak absorption \( k_0 \) of the unexcited material \( N_1 = N_0, N_2 = 0 \), then it follows from (6.8) and (6.10) that

\[
\frac{k_m}{k_0} = \frac{N_1 - (g_1/g_2)N_2}{N_0},
\]

(2.4)

provided that the lineshape remains the same.

Consider now an aggregate of atoms in which the natural population distribution has been inverted and which has the property of negative absorption in some frequency range. The necessary condition is

\[
\frac{N_i}{g_i} < \frac{N_j}{g_j},
\]

(2.5)

and \( E_i < E_j \) for some pair of levels \( i \) and \( j \). In such a material \( k_\gamma \) is negative in the vicinity of the frequency \( v_0 = (E_j - E_i)/h \); the intensity of a parallel beam of the proper frequency will grow according to the equation

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

(2.6)

where \( \alpha = -k_\gamma \). Under certain circumstances this amplification of light may lead to light generation or laser action.

We have seen that the laser as a device consists of a pair of parallel mirrors, between which is a piece of material that is brought into a condition of negative absorption for some frequencies. This device is represented schematically in Fig. 17, where the reflectors are shown detached from the amplifying material (ruby). However, this separation is not necessary. In order to obtain a power output from the laser, at least one of the reflectors is made partially transmitting. Transmissivity \( t \), reflectivity \( r \), and loss \( q \) are connected by the equation \( r + t + q = 1 \). When good dielectric multilayer reflectors are used, \( q \) may be neglected, and we may write \( t = 1 - r \).
As a result of spontaneous and stimulated emission, light is generated within the laser. Light that does not pass out through the sides travels the length of the laser and is reflected back and forth between the mirrors. At the time of each reflection a fraction $1 - r$ of the energy is lost. Therefore oscillations may be sustained only if the gain of the radiation passing through the full length of the crystal is sufficient to compensate for the energy loss at the end and other losses due to secondary causes. Starting at one point, the radiation will suffer two reflections before it passes the same point in the original direction. In each passage through the material the intensity gains by a factor of $\exp \alpha L$. If the reflection coefficients at the end faces are $r_1$ and $r_2$, the energy of the wave will have changed by a factor of $F = r_1 r_2 e^{2\alpha L}$.

We introduce $r$, the geometric mean of the reflection coefficients $r_1$ and $r_2$, and also the loss coefficient $\gamma = -\log_e r$. The gain in one complete (round trip) passage is $F = e^{2(\alpha L - \gamma)}$. When $F$ is not less than 1, oscillations will build up, starting from a small disturbance; when $F$ is less than 1, they will die out. Clearly, if somehow the situation $\alpha L > \gamma$ is brought about, the intensity of radiation of the proper frequency will build up rapidly until it becomes so large that the stimulated transitions will deplete the upper level and reduce the value of $\alpha$. This is a dynamic situation that will be encountered again in connection with pulsations and giant pulses. In order to determine what may happen in a stationary or steady-state situation, we recall that there is a steady generation of light of the proper frequency in the laser as a result of spontaneous emission which proceeds at the rate of, say, $n$ quanta per second. Let $t$ denote the transmission coefficient of the partially transmitting reflector. Then on every complete two-way passage of the light through the laser the fraction $t$ of the intensity incident on the reflector leaves the laser, and the intensity changes by the
factor \( \exp(2(\alpha L - \gamma)) \). Therefore light will be emitted from the laser at the following rate:

\[
p = n\hbar \nu t \left[ 1 + e^{2(\alpha L - \gamma)} + \cdots \right] = n\hbar \nu t \left[ 1 - e^{2(\alpha L - \gamma)} \right]^{-1}.
\]

The output would become infinite if \( \alpha L \) were equal to or greater than \( \gamma \). A steady output may be maintained, however, if the peak value \( \alpha_m \) of \( \alpha(\nu) \) remains just below the value \( \frac{\gamma}{L} \).

\[
\alpha_m = \frac{\gamma}{L} \quad (2.7)
\]

This equation is the threshold condition.

Since \( \alpha(\nu) \) is nearly equal to \( \alpha_m \) in a very narrow spectral range, it is clear that only in this range will the amplification be large enough to offset losses. Consequently, the output of the laser will be sharply peaked, and its linewidth will be much narrower than the atomic linewidth. We will see in Section II.4 how the linewidth of the laser output is related to the linewidth of the atomic spectrum, to the \( Q \) of the cavity formed by the reflectors of the laser, and also to the level of the power output.*

The threshold condition relates the peak of the amplification curve to the length of the laser and to the losses at the ends. The Füchtbauer-Ladenburg formula relates the integral of the amplification curve to the population inversion, or excitation. The threshold as a function of the excitation can be given if the shape of the amplification curve is known. Let \( P \) be the peak value of the amplification curve normalized so that its integral is 1. Then

\[
P_{kg_1} \left( \frac{N_2}{g_2} - \frac{N_1}{g_1} \right) = \frac{\gamma}{L} \quad (2.8)
\]

is the threshold condition. In special cases one may express \( P \) in terms of the linewidths and thus obtain

\[
\frac{N_2}{g_2} - \frac{N_1}{g_1} = \frac{\pi \Delta \nu}{2} \frac{\gamma}{\kappa g_1 L} \quad (2.9)
\]

for the Lorentz line, and

\[
\frac{N_2}{g_2} - \frac{N_1}{g_1} = \frac{\Delta \nu}{0.939} \frac{\gamma}{\kappa g_1 L} \quad (2.9')
\]

for the Gaussian line.

It is to be noted that the linewidth \( \Delta \nu \) of the Lorentz line is independent of the frequency of the radiation, but the linewidth of the Doppler-broadened line is proportional to the frequency (Chapter I, equation * For the definition of \( Q \) see p. 80.
The quantity $\kappa$ is proportional to $\nu^{-2}$ (Chapter I, equation (6.9)). Consequently the population inversion required for threshold is proportional to $\nu^2$ for Lorentz lines and to $\nu^3$ for Doppler-broadened lines. When other things are equal, it is therefore easier to fulfill the threshold condition at lower frequencies.

Equation (2.8) can be used to determine the minimum value of $N_2$ when $\kappa$ is known. When $k_0$ is known, we may turn to (2.4) and write the threshold condition in the form

$$\frac{g_1}{g_2} N_2 - N_1 = \frac{N_0 \gamma}{k_0 L} = \frac{\gamma}{\sigma_0 L}, \quad (2.10)$$

where ratio $k_0/N_0$ is $\sigma_0$, the peak absorption cross section per atom. According to the measurements of Maiman [1], the value of $\sigma_0$ for the ruby $R_1$ line is $2.5 \times 10^{-20}$ cm$^2$.

A general conclusion may be drawn from equation (2.10). Since the intensity of the excitation determines the population inversion, and this in turn has to exceed the minimal value $\gamma/k_0 L$, a certain tradeoff is possible between the reflection coefficient that determines $\gamma$ and the active length $L$ of the laser. Any deterioration of the reflector must be compensated by increased length or a penalty is paid in the form of increased threshold of excitation. It is also significant that the onset of oscillations can be prevented by lowering $r$ and thereby increasing $\gamma$.

If the negatively absorbing material is kept in a stationary state by means of steadily supplied excitation, the left-hand side of (2.9) or (2.10) may be related to the rate at which excitation is provided. We shall show how this calculation can be carried out in the case of the ruby. For the relevant levels in ruby $g_1 = g_2$; therefore the statistical weights may be canceled out. The reader is warned against the indiscriminate dropping of the factor $g_1/g_2$ in (2.9) or (2.10). Neglect of this factor where needed has led to erroneous conclusions in the literature.

The occupation of levels in a three-level system with a total of $N_0$ atoms is governed by the rate equations

$$\frac{dN_3}{dt} = W_{13} N_1 - (W_{31} + A_{31} + S_{32}) N_3,$$

$$\frac{dN_2}{dt} = W_{12} N_1 - (W_{21} + A_{21}) N_2 + S_{32} N_3,$$

$$N_0 = N_1 + N_2 + N_3. \quad (2.11)$$

Reference is made to Fig. 18, which shows such a system.

* Strictly speaking, it is not true that $g_1 = g_2$ for the ruby $R_1$ line. Practically, however, this is a good assumption because of the close coupling of the two levels, each of multiplicity 2, at which the $R_1$ and $R_2$ lines originate. The ground state of ruby has a multiplicity of 4.
Here the $W$’s denote probabilities of stimulated transitions which involve the radiation density; the $A$’s indicate probabilities of spontaneous transitions, $S_{32}$ the probability of radiationless transition from level 3 to level 2, and $N_0$ the constant total number of active atoms per unit volume. In a steady state the derivatives in (2.11) are 0, and we can solve for the ratio $N_2/N_1$. A simple solution is obtained under the assumption that $A_{31}$, the rate of spontaneous return from the top level, is low compared to all other processes. In this case

$$\frac{N_2}{N_1} = \left( \frac{W_{13}S_{32}}{W_{31} + S_{32}} + W_{12} \right) (A_{21} + W_{21})^{-1}. \tag{2.12}$$

Further simplification is possible by noting that $W_{13} = W_{31}$ and $W_{21} = W_{12}$, as a consequence of one of Einstein’s relations (5.4), and by restricting ourselves to materials in which the radiationless relaxation between levels 3 and 2 is fast. The last condition means that $S_{32} \gg W_{13}$. With this restriction,

$$\frac{N_2 - N_1}{N_0} = \frac{W_{13} - A_{21}}{W_{13} + A_{21} + 2W_{12}} \tag{2.13}$$

follows. The quantity $W_{13}$ is proportional to the incident exciting light, and $W_{12}$ is proportional to the laser action in progress. (At the threshold of oscillations $W_{12} = 0$.) Given $k_0$, $r$, and $L$, we may calculate the fraction $(N_2 - N_1)/N_0$ from (2.10). The minimum value of $W_{13}$ necessary to achieve this fraction is then computed from (2.13).
Clearly $W_{13}$ must be larger than $A_{21}$; equality is insufficient because of the unavoidable losses at the reflectors. Yet the condition $W_{13} = A_{21}$ is frequently used in estimating the threshold power required for oscillations. Actually, this is only the condition for the attainment of negative absorption in the material.

Let us estimate the exciting radiation required to produce negative absorption. If the excitation is accomplished by a monochromatic plane wave at the frequency $\nu_p$ and $n$ photons per second are incident on the unit surface area of the crystal, the following relations hold:

$$P = nh\nu_p \quad \text{and} \quad W_{13} = u_p B_{13} = \frac{1}{c} (nB_{13}h\nu_p\eta) = n\sigma_p,$$  \hspace{1cm} (2.14)

where the quantity

$$\sigma_p = \frac{1}{c} (B_{13}h\nu_p\eta)$$  \hspace{1cm} (2.15)

is the integrated absorption cross section for the pumping radiation. Then, in order to achieve the condition $W_{13} \geq A_{21}$, the incident radiative flux density $P$ must be at least

$$P_{\text{min}} = \frac{A_{21}h\nu_p}{\sigma_p}.$$  \hspace{1cm} (2.16)

Equation (2.16) shows the relevant factors that determine the required level of irradiation. The power required increases proportionally with the spontaneous transition probability; hence it varies inversely with the lifetime of the excited state.

Illumination by means of a plane wave is a mathematical artifice. A configuration more likely to be realized in practice is one in which the material is isotropically illuminated over most of its surface. Moreover, the illuminating source is likely to be a high-pressure discharge with a more or less continuous spectral output resembling the spectral distribution of the black-body radiator. It is therefore realistic to consider the laser crystal to be immersed in an isotropic black-body radiation and determine the minimum temperature of the black-body source which may produce population inversion in a three-level fluorescent solid. In view of (5.5) of Chapter I we may write

$$W_{13} = \frac{g_3}{g_1} B_{31} u(\nu_{13}),$$  \hspace{1cm} (2.17)

where $u(\nu_{13})$ is the energy density of black-body radiation. The expression for $u(\nu)$ valid in vacuum is (1.1) of Chapter I. In a solid this has to be modified by replacing $c$ with the appropriate velocity in the solid, namely
Combining the correct equation for \( u(\nu) \) with (5.6) of Chapter I, we obtain

\[
\frac{u(\nu)B_{31}}{e^{\frac{\nu}{kT}} - 1} = \frac{A_{31}}{e^{\nu/kT} - 1} \tag{2.18}
\]

for \( \nu = \nu_{13} \). The condition for population inversion is \( W_{19} > A_{21} \); it requires that

\[
\frac{g_3 A_{31}}{g_1 A_{21}} > e^{\nu/kT} - 1 \tag{2.19}
\]

for \( \nu = \nu_{13} \). Thus the minimum source temperature, \( T_s \), is given by

\[
T_s = \frac{E_3 - E_1}{k \log (1 + \frac{g_3 A_{31}}{g_1 A_{21}})} \tag{2.20}
\]

We note that these calculations are independent of the shape and width of the absorption curve as long as the absorption band is narrow enough so that the variation of the intensity of black-body radiation over the band may be neglected.

We may now make a rough estimate of the source temperature required for producing negative absorption in ruby, neglecting the known fact that what we designated as level 2 is actually a double level, and that only the lower one of these levels produces the 6943 \( \AA \) radiation. According to the measurements of Maiman [1, 2],

\[
E_3 - E_1 \approx 18,200 \text{ cm}^{-1} = 3.6 \times 10^{-12} \text{ erg},
\]

\[
A_{31} = 3 \times 10^8 \text{ sec}^{-1},
\]

\[
A_{21} = 232 \text{ sec}^{-1}.
\]

The multiplicities of the relevant levels are \( g_1 = 4 \) and \( g_3 = 12 \). (See Section III.1.) Therefore

\[
\frac{g_3 A_{31}}{g_1 A_{21}} = 3880.
\]

With these numerical values we obtain from (2.20) the temperature \( T_s = 3160 \text{°K} \). The actual temperature required for laser operation is higher than this because several complicating factors were neglected in the computation. Nevertheless this computation, which follows the line of reasoning employed by Maiman [1, 3], provides an indication of the magnitudes involved.

The preceding calculation is in principle applicable to any three-level solid laser excited by irradiation (optical pumping). The analysis of the kinetics of a similarly excited four-level solid laser requires equations involving the populations of all four levels and transitions of several types from every level to every other level. Approximate equations may be
obtained, however, by concentrating on the few dominant transitions shown in Fig. 19. The notation is similar to that for three-level lasers; $W$'s denote stimulated transitions, $A$'s spontaneous transitions, $S$'s non-radiative transitions. A few physical assumptions simplify the mathematical problem: It is assumed that the transition from level 4 to level 3 is very fast, which is equivalent to saying that level 4 is always practically empty and that the pumping takes place directly into level 3. Moreover, it is also assumed that transitions between levels 1 and 2 take place so fast that up to the time the stimulated emission begins, the occupancy of level 2 can be calculated by Boltzmann's formula

$$N_2 = N_1 e^{(E_1 - E_2)/kT}. \quad (2.21)$$

The essential variables of the problem are then $N_2$ and $N_3$, and the threshold condition is

$$\frac{g_2}{g_3} N_3 - N_2 = \frac{\gamma}{\sigma_{32} L}, \quad (2.22)$$

where $\sigma_{32}$ is an absorption cross section that is not easy to determine experimentally.

It must be remembered that the pumping radiation covers a broad spectral band; therefore it will cause transitions not only from the ground level but also from level 2 to level 4. The transitions take place into a band (level 4) which arises from the ionic levels under the influence of the crystal field. This structure of level 4 prevents the application of the selection rules, which in the case of a free atom forbid either the 1 to 4
or the 2 to 4 transition. The approximate rate equation is (before the onset of stimulated emission)

\[
\frac{dN_3}{dt} = W_{14}N_1 + W_{24}N_2 - (A_{31} + A_{32})N_3. \tag{2.23}
\]

We may eliminate \(N_1\) by means of (2.21). A stationary state is reached when \(\dot{N}_3 = 0\), i.e., when

\[
N_3 = \frac{W_{14} e^{(E_2 - E_1)/kT}}{A_{31} + A_{32}} + \frac{W_{24}}{A_{31} + A_{32}}. \tag{2.24}
\]

It is customary to introduce \(A_3 = A_{31} + A_{32}\), the total spontaneous decay rate of level 3. When levels 1 and 2 are separated by a very small gap it is also permissible to assume that \(W_{24}\) is equal to \(W_{14}\). In this case

\[
N_3 = \frac{W_{14}(1 + e^{(E_2 - E_1)/kT})}{A_3}. \tag{2.25}
\]

To obtain an equation of the type of (2.13) we introduce the total number of available active ions

\[
N_0 = N_1 + N_2 + N_3 = N_2[1 + e^{(E_2 - E_1)/kT}] + N_3. \tag{2.26}
\]

From (2.25) and (2.26) it follows that in the steady state

\[
\frac{N_3 - N_2}{N_0} = \frac{W_{14} - A_3[1 + e^{(E_2 - E_1)/kT}]^{-1}}{W_{14} + A_3}. \tag{2.27}
\]

This formula, as well as (2.13), was obtained by Maiman [4]. It is convenient for threshold calculations when the multiplicities of levels 2 and 3 are equal. In the general case one must fall back on equations (2.22), (2.25), and (2.26).

We emphasize that the above simplified calculations are applicable only when the population of the terminal level is determined by thermal processes and the transition rate between the ground level and the terminal level is high. This is definitely not the case in gaseous lasers whose terminal levels are generally very high above the ground level compared to \(kT\) and where the transition rate between the terminal level and the ground level is often the limiting factor of laser operation. The kinetics of such lasers is discussed in Chapter V. Moreover, the “stationary” distribution of population calculated according to (2.24) is applicable only before the onset of stimulated emission. In the case when laser operation is possible this distribution will not be stationary because stimulated emission will begin with a depletion of \(N_3 - N_2\). Therefore, equations (2.24) to (2.27) are useful only in that they help to decide whether or not laser operation is possible. Once such operation is possible the population of the levels
involved in a steady-state laser operation must be calculated from equations which take into account the stimulated transitions between levels 2 and 3. It is relatively easy to make input and output calculations for a ruby laser in a hypothetical steady radiating state. Such calculations are of little value, however, because of the large intensity fluctuations which seem inherent in the situation.

The power generated at frequency $v_{21}$ in a uniformly excited ruby laser of volume $V$ is

$$P_0 = W_{21}(N_2 - N_1)Vhv_{21}. \quad (2.28)$$

Here $W_{21}$ is proportional to the radiation density, and $N_2 - N_1$ depends on the radiation density as well as on the intensity of excitation. Starting with zero radiation density at frequency $v_{21}$ when the threshold is first reached, the radiation density builds up and depletes $N_2 - N_1$ until either a steady state is arrived at or the radiation density starts to fall again and an oscillation of intensity ensues. This oscillation of intensity is called pulsation. It has already been mentioned in Section II.1. Further material on pulsations is included in Chapter VI.

REFERENCES


3. THEORY OF OSCILLATION AND RADIATION MODES

Standing Waves

The laser consists of a vast number of atomic amplifiers placed between two partially reflecting mirrors which cause radiation to travel back and forth through the amplifying medium. Our concern at present is the structure of the electromagnetic field built up within the laser and the properties of the electromagnetic field radiated from the laser. The electromagnetic field within the laser may be regarded as a field in a cavity which is weakly coupled to the outside. The different types of electromagnetic oscillations of the laser regarded as an isolated cavity are called *modes of oscillation* or briefly *modes*. Simple optical theory tolerates the existence of uniform plane waves of finite extent. The field between the mirrors may be regarded as the superposition of plane waves traveling back and forth. The plane waves that travel longitudinally, that is,
perpendicular to the mirrors, form a standing wave pattern which leads to reinforcement when the distance of the mirrors is an integral multiple of the half-wavelength. Reinforcement takes place when

\[ n\lambda = 2L, \]  

(3.1)

where \( n \) is an integer, \( \lambda \) is the wavelength in the laser material, and \( L \) is the distance between the mirrors.* Similar relations are known for waves traveling in other directions.

For a ruby laser a few centimeters long, the integer \( n \) is of the order of 100,000. Each value of \( n \) corresponds to a frequency at which oscillations may occur provided that sufficient amplification is available at that frequency. A fixed value of \( n \) characterizes a resonance or a mode. Strictly speaking, two modes of radiation must be considered for each permissible frequency because of the possible variation of polarization.

From the equation

\[ \frac{1}{\lambda_0} = \frac{n}{2L\eta}, \]  

(3.2)

we find that two consecutive resonances (or modes) are separated by the wave number

\[ \Delta \left( \frac{1}{\lambda_0} \right) = \frac{1}{2L\eta}, \]  

(3.3)

and that the fractional change in wavelength from one axial mode to the next is

\[ \frac{\Delta \lambda_0}{\lambda_0} = \frac{1}{n}. \]  

(3.4)

Consecutive modes of this type are so closely spaced that many lie within one atomic linewidth. In other words, the bandwidth of the atomic amplifiers generally encompasses a number of laser modes.

The uniform plane wave theory correctly predicts the frequency separation of the principal resonances. However, other resonances may also occur. Moreover the assumption of uniform plane waves of finite extent is inconsistent with the principles of electromagnetic radiation. Finally observations of the radiation emerging from the laser offer a positive proof that the surface of the mirrors is not a uniform phase and amplitude surface. Consequently, a refinement of the theory is in order.

**Cavity Theory**

A solid-state laser, such as the ruby laser, in many ways resembles a metallic cavity in which a hole is cut so that some of the radiation may

* The symbol \( \lambda_0 \) is used for wavelength in vacuum; hence \( \lambda_0 = \lambda\eta \), where \( \eta \) is the index of refraction.
escape. The principal interest in the laser field is in the radiation which propagates nearly, or exactly, along the longitudinal direction. It is useful, although somewhat inexact, to think in terms of the ray theory and to note that radiation directed at large angles away from the longitudinal axis of the laser encounters the surface of the laser before the amplification within the laser can offset the inevitable loss at the surface. Therefore, at large angles away from the axial direction one will encounter negligible stimulated radiation. Nearly axial rays suffer total reflection at the polished side surfaces; therefore their fate is the same as it would be if the side surfaces were metalized. One of the end surfaces is generally completely reflecting and the other almost completely so. Therefore, for the purpose of describing the modes of the stimulated radiation, we may in a first approximation assume a conductive cavity and take the output into account as a perturbation of the cavity field.

It is easily shown that the vector potential in a rectangular cavity is the sum of terms of the form

\[ A_{k_1, k_2, k_3} \exp[(k_1x + k_2y + k_3z)], \]

(3.5)

where

\[ k^2 = k_1^2 + k_2^2 + k_3^2 = \frac{\omega^2 c^2}{\eta^2}. \]

(3.6)

The boundary conditions are \( k_1a = \ell \pi, \) \( k_2b = m \pi, \) \( k_3L = n \pi, \) where \( a, b, \) and \( L \) are the dimensions of the rectangular box and \( \ell, m, \) and \( n \) are integers. The resonant frequencies are then determined from the equation

\[ \left( \frac{\ell \pi}{a} \right)^2 + \left( \frac{m \pi}{b} \right)^2 + \left( \frac{n \pi}{L} \right)^2 = k^2 = \left( \frac{2\pi \eta c}{\ell \pi} \right)^2. \]

(3.7)

For a cylindrical laser of radius \( r, \) expression (3.5) is replaced by the cylinder functions, and (3.7) is replaced by

\[ \left( \frac{\ell \pi}{r} \right)^2 + \left( \frac{m \pi}{L} \right)^2 = k^2, \]

(3.8)

where \( \chi_{lm} \) is the \( m \)th zero of the Bessel function of order \( l \) or its derivative. The longitudinal or axial modes in either case are obtained by setting \( \ell = 0 \) and \( m = 0 \) and defining \( \chi_{00} \) as 0. Then we obtain equation (3.2). In other words, the cavity theory produces all the modes of the simpler theory plus additional, transverse modes, the latter being the modes for which \( \ell^2 + m^2 \neq 0. \) The first transverse mode in the cylindrical case corresponds to \( \ell = 0, m = 1. \) Let us calculate the separation of the first transverse mode from the axial mode for a large value of \( n \) using \( k_1 \) and \( k_0 \) to indicate the
wave numbers of the transverse and axial modes, respectively. Then from (3.8) it follows that
\[
k_1^2 - k_0^2 = \frac{\chi_{01}^2}{r^2},
\]
and the corresponding relative frequency difference is
\[
\frac{\nu_1 - \nu_0}{\nu_0} = \frac{\chi_{01}^2 \lambda^2}{8\pi^2 r^2}.
\]
Here \(\lambda\) is the average wavelength and \(\chi_{01} = 2.405\) is the first zero of the Bessel function \(J_0\). For a ruby rod 1 cm in diameter the quantity on the right of (3.10) is \(4.56 \times 10^{-10}\).

Actually, the solid laser is not a metallic box but a dielectric resonator, and the exact theory of the latter leads to more complicated field configurations within the cylinder than are obtained for a resonator enclosed by metal [1]. It is found, for example, that transverse electric and transverse magnetic fields exist only in the axially symmetric modes. However, the equation which determines the resonant frequencies is always of the form
\[
\kappa_{lm}^2 + \left(\frac{\pi n}{L}\right)^2 = k^2.
\]
While in the cylindrical resonator bounded by metal the quantities \(\kappa_{lm}\) are determined by the solutions of \(J_{\kappa}(\kappa r) = 0\) and \(J'_{\kappa}(\kappa r) = 0\), the equations determining the \(\kappa\)'s in the dielectric case are more complicated and lead to \(\kappa\)'s that are only approximately equal to those obtained in the simpler case. However, for the nearly axial modes of the laser the number \(n\) is very large, the second term predominates in (3.11), and a moderate error in \(\kappa\) does not significantly alter the value of the frequency. Where frequency differences of modes are involved one must be careful to use the correct values.

It is shown in electromagnetic theory that the cavity modes discussed above are orthogonal to each other. Physically this means that oscillations of a single mode may be excited without exciting others. This isolation of modes, however, is true only in the ideal case when the cavity walls are infinitely conducting and have the exact geometrical shape postulated. Deviations from these idealized conditions or inclusion of polarizable material in the cavity may result in the coupling of modes, i.e., the transfer of energy from one mode of oscillation to another.

The number of possible modes of oscillation in a laser is very large. In general, different modes have different frequencies, and it is of some interest to inquire concerning the distribution of the modes in frequency. It is easy to determine the density function \(p(\nu)\) which is defined as follows:
\[
p(\nu) \, d\nu \text{ is the number of modes per volume element in the frequency interval } \nu \text{ to } \nu + d\nu.
\]
Let us return to equation (3.7) and apply it to a cube of side $a$. When a triplet of integers $(l, m, n)$ satisfies the relation

$$l^2 + m^2 + n^2 = k^2 a^2 / \pi^2,$$

there is a pair of modes (two polarizations!) with a wave number $k$ that have the frequency $\nu = ck/2\pi$. Reflection on one of the sides changes $k_1$ into $-k_1$, or $k_2$ into $-k_2$, or $k_3$ into $-k_3$. Therefore a single mode in a reflecting cube consists of a combination of eight terms differing only in the signs of their $k$'s. When counting modes one should count only triplets with positive integers $l, m, n$. The number of such triplets that satisfy the inequality

$$l^2 + m^2 + n^2 \leq \left( \frac{2\nu a \eta}{c} \right)^2$$

is one-half the number of modes whose frequency does not exceed $\nu$. But the number of points with integral coordinates located in one octant of the sphere

$$l^2 + m^2 + n^2 \leq R^2$$

is $\pi R^3 / 6$; therefore the number of modes with frequencies not exceeding $\nu$ is

$$P(\nu) = \frac{8\pi \nu^3 a^3 \eta^3}{3c^3}.$$ 

Hence

$$p(\nu) = \frac{1}{a^3} \frac{dP}{d\nu} = \frac{8\pi \nu^2 \eta^3}{c^3} = \frac{8\pi}{\lambda^3 \nu}.$$ 

The expression for the mode density just derived plays an important role in connecting the spontaneous and the induced transition probabilities. See equation (5.6) of Chapter I.

We note that for ruby $\lambda = 4 \times 10^{-5}$ cm, $\nu = 4.3 \times 10^{14}$ sec$^{-1}$; therefore (3.15) gives a mode density in the relevant frequency region as approximately 1 per cycle per cubic centimeter volume. Only a very small fraction of these modes, however, lies in a narrow cone around the longitudinal axis, and the laser acts as a generator only in these axial or nearly axial modes. The principal problem in laser cavity design is to create a structure which has only a few high-$Q$ resonances within the linewidth of the atomic transition. This selectivity must be accomplished in spite of the large volume of active material required for amplification and for the attainment of adequate power levels.

**Diffraction Theory of the Fabry-Perot Interferometer**

The cavity theory serves as a basis for orientation and provides adequate solutions for solid-state lasers with reasonably high refractive indices. It is
not applicable to gas lasers which are open structures consisting of a pair of plane or curved mirrors at the ends of an amplifying column. In a typical case a pair of plane circular mirrors 2 cm in diameter may be located at a distance of 1 meter from each other. In a situation of this type diffraction loss may not be negligible, in fact it may be an important factor determining the distribution of energy in the interferometer during oscillation. For laser oscillations to occur the total loss in power from scattering, diffractive spill over, and incomplete reflection on the mirrors must be balanced with power gained by travel through the active medium. In the presence of decoupled or orthogonal modes of oscillation, the threshold condition must be satisfied for each mode in which oscillations are to occur.

The parallel, partially transparent mirrors of the laser form a Fabry-Perot interferometer. When such an instrument is operated as a passive device with uniform plane waves continuously supplied from the outside, the internal fields may also be essentially uniform plane waves. In a laser, however, where power is supplied only from within the interferometer, the loss of power from the “edges” of the wave by diffraction will cause a marked departure from uniformity.

What, then, are the modes of the Fabry-Perot interferometer? These modes may be defined and discussed in terms of self-reproducing field configurations over the surfaces of the reflectors. A field configuration is called self-reproducing or a transverse mode if, after propagation from one reflector to the other and back, the field returns in the same phase and amplitude pattern; i.e., the function representing the complex amplitude over the reflector is multiplied by a fixed complex number which gives the total phase shift and the loss of the round trip. For every such transverse mode, there is a sequence of longitudinal modes, for which the round-trip phase shift is an integral multiple of $2\pi$. The nomenclature and the general appearance of the simplest transverse modes of plane circular and rectangular reflectors are illustrated in Fig. 20. The modes of the Fabry-Perot interferometer are called TEM modes, presumably to indicate that the electric and magnetic fields are mostly perpendicular to the longitudinal axis of the interferometer. They resemble uniform plane waves which are truly transverse electromagnetic waves.

Fox and Li [2] calculated the most important self-reproducing field configurations for a variety of mirrors. Their calculations are based on the standard diffraction formula of physical optics, which gives field distribution at a point of observation in terms of the phase and amplitude distribution at a given aperture [3]. Starting with a uniform amplitude and phase distribution on one reflector, they computed the amplitude and phase distribution on the other numerically. The functions obtained served as a
starting point for the next calculation. In this manner a sequence of pairs of amplitude and phase distributions was generated which eventually converged toward a self-reproducing pair of amplitude and phase distributions. (A uniform decrease of the amplitude by a constant factor must be permitted in this calculation to allow for the inevitable losses.)

This self-reproducing pair of phase and relative amplitude distributions over the aperture may be regarded as the dominant resonant mode of the interferometer. The phase and amplitude configuration is a function of the parameter (Fresnel number) \( N = \frac{a^2}{\lambda L} \), where \( a \) is the radius of the circular apertures and \( L \) the distance between them. Figure 21 shows the amplitude and phase distribution in the dominant (TEM\(_{00}\)) mode for a pair of circular plane mirrors for Fresnel numbers \( N = 2, 5, \) and \( 10 \). The undulations on the curves are related to the number of Fresnel zones. Fox and Li also obtained interesting nonsymmetric distributions, and for several geometries calculated the diffraction losses associated with their modes as functions of the number \( N \). In Fig. 22 we reproduce their diffraction-loss data pertaining to the interferometer with plane circular mirrors. From this figure it appears that the loss due to diffraction in a single passage in the TEM\(_{00}\) mode is about 0.9 per cent for \( N = 10 \), a value comparable to the usual loss caused by incomplete reflection in a gas laser. The Fresnel number \( N = 10 \) is applicable to a laser 1 meter long and 7 mm in diameter, with a wavelength of 1.15 \( \mu \). It is interesting to note that the adjustment of phase and amplitude distribution over the
reflectors causes the diffraction loss to decrease in comparison to its value for a uniformly illuminated aperture.

The variation of diffraction loss from mode to mode is of interest because if there is not much difference in loss between the modes, a laser excited
Theory of Oscillation and Radiation Modes

well above threshold will oscillate in several modes simultaneously. Although losses associated with incomplete reflection at the mirrors may be larger than the diffraction losses, the former are constant for all modes; therefore the differences from mode to mode in the sum of all losses arise mainly from differences in diffraction losses. The threshold condition, as we know, involves the total gain along the path and the sum of all losses in a two-way passage. Kotik and Newstein [4] pointed out that in a laterally extended active Fabry-Perot interferometer an oblique mode may be excited before an axial one because the oblique ray passes through a longer amplifying path than an axial ray, and the losses of the modes are nearly identical. The maximum transverse dimension of the mirrors must be limited in terms of the mirror distance, mirror reflectivity, and wavelength in order that the axial mode be favored.

Although the original self-reproducing configurations were obtained by a long series of numerical computations involving about 300 iterations, the problem may be formulated as an integral equation whose eigenvalue is the fractional (diffraction) loss per transit, and whose solution is the self-reproducing complex amplitude distribution. Tang [5] obtained a variational solution for this formulation for the case of infinite plane strip.

Fig. 22. Diffraction loss per transit versus $N = a^2/L$ for circular plane mirrors. (Reproduced from the Bell System Technical Journal with the permission of the American Telephone and Telegraph Company.)
mirrors; Bergstein and Schachter [6] completely solved the physically more important case of flat circular mirrors by analytic techniques.

The frequency separation of different modes is of interest in connection with gas lasers, since they possess an extremely high resolution. In calculating the mode separations, we must proceed with the self-reproducing configurations, and not the cavity calculations which led to (3.9) and (3.10). Waveguide theory warns us not to assume that the velocity of phase

![Diagram of phase shift per transit](image)

**Fig. 23.** Phase shift per transit (leading relative to geometrical phase shift) versus $N = a^2/L\lambda$ for confocal spherical mirrors. (Dashed curves for circular plane mirrors are shown for comparison.)
propagation of the interferometer modes is equal to the velocity of light. Rather, it is to be expected that this velocity approaches that of light for large values of \( N = a^2 / \lambda L \). Fox and Li calculated the phase shift in one passage relative to the geometrical phase shift, that is, \( 2\pi L / \lambda \). Their results are shown in Fig. 23. The resonance condition for the \( n \)th mode of any kind is then

\[
\frac{2\pi L}{\lambda_n} + \varphi = \pi n,
\]

(3.16)

where \( \lambda_n = c/v_n \) and \( \varphi \) is the phase shift appropriate for the transverse configuration. From (3.16) it follows that

\[
v_n = \frac{c n}{2L} - \frac{c \varphi}{2\pi L},
\]

(3.17)

therefore the frequency difference of consecutive modes of the same type is

\[
v_{n+1} - v_n = \frac{c}{2L},
\]

(3.18)

which is the same as in the plane wave case. However, the frequencies of the TEM\(_{00n}\) and TEM\(_{10n}\) modes, which belong to different transverse configurations, namely TEM\(_{00}\) and TEM\(_{10}\), will be shifted relative to each other, and the shift is calculable from the phase shift data such as those shown in Fig. 23.

The radiation pattern of the laser is calculable from the phase and amplitude distributions of the self-reproducing configurations.

**Confocal Spherical Interferometers**

A plane parallel interferometer formed by two parallel plane mirrors is not the best multimode resonator. Considerable improvement may be obtained when two concave spherical reflectors are used in a confocal arrangement, that is, with the center of one sphere on the other reflector. Such spherical interferometers were proposed by Connes [7], who recognized that the spherical system has a greater resolving power than a plane system of similar dimensions. A schematic representation of the confocal interferometer is shown in Fig. 24.

Fox and Li [2] calculated the first few modes of a confocal spherical interferometer, using the technique already described. They observed the remarkable properties of the confocal system as contrasted with the plain one: In the confocal system the field is much more concentrated near the axis of the reflector and falls to a lower value at the edge than in the plane system. The amplitude distribution is smooth; the ripples of Fig. 21 are absent. The surface of the reflector is a phase front of the wave. The losses
are orders of magnitude lower in confocal systems than in comparable plane systems. The phase shifts per transit for each configuration are independent of \( N \) and are multiples of \( \pi/2 \). This is shown in Fig. 23. The frequency relations (3.17) and (3.18) are applicable to the confocal interferometer.

Boyd and Gordon [8] solved the integral equation pertaining to the confocal case. Their calculations show that the distribution of amplitude in the central part of the reflector is nearly Gaussian and that the surface of the reflector is an equiphasic surface, which is not so in the case of plane reflectors. On the reflectors the amplitude falls to \( 1/e \) of its peak value at the center at a distance

\[
w_s = \sqrt{\frac{L\lambda}{\pi}}, \tag{3.19}\]

where \( L \) is the radius of curvature of the mirrors (Fig. 24). In the space between the reflectors near the symmetry axis (\( z \)-axis) the transverse distribution of the amplitude follows the Gaussian curve with a spread that varies with the axial position. This is illustrated in Fig. 25. The variation of amplitude is approximately proportional to \( e^{-u} \), where \( u = (x^2 + y^2)/w^2 \), and

\[
w^2 = \frac{L\lambda}{2\pi} (1 + \xi^2). \tag{3.20}\]

Here \( \xi = 2z/L \) is the displacement from the focus measured in units of the focal length. At the reflector we have \( \xi = 1 \) and \( w = w_s \); at the focus we have \( \xi = 0 \) and \( w = w_s/\sqrt{2} \). Thus the beam at the focus narrows down to one-half its cross section at the reflectors.
To obtain the angular beam width of the radiation pattern, we take the ratio of the spot diameter obtained from (3.20), as $\xi$ tends to infinity, to the distance from the center of the resonator. The beam width between half-power points is given by

$$\theta = 2\sqrt{\ln 2/\pi} \sqrt{\lambda/L} = 0.939 \sqrt{\lambda/L} \text{ radian.} \quad (3.21)$$

The numerical factor arises from the half-power width of the Gaussian curve. It is interesting to note that $a$, the radius of the reflectors, does not enter into these formulas which govern the spread of the beam. However, (3.19), (3.20), and (3.21) are applicable only when $a > 3w_s$. It must also be emphasized that the Gaussian approximation is not valid near the edges of the reflector and that the diameter of the reflector is an important parameter for diffraction losses. According to Boyd and Gordon [8], the diffraction loss per transit of the fundamental mode of the confocal interferometer is $10.9 \times 10^{-4.94N}$, where $N = a^2/\lambda L$, as in the case of the plane interferometer. For the sake of simplicity we shall use $11 \times 10^{-5N}$. 

![Diagram of field strength distribution within the confocal resonator for the TEM$_{00q}$ mode.](image)
The following comparison may then be made between the diffraction loss of the plane and spherical interferometers of similar size.

\[
N = \frac{a^2}{\lambda L}
\]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss: plane</td>
<td>0.18</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>spherical</td>
<td>(11 \times 10^{-5})</td>
<td>(11 \times 10^{-10})</td>
<td>(11 \times 10^{-20})</td>
</tr>
</tbody>
</table>

All losses refer to the fundamental modes; the values for the plane interferometer are taken from the curves of Fox and Li reproduced in Fig. 22. Clearly the diffraction losses of the spherical system are orders of magnitude lower than those of the plane system. The confocal spherical system possesses several other significant advantages over the plane one. Not the least of these for the experimentalist is the noncritical nature of the adjustment of the reflectors.

Confocal resonators of equal size represent a special case of two spherically curved reflectors facing each other. Deviations either in size or in curvature of the mirrors as well as an axial displacement of one of the mirrors result in structures with rather complicated properties. These were analyzed in detail by Boyd and Kogelnik [9], who concluded that the true confocal system is an optimal one with respect to diffraction losses. A small deviation from equal curvature will, however, produce a disproportionate increase in loss in the confocal case; therefore in order to allow for manufacturing tolerances it may be advisable to deviate from the true confocal arrangement, or to imitate it by means of a spherical and a plane mirror.

It may be well to state that the incomplete reflectivity of the mirrors does not affect the mode structure of the plane or the confocal interferometer as long as the reflectivity is uniform over the mirror [10].

The Competition of Modes

The different modes of a resonant cavity may be regarded as almost independent oscillators. It is known that an oscillator is characterized by its resonant frequency and its dissipation. The dissipation is usually described in terms of the quality factor \(Q\), which is defined by the formula

\[
Q = \frac{2\pi v_0 E}{P_d},
\]  

(3.22)

where \(v_0\) is the resonant frequency, \(E\) the energy, and \(P_d\) the rate at which energy is dissipated in the oscillator. When the oscillator is excited by a periodic force of frequency \(\nu\), the power absorbed by the oscillator varies
with \( v \) in the manner described by the function \( g(v, v_0) \) of Section 1.6, with the full linewidth \( \Delta v \) replaced by \( v_0/Q \). Therefore we shall refer to \( v_0/Q \) as the linewidth of the oscillator or the linewidth of the cavity mode. A high \( Q \) represents low dissipation and narrow linewidth.

In a laser a complicated interaction takes place between an ensemble of atoms on the one hand and a resonant cavity with many available resonant frequencies on the other. The atomic system is characterized by a center frequency \( v_a \) and a linewidth \( \Delta v_a \), and the resonant cavity is characterized by the resonances centered at \( v_1, v_2, \ldots, v_n \), with quality factors \( Q_1, Q_2, \ldots, Q_n \), respectively. Before entering into a discussion of the general interaction problem it would be well to contemplate the interaction of the atomic system with a cavity possessing only one available mode in the frequency range of the atomic system. This is the standard problem encountered in connection with masers.

In this case the atoms and the cavity are equivalent to two coupled resonant circuits. The classical analysis of these systems shows that if the center frequencies of the coupled circuits are slightly different, the frequency of free coupled oscillations will be determined primarily by the circuit with the higher \( Q \). Let this be system 1, whose resonant frequency and linewidth are \( v_1 \) and \( \Delta v_1 \), respectively. The oscillation of the coupled circuit will not take place at the frequency \( v_1 \) but approximately at

\[
\nu_0 = v_1 + (v_2 - v_1) \frac{\Delta v_1}{\Delta v_2},
\]

(3.23)

or in terms of the \( Q \)'s, at

\[
\nu_0 = v_1 + (v_2 - v_1) \frac{Q_2}{Q_1},
\]

(3.24)

In an ammonia maser there is a cavity with a single resonance frequency and a low \( Q \) coupled to an atomic line with a high \( Q \). Consequently, the frequency of the system is very nearly that of the atomic line. In a laser there are many competing cavity modes. In the first approximation, coupling between them may be neglected. All derive their excitation from the same pool of atomic systems and all dissipate energy partly through incomplete reflection and partly through diffraction or escape of radiation to the sides that may occur when the modes are not axial. Axial modes have the least dissipation, the highest \( Q \).

As the excitation progresses to the point at which the excess \( N_2 - N_1 \) is large enough to cause a net stimulated emission, energy begins to rise in the various resonator modes. The driving force acting on these modes is the largest for frequencies near the center of the atomic resonance \( v_a \). As the excitation increases, the dominant fraction of the energy goes to a small
number of modes which have the highest $Q$'s and whose frequencies are near the peak of the atomic resonance. The mathematical theory of this process in a multimode cavity has been developed by Wagner and Birnbaum [11, 12], who calculated the frequency distribution of the output in terms of the atomic linewidth and the degree of excitation. Sidestepping the intricate mathematical analysis of Wagner and Birnbaum, we may state qualitatively that modes with more energy than others tend to grow faster and their growth rate increases with increasing $Q$. All feed on the same supply of excited atoms; therefore the rich modes get richer, the poor poorer, until almost all of the radiative energy is concentrated "in the hands of a privileged few." The two privileges are, of course, being near the center $v_a$, where the supply of quanta is the greatest, and having the highest $Q$, which enables the mode to grow faster than its competitors.

The significance of the relationships pertaining to diffraction losses introduced at the beginning of this section now becomes apparent. The $Q$ of each mode is determined according to formula (3.22), in which the denominator $P_d$ consists of the sum of all losses from that mode. The loss caused by incomplete reflection does not vary from mode to mode; it is determined by the nature of the reflecting layer. The diffraction loss, as we have seen, is variable. The effect this variation will have on the $Q$ depends on the ratio of the diffraction loss to the reflection loss. Once the diffraction loss of a mode is less than one-half the reflection loss, further reduction of the diffraction loss is of little consequence. Certainly this is the case for the dominant modes of the confocal spherical resonators. However, the diffraction losses of plane resonators may contribute significantly to the determination of $Q$ when the reflectors are small. The large loss of off-axis modes usually lowers their $Q$'s to the point at which their excitation becomes negligible and the laser may be correctly described entirely in terms of the axial modes.

The losses of the axial modes are about equal. We have already noted that these modes are located so close to each other in the frequency domain that the width of an atomic line will stretch over dozens or hundreds of these lines. Ordinarily quite a few of them lie so close to the top of the atomic line that they are excited simultaneously. However, even a small difference in the diffraction losses among modes of different types may provide a great discrimination in the excitation of these modes, and if the laser is operated sufficiently close to threshold, oscillations will occur only in modes of lowest loss.

**Relation of Mode Theory to Experiments**

The theory of the laser as a resonant structure was developed in this section on the basis of certain idealizations. The most important of these is
the assumption that the laser material is homogeneous and isotropic. Crystalline lasers are certainly not isotropic, and even if they are practically homogeneous in their unexcited state, the excitation necessary for laser action is not uniform over the material. Hence the assumption of an isotropic material with a uniform amplification and with no internal scattering leads to conclusions that are only partially verifiable with experiments on less perfect materials. These experiments will be discussed in Chapter VI. The discussion of this section serves as a guide to that material.

REFERENCES


4. LINEWIDTH PROBLEMS*

The output of the laser oscillator, like that of many electronic oscillators, is essentially amplified noise. In an electronic oscillator the noise is thermal; its energy is uniformly distributed over a wide band of frequencies, which includes the amplification band of the amplifier, whereas in the laser the noise is the spontaneous emission of radiation characterized by a center frequency and linewidth. In either case amplification is regenerative; a part of the output is returned as input. Consequently, the frequency corresponding to the peak of the amplifier gain predominates in the output. Under certain circumstances there may be several such peaks.

* The material in this section was contributed by R. W. Hellwarth.
Let us now consider a photon packet emitted spontaneously within the laser. This packet will have the frequency distribution characteristic of the natural atomic line: a center frequency $\nu_a$ and a linewidth $\Delta \nu_a$. It was explained in Section 11.2 that after $m$ passes through the laser the packet is amplified by the factor $r^m e^{r \Delta \nu_a L}$. As $m$ becomes large, the exponential factor tends to cause the spectral distribution to narrow around one frequency, which corresponds to the peak value of the amplification. If it were possible to add intensities, and if $r$ did not depend on the frequency, matters would be simple indeed. The fact that light amplitudes must be added has the consequence that reinforcement of light will take place only if light arrives back at its point of origin in a proper phase. Whether this takes place depends on the dimensions of the laser in terms of the wavelength. Clearly, we are dealing here with the problem of oscillations in a cavity combined with selective amplification.

In a moderately loss-free medium the different modes of a resonant cavity may be regarded as independent oscillators, each characterized by its resonant frequency and by its linewidth, or $Q$. The general problem encountered in lasers is the interaction of the atomic generators and amplifiers with all relevant modes of the cavity at the same time. We shall not attempt to solve that general problem here, but we shall find it instructive to discuss the interaction with one resonator mode chosen so that its resonant frequency approximately coincides with the peak of the atomic line. In lasers, the atomic line is wide compared with the closely spaced cavity lines, so that there are many cavity resonances near the peak of the atomic line. The calculation that follows pertains to the linewidth of any one of these modes. The analysis of the spectral distribution of radiation in a given mode is based on the assumption that the oscillations in one mode are independent of the oscillations in another mode in the following sense: The presence of electric oscillations at one frequency does not produce a polarization of laser atoms at another frequency. This assumption is unimportant in the case of single-mode operation, which has been obtained in both solid and gas lasers. In usual operation, which involves simultaneous oscillations in many modes, it is difficult to assess the accuracy of the results that are calculated on the basis of this assumption; some experiments indicate that there are instances when the assumption is incorrect. However, even in these cases the estimate of spectral purity based on the above assumption of independence, or linear amplification, is probably still reasonably accurate.

The expression we shall derive for linewidth was first stated by Schawlow and Townes [1] in 1958 as a modification of a corresponding calculation for the ammonia maser; it suggested extremely pure spectral output and spurred efforts at experimental realization of a maser in the optical region. The
modification of the original ammonia analysis consisted of reversing the inequalities: \( h\nu < kT \) and \( \Delta_{\text{cav}} \gg \Delta_{\text{mat}} \), where the last two symbols represent the linewidths of cavity and material, respectively.

The discussion of linewidth is invariably tied to the concepts of noise theory and power spectrum, since what we seek to determine is a distribution of the output energy over frequencies. The source of the original signal is the spontaneous emission of radiation, about which we have statistical knowledge: we know its energy distribution in frequency. In the mathematical treatment of such problems it is necessary to make transformations from time domain to frequency domain and back. It is convenient in this connection to use the angular frequency \( \omega = 2\pi\nu \) as a variable. Thus the chosen oscillator which represents the cavity has resonant frequency \( \omega_0 \) and linewidth \( \Delta_0 \). These quantities and the \( Q \) of the oscillation are connected by the relation \( \omega_0 = Q\Delta_0 \). Let the peak angular frequency of the atomic line also be \( \omega_0 \) and let the atomic linewidth \( \Delta_a \) be much larger than \( \Delta_0 \).\(^*\) The particular oscillation of the cavity is described by a coordinate \( X \), which may be the instantaneous electric field. It satisfies the differential equation

\[
\ddot{X} + \Delta_0 \dot{X} + \omega_0^2 X = F(t),
\]

where \( F(t) \) is the driving force. The second term on the left is the dissipative term. For a harmonic driving force \( F(t) = F_0 e^{i\omega t} \), the excitation of the oscillator depends on the amplitude \( F_0 \) and the frequency of the exciting force. The energy \( E \) stored in the oscillator and the rate \( P \) at which energy is dissipated by the oscillator are related by the equation

\[
P = \frac{\omega_0 E}{Q} = \Delta_0 E.
\]

This last relation is independent of the exciting force. In a cavity oscillation there may be several dissipative processes. One may be dissipation in the material within the cavity; others may permit energy to escape to the outside through various ports. When several independent processes of energy dissipation act simultaneously, there is associated with each of them a power \( P_i \) and a linewidth, so that

\[
P = \sum P_i, \quad \Delta_0 = \sum \Delta_i.
\]

The power dissipated through the \( i \)th port is then

\[
P_i = \Delta_i E.
\]

The energy in a steady state is determined by the linewidth \( \Delta_0 \) and by the power spectrum of the exciting force. Physically it is quite plausible that

\(^*\) This linewidth is \( 2\pi\Delta\nu \).
of $\hbar \omega_0$. In equilibrium the populations of these levels obey Boltzmann's equation. Therefore in the simplest case of equal multiplicities we have

$$e^{\hbar \omega_0 / kT} = \frac{N_1}{N_2}$$  (4.8)

Consequently, for an oscillator in equilibrium with an ensemble of two-level systems, (4.7) reduces to

$$\langle E \rangle \approx \frac{N_2 \hbar \omega_0}{N_1 - N_2}.$$  (4.9)

Therefore from (4.5) and (4.9) the noisy driving force associated with the two-level system is given by

$$\langle |f(\omega_0)|^2 \rangle = \frac{2A_1 N_2 \hbar \omega_0}{N_1 - N_2},$$  (4.10)

where $A_1$ is the net contribution to the cavity linewidth arising from absorption and stimulated emission by the two-level systems. In the case of nondegenerate levels the loss rate is proportional to the number of atoms $N_1$ that are available for absorption minus the number of atoms $N_2$ available for emission. Therefore, equation (4.10) indicates that the noise power $\langle |f(\omega_0)|^2 \rangle$ is proportional to the number of atoms in the upper level $N_2$. This conclusion agrees with our introductory statement that the noise arises from spontaneous emission. The constant of proportionality, $2A_1 \hbar \omega_0/(N_1 - N_2)$, is independent of $N_1$ and $N_2$; it remains the same when $N_2$ exceeds $N_1$, and the contribution to linewidth $A_1$ is negative.

Let us now return to the oscillation of the laser. Here real damping is provided by the radiation to the outside. The descriptive quantities are $P_L$, the power output of the laser, and the corresponding damping term $A_2$. The atomic ensemble provides the negative term $A_1$, so that the total damping term or linewidth is

$$A = A_1 + A_2.$$  (4.11)

Then, according to (4.4) and (4.5), the average power radiated is

$$\langle P_L \rangle = A_2 \langle E \rangle = \frac{A_2 \langle |f(\omega_0)|^2 \rangle}{2A_0}.$$  (4.12)

In view of (4.10), this becomes

$$\langle P_L \rangle = \frac{A_1 A_2}{A} \frac{N_2 \hbar \omega_0}{N_1 - N_2}.$$  (4.13)

We drop the brackets around the power output $P_L$, and we note that $A$ is the sum of a positive and a negative term which are nearly equal at
threshold because the losses balance the gains. In approximation, we write in the numerator \(-\Delta_1 \approx \Delta_2\) and replace \(\Delta_2\) by the symbol \(\Delta_e\) to indicate that it is the linewidth of the cavity determined solely by the rate of escape of radiation from the cavity. Then

\[
P = \frac{\Delta_e^2}{\Delta} \frac{N_2 \hbar \omega_0}{N_2 - N_1}.
\]  

(4.14)

Hence the final formula for the linewidth in angular frequency is

\[
\Delta = \frac{\Delta_e^2}{P} \frac{N_2}{N_2 - N_1} \hbar \omega_0.
\]  

(4.15)

Schawlow and Townes [1] derived this relationship for the special case \(N_1 = 0\) in terms of half-widths of the lines at half maximum power. The corresponding variables are related to ours as follows: \(\delta \nu = \Delta/4\pi\), \(\delta \nu_e = \Delta_e/4\pi\). Consequently, in their notation

\[
\delta \nu = \frac{4\pi(\delta \nu_e)^2}{P} \frac{N_2 \hbar \nu_0}{N_2 - N_1}.
\]  

(4.16)

It has been stated that the calculations starting with (4.4) are applicable in the case of equal multiplicities. In the general case, \(N_1\) and \(N_2\) are to be replaced by \(N_1/g_1\) and \(N_2/g_2\) respectively wherever they occur in equations (4.8) to (4.16).

We repeat, we have treated the special case in which the center of the cavity resonance coincides with the center of the broader atomic line. When this condition is not strictly true, (4.15) and (4.16) are still approximately valid; the oscillation frequency of the cavity mode is pulled toward the center of the atomic line as described in Section II.3. [Cf. (3.23), (3.24).] The competition of different cavity modes near the peak of the atomic line is also described qualitatively in Section II.3; for a more detailed discussion we refer to Wagner and Birnbaum [2].

What is observed in anything but the most refined experiments is not the linewidth given by (4.15) or (4.16) but an empirical linewidth that results from the simultaneous unresolved observation of many lines. This is one of the reasons that the early experiments with ruby lasers showed a measured linewidth of the order of 0.1 Å, which is strictly an engineering variable with little physical meaning. The real linewidth is much narrower and depends on the level of the power output. Even the measured linewidth of the most stable He-Ne laser is orders of magnitude larger than the intrinsic linewidth calculated from the equations derived here. This discrepancy indicates that the measured linewidth represents the effects of fluctuations in the laser structure.
The calculation for a typical He-Ne laser operating at 1.15 μm wavelength may run as follows: Assuming a length of 100 cm and a reflector loss of 2 per cent gives

\[ \Delta_c = c(1 - r)/L = 6 \times 10^6 \text{ rad/sec}. \]

For the laser in question \( h\nu = 1.72 \times 10^{-12} \text{ erg} \). Hence, for an assumed output of \( P = 1 \text{ mW} = 10^4 \text{ erg/sec} \), we obtain from (4.15)

\[ \Delta = 36 \times 10^{12} \times 10^{-4} \times 1.72 \times 10^{-12} \frac{N_2}{(N_2 - N_1)} = 0.0062 \frac{N_2}{(N_2 - N_1)}. \]

The factor \( \frac{N_2}{(N_2 - N_1)} \) depends on the operating conditions. From measurements of the decay times of the relevant levels we can conclude that \( N_2/N_1 < 8 \); therefore the factor \( N_2/(N_2 - N_1) \) is at least 8/7. Assuming \( N_2/(N_2 - N_1) = 2 \), we obtain

\[ \Delta = 0.0124 \text{ rad/sec} = 0.002 \text{ cps}. \]

Actual measurements indicate linewidths of tens of cycles per second for the most stable He-Ne laser. (Cf. Section V.4.)

REFERENCES