1 Electromagnetic waves, light, and lasers

1.1 The laser

A laser is an oscillator that operates at *optical* frequencies. These frequencies of operation lie within a spectral region that extends from the very far infrared to the *vacuum-ultraviolet* (VUV) or soft-X-ray region. At the lowest frequencies at which they operate, lasers overlap with the frequency coverage of masers, to which they are closely related, and millimeter-wave sources using solid-state or vacuum-tube electronics, such as TRAPATT, IMPATT, and Gunn diodes, klystrons, gyroklystrons, and traveling-wave tube oscillators, whose principles of operation are quite different [1]. In common with electronic-circuit oscillators, a laser is constructed using an amplifier with an appropriate amount of positive feedback. The positive feedback is generally provided by mirrors that re-direct light back and forth through the laser amplifier. The acronym LASER, which stands for *light* **amplification** *by stimulated emission of radiation*, is in reality therefore a slight misnomer.[1]

1.1.1 A little bit of history

The basic physics underlying light emission and absorption by atoms and molecules was first expounded by Albert Einstein (1879–1955) in 1917 [2]. Richard Chace Tolman (1881–1948) observed that stimulated emission could lead to "negative absorption." In 1928 Rudolph Walther Landenburg (1889–1953) confirmed the existence of stimulated emission and negative absorption. It is interesting to note that a famous spectroscopist, Curtis J. Humphreys (1898–1986), who for most of his career worked at the U.S. Naval Ordnance Laboratory in Corona, California, might have operated the first gas laser without knowing it. Humphreys and his colleagues, notably William F. Meggers (1888–1966), studied the spectra emitted by low-pressure xenon lamps beginning in the later 1920s [3], and if they had had access to photodetectors operating in the near and middle infrared would almost certainly have observed the anomalously strong lines from the pure-xenon gas laser, notably at 2.062 μm and 3.507 μm. These laser lines are so characteristic of low-pressure xenon gas discharges that it is almost impossible to stop them from exhibiting laser oscillation, even in short laser tubes with no mirrors. Later, namely in 1939, Valentin Aleksandrovich Fabrikant (1907–1991) predicted the actual use of stimulated emission in gas discharges to amplify light. He subsequently claimed that he had observed

[1] The more truthful acronym LOSER was long ago deemed inappropriate.

"negative absorption" experimentally, and in 1951 he and his students filed a patent, but their work had little impact. Attempts to replicate Fabrikant's results experimentally showed that they had arisen from an experimental artefact (C. E. Webb, private communication). In the early 1950s, Charles Hard Townes (born in 1915) and his colleagues at Columbia University worked on the development of the maser, a microwave amplifier using stimulated emission, and published their first report of a working device in 1954 [4, 5]. Independent researchers Joseph Weber (1919–2000), of the University of Maryland [6], and Nikolai Basov (1922–2001) and Aleksandr Prokhorov (1916–2002), of the P. N. Lebedev Physics Institute in Moscow, proposed similar ideas [7]. On May 16, 1960 Theodore H. Maiman at the Hughes Research Laboratories reported the successful operation of the first laser – the pulsed ruby laser [8]. On December 13, 1961, Ali Javan (born in 1926), William R. Bennett Jr. (1930–2008), and Donald Heriott (1928–2007) demonstrated the first continuous-wave laser – the helium–neon laser [9]. In 1964, Nikolai Basov, Charles Townes, and Aleksandr Prokhorov received the Nobel Prize for "fundamental work in the field of quantum electronics, which has led to the construction of oscillators and amplifiers based on the maser–laser principle." Subsequently, in 1981, Arthur Schawlow (1921–1999), who was Charles Townes's brother-in-law, and Nicolaas Bloembergen (born in 1920) received the Nobel Prize for "their contribution to the development of laser spectroscopy."

1.2 Light and electromagnetic waves

Light is one form of electromagnetic radiation, the many categories of which make up the electromagnetic spectrum. Electromagnetic radiation, which transports energy from point to point at the velocity of light, can be described in terms of both *wave* and *particle* "pictures" or "models." This is the famous "wave–particle" duality of all fields or particles in our model of the Universe. In the electromagnetic-wave picture, waves are characterized by their frequency ν, wavelength λ, and the velocity of light c, which are inter-related by $c = \nu\lambda$. A propagating electromagnetic wave is characterized by a number of field vectors, which vary in time and space. These include the electric field $\mathbf{E}$ (measured in volts/m), the magnetic field $\mathbf{H}$ (measured in amps/m), the displacement vector $\mathbf{D}$ (measured in coulombs/m^2), and the magnetic flux density $\mathbf{B}$ (measured in webers/m^2 or tesla). For a complete description the *polarization* state of the wave must also be specified. *Linearly polarized* waves have fixed directions for their field vectors, which do not re-orient themselves as the wave propagates. *Circularly* or *elliptically* polarized waves have field vectors that trace out circular, or elliptical, helical paths as the wave travels along. In the particle picture, electromagnetic energy is carried from point to point as quantized packets of energy called *photons*. The energy of a photon of frequency ν is $h\nu$, where h is Planck's constant, namely 6.626×10^{-34} J s. Photons have zero mass, and travel at the velocity of light, but carry both linear and angular momentum. The linear momentum of a photon of wavelength λ is $p = h/\lambda$, and the angular momentum depends on the equivalent polarization state of the corresponding wave. Circularly polarized photons have angular momentum $h/(2\pi) = \hbar$.

Our everyday experience of "light" generally encompasses only the small part of the electromagnetic spectrum to which the human eye is sensitive, a wavelength range running roughly from 400 nm to 700 nm. The full electromagnetic spectrum, going from low to high frequencies, is divided into radiowaves (0–1 GHz), microwaves (1–300 GHz), infrared waves (of wavelength $\lambda = 0.7-1000$ μm; 300 GHz to 430 THz),[2] visible light ($\lambda = 400-700$ nm), ultraviolet light ($\lambda = 10-400$ nm), X-rays ($\lambda = 0.1$–10 nm), and γ rays ($\lambda < 0.1$ nm).

1.3 Some basic electromagnetic theory

In this book we will frequently discuss the electromagnetic wave properties of laser radiation, so we begin with a brief sumary of some key concepts that are important in later chapters. A linearly polarized electromagnetic wave of angular frequency ω that is polarized in the x direction and is traveling in the z direction has electric and magnetic fields[3]

$$E_x = E_0 e^{j(\omega t - kz)} \tag{1.1}$$

$$H_y = H_0 e^{j(\omega t - kz)}. \tag{1.2}$$

A linearly polarized wave polarized orthogonally to this one would have field pairs (E_y, H_x). The parameter k is called the propagation constant of the wave. The corresponding vector **k**, the *wave vector*, represents the direction of propagation of the wave and has magnitude $|\mathbf{k}| = k = 2\pi/\lambda$. The *phase* velocity of the wave is $c = \omega/k$. The important relations between the **E** and **H** fields and related vectors of the wave are

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \tag{1.3}$$

and

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}), \tag{1.4}$$

where $\epsilon_0 = 8.854 \times 10^{-12}$ F m^{-1} (farads/m) is the *permittivity* of free space, **P** is the polarization produced in the medium through which the wave is traveling, $\mu_0 = 4\pi \times 10^{-12}$ H m^{-1} (henrys/m) is the *permeability* of free space, and **M** is the *magnetization* produced in the medium through which the wave is traveling. The polarization is produced by the electric field of the wave according to

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E}, \tag{1.5}$$

[2] 1 GHz (gigahertz) = 10^9 Hz; 1 THz (terahertz) = 10^{12} Hz. The infrared range can also be somewhat arbitarily divided up into the sub-millimeter region ($\lambda = 0.1$–1 mm), the far infrared ($\lambda = 20-100$ μm), the middle infrared ($\lambda = 3-20$ μm), and the near infrared ($\lambda = 0.7-3$ μm).

[3] The polarization direction of a wave is strictly speaking the direction of its **D** vector, which is parallel to **E** except in *anisotropic media*, which will be dealt with in Chapter 17.

where χ is the (electric) *susceptibility* of the medium. The magnetization produced by the magnetic field of the wave is

$$\mathbf{M} = \chi_{\mathrm{m}}\mathbf{H}, \tag{1.6}$$

where χ_{m} is the magnetic susceptibility of the medium. Using Eqs. (1.5) and (1.6), Eqs. (1.3) and (1.4) can also be written, respectively, as

$$\mathbf{D} = \epsilon_0\epsilon_{\mathrm{r}}\mathbf{E} \tag{1.7}$$

and

$$\mathbf{B} = \mu_0\mu_{\mathrm{r}}\mathbf{H}, \tag{1.8}$$

where ϵ_{r} is the relative permittivity[4] and μ_{r} is the relative permeability of the medium. Clearly $\epsilon_{\mathrm{r}} = 1 + \chi$, and $\mu_{\mathrm{r}} = 1 + \chi_{\mathrm{m}}$. One of the triumphs of James Clerk Maxwell in his 1873 formulation of electromagnetic theory [10] in terms of the famous *Maxwell* equations was to show that the velocity of light was related to the fundamental permittivity and permeability quantities of a medium by

$$c = \frac{1}{\sqrt{\mu_0\mu_{\mathrm{r}}\epsilon_0\epsilon_{\mathrm{r}}}}, \tag{1.9}$$

which was first verified experimentally by Weber and Kohlrausch in 1857 [11]. In 1907 Rosa and Dorsey [12] obtained a value for the velocity of light of 299,788 km s^{-1} in this way. This was the most accurately known value at that time.[5] The propagation constant is $k = \omega\sqrt{\mu_0\mu_{\mathrm{r}}\epsilon_0\epsilon_{\mathrm{r}}}$.

An electromagnetic wave transports energy at the velocity of light.[6] The local energy flux is described in terms of the *Poynting vector* $\mathbf{S} = \mathbf{E} \times \mathbf{H}$. The time-averaged value of the Poynting vector at a point in space is frequently called the *intensity* and is measured in W m^{-2} (watts/m^2). The direction of the Poynting vector at a point in space is the direction of the *light ray*. If all the field components of an electromagnetic wave are the same in planes perpendicular to the **k** vector then this is a *plane wave*. An ideal *spherical wave* has equal field components on spherical surfaces, whose centers correpond to the point source of the waves. The Poynting vector of a spherical wave is directed radially outwards at any point on the surface of these spheres. A *beam wave*, of which a laser beam is an example, has its energy flux localized in a cone around some axial direction, and has field components that vary with the radial distance from this axis.

[4] This is frequently called the *dielectric constant* of the medium.
[5] The velocity of light has now been measured so accurately that it has been *defined* to have a value of 2.99792458×10^8 m s^{-1}. Consequently in the S.I. system of units the meter is now a derived unit: it is the distance traveled by light in vacuum in 3.335640952 ns.
[6] A more detailed discussion is provided in Appendix 5.

1.4 The polarization state of an electromagnetic wave

Description of the polarization state of an electromagnetic wave, whether this be a plane wave, spherical wave, or beam wave, involves specification of the magnitude, and phase, of orthogonal field components tranverse to the propagation direction **k**. For propagation in the z direction linearly polarized light propagates with a fixed **D** direction, and circularly polarized light has orthogonal D_x and D_y components of equal amplitude, but 90° out of phase. The resultant **D** traces out a helical path of circular cross-section as the wave propagates. If the **D** vector rotates in a clockwise direction when viewed in the propagation direction, the light is said to be *left-hand* circularly polarized.[7] Elliptically polarized light has orthogonal D_x and D_y components of unequal amplitude and arbitrary phase difference. The polarization state can be specified by the use of the *Stokes*[8] parameters, I, Q, U, and V, which four parameters are often also written as S_0, S_1, S_2, and S_3. The parameter I is related to the total intensity of the wave and can be written, in the case of a right-handed coordinate system with the wave propagating in the $+z$ direction, as

$$I = \langle E_x^2 \rangle + \langle E_y^2 \rangle,$$

where the $\langle\rangle$ indicate time-averaging. The Q parameter is

$$Q = \langle E_x^2 \rangle - \langle E_y^2 \rangle,$$

so Q describes the extent of linear polarization, since for pure x or y polarization $Q = I$. If we write the time-harmonic behavior of the electric-field components as

$$E_x = E_{0x}(t)\cos(\omega t + \phi_x)$$

and

$$E_y = E_{0y}(t)\cos(\omega t + \phi_y),$$

then

$$U = 2\langle E_{0x}E_{0y}\cos(\phi_x - \phi_y)\rangle$$

and

$$V = 2\langle E_{0x}E_{0y}\sin(\phi_x - \phi_y)\rangle.$$

U also describes the linearly polarized character of the wave, while V describes its circular polarization character, since V is maximum for a phase difference of 90° between E_x and E_y. The polarization state of an electromagnetic wave is often represented on a diagram called the *Poincaré sphere*.[9] Figure 1.1 shows a representation of the Poincaré sphere. The Stokes parameters U, V, and Q are the Cartesian coordinates of a point on the sphere. At any point P on the surface of the sphere the intensity is $I = \sqrt{U^2 + V^2 + Q^2}$.

[7] For a picture see Fig. 17.16.
[8] Sir George Gabriel Stokes (1819–1903), British mathematician and physicist.
[9] This is named for the French mathematician and physicist Jules Henri Poincaré (1854–1912).

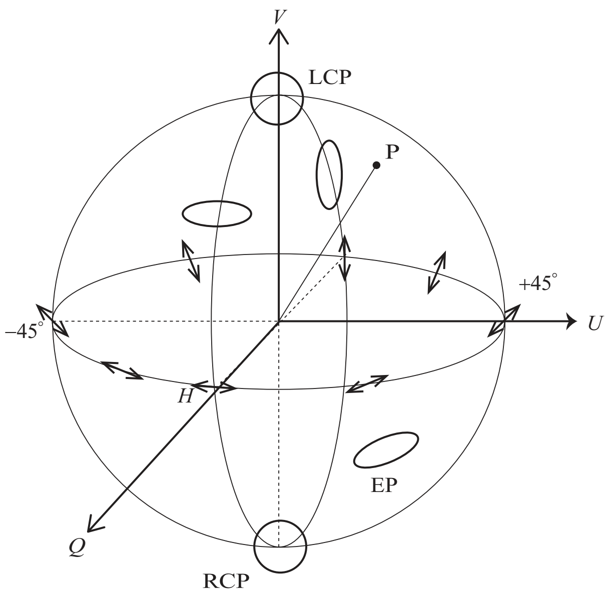

Fig. 1.1 The Poincaré sphere.

1.5 Electromagnetic waves and their propagation through matter

The values of the relative permittivity and relative permeability characterize a medium in terms of its difference from a vacuum. Most materials that are important in a discussion of lasers and optical devices are not strongly magnetic, and it is generally legitimate to assume that for such materials $\mu_r = 1$. It is important to note that the dielectric properties of a medium depend on frequency, even though this has not been explicitly stated in our preceding brief discussion. In addition, to include the properties of real, rather than ideal, matter in our discussion we allow the possibility that the dielectric can be represented by a complex number, so the frequency-dependent dielectric constant can be written as

$$\epsilon_r(\omega) = \epsilon'(\omega) - j\epsilon''(\omega). \tag{1.10}$$

Along with this goes a corresponding definition for the complex susceptibility, which is

$$\chi(\omega) = \chi'(\omega) - j\chi''(\omega). \tag{1.11}$$

Clearly, $\chi'(\omega) = \epsilon'(\omega) - 1$ and $\chi''(\omega) = \epsilon''(\omega)$. The minus sign in Eqs. (1.10) and (1.11) is a sign convention. We shall see that, with this sign convention, when $\chi''(\omega)$ is positive the medium absorbs energy from the wave.

Although a laser is a device that requires quantum mechanics for a complete description of its behavior, much can be learned from a classical analysis. In this context we describe a medium that has gain or absorption as one whose dielectric properties have been modified in a special way. When an electromagnetic wave propagates in the z direction through a medium with complex dielectric constant $\epsilon_r = \epsilon' - j\epsilon''$ the variation of the fields of the wave is, taking the electric field as an example,

$$\mathbf{E} = \mathbf{E}_0 e^{j(\omega t - k'z)}, \tag{1.12}$$

where the new propagation constant is

$$k' = \omega\sqrt{\mu_0\mu_r\epsilon_r\epsilon_0}. \tag{1.13}$$

Both ϵ' and ϵ'' vary with frequency. If we make the usual assumption that optical media are generally not strongly magnetic then we can take $\mu_r = 1$ and write

$$k' = \frac{\omega}{c_0}\sqrt{\epsilon' - j\epsilon''}. \tag{1.14}$$

For a medium that does not have large absorption or gain, we can make the assumption that $\epsilon'' \ll \epsilon'$, and Eq. (1.14) becomes

$$k' = \frac{\omega}{c}\left(1 - j\frac{\epsilon''}{2\epsilon'}\right), \tag{1.15}$$

where $c = c_0/\sqrt{\epsilon'}$ is the modified velocity of light in the medium. The variation of the amplitude of the electric field of the wave as it propagates is now

$$\mathbf{E} = \mathbf{E}_0 e^{-\frac{\alpha}{2}z} e^{j(\omega t - kz)}, \tag{1.16}$$

where $k = \omega/c$ and $\alpha = k\epsilon''/\epsilon'$. If we write $n = \sqrt{\epsilon'}$ and note from our previous discussion that $\epsilon'' = \chi''$, then

$$\alpha = \frac{k\chi''}{n^2}. \tag{1.17}$$

If the complex susceptibility were to become *negative* then we could write

$$\gamma = -\frac{k\chi''}{n^2}, \tag{1.18}$$

and the wave would increase in field amplitude as it propagated according to

$$\mathbf{E} = \mathbf{E}_0 e^{\frac{\gamma}{2}z} e^{j(\omega t - kz)}, \tag{1.19}$$

meaning that the medium has become an amplifier.

1.6 Spontaneous and stimulated transitions

In this chapter we shall consider the fundamental processes whereby amplification at optical frequencies can be obtained. These processes involve the fundamental atomic nature of matter. At the atomic level matter is not a continuum, but is composed of discrete particles – atoms, molecules, or ions. These particles have energies that can have only certain discrete values. This discreteness, or *quantization*, of energy is intimately connected with the duality that exists in nature. Light sometimes behaves as if it were a wave and in other circumstances as if it were composed of particles. These particles, called *photons*, carry the discrete packets of energy associated with the wave. For light of frequency ν the energy of each photon is $h\nu$, where h is Planck's constant, 6.6×10^{-34} J s. The energy $h\nu$ is the *quantum* of energy associated with the frequency ν. At the microscopic level the amplification of light within a laser involves the emission of these quanta. Thus, the term *quantum electronics* is often used to describe the branch of science that has grown from the development of the maser in 1954 and the laser in 1960.

The widespread practical use of lasers and optical devices in applications such as communications, and increasingly in areas such as signal processing and image analysis, has led to the use of the term *photonics*. Whereas electronics uses electrons in various devices to perform analog and digital functions, photonics aims to replace the electrons with photons. Because photons have zero mass, do not interact with each other to any significant extent, and travel at the speed of light, photonic devices promise small size and high speed.

1.7 Lasers as oscillators

In "conventional" electronics, where by the word "conventional" for the present purposes we mean frequencies at which solid-state devices such as transistors or diodes will operate, say below 10^{11} Hz, an oscillator is conveniently constructed by applying an appropriate amount of positive feedback to an amplifier. Such an arrangement is shown schematically in Fig. 1.2. The input and output voltages of the amplifier are V_i and V_o, respectively. The voltage gain of the amplifier is A_0, and, in the absence of feedback, $A_0 = V_\mathrm{o}/V_\mathrm{i}$. The feedback circuit returns part of the amplifier output to the input. The feedback factor $\beta = |\beta|e^{j\phi}$ is in general a complex number with amplitude $|\beta| \leq 1$ and phase ϕ:

$$V_\mathrm{o} = A_0(V_\mathrm{i} + \beta V_\mathrm{o}), \tag{1.20}$$

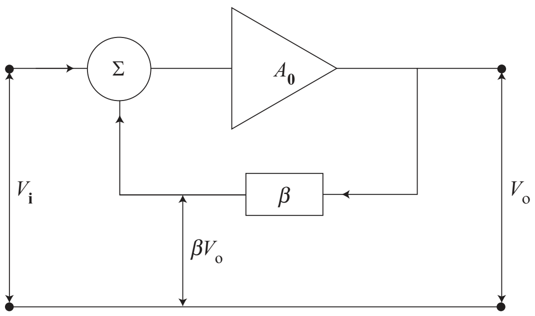

Fig. 1.2 The circuit diagram of a simple amplifier with feedback.

so

$$V_o = \frac{A_0 V_i}{1 - \beta A_0}, \tag{1.21}$$

and the overall voltage gain is

$$A = \frac{A_0}{1 - \beta A_0}. \tag{1.22}$$

As βA_0 approaches $+1$ the overall gain of the circuit goes to infinity, and the circuit would appear capable of generating a finite output without any input. This corresponds to the output of the amplifer being fed back being in phase with the input, which is referred to as *positive* feedback. In practice, electrical "noise," which is a random oscillatory voltage generated to a greater or lesser extent in all electrical components in any amplifier system, provides a finite input. Because βA_0 is generally a function of frequency the condition $\beta A_0 = +1$ is generally satisfied only at one frequency. The circuit oscillates at this frequency by amplifying the noise at this frequency which appears at its input. However, the output does not grow infinitely large, because, as the signal grows, A_0 falls – this process is called saturation. This phenomenon is fundamental to all oscillator systems. A laser (or maser) is an optical (microwave)-frequency oscillator constructed from an optical (microwave)-frequency amplifier with positive feedback, as shown schematically in Fig. 1.3. Light waves that become amplified on traversing the amplifier are returned through the amplifier by the reflectors and grow in intensity, but this intensity growth does not continue indefinitely because the amplifier saturates. The arrangement of mirrors (and sometimes other components) which provides the feedback is generally referred to as the laser *cavity* or *resonator*.

The characteristics of the device consisting of amplifying medium and resonator will be covered later; for the moment we must concern ourselves with the problem of how to construct an amplifier at optical frequencies, which range from 10^{11} Hz to beyond 10^{16} Hz. The operating frequencies of masers overlap this frequency range at the low-frequency end, the fundamental difference between the two devices being primarily one of scale. If the length of the resonant cavity that provides feedback is L, then for $L \gg \lambda$, where λ is the wavelength at which oscillation occurs, we generally have a laser. For $L \sim \lambda$ we usually have a maser, although the development of *microlasers*, which have small cavity lengths, has removed this easy way of distinguishing lasers from masers.

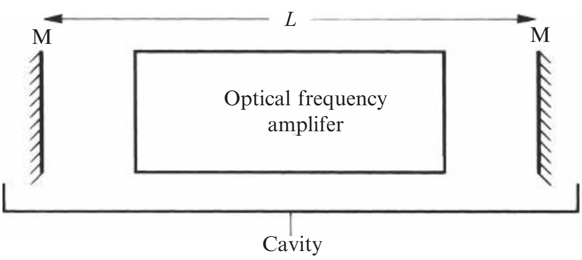

Fig. 1.3 A schematic diagram of a basic laser structure incorporating an amplifying medium and two feedback mirrors, M.

1.8 The energy levels of atoms, molecules, and condensed matter

All particles in nature have distinct states[10] that they can occupy. These states in general have different energies, although it is possible for particles in different states to have the same energy. The term “energy level” is used to describe a particle with a specific, distinct energy, without implying any particular information about its (quantum) state. The lowest energy state, in which a particle is stable, is called the *ground state*. All higher energy states are called *excited* states. Excited states are intrinsically unstable, and a particle occupying one will eventually lose energy and fall to lower energy states. When a particle falls from a higher energy state to a lower, energy is conserved. The energy ΔE lost by the particle can be emitted as a photon with energy $h\nu = \Delta E$: this is radiative energy loss. The particle can also lose energy *non-radiatively*, in which case the energy is dissipated into heating. Atomic systems have only electronic states, which in the simple Bohr model of the atom correspond to different configurations of electron orbits. The types of energy state that exist in a molecular system are more varied, and include electronic, vibrational, and rotational states. In a molecule, changes in the internuclear separation of the constituent atoms give rise to *vibrational* energy states, which have quantized energies. The various characteristic vibrational motions of a molecule are called its *normal modes*, which for a molecule with N atoms number $3N - 6$, unless the molecule is linear, in which case they number $3N - 5$. The quantized energies of a normal mode can be written as (see Chapter 10 for more details)

$$E_{\text{vib}} \simeq \left(n + \tfrac{1}{2}\right) h\nu_{\text{vib}}, \tag{1.23}$$

and form a ladder of (almost) equally spaced energy levels. Molecules also have quantized rotational energy levels, whose energies can be written as

$$E_{\text{rot}} \simeq BJ(J + 1), \tag{1.24}$$

where B is a rotational energy constant, and J is called the rotational quantum number. The overall energy state of a molecule thus has electronic, vibrational, and rotational components. A molecule in a particular combination of electronic and vibrational states is described as being in a *vibronic* state. A state with a specific combination of vibrational and rotational energies would be described as being in a *vibrot* state. As a rough rule of thumb, transitions between different vibronic states where the electronic state changes lie in the visible spectrum with energy spacings[11] $\sim$20,000 cm^{-1}, which corresponds to an energy spacing of $3 \times 10^{10} h/\lambda$ J. Transitions between vibrot states where the electronic energy does not change, but the vibrational state changes, are typically of energy $\sim$1,000 cm^{-1}. Transitions between different rotational states where the electronic and vibrational states do

[10] The term “state” in quantum mechanics corresponds to a configuration with a particular “state function,” which often corresponds to a specific set of quantum numbers that identify the state.

[11] The cm^{-1} unit is often used to describe energy spacings. A transition at wavelength λ (cm) between two levels has an energy spacing characterized by λ cm^{-1}.