Paper C2: Laser Science and Quantum Information Processing - Laser Physics

Prof. Simon Hooker

Michaelmas Term 2013

Contents

Lecture 1

Basic Laser Physics

1.1 Introduction

In these lectures we build on the basic laser physics covered last year in Paper BIII. All of the material covered in those lectures will be required for this course, and hence you should revise all of that work.

I would be very grateful if you were to bring to my attention any errors or unclear passages (email: simon.hooker@physics.ox.ac.uk).

1.2 Recommended texts

The lectures and notes should give you a good base from which to start your study of the subject. However, you will need to do some further reading. The following books are at about the right level, and contain sections on almost everything that we will cover:

1. "Laser Physics," Simon Hooker & Colin Webb, Oxford University Press.
2. "Principles of Lasers," Orazio Svelto, fourth edition, Plenum Press.
3. "Lasers and Electro-Optics: Fundamentals and Engineering," Christopher Davies Cambridge University Press.
4. "Laser Fundamentals," William Silfvast, Cambridge University Press.

You may also want to consult:

5. "Lasers," Anthony Siegman, University Science Books. This book approaches laser physics from more of an engineering viewpoint, but contains a wealth of information. It is particularly strong on cavity modes and Gaussian beams.

1.3 Lineshapes

As we saw on the BIII course, a variety of mechanisms can lead to broadening of the frequency spectrum of an optical transition. These mechanisms can be classified as being **homogeneous** or **inhomogeneous**.

The key property of homogeneous broadening mechanisms is that the broadening mechanism affects all atoms in the sample in the same way. As a consequence all atoms in the same energy level interact with radiation of frequency ω with the same strength. In contrast, inhomogeneous broadening arises from processes which alter

the frequency of an atomic transition by an amount that depends on some property of the atom, such as its velocity or position. In this case, the atoms may be considered to belong to different **classes**, depending on the frequency the frequency of the transition. The strength with which an atom will interact with radiation of frequency ω then depends on the class to which it belongs.

1.3.1 Homogeneous broadening

Homogeneous line broadening was discussed in detail as part of the BIII course. Examples include: lifetime, or natural, broadening; pressure broadening; and phonon broadening. In each case it was found that the spontaneous emission from a sample of atoms in some energy level E_2 on a transition to a lower level E_1 gave rise to a spectrum proportional to $g_{\rm H}(\omega-\omega_0)$, where ω_0 is the central frequency of the transition, $\hbar\omega_0 = E_2 - E_1$. The function $g_{\rm H}(\omega-\omega_0)$ is normalized[1] such that $\int_0^\infty g_{\rm H}(\omega-\omega_0)d\omega = 1$. For example, in the case of lifetime broadening we find that $g_{\rm H}(\omega-\omega_0)$ is a Lorentzian function:

$$g_{\rm H}(\omega-\omega_0) = \frac{1}{\pi}\frac{\gamma/2}{(\omega-\omega_0)^2+(\gamma/2)^2}, \tag{1.1}$$

where γ is the full-width at half maximum of the lineshape. This is equal to the sum of the contributions from the upper and lower levels:

$$\gamma = \frac{1}{\tau_2^{\rm rad}} + \frac{1}{\tau_1^{\rm rad}} \tag{1.2}$$

where $\tau_i^{\rm rad}$ is the radiative lifetime of level i.

1.3.2 Inhomogeneous broadening

Inhomogeneous broadening mechanisms cause the transition frequency of different atoms to be shifted by different amounts. Such broadening mechanisms are said to be **inhomogeneous**.

Inhomogeneous broadening will always be associated with some homogeneous broadening since natural broadening is always present, and the atoms may also be subject to other homogeneous broadening mechanisms such as phonon broadening. The transition lineshape observed from a sample subjected to both homogeneous and inhomogeneous broadening processes is given by the convolution of the homogeneous and inhomogeneous lineshapes. It should be clear, however, that when the frequency width of the inhomogeneous distribution is much greater than that of the homogeneous distribution the observed transition lineshape will simply be the inhomogeneous lineshape.

Doppler broadening

For gas lasers the most important line broadening mechanism is usually **Doppler broadening**, which arises from a combination of the Doppler effect and the thermal motion of the atoms: the observed frequency ω emitted by atoms moving towards the observer with velocity v_z will be $\omega-\omega_0 = (v_z/{\rm c})\omega_0$, and hence any distribution of velocities will give rise to a distribution of observed frequencies. If the distribution is Maxwellian then the proportion $P(v_z)dv_z$ of atoms with velocities in the range v_z to v_z+dv_z is given by :

$$P(v_z)dv_z = \sqrt{\frac{M}{2\pi k_B T}}\exp\left[-\frac{Mv_z^2}{2k_{\rm B}T}\right]\,dv_z, \tag{1.3}$$

[1] Actually the function $g_{\rm H}(x)$ is normalized such that $\int_{-\infty}^{\infty} g_{\rm H}(x)dx = 1$. In eqn (1.1) the argument of $g_{\rm H}(x)$ extends only to $-\omega_0$. However, since the linewidth is always very small compared to ω_0, extending the lower limit to $-\infty$ will have negligible effect on the integral.

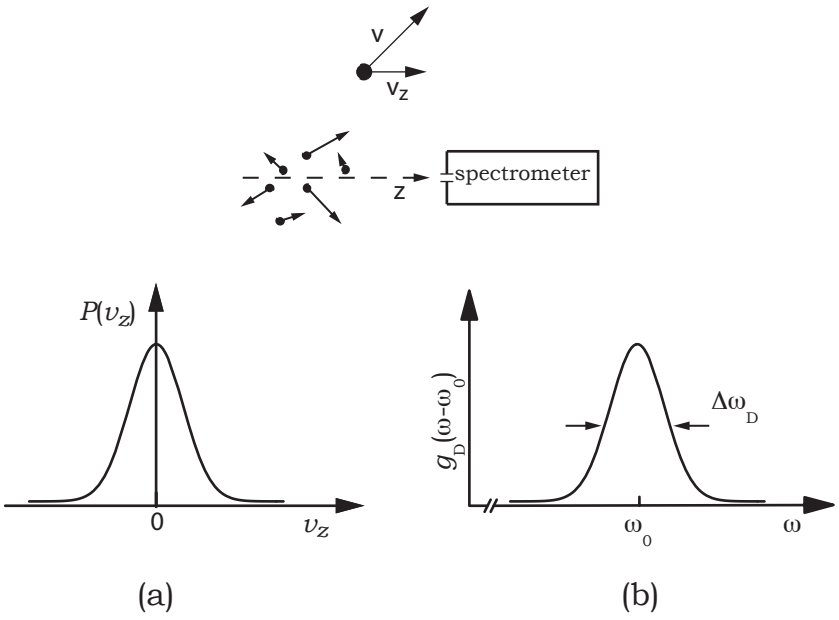

Figure 1.1: The formation of an inhomogeneously broadened spectral line by Doppler broadened gas atoms. Radiation emitted by the atoms in the z-direction is detected by the spectrometer. The distribution of atomic velocities in the z-direction (a) leads, via the Doppler effect, to a distribution of frequencies measured by the spectrometer, and hence broadening of the spectral line (b).

where M is the mass of each atom, and $k_{\rm B}$ the Boltzmann constant. The observed lineshape $g_D(\omega-\omega_0)$ is then found by making the substitution $v_z/c = (\omega-\omega_0)/\omega_0$, to yield

$$g_{\rm D}(\omega-\omega_0)d\omega = \sqrt{\frac{M}{2\pi k_{\rm B}T}}\exp\left[\frac{-Mc^2}{2k_{\rm B}T}\left(\frac{\omega-\omega_0}{\omega_0}\right)^2\right]\frac{c}{\omega_0}d\omega. \tag{1.4}$$

This is a Gaussian distribution with a full-width at half maximum, in this context known as the **Doppler width** $\Delta\omega_D$, of

$$\Delta\omega_D = \sqrt{8\ln 2}\sqrt{\frac{k_BT}{Mc^2}}\omega_0. \tag{1.5}$$

It is often convenient to re-write the lineshape in terms of $\Delta\omega_D$, *i.e.*

$$g_{\rm D}(\omega-\omega_0) = \frac{2}{\Delta\omega_D}\sqrt{\frac{\ln 2}{\pi}}\exp\left\{-\left[\frac{\omega-\omega_0}{(\Delta\omega_D/2)}\right]^2\ln 2\right\}. \tag{1.6}$$

This lineshape is normalized so that $\int_{-\infty}^{\infty} g_{\rm D}(x)dx = 1$.

Broadening in amorphous solids

In a number of important laser systems the lasing species are ions doped into a solid. If the solid is a good quality crystal, the laser transition will be broadened homogeneously by natural and phonon broadening. However if the solid is non-uniform, ions in different regions will experience different local environments. Of particular importance is the local value of the strain of the crystal lattice since this affects the electric field experienced by the active ion, which in turn affects the energy levels of the ion through the Stark effect. Other aspects of the local environment which can affect the transition frequencies of an active ion are the presence of impurity ions, or variations in the orientation of the crystal lattice. Since such effects change the centre frequency of the active ions according to their location in the medium, this contribution to the broadening of the frequency response of a macroscopic sample is inhomogeneous.

These effects are particularly important for ions doped in glasses, such as the Nd^{3+} ions in the Nd:glass laser, since in a glassy material the local environment varies significantly with position, leading to substantial inhomogeneous broadening. Very often the distribution of centre frequencies is found to follow a Normal, *i.e.* Gaussian, distribution, in which case the transition lineshape is also Gaussian.

Treatment of inhomogeneous broadening

We can develop a mathematical treatment of inhomogeneous broadening by considering that atoms with different centre frequencies behave as if they were different 'types' of atom. The corollary of this is that atoms with similar centre frequencies are of the same type, or **class**. The frequency range over which each class can interact appreciably is determined by the (smaller) degree of homogeneous broadening experienced by the class. For example, the strength with which the class of atoms with centre frequency ω_c interacts with radiation of frequency ω is determined by the homogeneous optical cross-section $\sigma_{21}(\omega - \omega_c)$. Note that the strength with which each class of atoms interacts with radiation of frequency ω depends on the detuning of ω from ω_c, not on the detuning from ω_0, the average centre frequency of the entire inhomogeneous lineshape.

In treating the interaction of inhomogeneously broadened atoms with radiation we can treat each class separately. We can define the number density of atoms in the upper and lower levels with centre frequencies between ω_c and $\omega_c + \delta\omega_c$ as $\Delta N_2(\omega_c)\delta\omega_c$ and $\Delta N_1(\omega_c)\delta\omega_c$ respectively. For example, suppose inhomogeneously broadened atoms interact with radiation of total intensity I in a narrow band near ω_L. The rate equation for atoms in the upper level would then be of the form:

$$\frac{d\Delta N_2(\omega_c)}{dt} = R_2(\omega_c) - \Delta N^*(\omega_c)\sigma_{21}(\omega_L - \omega_c)\frac{I}{\hbar\omega_L} - \frac{\Delta N_2(\omega_c)}{\tau_2} + \ldots, \tag{1.7}$$

where $R_2(\omega_c)$ is the pump rate of atoms of class ω_c into the upper level, $\Delta N^*(\omega_c) = N_2(\omega_c) - (g_2/g_1)N_1(\omega_c)$ is the population inversion density for atoms with a centre frequency ω_c, and τ_2 is the fluorescence lifetime of the upper level (assumed to be the same for all classes). In principle the rate equations for each class can then be solved separately, and the results integrated over ω_c to give the total population densities in the upper and lower levels.

1.3.3 Comparison of homogeneously and inhomogeneously broadened laser transitions

Imagine passing a probe beam of narrow-band, tunable radiation through a sample of atoms, as shown schematically in Figure 1.2. If we tune the frequency of the probe beam in the region around one of the absorption transitions of the atoms, we can record the fluorescence spectrum — that is, the spectrum of radiation emitted spontaneously from the excited level[2].

[2]Note that it is important to ensure that we record the spectrum of spontaneous emission rather than scattered radiation. In principle this can be achieved by using a pulsed probe beam and turning the detector on — 'gating' it — a short time after the

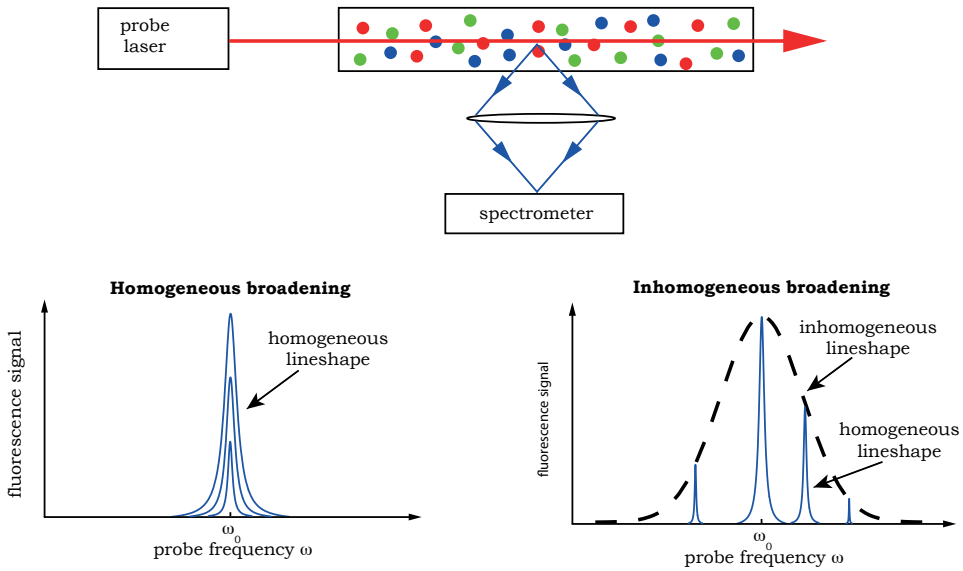

Figure 1.2: Fluorescence spectra recorded as a function of the frequency ω of a probe laser for the cases of homogeneous and inhomogeneous broadening.

The behaviour of homogeneously and inhomogeneously broadened transitions is compared in Table 1.1.

1.4 The optical gain cross-section

In the BIII course it was shown that the growth of a beam of radiation of spectral intensity $I(\omega, z)$ interacting with atoms via a homogeneously broadened transition is given by,

$$\frac{\partial I}{\partial z} = N^* \sigma_{21}(\omega - \omega_0) I(\omega, z), \tag{1.8}$$

where the **population inversion density** is defined as,

$$N^* = N_2 - \frac{g_2}{g_1} N_1, \tag{1.9}$$

and the **optical gain cross-section** is given by,

$$\sigma_{21}(\omega - \omega_0) = \frac{\hbar \omega_0}{c} B_{21} g_{\mathrm{H}}(\omega - \omega_0) = \frac{\pi^2 c^2}{\omega_0^2} A_{21} g_{\mathrm{H}}(\omega - \omega_0). \tag{1.10}$$

Note that it is often useful to define the **gain coefficient** $\alpha_{21}(\omega - \omega_0)$:

passage of the probe pulse. Scattered radiation, which is prompt and (in the absence of nonlinear effects) has the *same* frequency as the incident radiation, is not then detected.

Homogeneous transition	**Inhomogeneous transition**
For a given frequency ω of the probe laser, all the atoms in the sample will interact with the probe beam with the same strength	For a given frequency ω of the probe beam, only those atoms with centre frequencies ω_c close to ω will be excited.
The strength of the interaction increases as ω is tuned close to the centre frequency ω_0 of the transition since the homogeneous cross-section increases	The strength of the interaction increases as ω is tuned close to ω_0 since more atoms have this centre frequency
Atoms excited to the upper level will emit radiation with the same spectrum — described by $g_H(\omega - \omega_0)$	Atoms excited to the upper level will emit a homogeneously broadened spectrum centred at $\omega_c = \omega$
The *shape* and centre frequency of the fluorescence spectrum will be unchanged as the frequency of the probe laser is tuned across the transition	The shape of the fluoresence spectrum is unchanged, but the centre frequency tunes with ω.
The *intensity* of the fluorescence changes as the probe frequency is tuned; the fluorescence will be brightest when ω is tuned close to ω_0, and be negligibly weak when tuned many homogeneous linewidths away from ω_0	The intensity of the fluorescence changes as the probe frequency is tuned; the intensity of the spectrum will be greatest for probe frequencies close to ω_0 since more atoms can interact with the probe beam.

Table 1.1: Comparison of classes of line broadening

$$\alpha_{21}(\omega - \omega_0) = \frac{1}{I}\frac{\partial I}{\partial z} = N^* \sigma_{21}(\omega - \omega_0). \tag{1.11}$$

1.4.1 Laser rate equations in terms of the optical cross-section

We will often need to write rate equations to describe the transfer of population between levels of an atom. In terms of the Einstein coefficients, the rate equation for (say) the upper laser level can be written:

$$\frac{dN_2}{dt} = R_2 - (N_2 B_{21} - N_1 B_{12}) \int_0^\infty g_H(\omega - \omega_0)\rho(\omega) d\omega + \ldots \tag{1.12}$$

$$= R_2 - N^* \int_0^\infty \sigma_{21}(\omega - \omega_0) \frac{I(\omega)}{\hbar\omega} d\omega + \ldots \tag{1.13}$$

where R_2 is the rate of pumping of the upper laser level, $\rho(\omega)$ is the spectral energy density of the beam, and the $+\ldots$ indicate that in general there may be other processes to consider. The second line follows from the relations between the Einstein B-coefficients and the definition of the optical gain cross-section.

For narrow-band radiation the gain cross-section varies slowly over the spectral width of the radiation and so $I(\omega)$ acts like a Dirac delta function: $I(\omega) = I_T \delta(\omega - \omega_L)$ where I_T is the total intensity and ω_L the centre frequency of the beam. We then have:

$$\frac{dN_2}{dt} = R_2 - N^* \sigma_{21}(\omega_L - \omega_0) \frac{I_T}{\hbar\omega_L} + \ldots, \tag{1.14}$$

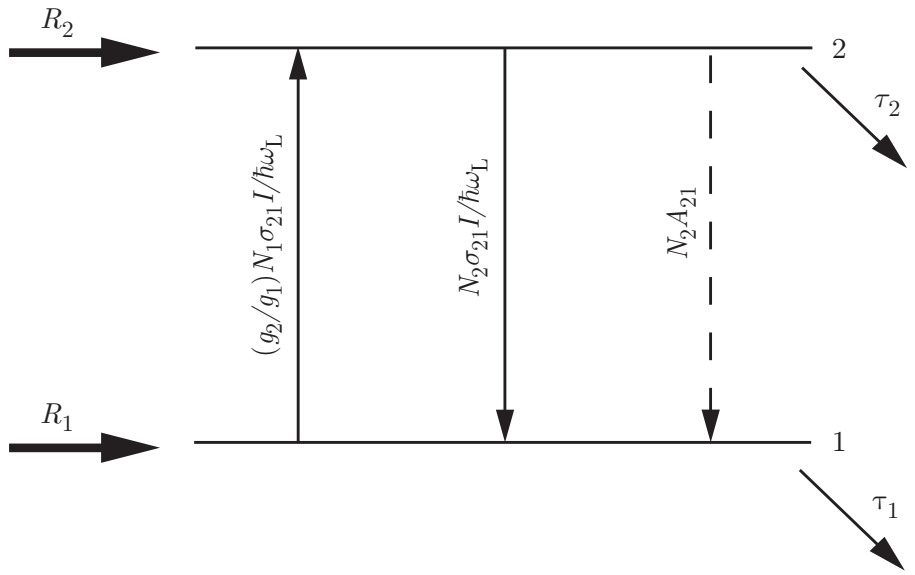

Figure 1.3: Processes affecting the upper and lower laser levels in a laser operating under steady-state conditions in the presence of an intense radiation beam.

In a similar way, we can integrate eqn (1.8) over the bandwidth of the beam to find the growth equation for the total intensity.

$$\frac{dI_T}{dz} = N^*\sigma_{21}(\omega_L - \omega_0)I_T = \alpha_{21}(\omega_L - \omega_0)I_T. \tag{1.15}$$

Equations (1.14) and (1.15) will be used frequently in our analysis of laser systems.

1.5 Gain saturation

Gain saturation refers to the phenomenon whereby the optical gain on a laser transition depends on the intensity of the radiation interacting with it. Figure 1.3 shows schematically the processes which affect the level populations of a laser operating under steady-state conditions in the presence of an intense beam of radiation. In the absence of radiation on the laser transition, pumping of the upper and lower levels at rates R_2 and R_1, respectively, will form a population inversion $N^*(0)$. However the populations in the upper and lower laser levels will be altered by the presence of radiation on the laser transition owing to stimulated transitions from $2 \rightarrow 1$ and absorption from $2 \leftarrow 1$. Hence the population inversion in general will be a function of the intensity[3] I of the (assumed narrow-band) radiation and the frequency detuning of this radiation from the centre frequency of the transition.

[3]We will usually drop the subscript T, and distinguish spectral intensity by writing it in the form $I(\omega)$.

1.5.1 homogeneous broadening

For a homogeneously broadened laser transition the rate equations for the laser levels may be written as,

$$\frac{dN_2}{dt} = R_2 - N^*\sigma_{21}(\omega_L - \omega_0)\frac{I}{\hbar\omega_L} - \frac{N_2}{\tau_2} \tag{1.16}$$

$$\frac{dN_1}{dt} = R_1 + N^*\sigma_{21}(\omega_L - \omega_0)\frac{I}{\hbar\omega_L} + N_2A_{21} - \frac{N_1}{\tau_1}. \tag{1.17}$$

In last year's lectures we showed how these may be solved straightforwardly to give the steady-state population inversion:

$$N^*(I) = \frac{N^*(0)}{1 + I/I_s(\omega_L - \omega_0)}. \tag{1.18}$$

where I_s is the **saturation intensity** (Units $\mathrm{W\,m^{-2}}$), and is given by

$$I_s(\omega_L - \omega_0) = \frac{\hbar\omega_L}{\sigma_{21}(\omega_L - \omega_0)\tau_R}, \qquad \textbf{Saturation intensity} \tag{1.19}$$

where we have emphasized that the saturation intensity depends (through the optical cross-section) on the detuning of the intense radiation from the centre-frequency of the transition[4]. The *recovery time* τ_R is given by

$$\tau_R = \tau_2 + \frac{g_2}{g_1}\tau_1[1 - A_{21}\tau_2]. \tag{1.20}$$

The recovery time depends on the lifetimes of both the upper and lower laser levels. It can usefully be thought of as giving an approximate measure of the time taken for the population inversion to re-establish itself after a perturbation, such as the passage of an intense laser pulse.

The **saturated gain coefficient** can be found directly from the saturated population inversion,

$$\alpha_I(\omega - \omega_0) = N^*(I)\sigma_{21}(\omega - \omega_0) = \frac{\alpha_0(\omega - \omega_0)}{1 + I/Is}. \tag{1.21}$$

Note that we can calculate the saturation intensity for absorption in a similar way. The result is of the same form as eqn (1.19), but with a different expression for the recovery time τ_R.

To illustrate how the gain coefficient saturates in a homogeneously broadened laser medium, we imagine measuring the gain coefficient under the following conditions:

1. Using a weak, tunable probe beam in the absence of any other radiation;
2. Using a weak, tunable probe beam in the presence of a second (intense) beam of radiation at a frequency ω_L of intensity $I_s(\omega_L - \omega_0)$.
3. An intense, tunable probe beam of constant intensity $I_s(0)$.

[4]The saturation intensity is *not* a spectral quantity; it has units of intensity! However, just like the optical gain cross-section itself, it *does* depend on the frequency detuning. In order to avoid clutter, we will often drop the explicit frequency dependence of the saturation intensity and just write I_s.

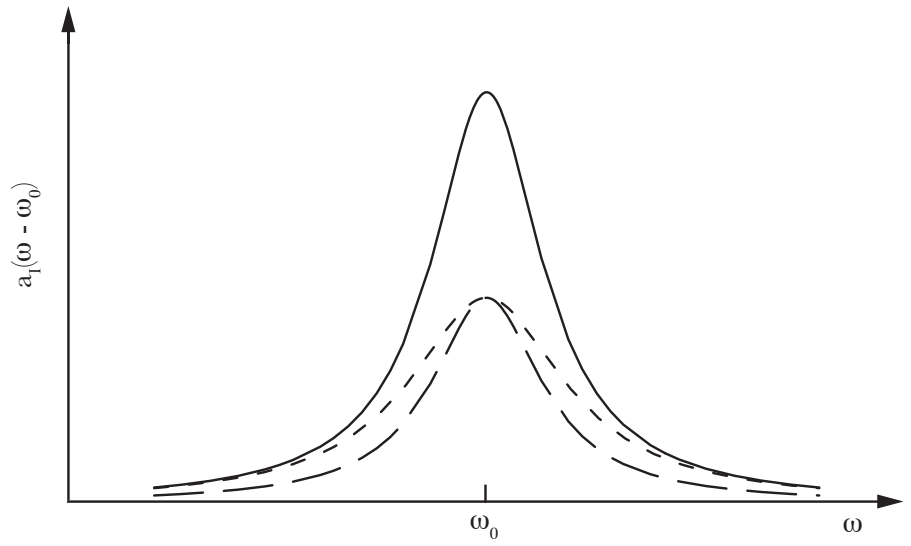

Figure 1.4: The gain coefficient measured for a homogeneously broadened transition by: a weak probe beam (solid line); a weak probe beam in the presence of an intense beam of intensity $I_s(\omega_L - \omega_0)$ at frequency ω_L (long-dashed line); an intense probe beam of intensity $I_s(0)$ (short-dashed line).

Clearly in the first case the gain coefficient measured as a function of the frequency ω of the probe beam will simply be the small-signal gain coefficient $\alpha_0(\omega - \omega_0)$. In the second case, however, the presence of the intense beam will decrease the population inversion density. Since the intense beam has an intensity equal to the saturation intensity (by definition, whatever the frequency ω_L of the intense beam), the population inversion will burnt down to *half* that generated by the pumping in the absence of the intense beam. As the frequency of the probe beam is tuned across the laser line, the gain measured will therefore always be half of the small-signal gain. Finally, for the case of the intense probe beam, the degree of saturation of the population inversion depends on the detuning of the probe from the centre frequency of the transition. In this case the intensity of the probe is equal to the saturation intensity when tuned to the line-centre. Hence when $\omega = \omega_0$ the measured gain coefficient will be half the small-signal gain coefficient. At large detuning, however, the intensity of the probe beam will be small compared to the saturation intensity; here the beam interacts only weakly with the atoms and hence perturbs their populations only slightly. The measured gain coefficient must, then, tend to the small-signal value at large detuning. The three measured gain profiles are shown in Figure 1.4.

1.5.2 Inhomogeneously broadened transitions

The phenomenon of gain saturation is considerably more complicated when the laser transition is inhomogeneously broadened since we must now consider separately the extent to which the population inversion of each frequency class is saturated, and the contribution by each such class to the overall gain.

We will restrict ourselves to a qualitative discussion.[5] Consider an inhomogeneously broadened gain medium subjected to a beam of intense, narrow-band radiation with frequencies close to ω_L. The extent to which each class of atoms is saturated — that is, the extent to which, for each class of atoms, the population density of the upper level is reduced — depends on the detuning of the intense beam from the centre frequency of the class. It should be clear that significant saturation will only occur for those classes of atoms with centre frequencies such that $|\omega_c - \omega_L| \ll \Delta\omega_H$; for these atoms the population inversion will be strongly burnt down. However, those

[5] A mathematical treatment may be found in Chapter 5 of H&W.

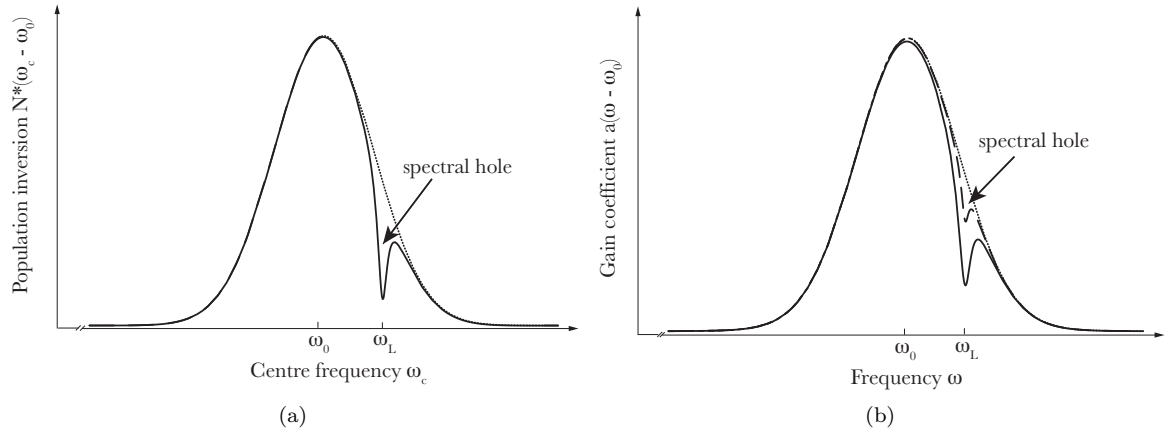

Figure 1.5: Saturation effects on inhomogeneously broadened transitions. (a) The effect of a beam of intense, narrow-bandwidth radiation of total intensity I with frequencies close to $\omega_{\rm L}$ on the distribution of the population inversion over centre frequencies. The dotted curve shows the unperturbed distribution for $I = 0$, and the solid curve the distribution for $I \approx I_{\rm s}(0)$. For this plot $\Delta\omega_{\rm H}/\Delta\omega_{\rm D} = 0.1$. (b) Spectral hole burning in the gain profile of an inhomogeneously broadened laser transition by a narrow-bandwidth beam of intensity I and with frequencies close to $\omega_{\rm L}$. The gain profile measured by a weak probe beam is shown for $I = 0$ (dotted), $I = I_{\rm s}(0)$ (dashed), $I = 10I_{\rm s}(0)$ (solid). For this plot $\Delta\omega_{\rm H}/\Delta\omega_{\rm D} = 0.03$.

classes with centre frequencies significantly different from that of the radiation (on the scale of $\Delta\omega_{\rm H}$) will not be saturated to any appreciable extent. As such, a **spectral hole** will be burnt into the population inversion, as illustrated in Fig. 1.5(a). The width of this spectral hole will be approximately $\Delta\omega_{\rm H}$.

The gain coefficient measured by a weak probe is given by integrating the contribution to the gain from each class. Hence we expect the measured gain coefficient in this case to exhibit a spectral hole, reflecting that in the population inversion. This phenomenon is shown in Figure 1.5(b).[6]

Intense probe The saturation of the gain as measured by an intense probe (with no other beam of radiation present) is very different from that shown in Figure 1.5(b) since as the probe is tuned across the transition the class of atoms which provide gain are also the class which experiences saturation. Hence in this situation we do not expect a spectral hole to be burnt into the gain profile.

It can be shown that for the case of extreme inhomogeneous broadening (i.e. $\Delta\omega_{\rm D} \gg \Delta\omega_{\rm H}$), the gain coefficient measured by an intense probe is equal to a constant fraction of that measured by a weak probe. Further, if the homogeneous lineshape is Lorentzian, it may be shown that the saturated gain coefficient is,

$$\alpha_I^{\rm D}(\omega - \omega_0) = \frac{\alpha_0^{\rm D}(\omega - \omega_0)}{\sqrt{1 + I/I_{\rm s}(0)}}. \qquad \text{(Special case)} \tag{1.22}$$

Notice that this looks rather similar to eqn (1.21), which was obtained for homogeneous broadening. However,

[6] It can be shown that in the limit of extreme inhomogeneous broadening — i.e. the inhomogeneous linewidth is much greater than the homogeneous linewidth — and if the homogeneous lineshape is Lorentzian, the spectral hole burnt in the gain profile is also Lorentzian, with a full-width at half maximum of $\Delta\omega_{\rm hole} = \Delta\omega_{\rm H}\left[1 + \sqrt{1 + I/I_{\rm s}(0)}\right]$. For this case we see that if the intensity of the saturating beam is small compared to the saturation intensity $\Delta\omega_{\rm hole} \approx 2\Delta\omega_{\rm H}$. This result is consistent with our expectation that the spectral hole burnt in the population inversion is approximately $\Delta\omega_{\rm H}$ wide, and the range of frequency classes with which the probe interacts is also approximately equal to $\Delta\omega_{\rm H}$.

this similarity is very mis-leading since eqn (1.21) refers to the gain measured for a homogeneously- broadened transition by a weak probe in the presence of a saturating beam at $\omega = \omega_{\rm L}$; eqn (1.22) refers to the gain measured for an extreme inhomogeneously broadened transition by an intense probe in the absence of any other saturating radiation. The situations couldn't be more different!

1.5.3 Summary

To summarize: for homogeneously broadened gain media an intense beam of radiation at $\omega = \omega_{\rm L}$ causes the entire population inversion to be burnt down by a factor $[1 + I/I_{\rm s}(\omega_{\rm L} - \omega_0)]$ such that the gain profile measured by a weak probe beam has the same shape as that measured in the absence of the saturating beam, but with the gain being reduced by the factor $[1 + I/I_{\rm s}(\omega_{\rm L} - \omega_0)]$; for inhomogeneously broadened gain media a spectral hole is burnt into the population inversion, and the gain measured by a weak probe beam is only reduced for frequencies close to $\omega_{\rm L}$. For homogeneously broadened systems the gain measured by an intense probe beam is reduced by a greater extent at frequencies close to the line centre, leading to a flattening and broadening of the measured gain profile; for strongly inhomogeneously broadened media the gain is everywhere reduced by the same factor $\sqrt{1 + I/I_{\rm s}(0)}$, and the measured gain profile has the same shape as measured by a weak probe beam.

1.6 Threshold behaviour

We now examine how lasers operating on homogeneously and inhomogeneously broadended transitions behave as they are brought above threshold.

Let us suppose that the gain medium is inserted into a optical cavity of length L. In general the modes of an optical cavity are described by the set of integers (l, m, p). The integers l and m determine the transverse profile of the right- and left-going beams in the cavity, and are said to determine the **transverse modes** of the cavity; the integer p determines the number of half-wave loops of the radiation field along the axis of the cavity, and so describes the **longitudinal modes**. The frequencies of the cavity modes are determined primarily by p; for example, for the case of a plane-plane cavity the longitudinal modes are separated in frequency by an amount,

$$\omega_{p,p-1} = \frac{\pi {\rm c}}{L}. \tag{1.23}$$

As discussed in the BIII course, in order for laser oscillation to occur on a cavity mode the round-trip gain must equal the round-trip loss.

1.6.1 Homogeneously broadened laser transitions

As the pumping is increased from zero the population inversion N^*, and hence the small-signal gain coefficient $\alpha_0(\omega)$, will increase. For low levels of pumping, spontaneous emission will increase the energy in all of the cavity modes with frequencies lying within the linewidth of the laser transition. The intensity will be low, and the number of photons per mode will typically be much less than unity.

With further increases in pumping the small-signal gain coefficient will increase until the threshold condition is met for the cavity mode closest to the line centre, causing the onset of oscillation on that cavity mode. The onset of oscillation will be accompanied by a very large (by a factor of order 10^{15}) increase in the energy density and number of photons in the oscillating mode.

Further increases in the pumping do not increase either the gain coefficient or the population inversion. In fact this must be the case since in the steady-state the round-trip gain must always be balanced by the round-trip loss. Instead, as the pumping is increased the intensity of the oscillating mode increases — this,

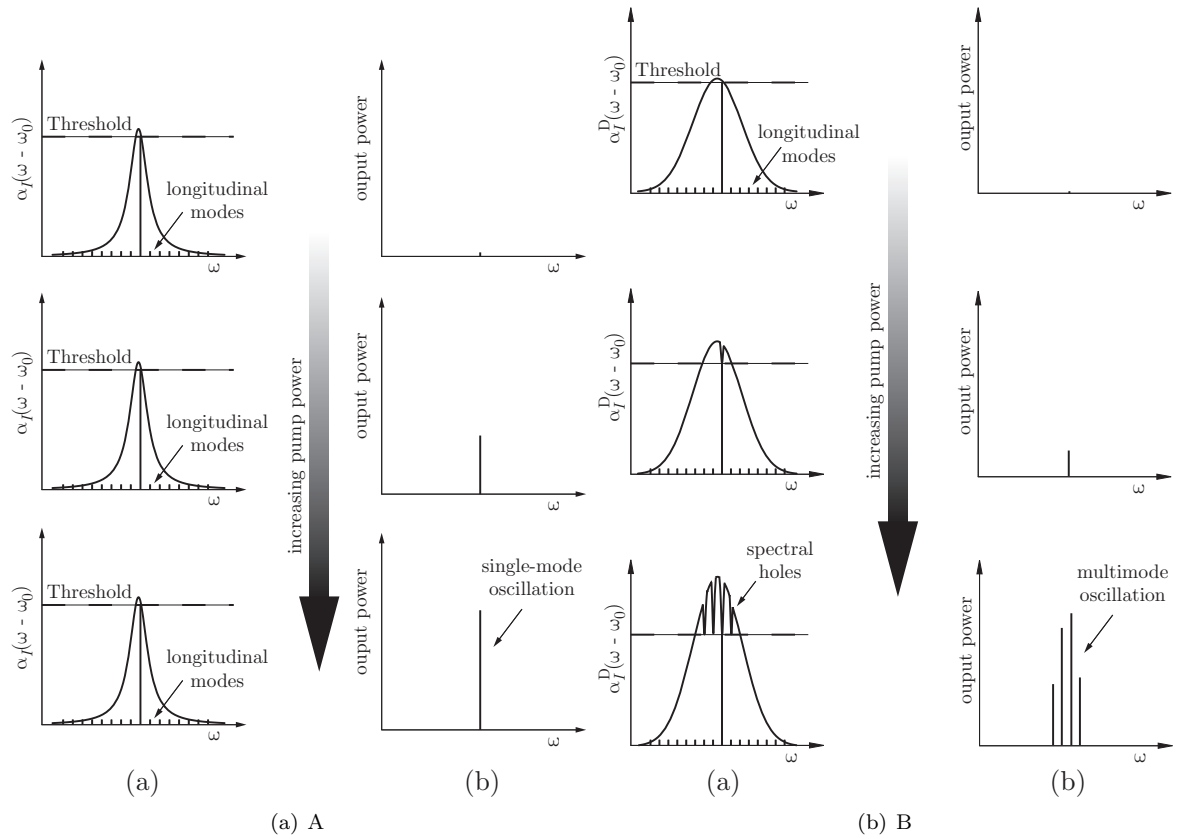

Figure 1.6: Behaviour of the small-signal gain (a) and output power (b) as the pumping power is increased above the threshold value for lasers operating on : (A) homogeneously; and (B) inhomogeneously broadened transitions

after all, is what we are trying to achieve! The increased intensity of radiation circulating within the cavity reduces, or burns down, the population inversion to the threshold value. In other words, the laser transition is saturated. For a given level of pumping the intensity of the oscillating mode is determined by the condition that the saturated round-trip gain is equal to the round-trip loss, or, equivalently, that the population inversion is burnt down from the value, N_0^*, that would be achieved at that level of pumping, to the threshold value $N_{\rm thresh}^*$.

The behaviour of the saturated gain coefficient $\alpha_I(\omega - \omega_0)$ and the output power of the laser as the level of pumping is increased is illustrated schematically in Figure 1.6(a).

Spatial hole burning

In principle, for a homogeneously broadened laser system only the cavity mode with the largest gain can ever oscillate; since the population inversion is clamped to the threshold value, and all atoms interact with radiation of a given frequency with the same strength, the other modes will always be below the threshold for oscillation. However, in practice in some circumstances more than one mode can oscillate. This behaviour arises from the fact that the spatial distribution of the intensity of the oscillating mode is not uniform. In particular the oscillating mode will form a standing wave within the cavity. Near the anti-nodes of the standing wave the population inversion will be burnt down to threshold value; near the nodes the intensity will be low, and the population inversion essentially unsaturated. The non-uniform burning down of the population inversion in

a homogeneously broadened laser system is known as **spatial hole-burning**, by analogy with the spectral hole-burning discussed in the next section. Spatial hole-burning allows other modes, with slightly different frequencies, to feed off regions of unsaturated population inversion and reach the threshold for oscillation. In such circumstances two or more cavity modes can oscillate, leading to **multimode oscillation** at several frequencies within the linewidth of the transition.

1.6.2 Inhomogeneously broadened systems

The above-threshold behaviour of inhomogeneously broadened lasers is quite different since different classes of atoms interact with different frequencies, and hence different cavity modes. Once again, as the pumping is increased to the threshold value the mode with the largest gain will start to oscillate. Now, however, as the pumping is increased the population inversion is only clamped to the threshold value for those classes of atoms which can interact with the frequency of the oscillating mode; other classes will be essentially unaffected, and for these classes the population inversion will continue to increase as the pumping increases. As a consequence the gain coefficient $\alpha_I^{\mathrm{D}}(\omega - \omega_0)$ develops a spectral hole at the frequency of the oscillating mode.

Further increases in the pumping level will allow other modes, of different frequencies, to reach the threshold for oscillation. Thus, in general as the pumping level is increased in an inhomogeneously broadened laser oscillator, the number of oscillating modes increases as well as the output power of each oscillating mode. The process is illustrated schematically in Figure 1.6(b).

Lecture 2

Solid State Laser Materials

A large number of technologically and scientifically important lasers operate by optically pumping transitions within ions doped as an impurity species into a variety of crystalline or glass hosts. The term **solid state lasers** is usually reserved for systems of this type, lasers operating between levels of the electron band structure found in semiconductors being referred to as **semiconductor** or simply **diode** lasers.

Solid state laser materials offer advantages over liquid and gas lasers in that they are robust, chemically inert, require no special handling, and do not degrade or become contaminated with use.

The spectroscopy of ions doped into solid hosts is a complex subject. Fortunately for our purposes we need only understand a few key concepts in order to appreciate how different laser systems work, and why they behave as they do. In particular we would like to know the energy level structure of the ions, the lifetimes of the levels, and the linewidth of the transitions between them.

2.1 General considerations

Ions doped into a solid host differ from free ions in that they are subjected to the electric field — the **crystal field** — arising from the ions of the solid host. The interaction with the crystal field leads to an additional term in the Hamiltonian of the form,

$$H_{\mathrm{c}} = -\mathrm{e}V_{\mathrm{c}}, \tag{2.1}$$

where V_{c} is an electrostatic potential describing the crystal field. In general the effect of this interaction is to split and shift the energy levels from their positions in the absence of the crystal field, an effect known as the Stark effect[1].

The crystal field has a symmetry reflecting that of the crystal lattice. Consequently a proper treatment of the effect of the crystal field involves detailed consideration of this symmetry; indeed in some circumstances these symmetry properties are used to label the energy levels arising from the interaction with the crystal field.

The nature of the influence of the crystal field on the energy level structure depends critically on its strength relative to the other terms in the Hamiltonian. Three limiting cases may be identified:

- **Weak field**: $|H_{\mathrm{c}}| \ll |H_{\mathrm{s-o}}| \ll |H_{\mathrm{re}}|$. In this case the interaction of the crystal field is weak compared to the spin-orbit and residual electrostatic interactions (as well the interaction with the central field) so that the energy levels are changed only slightly from those in the free ion. The effect of the crystal field in this case is to cause a small splitting and shifting of the levels of the free ion to form a so-called **manifold** of

[1] Named after the German physicist Johannes Stark (1874-1957).

closely-spaced levels. Since the energy shifts are small, the labelling of the manifolds is simply that of the energy levels of the free ion. Further, in the weak field case the energy level structure of an ion will be almost independent of the crystal host in which it is embedded. A good example of the weak field case are the energy levels formed by the $4f^n$ configurations in trivalent rare earth ions discussed in Section 2.2.

- **Intermediate field**: $|H_{s-o}| \ll |H_c| \ll |H_{re}|$. In this situation the crystal field interaction must be considered as a perturbation acting on the terms formed by the residual electrostatic interaction before the spin-orbit interaction is taken into account; the crystal field acts on the terms formed by the residual electrostatic interaction and splits and shifts them. The spin-orbit interaction leads to further splitting of the energy levels. For intermediate crystal fields it is no longer possible to label the energy levels with the quantum numbers L, S, and J. The first-row transition metal ions, such as Cr^{3+} frequently exhibit crystal fields of intermediate strength. It is more difficult to calculate the energy levels for intermediate crystal fields than for weak- or strong-fields. Frequently, therefore, the energy levels of ions experiencing intermediate crystal fields are labelled with either a weak- or strong-field notation.

- **Strong field**: $|H_{s-o}|, |H_{re}| \ll |H_c|$. Here the spin-orbit or residual electrostatic interactions may be ignored (in the first approximation) and the crystal field acts on the single-electron orbitals of the central potential to give single-electron 'crystal-field orbitals' which reflect the symmetry of the crystal field. As discussed below, the energies of the orbitals depends on the spatial distribution of the orbital wave function with respect to the crystal lattice. The energy levels are labelled by the symmetry properties of the electron wave function using a notation derived from group theory. Strong crystal fields can arise in the second- and third-row transition-metal ions.

For each electron the relative strengths of the residual electrostatic, spin-orbit, and crystal field interactions depends very strongly on the radial distance r_i of the electron from its nucleus, and the distance d from the active ion to the neighbouring ions. For example, for neighbouring ions distributed with octahedral symmetry the energy of interaction is proportional to,

$$\frac{\langle r_i^4 \rangle}{d^5}$$

where $\langle r_i^4 \rangle$ is the expectation value of r_i. In contrast, the energy shift arising from the spin-orbit interaction is proportional to,

$$\left\langle \frac{1}{r_i^3} \right\rangle$$

Hence the ratio of the interaction with the crystal field to the spin-orbit interaction varies approximately as,

$$\frac{\Delta E_c}{\Delta E_{s-o}} \propto \frac{\langle r_i^7 \rangle}{d^5}. \tag{2.2}$$

As a consequence very small changes in the relative size of the mean electron radius or the nearest-neighbour distance can change the relative strength of the interaction with the crystal field enormously.[2]

[2]Frequently it is said that the relative strength of an electron's interaction with the crystal field depends on the extent to which it is shielded by other electrons in the ion. For example, the 4f electrons in trivalent rare earth ions are often said to be shielded by the 'outer' 5s and 5p electrons. Shielding effects of this type do occur, but calculation shows them to be approximately an order of magnitude smaller than the effects of changing mean electron radius and nearest-neighbour distance. For a discussion of this point see G. Burns, *Solid State Physics*, Academic Press.

2.1.1 Radiative transitions

For all the solid-state laser transitions discussed in this lecture, the laser transition occurs between two levels *within the same electron configuration*. The parity of the electron wave function does not change in such transitions, and hence they are forbidden by the selection rules of electric dipole radiation. Whilst this would indeed strictly be the case for transitions between the equivalent levels in a free ion, for ions doped into a solid, these electric dipole transitions may occur for two reasons. The first is simply that the presence of the crystal field means that there is no longer inversion symmetry at the site of the active ion: in other words the electron wave function no longer has a definite parity and under the transformation $\boldsymbol{r}_i \rightarrow -\boldsymbol{r}_i$ the electron wave function can change magnitude (as well as possibly change sign). A *pure* electron configuration *must* have a well-defined parity since it is a solution of the Schrödinger equation in a spherically-symmetric central field. However, the crystal field interaction is not spherically symmetric and therefore introduces small admixtures of configurations of opposite parity to form a wave function which is no longer spherically symmetric, and no longer has a well-defined parity. These admixtures of configurations of opposite parity can allow transitions to occur between two levels which nominally have the same electron configuration.

A second mechanism by which electric dipole transitions may occur between two levels of the same configuration is a dynamically-induced transition strength through vibrations of the crystal lattice which destroy the inversion symmetry at the site of the active ion. This second mechanism becomes important when the active ion is located at a crystal site exhibiting a high degree of symmetry. As might be expected, in such cases the radiative lifetime of the level depend strongly on the temperature of the crystal.

Notwithstanding the above, the dipole-forbidden nature of transitions within a single configuration means that the transition rates are significantly slower than fully dipole-allowed transitions. The *radiative* lifetimes of the upper levels of visible transitions within ions doped into a solid are 3 to 6 orders of magnitude longer than fully dipole-allowed transitions in free atoms and ions.

Finally we note that the strength of a radiative transition can depend on the orientation of the polarization of the radiation with respect to the axes of the crystal. This is true for both absorption and emission. In such cases the orientation of the crystal axes with respect to the axis of the laser cavity and the direction and polarization of the pump radiation can be important, and the laser output can be partially or totally polarized.

2.1.2 Non-radiative transitions

Non-radiative transitions play a very important role in the operation of solid-state lasers: rapid non-radiative decay can provide an efficient route for feeding population from one or more excited levels into the upper laser level, and similarly can help to keep the population of the lower laser level low.

Phonon de-excitation

The energy levels of an active ion are coupled to the vibrations of the crystal lattice via the crystal field interaction. The lattice vibrations have a spectrum of normal modes,[3] and their excitation is quantized; the unit of excitation being known as a **phonon**. Transitions between energy levels may therefore occur not by emission of radiation, but by the emission or absorption of one or more lattice phonons. The rate of such **phonon de-excitation** is found to decrease very rapidly with the number of phonons involved. For example, for the case of weak coupling it may be shown that the rate of de-excitation on a transition of energy E_{21} by phonons of energy E_{p} varies as $\exp[-\beta(E_{21}/E_{\mathrm{p}})]$, where β is a positive constant which depends on the transition. In practice the de-excitation rate is found to be small if the number of phonons required is greater than approximately 5, and consequently the dominant contribution to the decay rate is from the phonons with the largest possible energy.

[3] There are two classes of these modes, known as acoustic and optical modes. A discussion of normal modes of crystals may be found in textbooks on solid-state physics.

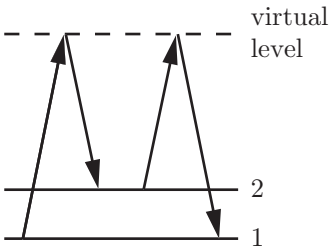

Figure 2.1: Schematic diagram of the two-phonon Raman process leading to line broadening. In this process phonons of two different energies transfer ions from level 1 to level 2 via a virtual level, followed by the reverse process which returns the ion to level 1. The net result is no change in either the level occupied by the ion or the phonon spectrum but a broadening of the energy level by an amount proportional to the rate at which this process occurs.

The rate of phonon de-excitation depends strongly on the strength of coupling between the electrons of the ions and the crystal field, and on the phonon spectrum of the lattice — and in particular on the maximum phonon energy. The de-excitation rates therefore vary widely for different combinations of impurity ions and hosts; the rate of phonon de-excitation also increases rapidly with temperature which reflects the increasing density of phonons in a given mode with temperature (determined by the Bose-Einstein distribution). For small energy differences the rate of non-radiative transitions can be extremely high: at room temperature the rate is of order 10^{11} - 10^{12} s^{-1}.

The strong dependence of the rate of phonon de-excitation on the energy gap between the initial and final levels of the ion means that if a level is to be suitable as an upper laser level it should lie significantly above the nearest lower-lying level. This is necessary if the level is to have a long fluorescence lifetime, which allows a large population to build up in the level; and ensures that it decays predominantly by emission of radiation, which enables the population in the upper laser level to be used efficiently.

2.1.3 Line broadening

The natural linewidths of transitions between the levels of an impurity ion are typically very small, owing to the long radiative lifetimes of the levels. In addition a temperature-dependent lifetime broadening arises from the increased rate of decay of the levels caused by non-radiative transitions. In practice, however, the measured linewidths of transitions are significantly broader than can be accounted for by the measured lifetimes of the upper and lower levels, and hence other processes must be responsible for the additional broadening.

A variety of mechanisms can cause this additional broadening. One of these, two-phonon Raman scattering is illustrated in Fig. 2.1. In this process two phonons of different energy are successively absorbed and emitted so as to remove, and then return an ion from its energy level. Since the ion is returned to its initial level there is no change in the lifetime of the level, but a broadening occurs proportional to the rate of scattering. As might be expected the rate of scattering, and hence its contribution to the linewidth, depends strongly on temperature.

Phonons may also play a more direct role in determining the linewidth of a radiative transition. For some ion-host combinations radiative transitions can be accompanied by the emission or absorption of one or more phonons. These so-called vibronic transitions (discussed in Section 2.3.1) have very broad, temperature-dependent linewidths, as discussed below.

2.2 Trivalent rare earths, $4f^n - 4f^n$ transitions

2.2.1 Energy level structure

A large number of solid-state lasers operate on $4f^n - 4f^n$ transitions in trivalent rare earth ions, and in particular the lanthanide series — the actinides being radioactive, and hence difficult to work with. Important trivalent rare-earth laser ions include: Ce^{3+}, Nd^{3+}, Ho^{3+}, and Er^{3+}.

Neutral atoms of these rare earths have a ground state electronic configuration of the form $[Xe]\,4f^{n+1}\,6s^2$, or $[Xe]\,4f^n\,5d\,6s^2$ where $n \geq 0$ and [Xe] represents the ground state configuration of xenon:

$$1s^2\,2s^2 2p^6\,3s^2 3p^6 3d^{10}\,4s^2 4p^6 4d^{10}\,5s^2 5p^6.$$

In forming the trivalent ion, the three loosest-bound electrons are lost:[4] the two 6s electrons and either one of the 4f electrons or the 5d electron, to give a ground-state configuration which is always of the form $[Xe]\,4f^n$. The absorption spectrum of the ions corresponds to transitions within the 4f configuration ($4f^n \leftarrow 4f^n$ transitions), or $4f^{n-1}\,5d \leftarrow 4f^n$ transitions to the lowest-lying empty orbital, 5d.[5]

Of key importance to the spectroscopy of the trivalent rare earth ions in crystalline media is the fact that the wave function of the 4f orbital is more compact than those of the 5s and 5p orbitals. As such the mean radius of the 4f electrons are relatively small compared to the size of the ions, and hence to the nearest-neighbour distance. The interaction with the crystal field is therefore weak, producing a series of manifolds which are only slightly perturbed from those of the levels of the free ion. As a consequence the energy level structure of the 4f configurations in the trivalent rare earths is approximately independent of the crystal host, although the optical cross-sections of transitions may vary significantly.

The manifolds are labelled by the levels of the free ion from which they arise, i.e. ${}^{2S+1}L_J$. If the total spin quantum number S is an integer, the maximum number of levels within the manifold is given by $(2J + 1)$; if S is half-odd-integer the levels of the crystal field are all doubly degenerate,[6] and consequently the maximum number of levels within the manifold is reduced to $(2J + 1)/2$. It should be emphasized that these are the maximum number of non-degenerate levels; the actual number formed depends on the symmetry of the crystal field.

2.2.2 Transition linewidth

The natural linewidth of the $4f^n - 4f^n$ transitions is very narrow owing to the long radiative lifetimes of the levels. This linewidth is increased substantially by phonon collisions to give a homogeneously-broadened transition with a linewidth which depends on temperature. At room temperature the linewidth is typically in the range 0.1 - 3 THz (10 - 100 cm^{-1}).

Inhomogeneous broadening can arise if the crystal field varies with position in the crystal. For example, the presence of crystal defects, strain, or impurity ions can cause variation in the crystal field and hence local shifts of the ion energy levels. For good quality laser crystals these effects should be small, in which case the line broadening will be predominantly homogeneous. In contrast, ions doped into glasses experience a very wide range of local environments. The energies, broadening, and even the number of levels varies significantly from site to site and consequently all transitions will be strongly inhomogeneously broadened. The inhomogeneous linewidth of the $4f^n - 4f^n$ transitions in ions doped into a glass host are of order 30 THz (1000 cm^{-1}).

2.2.3 Nd:YAG laser

Perhaps the most important example of a laser based on a $4f^n - 4f^n$ transition in a trivalent rare earth ion is the Nd:YAG laser. Neodymium-based lasers are very widely used in science and industry. They are frequently used

[4] In the free ion the electrons are removed from the atom, in a solid they form bonds with the atoms of the host lattice.
[5] The 5s and 5p orbitals already being occupied.
[6] This is known as Kramers' degeneracy.

as pump lasers for dye lasers, or other solid-state lasers such as Ti:sapphire. In medicine they find applications in removing secondary cataracts or tissue removal. High-power Nd:YAG lasers can be used in laser drilling and welding.

Crystal properties

In Nd:YAG the Nd^{3+} ion replaces the Y^{3+} ion. Since the size of the Nd ion, which has a radius of 98 pm, is greater than that of the Y ion, of radius 90 pm, it is not possible to introduce Nd ions with an atomic concentration much above 1.5% without straining the crystal lattice unduly. It is worth noting that an atomic concentration of Nd ions equal to 1% corresponds to a density of Nd ions of $1.386 \times 10^{20}\,\mathrm{cm}^{-3}$.

Energy levels

Figure 2.2 shows the energy levels of the Nd^{3+} ion and how these are split into manifolds by the crystal field of YAG. Broad pump bands are provided by several closely-spaced manifolds with energies between approximately 12 000 and 33 000cm^{-1} (300 - 800 nm). Particularly strong pumping occurs at 810 nm (12 300 cm^{-1}) and 750 nm (13 300 cm^{-1}), which can be accessed by diode laser pumping.

The manifolds excited by optical pumping on these bands are relatively closely spaced compared to the maximum phonon energy in YAG of $850\,\mathrm{cm}^{-1}$. Consequently excited ions cascade down through these manifolds by very rapid non-radiative transitions until the two levels (denoted R_1 and R_2) of the $^4\mathrm{F}_{3/2}$ manifold is reached. The levels of the $^4\mathrm{F}_{3/2}$ manifold are metastable because the nearest lower-lying level — the top of the $^4\mathrm{I}_{15/2}$ manifold — is separated by some $4\,698\,\mathrm{cm}^{-1}$, corresponding to more than 5 phonons. The rate of non-radiative decay of the $^4\mathrm{F}_{3/2}$ manifold is therefore slow, and it decays instead almost entirely by radiative transitions to levels of the ^{4}I manifolds, with a fluorescence lifetime of 230 μs. In contrast, the levels of the ^{4}I manifolds, are separated by less than $1\,500\,\mathrm{cm}^{-1}$, and consequently decay non-radiatively with lifetimes of order 100 ps.

It can be seen, therefore, that the energy levels of Nd:YAG are very well suited to achieving laser oscillation: the levels of the $^4\mathrm{F}_{3/2}$ manifold may be populated efficiently by cascade from broad pump bands, and they are metastable — ideal properties of an upper laser level. Suitable lower laser levels exist within the ^{4}I manifolds; these lie significantly above the ground state and decay very rapidly to it. The system is therefore an almost ideal example of a four-level laser system. Figure 2.3 shows a simplified energy level scheme for the strongest Nd:YAG laser transition; transitions to the levels of other ^{4}I manifolds are analogous, with the exception of those to the lowest-lying manifold, $\mathrm{I}_{9/2}$. Lasers operating on transitions to this manifold behave quite differently than those operating on transitions to the higher-lying manifolds, since the levels of the $\mathrm{I}_{9/2}$ manifold can have a significant thermal population. As such, these lasers behave more like three-level systems.

Lasing has been achieved from the $^4\mathrm{F}_{3/2}$ manifold on more than 20 transitions to the $\mathrm{I}_{13/2}$, $\mathrm{I}_{11/2}$, and $\mathrm{I}_{9/2}$ manifolds, with wavelengths near 1319 nm, 1064 nm, and 946 nm respectively. Of these, the strongest — and certainly the most commonly used — transition is the $^4\mathrm{F}_{3/2} \rightarrow {}^4\mathrm{I}_{11/2}$ transition.[7]

The R_1 and R_2 levels of the $^4\mathrm{F}_{3/2}$ manifold are separated by $84\,\mathrm{cm}^{-1}$. Rapid non-radiative transitions ensure that the populations of these two levels is maintained in thermal equilibrium and, importantly, the rate at which the population in these two levels is mixed is sufficiently fast that this remains the case even during laser oscillation. Hence at room temperature the relative populations of these levels are approximately 60% and 40% respectively. Of the possible $^4\mathrm{F}_{3/2} \rightarrow {}^4\mathrm{I}_{11/2}$ transitions, the strongest is the ℓ_2 transition between the R_2 and Y_3 levels with a vacuum wavelength of 1064.15 nm. At room temperature this transition dominates the laser output, since the population of the R_2 levels is sufficiently large that the gain on this transition is largest. In contrast, if the crystal is cooled to low temperatures almost all the population of the $^4\mathrm{F}_{3/2}$ manifold resides in the R_1 level and lasing occurs from this level on the ℓ_1 transition from R_1 to the Y_2 level of $^4\mathrm{I}_{11/2}$ at a vacuum wavelength of 1064.40 nm.

[7] Indeed, in order to obtain lasing to the other manifolds it is necessary to employ frequency-selective cavities which suppress lasing to the $^4\mathrm{I}_{11/2}$ manifold and enhance the feedback on the desired transition.

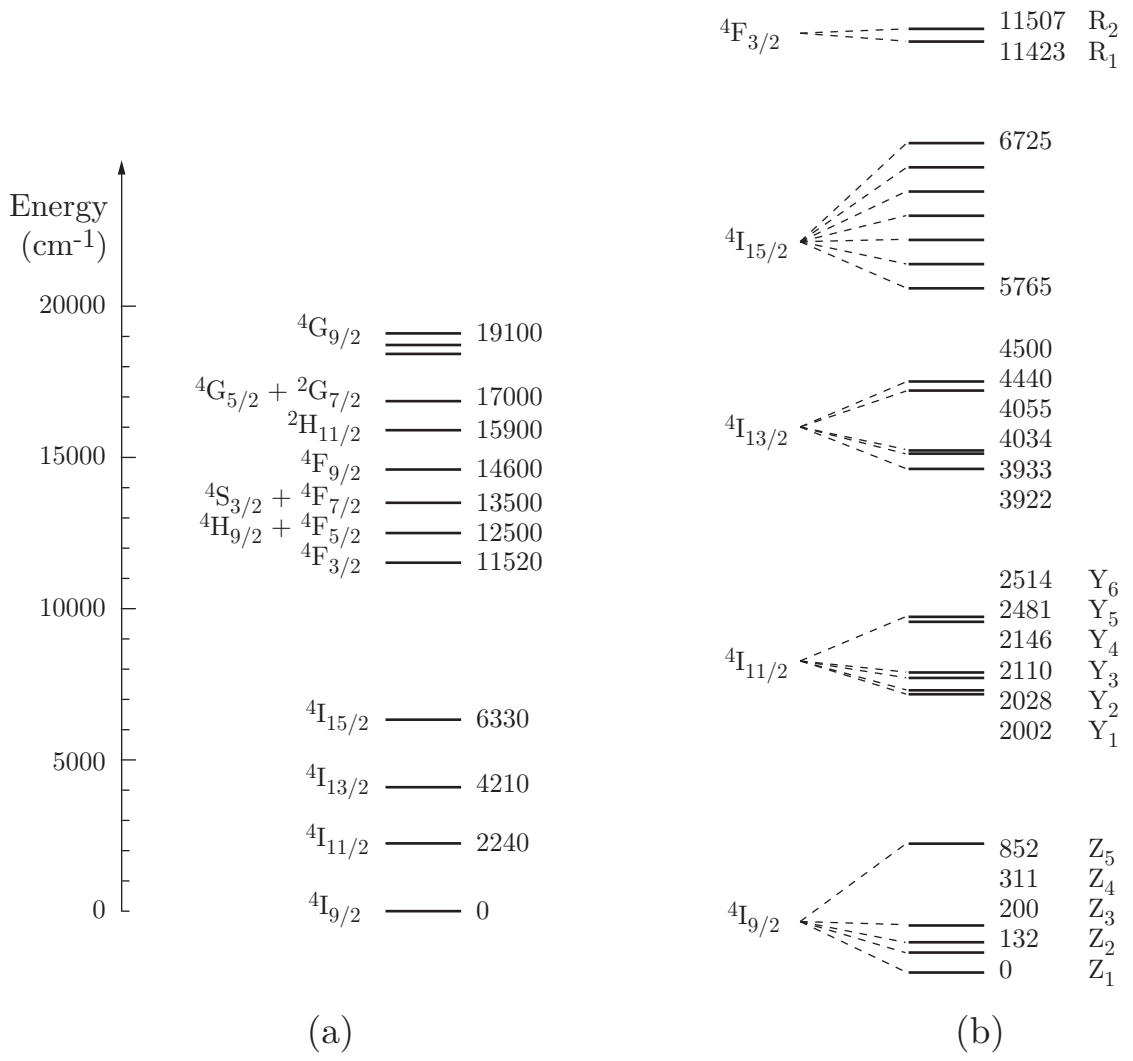

Figure 2.2: Energy level diagrams for the Nd^{3+} ion in YAG. (a) shows the lowest-lying levels of the free ion formed from the [Xe] $4f^3$ configuration; (b) illustrates how these levels are split by the crystal field in YAG to form manifolds of closely-spaced levels. Note that in (b) the levels of each manifold are shown on an expanded energy scale, and the spacing between manifolds is not drawn to scale. The energies of the levels is given in cm^{-1}.

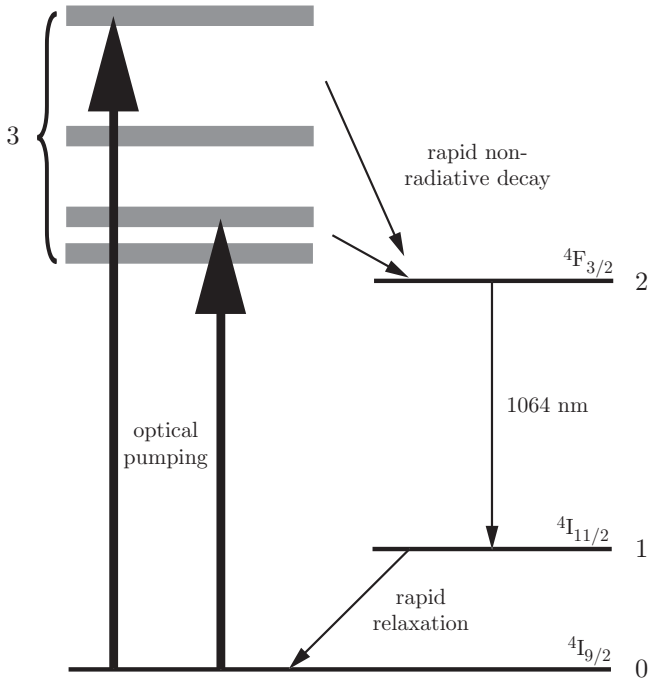

Figure 2.3: Simplified energy level diagram of the $^4F_{3/2} \rightarrow ^4 I_{11/2}$ laser transition in Nd:YAG showing the four-level nature of the laser scheme.

Broadening

In Nd:YAG the $^4F_{3/2} \rightarrow {}^4I$ transitions are homogeneously broadened by phonon collisions, with a full-width at half-maximum of approximately 190 GHz (6.5 cm^{-1}). Note that this is much larger than the natural broadening — dominated by the short lifetime of the lower laser level — which is of order 1 GHz.

Effective gain cross-section

Calculation of the optical gain is complicated by the fact that both the ℓ_1 and ℓ_2 transitions in Nd:YAG can contribute to the gain. In such cases it is useful to use a single **effective cross-section** which takes into account the contribution to the gain from all the transitions involved. The effective cross-section is defined in such a way that the gain coefficient is given by multiplying it by the *total* population of the manifold of levels, rather than the population of any single level.

The peak optical gain cross-sections of the ℓ_1 and ℓ_2 transitions are $1.9 \times 10^{-19}\,cm^2$ and $7.1 \times 10^{-19}\,cm^2$ respectively. Hence the dominant contribution comes from the ℓ_2 transition; if we neglect the contribution of the gain from the ℓ_1 transition, the peak value of the effective gain cross-section is given by multiplying the peak cross-section of the ℓ_2 transition by the proportion of population in the R_2 level (39%): $\sigma_{\rm eff}(0) \approx f_{R_2}\sigma_{R_2}(0) = 2.8 \times 10^{-19}\,cm^2$. In fact, contribution from the spectral wings of the ℓ_1 transition increases the effective cross-section to $3.7 \times 10^{-19}\,cm^2$.

Practical implementation

Nd:YAG lasers are one of the most common types of laser in use, and can be operated in a wide variety of different configurations: they can run continuously or in a pulsed mode; they can be pumped by flashlamps or laser diodes; and they can provide mean output powers from a few milliwatts up to several kilowatts. Very often Nd:YAG lasers are frequency doubled, tripled, or quadrupled to generate radiation of wavelength 532, 355, or 266 nm respectively.

Flashlamp-pumped systems employ pump chambers with single or multiple flashlamps. In such systems the laser rod is typically of about 5 mm diameter and 30 - 150 mm long, and the optical cavity is of order 0.5 m in length. With pulsed pumping the laser is often Q-switched at a pulse repetition of 10 - 50 Hz to yield output pulses of 5 - 10 ns duration and energies of a few tens to several hundred millijoules. The addition of one or more Nd:YAG amplifier stages can increase the output pulse energy to several Joules.

Flashlamp-pumped systems can generate pulsed output with a much higher pulse repetition frequency by employing continuous pumping and employing acousto-optic modulators to Q-switch the cavity at a repetition rate of 10 - 20 kHz. The mean output power of this type of system can be as high as 15 W.

ND:YAG lasers may be pumped by GaAs diode lasers operating at 808 nm which closely matches the peak absorption corresponding to excitation of levels of the $^2H_{9/2}$ and $^4F_{5/2}$ manifolds. The excited levels lie only 900 cm^{-1} above the upper laser levels, and consequently much less energy is deposited in the crystal in the form of heat (phonons) than is the case for flashlamp pumping which excites all manifolds up to some 10 000 cm^{-1} above the upper laser levels.

Diode-pumped Nd:YAG lasers can generate continuous-wave output with powers of: 10 W or more with end-pumping; above 100 W with side-pumping; and of order 1 kW with slab-pumping. Q-switching and modelocking is also possible. The slope efficiencies of diode-pumped Nd:YAG can be as high as 50 - 60% for end-pumped configurations, and 25 - 40% for side-pumping. These figures are much higher than for flashlamp-pumped lasers, for which the slope efficiency is typically 3%.

2.2.4 Other crystalline hosts

Neodymium ions have exhibited laser oscillation in several crystalline hosts including lithium yttrium fluoride ($LiYF_4$ — usually written as YLF) several vanadates such as YVO_4 and $GdVO_4$. Vanadates are particularly

Table 2.1: Important parameters of Nd:YAG, Nd:YLF, Nd:$GdVO_4$, and Nd:Glass (Hoya LHG-5 phosphate) lasers.

	Nd:YAG	Nd:YLF	Nd:$GdVO_4$	Nd:Glass
λ (nm)	1064	1053	1063	1054
τ_2 (μs)	230	450	90	290
τ_1 (ps)	100s	100s	100s	100s
$\Delta\nu$ (GHz)	160	380	180	5 000
σ_{21} (10^{-20} cm^{-2})	37	18.7	76	4.1
κ (W m^{-1} K^{-1})	13.0	6.0	12.3	1.19

useful in conjunction with diode-pumping since in these hosts the pump bands are some 80% broader, thereby reducing the tolerance required on the wavelength of the diode laser.

2.2.5 Nd:Glass laser

A wide variety of oxide-, fluoride-, and sulphide-based glasses have been developed as hosts for Nd ions. The wavelength of the $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition varies from approximately 1054 to 1062 nm depending on the glass host. The lifetime of the upper laser level is similar to that in YAG.

In glass hosts the laser transition is strongly inhomogeneously broadened, the linewidth increasing by a factor of approximately 50 to typically 6.5 THz ($\Delta\lambda \approx 25$ nm) compared to that found in Nd:YAG. Largely as a result of the greater linewidth, the gain cross-section of Nd:Glass is an order of magnitude smaller than that in Nd:YAG.

The small optical gain cross-section of Nd:Glass allows a large population inversion density to be generated without the onset of amplified spontaneous emission. The material is therefore frequently employed to amplify laser pulses to large energies. The low thermal conductivity of Nd:Glass restricts the pulse repetition rate of amplifiers based on this material to no more than a few pulses per second; and for very high energy amplifiers, the maximum pulse repetition rate is much lower. Finally we note that the large linewidth supports amplification of pulses as short as 100 fs.

2.2.6 Erbium lasers

Laser action has been achieved in erbium ions doped into a variety of garnet and fluoride crystalline hosts, as well as in several types of glass. There are two transitions of interest, with wavelengths of approximately 2.9 μm and 1.5 μm. Radiation at 2.9 μm is absorbed very strongly by water, and consequently erbium lasers operating on this transition have found applications in medicine. The 1.5 μm transition matches the third transparency window of optical fibres, and consequently is important for optical communications, particulary in the form of the **erbium-doped fibre amplifier (EDFA)**, the construction of which is discussed in Lecture 6. Further, this wavelength falls into the so-called 'eye-safe' window,[8] and consequently erbium lasers may also be used in telemetry and laser-ranging applications.

Figure 2.4 shows the energy levels of Er^{3+} ions in a glass host, and the important $^4I_{13/2} \rightarrow {}^4I_{15/2}$ transition at 1.54 μm. The upper and lower manifolds of levels are split by the crystal field into 7 and 8 degenerate levels respectively. In a glass host the 56 possible transitions between these levels are broadened homogeneously by

[8] Lasers operating between 1.45 and 1.70 μm are sometimes known as 'eye-safe' since radiation in this region is strongly absorbed by the cornea of the eye, and so cannot reach the retina. The term, however, is something of a misnomer since a sufficiently high-power laser operating in the 'eye-safe' region could still cause damage to the eye.

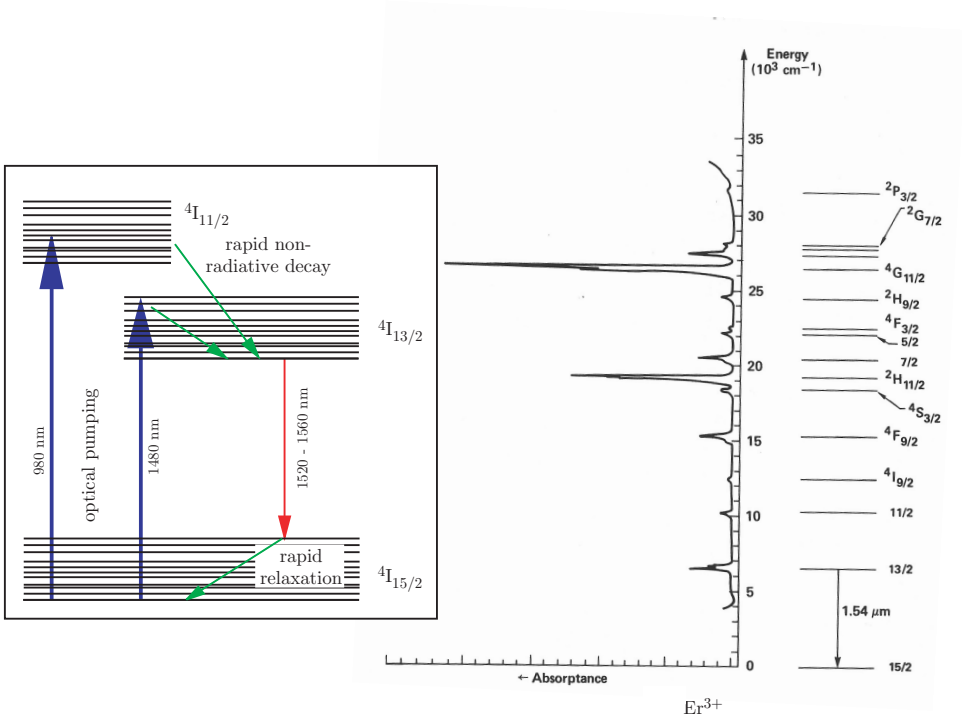

Figure 2.4: Energy level structure of the Er^{3+} ion in a glass host showing some possible laser transitions. Inset: simplified energy level diagram for the important transition at $\approx$ 1550 nm used in the EDFA.

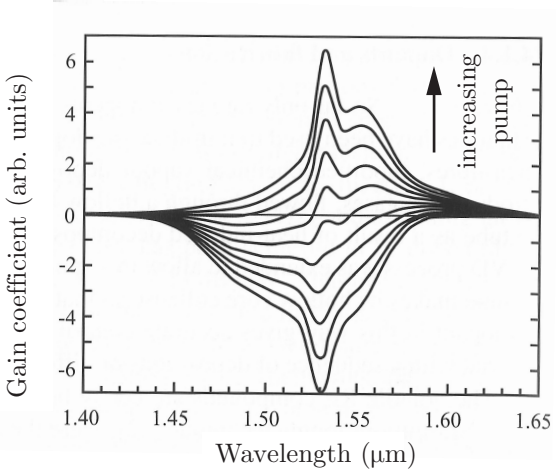

Figure 2.5: Measured gain coefficient as a function of wavelength for the $^4I_{13/2} \rightarrow ^4 I_{15/2}$ Er:Glass laser transition for different pump laser intensities.

lifetime and phonon broadening, and inhomogeneously by spatial variations in the local environment. The broadening characteristics of the laser transition are therefore complex, and to some extent the relative importance of homogeneous and inhomogeneous broadening depends on the properties of the glass host employed. However, the linewidth of the laser transition is approximately an order of magnitude greater than in a crystalline environment and hence the transition may be considered to be predominantly inhomogeneously broadened.

The most important pump band occurs at 980 nm (10 200 cm^{-1}) to the $^4I_{11/2}$ manifold since this can be accessed by InGaAs/GaAs diode lasers. In-band pumping can also occur at 1480 nm (6 800 cm^{-1}) to the higher levels of the $^4I_{13/2}$ manifold. For either pump band, excitation is followed by rapid non-radiative decay to the lowest level of the $^4I_{13/2}$ manifold, which acts as the upper laser level.

As for the case of Er:YAG, since the $^4I_{13/2}$ level lies well above the next lower level, those of the ground manifold, the rate of multiphonon decay of this level is low and the level decays predominantly radiatively with a fluorescence lifetime of approximately 8 ms. Lasing occurs from the bottom of the $^4I_{13/2}$ manifold to the levels of the ground-state manifold, $^4I_{15/2}$. Note that for the $^4I_{13/2} \rightarrow ^4 I_{15/2}$ transition, up-conversion[9] depletes the upper laser level; as a consequence the concentration of Er^{3+} ions must be kept low, generally to below 1×10^{20} cm^{-3}.

The Er:Glass laser is interesting in that it exhibits both three- and four-level characteristics. Figure 2.5 shows the measured gain coefficient as a function of wavelength for different pump intensities. It is seen that at long wavelengths positive gain is achieved at low levels of pumping, i.e. four-level laser behaviour. In contrast, at shorter wavelengths gain only occurs for high pump intensities, which is characteristic of a three-level laser. This complex behaviour reflects the fact that the lower manifold of states extends over approximately 850 cm^{-1}, compared to $k_B T \approx 210$ cm^{-1}. Lasing at long wavelengths occurs to the top of the $^4I_{15/2}$ manifold which has only a low thermal population and which undergoes rapid relaxation. However, lasing at shorter wavelengths occurs to lower-lying Stark levels which have a large thermal population; this population must be significantly reduced by the pumping before a population inversion is achieved.

The properties of some Er lasers are listed in Table 2.2.

[9]Up-conversion proceeds through the same ion-ion interactions that cause concentration quenching. In up-conversion ion A decays from an intermediate level to the ground state, causing ion B to be excited from an intermediate level to a higher-lying level. If the two ions are of the same species this process is known as up-conversion, and can proceed very rapidly if the energies of the two transitions are closely-matched. As for concentration quenching, up-conversion transitions are non-radiative.

Table 2.2: Important parameters of Er:YAG and Er:Glass lasers.

	Er:YAG		Er:Glass	
	$^4I_{11/2} \rightarrow ^4 I_{13/2}$	$^4I_{13/2} \rightarrow ^4 I_{15/2}$	$^4I_{11/2} \rightarrow ^4 I_{13/2}$	$^4I_{13/2} \rightarrow ^4 I_{15/2}$
λ (nm)	2940	1646		1540
τ_2 (μs)	100	7700	100	8000
τ_1 (μs)	7700	∞	8000	∞
$\Delta\nu$ (GHz)				3500
σ_{21} (10^{-20} cm^{-2})	2.6	0.5		0.7
κ (W m^{-1} K^{-1})	13	13	1	1

Er:Glass lasers have been operated in a variety of formats and in both continuous and pulsed mode. Q-switched operation is able to generate pulses of nanosecond duration, which is of interest in eye-safe laser-ranging. Alternatively, modelocked microlasers can generate pulses of approximately 20 ps duration with a pulse repetition rate as high as several GHz.

2.3 Trivalent iron group, $3d^n - 3d^n$ transitions

2.3.1 Energy level structure

The ground-state configuration of neutral atoms in the iron group of the transition metals is of the form $[Ar]\,3d^{n+1}\,4s^2$ or $[Ar]\,3d^{n+2}\,4s$, and hence that of the trivalent ions is $[Ar]\,3d^n$. The spectroscopy of these ions is very different from that of the $4f^n$ configuration in the trivalent rare earths owing to the fact that the size of the 3d orbitals are larger than those of the 4f orbitals in the trivalent rare earths, and the ionic size (and hence d) is smaller for the iron group than the rare earths. As a consequence the interaction with the crystal field is more than an order of magnitude bigger for the 3d electrons in the iron group than the 4f electrons in the rare earths, to the extent that the interaction with the crystal field is stronger than the spin-orbit interaction; the LS coupling scheme no longer applies.

It is highly instructive to consider in outline the form of the wave functions and energy levels in the strong crystal field limit. In this case the effect of the crystal field must be considered before the residual electrostatic and spin-orbit interactions, and consequently we determine the effect of the crystal field acting on the single-electron orbitals formed in the central potential of the active ion.

The crystal field removes the spherical symmetry of the free ion, and hence ℓ is no longer a good quantum number. Instead the modified electron states are labelled by its symmetry properties, and by the electron spin, which is unchanged by the interaction with the crystal field.

As an example, we consider the case of a d-electron moving in a crystal field with octahedral symmetry.[10] Octahedral symmetry of the crystal field arises when the active ion is located at the centre of a cube, and identical neighbouring ions are located on each face of the cube, as illustrated in Fig. 2.6. In the absence of the crystal field the d-electron is five-fold degenerate, corresponding to $m = -2, -1, 0, 1, 2$. A crystal field with octahedral symmetry partially lifts this degeneracy to form a triply-degenerate crystal field orbital labelled t_2, and a higher-lying doubly-degenerate orbital e.[11] The spatial distributions of the t_2 and e orbitals are very

[10] We note that other, generally lower, symmetries of the crystal field are often treated by considering the difference between the true symmetry and octahedral symmetry as a perturbation which acts on the energy levels found for the case of octahedral symmetry.

[11] Just as in molecular spectroscopy, a subscript 'g' (standing for 'gerade') may also be included in the label to indicate that the orbitals have even parity, i.e. e_g and t_{2g}. Orbitals with odd parity can be labelled with a subscript 'u', standing for 'ungerade'.

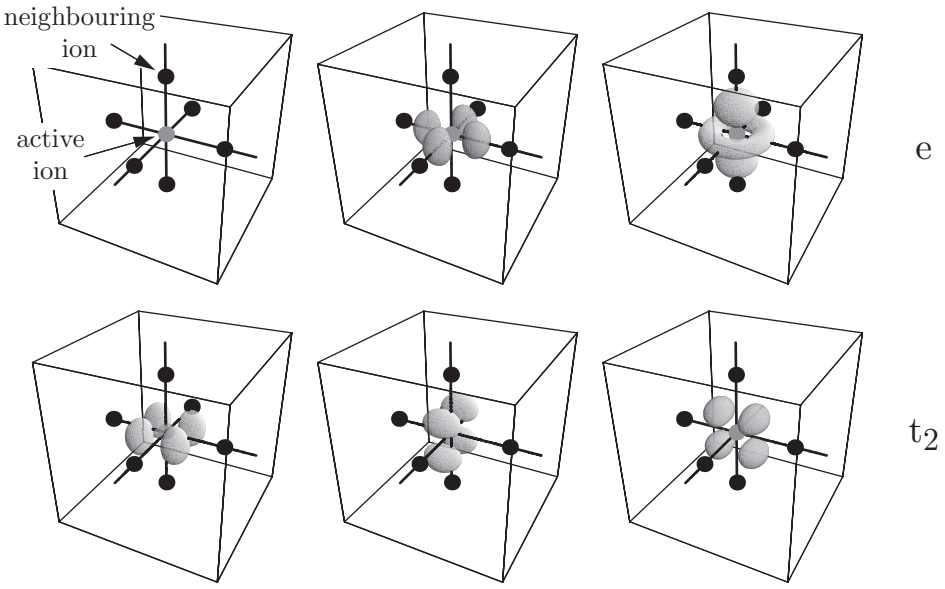

Figure 2.6: Crystal field orbitals of a d-electron in a field with octahedral symmetry. The top row shows the relative positions of the active ion and the neighbouring ions, and a surface of constant probability density for the two degenerate wave functions corresponding to the e_g orbitals. The bottom row shows equivalent plots for the three degenerate wave functions corresponding to the t_2 orbital. Notice that the wave functions of the e orbitals have lobes pointing towards the neighbouring ions, and hence have higher energy than the t_2 orbitals for which the lobes in the wave function point between the neighbouring ions.

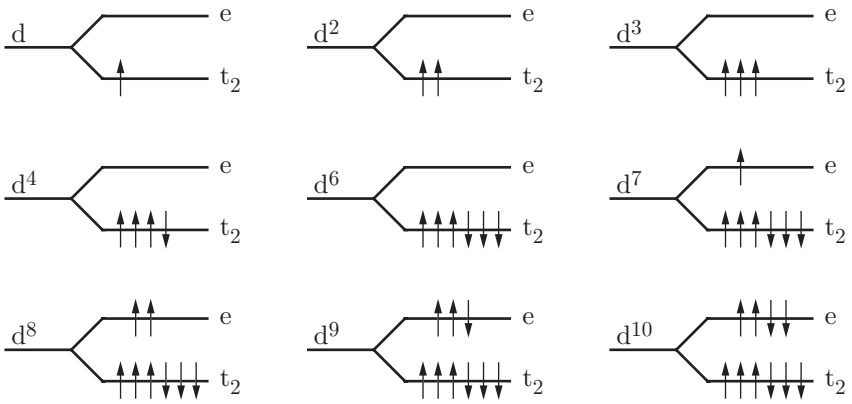

Figure 2.7: Schematic diagram showing the ground-state configurations arising from splitting of a d^n configuration a strong crystal field with octahedral symmetry. The arrows indicate electron spin.

different, as illustrated schematically in Fig. 2.6. The e wave functions have regions of high probability density centred on the axes connecting the active ion with the neighbouring ions, whereas for the t_2 orbitals these regions are centred between the axes. Consequently electrons in the e orbitals are on average closer to the neighbouring ions than those in the t_2 orbitals and therefore their interaction with the neighbouring ions — that is, with the crystal field — causes their energy to be increased from the unperturbed energy by a greater amount than electrons in the t_2 orbitals.

We may now consider the configurations formed in the crystal field as the number of d-electrons is increased. The ground state configuration of a single d-electron is simply t_2. Adding a second and third d-electron to the system gives rise to ground-state configurations comprised of two or three t_2 electrons respectively, since the t_2 orbital is three-fold degenerate. Since the spin-orbit interaction is minimized if the number of aligned spins is maximized,[12] the spins of the electrons will all point in the same direction.[13] A fourth d-electron added to the system may also occupy the t_2 orbital, but its spin must point in the opposite direction to satisfy the Pauli exclusion principle.[14] Considerations of this kind lead straightforwardly to the ground-state configurations illustrated schematically in Fig. 2.7.

The possible configurations of excited levels may be identified in a similar way. For example, for the case of three d-electrons two different excited configurations may be formed: one of the spins of the t_2 electrons may be reversed to give a configuration lying above the ground-state owing to the larger spin-orbit interaction; alternatively, one of the electrons can be promoted to the e orbital. The two types of excited configurations give rise to two types of transition, as illustrated in Fig. 2.8. In a **spin-flip** transition all the electrons remain in the same orbital, but the spin of one of the electrons is reversed, as illustrated in 2.8(a); in a **configuration transition** one electron makes a transition from an e orbital to a t_2 orbital.[15]

As we shall explore further below, these two types of transition have very different characteristics. In a spin-flip transition the spatial distribution of the total electronic wave function is unchanged. As a consequence the strength of the interaction of the active ion with the neighbouring ion is unchanged and, as we will see, this results in narrow-band transitions in which there is no change in the vibrational energy of the crystal lattice.

[12] This is 'Hund's rule'.

[13] Hence the ground-state configurations of t_2^2 will have $S = 1$ (a triplet), and that of t_2^3 will have $S = 3/2$ (a quartet).

[14] We note that if the crystal field were weak, the ground-state configuration would be t_2^3e, with all spins parallel since this minimizes the spin-orbit interaction; in other words the increase in energy associated with promoting an electron to the e orbital would be smaller than the increase in energy caused by reversing the spin of the electron.

[15] The spin of the electron may, or may not, also change.

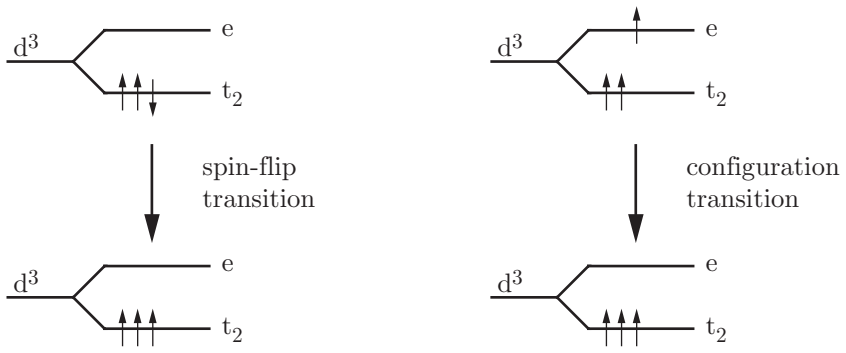

Figure 2.8: Spin-flip and configuration transitions of a d-electron in a crystal field with octahedral symmetry.

Transitions of this type are known as **zero-phonon** transitions.[16]

In contrast, in a configuration transition the spatial distribution of the total electron wave function changes significantly. Since the interaction of the active ion with the neighbouring ions is very different for the upper and lower levels, the neighbouring ions move during the transition. Configuration transitions are therefore associated with changes in the vibrational motion of the lattice, and are characterized by broad-band, temperature-dependent lineshapes — the broad bandwidth arising from the emission or absorption of varying numbers and energies of lattice phonons. Transitions of this type are therefore known as **phonon-assisted** or **vibronic transitions** and are the basis of an important class of widely-tunable solid-state laser.

Configuration co-ordinate diagrams

Further insight into the properties of energy levels in the strong-coupling limit is provided by the **configuration co-ordinate diagram**. In this model the active ion and the surrounding ions are treated as a single system. To simplify matters this multi-dimensional problem[17] is reduced to one-dimension, and the distances between the active ion and the surrounding ions is parameterized by a 'configuration co-ordinate' Q. We can think of Q as representing some sort of average value of the distance between active and neighbouring ions.[18]

Within the configuration co-ordinate picture the Schrödinger equation for the active ion may be written in the form,

$$[T_{\rm n} + T_{\rm e} + V(Q, \boldsymbol{r})]\, \Psi(Q, \boldsymbol{r}) = E\Psi(Q, \boldsymbol{r}), \tag{2.3}$$

where: $\boldsymbol{r}$ represents the co-ordinates of the electrons; $T_{\rm n}$ is the kinetic energy operator of the nuclei of the active and surrounding ions; $T_{\rm e}$ is the sum of the kinetic energy operators of the electrons in the active ion[19]; and $V(Q, \boldsymbol{r})$ is the electrostatic potential energy, comprised of terms arising from the attraction of the electrons to the nuclei of the active and neighbouring ions, mutual repulsion between the electrons of the active and neighbouring ions, and interactions between the nucleus of the active ion and the electrons and nuclei of the neighbouring ions.

[16] The $4f^n - 4f^n$ transitions considered above are zero-phonon transitions since the interaction with the crystal field is weak for both the upper and lower levels.

[17] If there are N neighbouring ions there are in principle N different values of the spacing between the active ion and the nearest neighbours. Whether all N values are different depends on the nature of the lattice vibrations, as discussed in footnote 20.

[18] Although see Note 20.

[19] We will ignore the electrons in the neighbouring ions since they will not be involved in transitions.

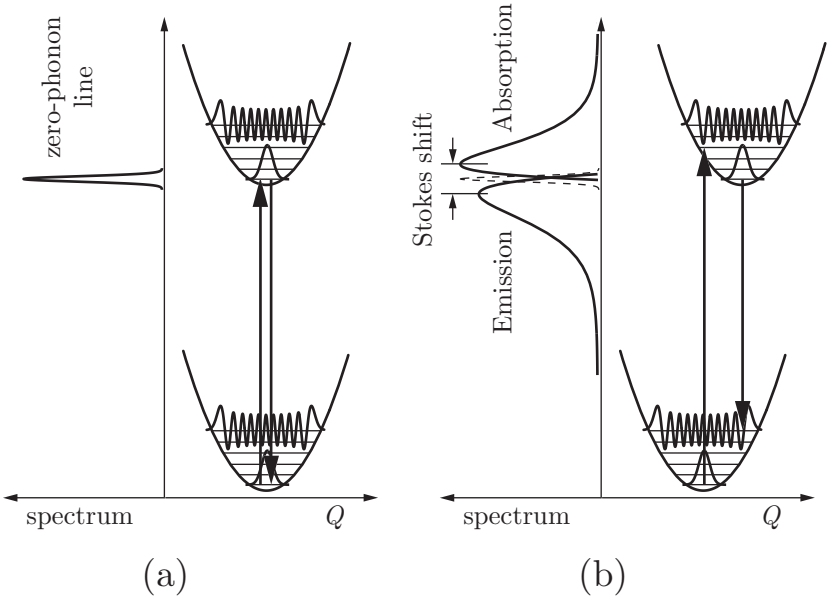

Figure 2.9: Transitions between electronic levels: (a) when the equilibrium configuration co-ordinate is essentially equal for the upper and lower electronic levels; (b) when the equilibrium configuration coordinates are different for the two levels, causing the potential curves to be displaced. Vibrational wave functions are illustrated schematically for the lowest vibrational level and an excited vibrational level. In (b) the position of the zero-phonon line is shown by the dotted curve.

Just as for the case of diatomic molecules discussed last year, we may make the **Born-Oppenheimer approximation** and assume that the total wave function may be written as a product of an electronic wave function $\phi_{\rm e}(\boldsymbol{r})$ and a nuclear (or 'vibrational') wave function $\phi_{\rm n}(Q)$. We assume also that the electronic wave function depends only weakly on Q. It is straightforward to show that the Schrödinger equation may then be separated:

$$[T_{\rm e} + V(Q, \boldsymbol{r})]\,\phi_{\rm e}(\boldsymbol{r}) = E(Q)\phi_{\rm e}(\boldsymbol{r}) \tag{2.4}$$

$$[T_{\rm n} + E(Q)]\,\phi_{\rm n}(Q) = E\phi_{\rm n}(Q). \tag{2.5}$$

In principle, these equations may be solved as follows. The Schrödinger equation for the electronic wave function can be solved *for fixed* Q to yield an energy eigenvalue, $E(Q)$, that is a function of the configuration co-ordinate. We see from eqn (2.5) that this energy eigenvalue then acts as a potential well within which the nuclei of the ions move. The energy of the nucleus moving in this potential is the *total energy of the system* (i.e. the energy of the active ion and the vibrational motion of the lattice).

The potential well in which the ions move resembles that in which the nuclei of a diatomic molecule move, and hence in a similar way the ion will exhibit a ladder of vibrational energy levels as shown.

The strength of an optical transition between two electronic levels is proportional to the square modulus of the dipole (or higher order) matrix element. Since the dipole operator does not operate on the vibrational wave function $\phi_{\rm n}(Q)$, the strength of optical transitions between two electronic levels is proportional to the square of the overlap integral of the vibrational wave functions of the upper and lower level — just as in the case of transitions in a diatomic molecule.

An important parameter of the potential well is the value of the configuration co-ordinate, Q_0, for which the potential is a minimum; this is the equilibrium value of the configuration co-ordinate.[20] If the spatial distribution of the electronic wave functions is similar for the upper and lower electronic levels then the interaction of the ion with the neighbouring ions will be similar for the two levels, and hence the upper and lower potential wells will have similar values of the equilibrium values of the configuration co-ordinate and so lie vertically above one another, as shown in 2.9(a). In contrast, if the spatial distributions of the wave functions for the upper and lower levels is very different, the interaction of the electrons of the active ion will be different, and consequently Q_0 will be different for the two levels. This situation is shown in 2.9(b).

We may now use the configuration co-ordinate picture to deduce the form of the absorption and emission spectra of active ions for the two types of transition, as illustrated in Fig. 2.9. Suppose that the lower electronic level corresponds to the electronic ground-state, and let us consider optical excitation to the upper electronic level. The vibrational wave functions for the lowest vibrational level generally resemble a Gaussian curve, whilst those of high-lying levels are strongly peaked near the classical turning points.[21] Absorption from the lower electronic level will be dominated by transitions from the lowest vibrational level, since this will have by far the largest thermal population. The strongest absorption will therefore by from here to the lowest vibrational level of the upper electronic level, as shown in 2.9(a), since the overlap integral for these two vibrational wave functions will be large. The overlap integral will be comparatively small for transitions to high-lying vibrational levels since the vibrational wave function of such levels show rapid oscillations and are only large near the classical turning points, where the amplitude of the wave function of the lowest vibrational level of the ground state is small. In any case, excitation to excited vibrational levels of the upper electronic level will be followed by extremely rapid de-excitation to the lowest vibrational level by phonon de-excitation. The emission spectrum will therefore be dominated by emission from the lowest vibrational level of the upper electronic level. Just as for the absorption spectrum, and for the same reason, the emission will be dominated by transitions to the lowest vibrational level of the electronic ground-state. In other words, when the equilibrium co-ordinates of the upper and lower electronic levels are similar — or, equivalently, when the spatial distribution of the wave functions of the upper and lower levels are similar, such as in spin-flip transitions — the absorption and emission spectra are dominated by a single, zero-phonon transition as we would expect from our earlier considerations.

In contrast, consider the case shown in 2.9(b) in which the spatial distribution of the electronic wave functions are different for the upper and lower levels, and consequently the equilibrium configuration co-ordinates are different. Now absorption from the lowest vibrational level of the ground electronic state is distributed over many vibrational levels of the upper electronic level, leading to absorption extending to higher frequencies than that of the zero-phonon line. Absorption at frequencies above that of the zero-phonon line corresponds to electronic excitation accompanied by excitation of lattice vibrations. Following absorption, ions in excited vibrational levels of the upper electronic level will undergo rapid phonon de-excitation to the lowest-lying vibrational level. Just as for the absorption, emission will then occur to a wide range of vibrational levels of the ground electronic state, producing a broad emission spectrum shifted to frequencies below that of the zero-phonon line. The absorption and emission spectra are widely separated, the separation in frequency between the peaks of the two spectra being known as the Stokes shift. Hence, as expected from our earlier discussion, transitions between electronic levels with different spatial distributions (and hence different values of Q_0), such as configuration transitions, are broad-band, vibronic transitions. Note that the width of the absorption and

[20] We see now that Q need not be thought of as the average value of the distance between active and neighbouring ions. Rather, when $Q = Q_0$ all the neighbouring ions are in their equilibrium positions; when $Q = Q_0 + \delta Q$ all the neighbouring ions have increased their separation from the active ion by a distance δQ. The configuration co-ordinate model is therefore effectively considers the lattice vibrations to be such that all the neighbouring ions change their separation from the active ion by the same amount, and consequently can be represented by a single value of Q. Vibrational modes of this type are known as 'breathing modes.'

[21] For a given vibrational energy level the classical turning points are the values of Q for which the potential curve has an energy equal to the total vibrational energy. In classical mechanics the kinetic energy of a particle is zero at such points, and hence they represent the extrema of its oscillatory motion. Classically, the probability of finding a particle is greatest at the turning points, since there it is stationary. The Correspondence Principle therefore tells us to expect the wave functions of high-lying (i.e. with large energy) vibrational levels to be strongly peaked at these points.

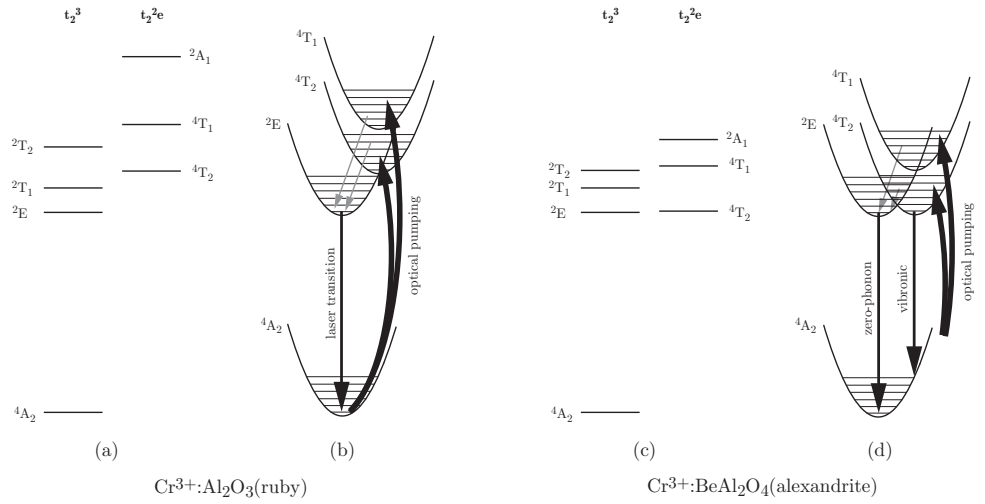

Figure 2.10: The crystal field configurations and energy levels (a and c) and simplified configuration co-ordinate diagrams (b and d) for ruby and alexandrite. Note that the diagrams are not to scale, and for clarity not all the energy levels shown in (a) and (c) are shown in the corresponding configuration co-ordinate diagram.

emission bands, and the magnitude of the Stoke's shift, both increase with the magnitude of the difference in the equilibrium configuration co-ordinates of the two levels.

2.3.2 The ruby laser

The ruby laser is of enormous significance since it was the first optical maser. It is much less commonly used today than it once was, but it still finds applications in holography (where the relatively short wavelength for a solid-state laser is better suited to the response of photographic emulsions), pulsed interferometry, and some medical applications. It is also the best known example of a three-level laser.

Crystal properties

Ruby is formed by doping Cr^{3+} ions into a sapphire (Al_2O_3) host with a doping level[22] of approximately 0.05 wt%, corresponding to an active ion density of $1.58 \times 10^{19}\,\mathrm{cm}^{-3}$. The chromium ions, of radius 62 pm, substitute for the aluminium ions, of radius 53 pm, with little distortion of the host lattice. Ruby may be grown by the Czochralski method, and has excellent mechanical, thermal, and chemical properties. Ruby laser rods typically have a diameter of between 3 and 25 mm and a lengths as long as 200 mm.

Energy levels

Figure 2.10 shows the energy levels of Cr^{3+} ions doped in Al_2O_3 together with a simplified configuration co-ordinate diagram. In ruby the crystal field is strong, and hence, as discussed in Section 2.3.1 the crystal field configuration of the ground state is t_2^3. This configuration forms crystal field terms labelled[23] as $^4A_2$ (the ground-state), ^2E, $^2T_1$, and $^2T_2$ where the superscripts are equal to $2S + 1$. The lowest-lying excited configuration is

[22] In ruby gemstones the doping level of chromium ions is much higher, of order 1%, which gives them a rich red colour. Ruby laser rods typically appear pink.

[23] Just as for levels in a free atom or ion, single-electron orbitals are labelled with lower-case letters, crystal field terms are labelled with upper-case letters.

t_2^2 e, which forms the levels $^4T_2$, $^4T_1$ and $^2A_1$. Of particular importance is the fact that since the crystal field is strong the $^4T_2$ and $^4T_1$ terms (which arise from the t_2^2e configuration) lie above the ^2E term (which arises from the t_2^3 configuration.[24]

Following the argument above, the wave functions of the levels of the t_2^2 e configuration have a very different spatial distribution than those of the t_2^3 configuration, and consequently, as shown in Figure 2.10 the equilibrium configuration co-ordinate of the $^4T_2$ and $^4T_1$ levels is quite different from that of the $^4A_2$ and ^2E levels.

The absorption spectrum of ruby is dominated by the $^4T_2 \leftarrow {}^4A_2$ and $^4T_1 \leftarrow {}^4A_2$ transitions since these are fully spin-allowed. Since these are configuration transitions, as discussed in Section 2.3.1, they form broad, vibronic pump bands some 50 nm wide and centred at approximately 554 nm and 404 nm respectively.[25] These may be pumped conveniently by flashlamp radiation. Absorption to the doublet levels are much weaker since they are spin-forbidden.

Absorption on the $^4T_{2,1} \leftarrow {}^4A_2$ pump bands is followed by rapid non-radiative relaxation to the lowest-lying excited level: the ^2E level. This level is metastable since it can only decay on the spin-forbidden $^2E \rightarrow {}^4A_2$ transition; in fact it has one of the longest fluorescence lifetimes of any level in the solid state, approximately 3 ms. Since the $^2E \rightarrow {}^4A_2$ transition involves no change in the crystal field orbitals, it is a narrow-band, zero-phonon transition.

The ^2E level is actually split[26] into two levels separated by only 29 cm^{-1}. Population is rapidly transferred between these two levels by phonon collisions, so that the levels may be considered to be in thermal equilibrium, with a slightly larger population in the lower of the two ^2E levels. The transitions from the two levels of ^2E to $^4A_2$ are denoted R_1 and R_2.

Laser parameters

The R_1 and R_2 transitions are homogeneously broadened by phonon interactions to give a temperature-dependent linewidth, which at room temperature is approximately equal to 450 GHz. The optical gain cross-section of the R_1 transition in ruby is approximately 2.5×10^{-20} cm^{-2}.

The energy levels of ruby therefore have many desirable features for a laser transition: broad pump bands allow for efficient absorption of broad-band pump radiation; almost all the ions excited on these transitions are transferred by rapid non-radiative transitions to the upper laser level; the upper laser level has a long fluorescence lifetime, and therefore acts as a storage level for excitation. The main difficulty, of course, is that the laser transition occurs to the ground-state and consequently the laser is a three-level laser system as illustrated schematically in Fig. 2.11.

In order to achieve a population inversion the optical pumping must be sufficiently strong to deplete the population of the ground-state by approximately one-half, as discussed in the course last year. The splitting of the ^2E level causes the lasing transition to be split into two, closely-spaced lines denoted R_1 and R_2 with wavelengths of 694.3 nm and 692.8 nm respectively. Of these two transitions, the R_1 line at 694.3 nm usually dominates owing to the fact that it arises from the lower of the ^2E levels, and consequently has a higher thermal population.

Practical implementation

Since ruby is a three-level laser, it is almost invariably operated in a pulsed mode. Optical pumping is usually provided by a medium-pressure xenon flashlamp.

[24] The $^4T_2$ and $^4T_1$ levels may also be denoted in an alternative notation as $^4F_2$ and $^4F_1$ respectively.

[25] It is these strong absorption bands in the green and violet part of the spectrum which are responsible for the striking colour of ruby gemstones.

[26] The splitting arises from a combination of the spin-orbit interaction and lowering of the symmetry of the crystal field from octagonal symmetry. The $^4A_2$ ground state is also split by the same interactions, but its splitting is even smaller. The split levels of the ^2E level are denoted $^2E_{3/2}$ and $^2E_{1/2}$, the latter being the lower. In the alternative notation mentioned in Note 2.3.2 these levels are denoted $\bar{E}$ and $2\bar{A}$.

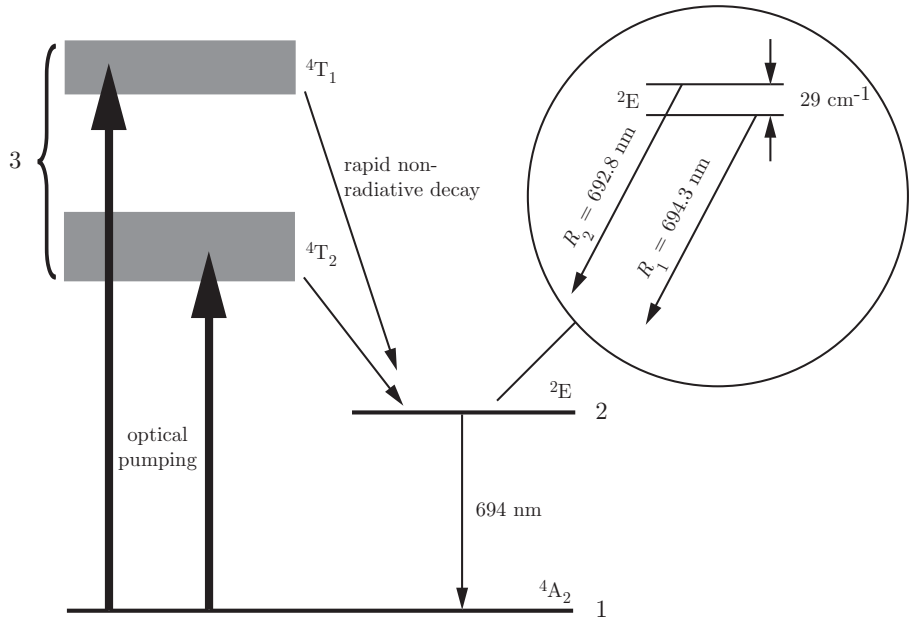

Figure 2.11: Simplified energy level scheme for the ruby laser. Inset: Splitting of the upper laser level.

The long lifetime of the upper laser level means that it is straightforward to Q-switch the laser, and indeed this is necessary if strong laser spiking is to be avoided. Q-switched ruby lasers typically generate pulses with an energy of several Joules, and a duration between 10 and 50 ns. Ruby lasers may also be actively or passively modelocked to give output pulses of 5 to 10 ps duration.

2.3.3 Alexandrite laser

Alexandrite, or chromium-doped chrysoberyl (Cr^{3+}:$BeAl_2O_4$), was the first commercially important tunable solid-state laser. Alexandrite lasers have proved useful in applications where their wide tunability is useful, such as detection of pollutant species in the atmosphere and in several medical applications.

Table 2.3: Important parameters of the ruby, alexandrite, and Ti:sapphire laser transitions.

	Ruby	Alexandrite	Ti:sapphire
λ (nm)	694.3	680.4, 700 - 820	660 - 11 80
τ_2 (μs)	3000	260 (vibronic)	3.8
τ_1	∞	$\approx$ 100 ps (vibronic)	$\approx$ 100 ps
σ_{21} (10^{-20}cm^{-2})	2.5	1	0.34
$\Delta\nu$ (GHz)	330	330,26 000	100 000
Doping level (%)	0.05	0.1	0.1
Pump bands (nm)	375 - 425, 525 - 575	380 - 630	380 - 620

Crystal properties

The Aluminium ions of the host material occupy two types of sites within the lattice: one in which the crystal field has a mirror symmetry, and one for which the field exhibits inversion symmetry. Typical doping levels are 0.04 - 0.1 at.%, the chromium ions substituting for aluminium ions, with a significant majority (78%) going into the sites with mirror symmetry.

The host is optically biaxial, leading to a strong polarization dependence on its optical properties which dominates thermal or stress-induced birefringence. The host has a very high thermal conductivity, a high optical damage threshold, is chemically stable, and has good mechanical strength and hardness. Laser rods of high optical quality may be produced with diameters of about 5 mm diameter and lengths up to 100 mm.

Energy levels

Figure 2.10(b) and (d) shows the energy levels of alexandrite. In this material the crystal field is of intermediate strength and consequently the $^4T_1$ and $^4T_2$ levels are closer to the ^2E level than is the case for Cr^{3+} ions in Al_2O_3 (i.e ruby). For example, in alexandrite the $^4T_2$ level lies only 800 cm^{-1} above the ^2E level, compared to approximately 1500 cm^{-1} in ruby.

As in the case of ruby, broad pump bands are provided by the vibronic $^4T_2 \leftarrow {}^4A_2$ and $^4T_1 \leftarrow {}^4A_2$ transitions lying at 550 - 650 nm and 350 - 450 nm respectively. These may be pumped by xenon or mercury flashlamps, or by argon and krypton ion lasers.

Absorption on the $^4T_2 \leftarrow {}^4A_2$ pump bands is followed by rapid non-radiative relaxation to the lowest vibrational level of the $^4T_2$ level and to the ^2E level. Since the $^4T_2$ level lies only 800 cm^{-1} above the ^2E level, population is transferred between these levels by rapid non-radiative transitions to form a thermal distribution over the two levels. The radiative lifetime of the ^2E level is similar to that of the same level in ruby: 2.3 ms; that of the $^4T_2$ level is somewhat shorter, 7.1 μs, since it may decay to the ground state via a spin-allowed transition. Since the population in these levels is shared, the effective lifetime of ions in either level depends on the relative population in each level, and is therefore strongly temperature dependent. At room temperature the effective lifetime of either level is approximately 300 μs.

Laser parameters

Lasing can occur in alexandrite on the analogous transition to the ruby laser transition: the zero-phonon $^2E \rightarrow {}^4A_2$ transition at 680.4 nm, which operates as a three-level scheme. However, of more importance is the $^4T_2 \rightarrow {}^4A_2$ since this is a vibronic transition and hence is widely tunable between approximately 700 nm and 818 nm.

The vibronic laser transition operates as a four-level laser: Pumping on the broad-band $^4T_{2,1} \leftarrow {}^4A_2$ transitions is followed by rapid non-radiative relaxation to the bottom of the $^4T_2$ level and to the ^2E level; the ^2E level acts as a population storage level, increasing the effective lifetime of the upper laser level considerably; lasing can then occur to excited vibrational levels of the $^4A_2$ ground electronic level, followed by rapid non-radiative decay to the lowest vibrational level of the ground electronic state. A simplified energy level diagram of the operation of the vibronic laser transition in alexandrite is shown in Fig. 2.12.

Since the populations of the $^4T_2$ and ^2E levels are closely-coupled, it is appropriate to use an effective optical gain cross-section together with the total population in the two levels, as discussed on page 23 for the case of the Nd:YAG laser. The effective cross-section for the vibronic transition in alexandrite is approximately 1×10^{-20} cm^2, although this is strongly dependent on temperature.

The main parameters of the laser transitions in ruby and alexandrite are given in Table 2.3

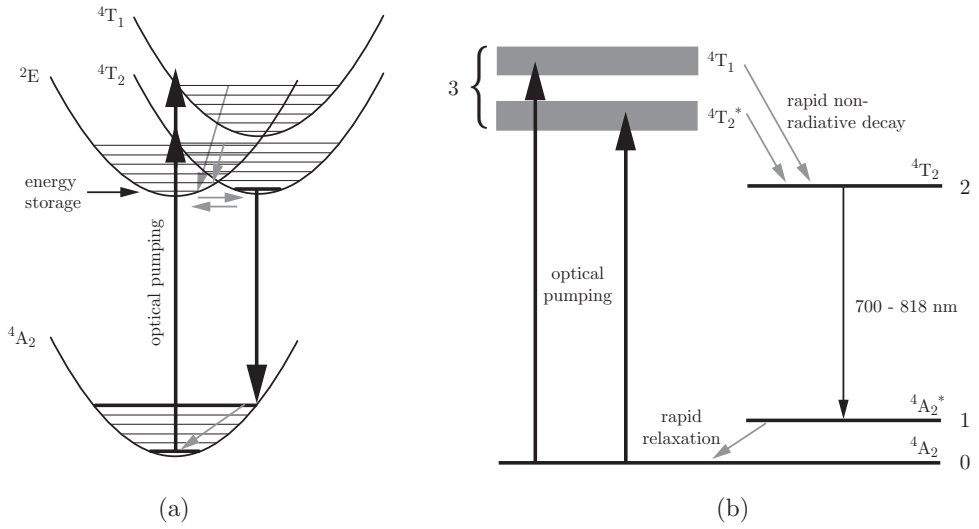

Figure 2.12: Simplified energy level scheme for the vibronic laser transition in alexandrite. (a) Simplified configurational co-ordinate diagram illustrating how the lowest vibrational level of the ^2E level acts as a storage level and is in thermal equilibrium with the lowest vibrational level of $^4T_2$; (b) Energy level scheme showing the four-level nature of the laser transition. Note in (b) an asterisk indicates that the level is in an excited vibrational level.

Practical implementation

Alexandrite lasers operate in a similar configuration to Nd:YAG lasers. The main difference is that the laser rod may be heated to between 50 and 70 °C in order to increase the relative population in the $^4T_2$ level.[27]

Alexandrite lasers are usually flashlamp-pumped. Although they may in principle be operated continuously, the relatively low gain cross-section means that the threshold population inversion density is high, and so pulsed operation is more usual. Pulse repetition rates between 10 and 100 Hz are typical, with output pulse energies of up to about 1 J.

Alexandrite lasers naturally produce output pulses with a duration close to the effective lifetime of the upper laser level, i.e. of 200 - 300 μs. Q-switching reduces the duration of the output pulse to typically 50 ns.

2.3.4 Ti:sapphire

Crystal properties

In titanium-doped sapphire, Ti:Al_2O_3, titanium ions (ionic radius 74.5 pm) substitute for aluminium ions (ionic radius 53 pm) with a typical concentration of 0.1 wt%. High-quality crystals may be grown by the Czochralski technique.

Energy levels

The Ti^{3+} ion has only a single d-electron, and consequently its energy level structure is rather simple: the ground crystal field configuration is t_2, which forms a $^2T_2$ term; the only other possible crystal field orbital for

[27] Heating to higher temperatures would increase the population of the $^4T_2$ level still further. However, raising the temperature of the laser rod also increases the thermal population of the lower laser level, and hence the optimum temperature is below that which would maximize the thermal population of the upper laser level.

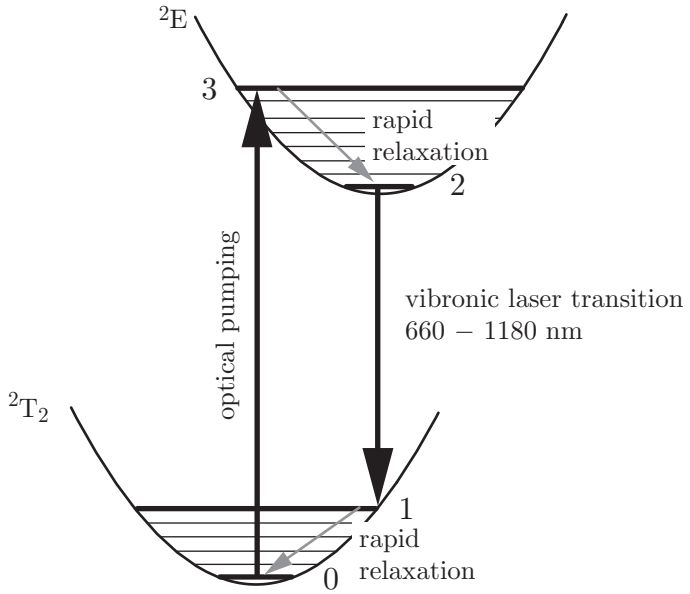

Figure 2.13: Simplified energy level structure and laser transitions of the Ti^{3+} ion in Al_2O_3 (Ti:sapphire).

a d-electron is e, which gives rise to a ^{2}E term, as shown in Figure 2.13.

Radiative transitions between these two electronic levels obviously involve a change in the crystal field configuration, and are therefore broad, vibronic transitions. It is worth noting that the sparse energy level structure of Ti:sapphire means that losses arising from excited-state absorption are negligible.

Figure 2.14 shows the absorption and emission spectra of Ti:sapphire. The very broad bands, and their wide separation is a consequence of the large difference in equilibrium configuration co-ordinates for the upper and lower levels. Absorption from the ground electronic level occurs on the strong vibronic $^2\mathrm{E} \leftarrow {}^2\mathrm{T}_2$ transitions to the excited vibrational levels of the ^{2}E level, with wavelengths between 630 nm and 400 nm. Ions in excited vibrational levels decay rapidly by non-radiative transitions to the lowest vibrational level of the ^{2}E level, which forms the upper level of the laser transition. At room temperature ions in this level decay predominantly radiatively to the vibrational levels of the $^2\mathrm{T}_2$ level, with a fluorescence lifetime of approximately $3.2\,\mu$s, the relatively short lifetime reflecting the fact that the transition to the ground state is spin-allowed.

Laser parameters

The Ti:sapphire laser is a four-level laser: lasing may occur from the lowest vibrational level of the ^{2}E level to excited vibrational levels of the $^2\mathrm{T}_2$ level, at wavelengths tunable between 660 nm and 1180 nm. Indeed Ti:sapphire has the largest bandwidth of any laser.

Although the bandwidth of the laser transition is very large, the large radiative decay rate of the laser transition ensures that the optical gain cross-section remains relatively high for a solid-state laser: approximately $3 \times 10^{-19}\,\mathrm{cm}^2$, which is comparable to that in Nd:YAG

Practical implementation

A major difference between Ti:sapphire and all the lasers discussed so far in this chapter is the short fluorescence lifetime of the upper laser level. Flashlamp pumping is therefore not usually very efficient owing to the much

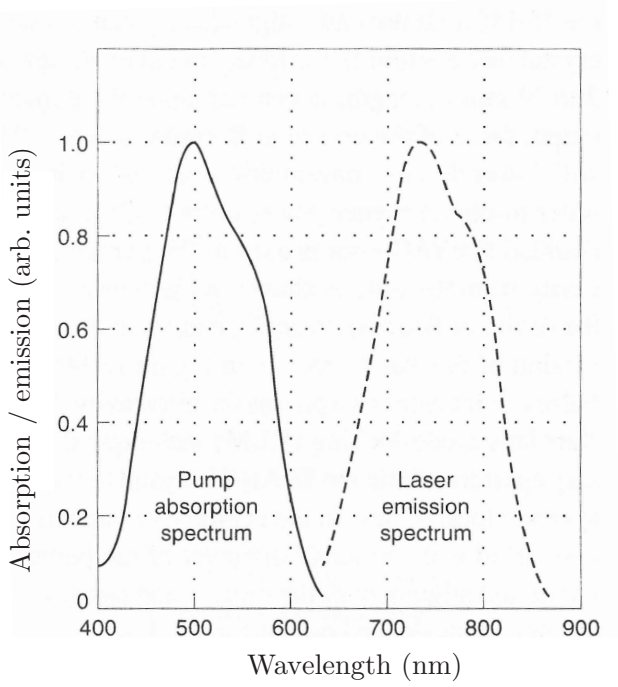

Figure 2.14: Absorption and emission bands of the Ti:sapphire laser.

longer duration of the flashlamp pulse. Instead, laser pumping is usually employed, the most commonly employed pump lasers being frequency-doubled Nd lasers (YAG, YLF or vanadate).

Figure 2.15 shows the construction of two common types of Ti:sapphire laser oscillator. In both designs the Ti:sapphire laser rod is 2 - 10 mm long, with end faces cut at Brewster's angle so as to eliminate reflection losses for one polarization. The z-folded cavity compensates for astigmatism. This arises because it is necessary to focus the pump radiation to a small cross-sectional area in order to achieve sufficient population inversion. The design of the optical cavity design must therefore be such that the transverse modes are also brought to a focus at the Ti:sapphire crystal. However, the Brewster-cut faces of the crystal introduce astigmatism to the intra-cavity beam — i.e. the divergence of the beam becomes different in the horizontal and vertical planes. In the cavity shown in Fig. 2.15 this is corrected by the two concave mirrors used away from normal incidence, which introduce astigmatism of the opposite sign to that caused by refraction at the laser rod.

Continuously-tunable c.w. Ti:sapphire lasers may be tuned between approximately[28] 700 nm and 1000 nm, with output powers of 1 - 2 Watts. Wavelength tuning may be accomplished by a variety of techniques. In the design illustrated in Figure 2.15(a), tuning is achieved by adjusting a birefringent filter.[29] Other techniques,

[28]The upper and lower limits to the tuning curve may be extended to approximately 100 nm and 675 nm respectively by employing cavity mirrors optimized for these wavelengths. However, it is not possible continuously to tune from 675 nm to 1100 nm without changing the cavity mirrors.

[29]At its simplest a birefringent filter is simply a plate of birefringent material with its optical axis in the plane of the plate. The retardance introduced by the plate depends on the angle between the optic axis and the electric field of the incident light, and on the wavelength of light. As the plate is rotated the wavelength(s) for which the retardance is equal to an integer times 2π varies. The polarization of these wavelength(s) is unchanged, unlike other wavelengths which in general become elliptically polarized, and consequently experience lower losses in passing through a linear polarizer (or any other polarization-sensitive element, such as the laser rod). In order to reduce losses, the filter can be mounted at Brewster's angle, and tuning achieved by rotating the filter about an axis parallel to its normal.

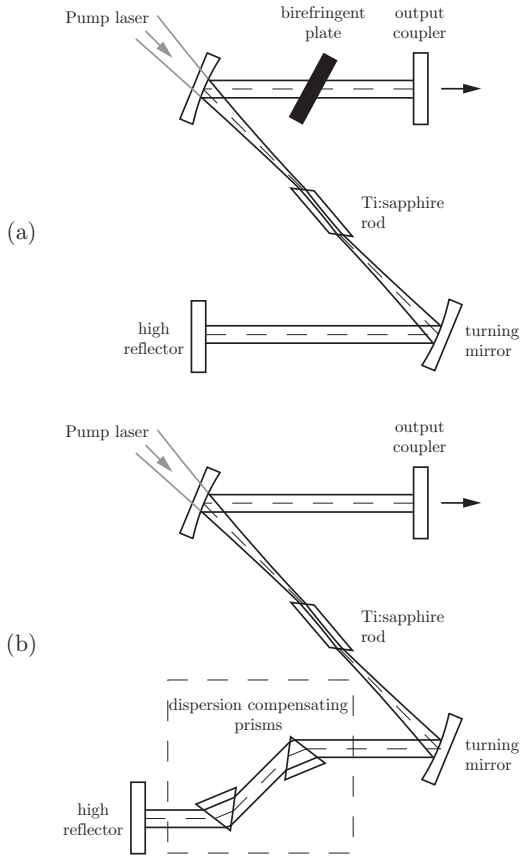

Figure 2.15: Two types of Ti:sapphire laser oscillator: (a) a wavelength-tunable oscillator; (b) a modelocked oscillator incorporating intra-cavity dispersion control.

such as prisms or gratings, may be employed for selecting the output wavelength. For the oscillator shown in Fig. 2.15(a) the linewidth of the laser output is typically about 40 GHz; this may be reduced to better than 1 GHz by employing an intra-cavity etalon. Further reduction of the linewidth to around 10 MHz is possible by employing a ring cavity, and by locking the operating frequency to an external cavity the linewidth can be reduced to as little as 50 kHz.

Ti:sapphire lasers are widely used to generate very short optical pulses by the Kerr modelocking technique discussed in Lecture 4. The cavity design shown[30] in 2.15(b) is typical of modelocked Ti:sapphire lasers, and incorporates a pair of prisms for dispersion control as discussed in Lecture 5. The train of modelocked laser pulses from such a system typically has a pulse repetition rate of 80 MHz, a mean power of approximately 0.5 W (i.e. $\approx$ 5 nJ per pulse), with pulse durations of 35 - 50 fs. Modelocked oscillators of this type can also be wavelength tuned, although the tuning range is more restricted: 780 - 820 nm is common.

[30]In designing suitable cavities it is desirable to maximize the ratio of the round-trip losses under c.w. operation to those when Kerr lens modelocking occurs.

2.3.5 Host materials

The host into which the active ion of a solid-state laser is doped plays an important role in the operation of the laser through its physical properties and its influence on the energy level structure of the ion. Desirable physical properties include: hardness, so that the material surface may be polished to an optical-quality finish; high thermal conductivity, so that waste heat may be conducted away rapidly; and ease of growth, so that the material may be grown to the required size whilst maintaining high optical quality throughout the sample. It is also desirable that the crystal exhibits low thermal lensing[31] and low thermally-induced birefringence.[32] At the microscopic level, the host material must be capable of accepting the active ion into its structure.

Solid state hosts fall into three classes: crystals, glasses, and ceramics, as discussed below.

Crystalline hosts

Crystalline hosts have thermal conductivities which are typically an order of magnitude greater than in glass hosts, and can therefore sustain higher mean pumping powers before the thermal lensing of the laser rod becomes problematic. Laser transitions in crystal hosts are usually homogeneously broadened, which makes single mode operation easier to achieve, and gives narrower linewidths for non-vibronic transitions (see Section 2.3.1). Disadvantages of crystal hosts are that they are often birefringent, making them polarization sensitive, and it can be difficult to grow large crystals of high optical quality.

Glass hosts

Glass hosts have several advantages over crystals: glass can be cast into many forms — from optical fibres to metre-sized discs — with good homogeneity and high optical quality; glass is inexpensive; and it has low, or zero, birefringence. Further, to some extent, the physical properties of the glass and the properties of the laser transition can be controlled by adjusting the composition of the glass. The biggest drawback of glass is the relatively low thermal conductivity, which makes thermal management more difficult. Optical transitions within ions doped into a glass host are always inhomogeneously broadened, with linewidths several orders of magnitude larger than non-vibronic transitions in crystalline hosts. This offers advantages in providing broad pump bands, but means that the laser transition will also be broad. As a consequence, the optical gain cross-section will be low, making it more difficult to achieve laser oscillation in an oscillator owing to the lower round-trip gain (although this restriction does not apply to fibre lasers). The low gain cross-section has significant advantages, though, in laser *amplifiers*. In an amplifier an upper limit on the single-pass gain is set by the onset of parasitic laser oscillation caused by small Fresnel reflections from the ends or edges of the laser rod, or depletion of the gain by strong amplified spontaneous emission. The low gain cross-section obtained with glass laser materials allows a larger population inversion to be generated before this is a problem, and this allows more energy to be extracted by an injected laser pulse.

To summarize: the higher gain and good thermal conductivity of crystalline hosts means that they are often employed in laser oscillators and lower-power amplifiers; glass hosts are frequently used in high-energy pulse amplifiers (with low mean power) owing to the low optical gain cross-section and ease with which large rods and discs may be fabricated.

[31] Thermal lensing refers to a spatial variation of the refractive index of the laser gain medium arising from spatial variations in temperature and thermal stresses within the material. Frequently the induced variation in refractive index – and, for a solid material, possible changes in the surface figure of the laser rod or disk – causes the gain medium to act as a thin lens of focal length f. For solid-state lasers f may be positive or negative, with a magnitude that can be as small as a few 10s of centimetres. Further details may be found in Koechner and Bass *Solid-State Lasers: A Graduate Text*, Springer (2003).

[32] The change in refractive index resulting from thermally-induced strain within the crystal can be different for different polarizations. For example, in a laser rod the change in refractive index will in general be different for the radial and azimuthal polarization components.

Lecture 3

Q-Switching

3.1 Laser Spiking and relaxation oscillations

Two types of transient behaviour are frequently observed in the output of laser systems:

1. **Laser spiking:** This can occur when a laser is first turned on, and comprises large amplitude, somewhat irregularly-spaced, pulses of laser radiation.

2. **Relaxation oscillations:** These are more regular oscillations in the output power of the laser, typically with an exponentially-decaying amplitude. Relaxation oscillations can occur in the tail of a series of laser spikes, or can follow a small perturbation to the laser system such as disturbance of a cavity mirror, or a change in the pumping power.

The two phenomena are related; both arise from the fact that the population inversion density and the intra-cavity laser intensity are coupled together via stimulated emission and feedback from the cavity mirrors, and that the inversion density and intra-cavity intensity respond to changes in the operating conditions of the laser on different time scales.

Consider first the case of laser spiking. Immediately after the pumping is switched on, the population in the upper laser level will grow from zero in a time of order the upper level lifetime τ_2. At this point the density of photons in the cavity will be very low since the photons arise from spontaneous emission only. At some point the population inversion density will reach the threshold for laser oscillation, $N_{\rm th}^*$, but laser oscillation will not commence since the density of photons in the relevant cavity mode will still be low. As pumping continues, the population inversion grows beyond the threshold value, and the density of photons in the laser mode will build up by stimulated emission. Eventually the density of photons in the laser mode will be sufficiently high to burn the population inversion down towards $N_{\rm th}^*$, further increasing as it does the rate of stimulated emission and the photon density. Inevitably, the population inversion will be driven well below the threshold value, leading to reduction in the photon density and the cessation of lasing. The dramatic growth and rapid decline of the intra-cavity photon density correspond to a spike in the output power of the laser. As pumping continues, the population inversion will start to grow again and the cycle can repeat. Laser spiking of this type will not usually continue indefinitely, even in continuously-pumped lasers, owing to the fact that the population inversion and photon densities do not decay to zero after the spike. Consequently, after each spike these quantities tend to end up closer to their steady-state values, in which case the spiking will decay. Figure 3.1 shows a typical example of laser spiking from the first ever ruby laser.

It should be noted that laser spiking is often associated with rapid jumps in the longitudinal and/or transverse mode of the laser. This occurs for essentially the same reasons as spatial hole-burning in lasers operating on homogeneously broadened transitions: immediately after a laser spike, significant population inversion may

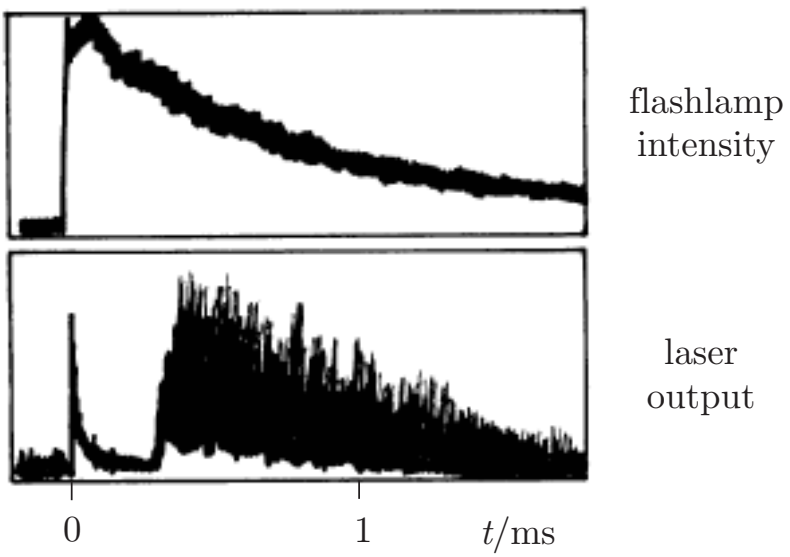

Figure 3.1: Laser spiking in the first ruby laser. The top trace shows the flashlamp pulse, the lower the output from the ruby laser as recorded by a photodiode. Note that the sharp feature at $t = 0$ is caused by flashlamp light reaching the photodiode.

remain in regions near nodes of the longitudinal nodes, or in regions where the transverse profile of the mode of the previous pulse was low in intensity. This gain can be exploited by other modes which, as pumping continues, may therefore reach threshold sooner than the mode (or modes) of the previous spike.

Laser spiking tends to occur when the time scale on which the population inversion responds to the pumping, typically of order τ_2, is slow compared to that in which the photon density changes in the cavity, which is of order the cavity lifetime τ_c. Hence we expect laser spiking to occur if,

$$\tau_2 \gg \tau_c \quad \text{Approximate condition for laser spiking.} \tag{3.1}$$

3.1.1 Rate equations

For a homogeneously broadened laser the rate equations for the upper and lower laser level may be written simply as[1],

[1] In writing the rate equations in this form we have made several simplifying assumptions: we assume that the laser oscillates on a single longitudinal and transverse mode; and we ignore transverse variations in the intra-cavity laser intensity and in the population inversion.

$$\frac{dN_2}{dt} = R_2(t) - N^*(t)\sigma_{21}\frac{I(t)}{\hbar\omega_{\rm L}} - \frac{N_2(t)}{\tau_2} \tag{3.2}$$

$$\frac{dN_1}{dt} = R_1(t) + N^*(t)\sigma_{21}\frac{I(t)}{\hbar\omega_{\rm L}} - \frac{N_1(t)}{\tau_1} + N_2(t)A_{21}, \tag{3.3}$$

where R_i is the pump rate of level i, and $\omega_{\rm L}$ and $I(t)$ are the angular frequency and total intensity of the oscillating laser mode.

Notice that in the absence of stimulated emission ($I(t) = 0$) the population of the upper laser level grows in a time of order τ_2.

The rate per unit volume at which photons are stimulated into the oscillating mode is simply $N^*\sigma_{21}I/\hbar\omega_{\rm L}$. Hence if we suppose that the fraction of the volume of the oscillating mode lying within the gain medium is equal to $f_{\rm c}$, the rate equation for the density of photons is simply,

$$\frac{dn}{dt} = f_{\rm c}N^*(t)\sigma_{21}\frac{I(t)}{\hbar\omega_{\rm L}} - \frac{n(t)}{\tau_{\rm c}}. \tag{3.4}$$

The last term gives, in terms of the **cavity lifetime** $\tau_{\rm c}$,the rate of loss of photon density by coupling through the output mirror (plus any other losses). We derive an expression for $\tau_{\rm c}$ in Section 3.1.2.

This equation may be simplified by noting that $I(t) = n(t)\hbar\omega_{\rm L}c$. Then,

$$\frac{dn}{dt} = \left(\frac{N^*}{N^*_{\rm th}} - 1\right)\frac{n}{\tau_{\rm c}}, \tag{3.5}$$

where,

$$N^*_{\rm th} = \frac{1}{f_{\rm c}\sigma_{21}c\tau_{\rm c}} \tag{3.6}$$

and we have dropped the time dependence of $N^*(t)$ and $n(t)$ to avoid clutter. We see that $N^*_{\rm th}$ is the threshold population inversion density since we must have $N^* > N^*_{\rm th}$ for the photon density to grow (by stimulated emission). Equation (3.5) also shows that the time scale in which the photon density changes is $\tau_{\rm c}$ divided by a factor which depends on the extent to which the population inversion exceeds $N^*_{\rm th}$. As such we expect the photon density to change in a time of order $\tau_{\rm c}$, as stated earlier.

3.1.2 Cavity lifetime

To calculate the cavity lifetime we consider some photons circulating within a laser cavity. In the absence of further photons being supplied by any source (an injected beam, or by spontaneous or stimulated emission), the number of photons within the cavity will decrease with time. Suppose that at some moment the cavity contains n photons per unit volume. The flux — i.e. the number of photons per unit area, per unit time — of photons striking each mirror is $\frac{1}{2}nc$, the factor of one-half arising since half the photons travel to the left, and half to the right. For each mirror i the proportion of these photons which are transmitted is equal to T_i. Hence, ignoring any other losses, the density of photons within the cavity will decay according to

$$\frac{{\rm d}n}{{\rm d}t} = -\frac{\frac{1}{2}ncA_{\rm mode}\,(T_1 + T_2)}{V_{\rm mode}},$$

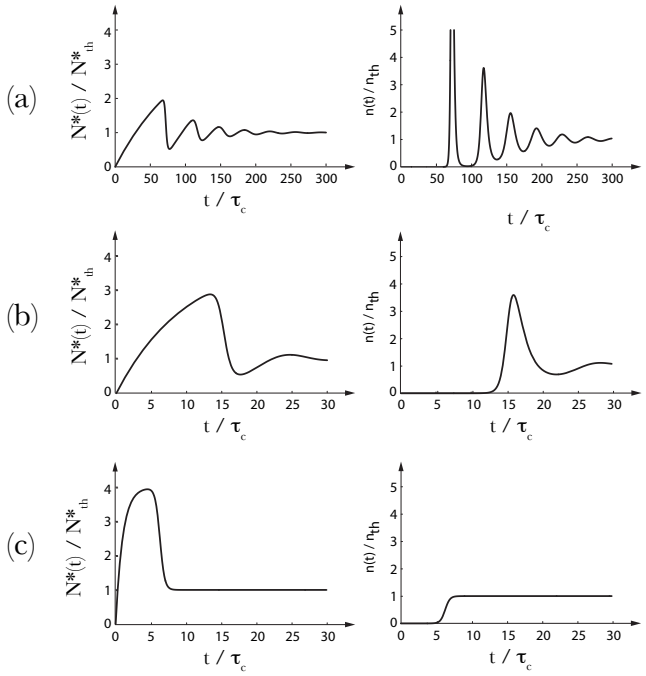

Figure 3.2: Numerical analysis of laser spiking in an ideal four-level laser with a filling factor f_c of unity pumped with an over-pumping ratio $r = 4$ and: (a) $\tau_2/\tau_c = 100$; (b) $\tau_2/\tau_c = 10$; (c) $\tau_2/\tau_c = 1$. In each case the plots show the temporal evolution of the population inversion density and the photon density normalized to the steady-state value, n_{th}.

where A_{mode} and V_{mode} are the area and volume of the oscillating mode. Hence we may write,

$$\frac{dn}{dt} = -\frac{n}{\tau_c},$$

where the cavity lifetime is given by

$$\tau_c = \frac{2V_{mode}}{A_{mode}c}\frac{1}{T_1 + T_2}.$$

Now $V_{mode} \approx A_{mode}L_c$, and hence

$$\tau_c = \frac{2L_c}{c}\frac{1}{T_1 + T_2} \approx \frac{L_c}{c}\frac{1}{1 - R} \qquad \text{Cavity lifetime}$$

where $R = \sqrt{R_1 R_2}$ and the approximation holds if $R_1 \approx R_2 \approx 1$.

3.1.3 Numerical analysis of laser spiking

Fig. 3.2 shows numerical solutions of these rate equations for a system with a filling factor f_c of unity and constant pumping at a rate corresponding to an over-pumping ratio (see Section 3.1.4) $r = 4$. In each of these

simulations the pumping would, in the absence of stimulated emission, produce a population inversion density which was a factor of r greater than the threshold value. However, once the population inversion rises above $N_{\rm th}$ the rate of stimulated emission exceeds the rate of loss of photons from the cavity, increasing the density of photons within the cavity, and causing in turn the population inversion to burn down towards the threshold value. In all the cases shown the population inversion and photon densities tend to their steady-state values $N_{\rm th}$ and $n_{\rm th}$ respectively at long times after the pumping is switched on at $t = 0$. Substantial laser spiking is observed when $\tau_2/\tau_{\rm c} \gg 1$, consistent with our discussion above, and the number and amplitude of the spikes in the laser output is seen to decrease as the ratio $\tau_2/\tau_{\rm c}$ is reduced. In particular, for the case $\tau_2/\tau_{\rm c} = 1$ there is essentially no spike in the photon density. Instead the population inversion density photon density reaches $rN_{\rm th}$ before the smoothly-rising photon density burns the population inversion down to the steady-state value.

3.1.4 Analysis of relaxation oscillations

Relaxation oscillations, being somewhat gentler than laser spiking, are amenable to algebraic analysis. We consider the simple example of an ideal homogeneously broadened, four-level laser system in which the population of the lower laser level is essentially zero at all times. In this case eqn (3.2) is also the rate equation for the population inversion density:

$$\frac{dN^*}{dt} = \frac{dN_2}{dt} = R_2 - N^*\sigma_{21}\frac{I}{\hbar\omega_{\rm L}} - \frac{N_2}{\tau_2} \tag{3.7}$$

$$= R_2 - \frac{1}{f_{\rm c}}\frac{N^*}{N^*_{\rm th}}\frac{n}{\tau_{\rm c}} - \frac{N^*}{\tau_2} \tag{3.8}$$

Notice that in the absence of any stimulated emission ($I = 0$) the population inversion would reach a steady-state value of $R_2\tau_2$.

Under steady-state conditions $N^*(t) = N^*_{\rm th}$. Substituting this into eqn (3.8) then yields an expression for the steady-state photon density,

$$n_{\rm th} = \left(R_2 - \frac{N^*_{\rm th}}{\tau_2}\right) f_{\rm c}\tau_{\rm c} \tag{3.9}$$

$$= (r-1)\frac{N^*_{\rm th} f_{\rm c}\tau_{\rm c}}{\tau_2}, \tag{3.10}$$

where the parameter $r = R_2\tau_2/N^*_{\rm th}$ is the ratio of the population inversion that would be produced in the absence of any stimulated emission to $N^*_{\rm th}$. It is also equal to the over-pumping ratio that will appear in our discussion of Q-switching.

We now suppose that having established steady-state conditions, the laser is subject to a small perturbation that shifts the population inversion and photon densities to:

$$N^*(t) = N^*_{\rm th} + \Delta N^*(t) \tag{3.11}$$

$$n(t) = n_{\rm th} + \Delta n(t) \tag{3.12}$$

We may then linearize the differential equations for $N^*(t)$ and $n(t)$ by substituting eqns (3.11) and (3.12) into (3.5) and (3.8), and ignoring the small product terms in $\Delta N^*(t)\Delta n(t)$. We find:

$$\frac{d\Delta N^*}{dt} = -r\frac{\Delta N^*}{\tau_2} - \frac{1}{f_{\rm c}}\frac{\Delta n}{\tau_{\rm c}} \tag{3.13}$$

$$\frac{d\Delta n}{dt} = (r-1)f_{\rm c}\frac{\Delta N^*}{\tau_2} \tag{3.14}$$

We now assume solutions of the form,

$$\Delta N^*(t) = a\exp(mt) \tag{3.15}$$
$$\Delta n(t) = b\exp(mt). \tag{3.16}$$

Substitution into eqns (3.13) and (3.14) then gives,

$$\left(m + \frac{r}{\tau_2}\right)a + \frac{1}{f_c\tau_c}b = 0 \tag{3.17}$$
$$(r-1)\frac{f_c}{\tau_2}a - mb = 0. \tag{3.18}$$

These two simultaneous equations can be solved by the usual matrix method: non-trivial solutions require the determinant to be zero, yielding,

$$m = -\frac{r}{2\tau_2} \pm \sqrt{\left(\frac{r}{2\tau_2}\right)^2 - \frac{r-1}{\tau_2\tau_c}}. \tag{3.19}$$

Clearly, relaxation oscillations will occur if m is complex, i.e. if

$$\left(\frac{r}{2\tau_2}\right)^2 < \frac{r-1}{\tau_2\tau_c} \tag{3.20}$$
$$\Rightarrow \tau_2 > \frac{r^2}{r-1}\frac{\tau_c}{4}. \tag{3.21}$$

In other words if the upper level lifetime is sufficiently short, relaxation oscillations do not occur and perturbations to $N^*(t)$ and $n(t)$ decay exponentially back to their steady-state solutions. Note that for $r \gg 1$ eqn (3.21) is in agreement with our earlier estimate (3.1).

If the above condition is not met relaxation oscillations will occur with an angular frequency

$$\omega_{ro} = \sqrt{\frac{r-1}{\tau_2\tau_c} - \left(\frac{r}{2\tau_2}\right)^2} \approx \sqrt{\frac{r-1}{\tau_2\tau_c}}, \tag{3.22}$$

where the approximation holds if $\tau_2 \gg \tau_c$. Notice that very often $r - 1 \approx 1$, in which case ω_{ro} is close to the geometric mean of the upper level and cavity decay rates.

The condition (3.21) explains why relaxation oscillations (and also laser spiking) are unusual in gas lasers. Such systems typically operate on allowed transitions with upper level lifetimes of 10s to 100s of nanoseconds[2]. This is reasonably short compared to a typical cavity lifetime of 10 ns. In contrast, the upper level lifetime for many solid state laser transitions is of the order of microseconds or even milliseconds. In such cases laser spiking and relaxation oscillations are to be expected.

[2]A notable exception is the CO_2 laser which has an upper level lifetime of 4 ms, and hence can exhibit relaxation oscillations.

3.2 Q-switching

Q-switching is a technique that generates pulses of laser light with a peak power very much greater than the mean power that can be produced in steady-state operation. The high peak power is particularly useful in applications that require a high peak intensity, such as: laser cutting or drilling; or in pumping other laser systems such as Ti:sapphire amplifiers.

In a Q-switched laser the cavity is deliberately spoiled — imagine, for the moment, that one of the cavity mirrors is blocked — whilst the pumping builds up the population inversion. Since the losses of the modified cavity are so high, the population inversion can build up to much greater levels than the threshold inversion for lasing in the normal cavity without any lasing occurring. Once the population inversion has grown to a large value the mirror is unblocked, thereby switching the 'Q' of the cavity from a low value to a high value. High rates of stimulated emission then cause the density of photons in the cavity to grow very rapidly, burning the population inversion density well below the threshold value. The output from the laser takes the form of a giant pulse of radiation — essentially a single laser spike of the type discussed above.

Note that the pumping in Q-switched lasers may be pulsed or continuous. With pulsed pumping the ratio of the peak population inversion to the threshold value can be larger than can be achieved with continuous pumping, leading to a higher peak output power (although the mean output power may be similar).

3.2.1 Techniques for Q-switching

A wide variety of techniques have been used to implement Q-switching.

Rotating mirror

Historically one of the earliest techniques was to rotate one of the cavity mirrors with a motor. The cavity would then be completed briefly whenever the mirror was rotated into alignment. It is best to use a roof prism as shown schematically in Figure 3.3, rather than a plane mirror, since any ray incident in a plane perpendicular to the axis of the roof is reflected in the anti-parallel direction to the incident ray. Hence for the system illustrated alignment in the vertical plane is automatic — this is useful given the inevitable vibration in a mechanical system. Alignment in the horizontal plane is achieved once per rotation.

The disadvantages of this approach are obvious. The switching is rather slow and the timing of the output pulse has a relatively high uncertainty. In order to estimate the switching time, we suppose that the mirror has to be aligned to an angular tolerance about equal to the divergence of the oscillating cavity mode. This divergence will typically be of the order of 1 mrad. Hence, for a motor running at 24 000 r.p.m. the switching time is about 400 ns. The disadvantages are such that this technique is now used only very rarely.

Electro-optic switching

A more modern approach is the use of an electro-optic switch, as illustrated schematically in Figure 3.4. The refractive index of an electro-optic material changes when it is subjected to an external electric field. In the **Pockels effect**[3], observed in non-centro-symmetric materials, the induced refractive index varies linearly with the applied electric field. Suitable materials are KD*P or lithium niobate for visible to near infrared wavelengths, or cadmium telluride for the mid-infrared. In the **Kerr effect**[4] the change in refractive index is proportional to the square of the electric field. We note that all materials exhibit the Kerr effect to some extent. Since the Kerr effect arises from higher-order processes than in the Pockels effect (via the third-order susceptibility rather than the second-order), larger electric fields are usually required. Since these must be switched quickly for Q-switching to be effective, it is more usual to use the Pockels effect in this application.

[3]Named after the German physicist F. C. A. Pockels (1865 - 1913).
[4]Named after the Scottish physicist John Kerr (1824 - 1907).

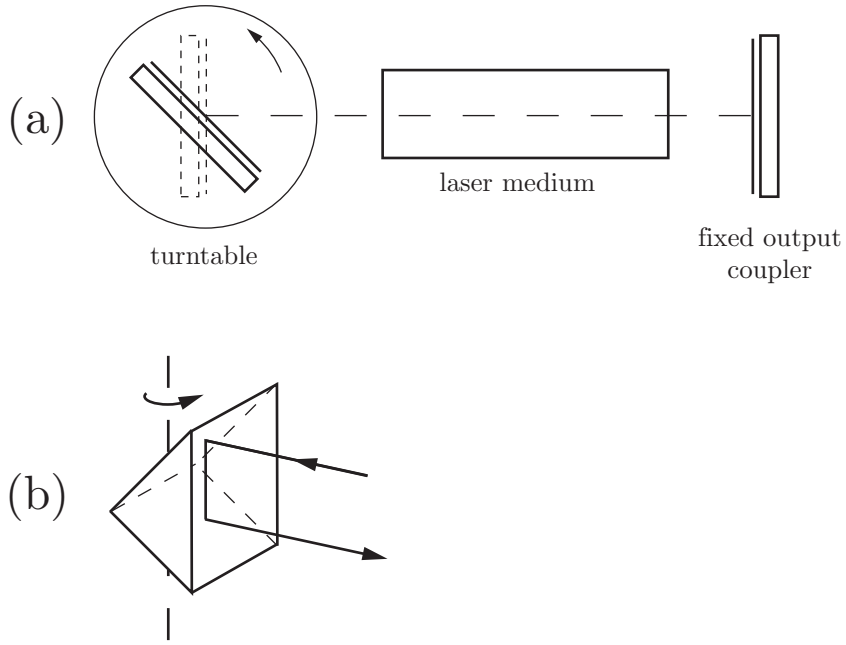

Figure 3.3: Q-switching with a rotating mirror: (a) schematic diagram of the layout of the optical cavity; (b) use of a roof prism as the rotating element. A roof prism has the advantage of automatically ensuring alignment of the reflected beam in one plane (in this case the vertical plane).

We can treat electro-optic devices of this type as wave plates with a birefringence which depends on the strength of the applied electric field. In the arrangement shown in Figure 3.4 a Pockels cell is oriented with its axis at 45^o to the axis of a polarizer. The cavity is held in a low-Q state by applying a certain voltage (the 'quarter-wave' voltage) to the Pockels cell, which then behaves as a quarter-wave plate. Consequently vertically-polarized light transmitted by the polarizer becomes circularly polarized after it transmission through the Pockels cell. Upon reflection from the end mirror of the cavity the handedness of the circular polarization changes, and hence on passing through the Pockels cell again it is converted to horizontally-polarized light and so reflected out of the cavity by the polarizer. Hence, with the quarter-wave voltage applied to the Pockels cell the end mirror of the cavity is effectively blocked.

The mirror can be unblocked by switching the voltage across the Pockels cell to zero, which reduces the birefringence of the Pockels cell to zero[5]. Vertically-polarized light can now pass through the polarizer and Pockels cell with essentially zero loss, and the Q-switched pulse can build up. A Kerr cell would be employed in essentially the same way.

Electro-optic switching has the advantages of fast switching times (of the order of 10 ns) and a high hold-off ratio, which allows the population inversion to be built up to many times the threshold value for the high-Q cavity. The disadvantages are that the technique requires several optics to be placed in the cavity (each with some insertion loss and subject to possible damage), and the devices and associated electronics are relatively expensive. However, the advantages are such that the technique is very widely used.

[5]It would also be possible to devise an arrangement in which the low-loss condition was reached by switching *on* the Pockels cell. However, it is usually technically easier to switch off voltages rapidly — a technique known as 'crowbaring' — than to switch them on quickly.

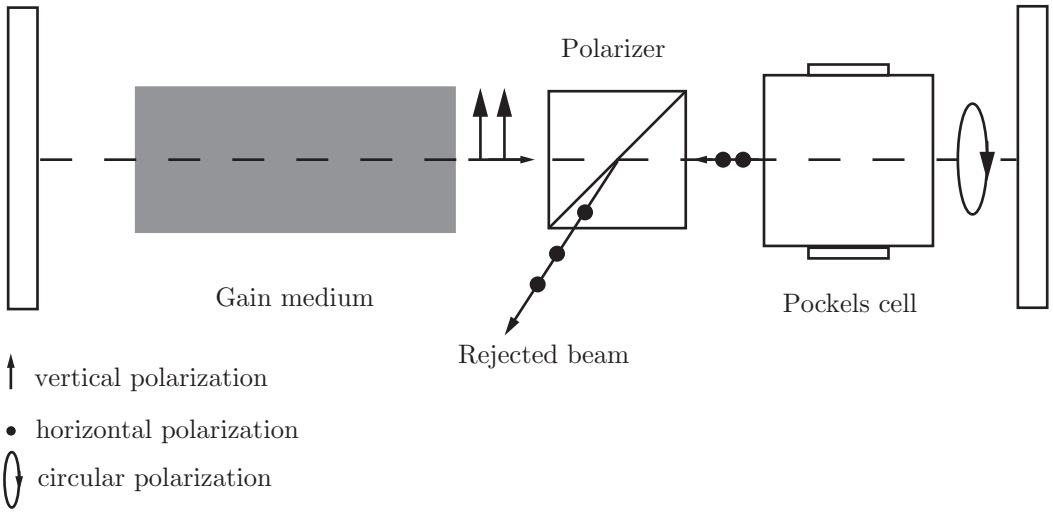

Figure 3.4: Electro-optic Q-switching. The system is shown with the Pockels cell voltage set so as to reject the beam reflected from the right-hand mirror.

Acousto-optic switching

If an acoustic wave is launched through an acousto-optic crystal the refractive index is changed slightly at the peaks and troughs as a result of the local expansion or contraction of the crystal lattice. The resulting periodic variation in the refractive index will cause incident radiation to be diffracted. We can think of this as Bragg reflection from the high- and low-index planes established in the crystal.

A wide variety of acousto-optic materials are available, including: fused silica, tellurium dioxide and lithium niobate for visible wavelengths; gallium arsenide for use in the infrared; and silicon for radiation in the 10 μm region.

Figure 3.5 shows an acousto-optic switch employed in a Q-switched laser. With the radio-frequency (RF) voltage applied to the crystal an acoustic wave propagates along the length of the crystal and is damped at the opposite end, the damping often being achieved by cutting the end of the crystal to form a wedge. Some fraction of the laser radiation incident on the crystal will be Bragg reflected away from the cavity axis and hence lost from the cavity mode. With the RF voltage applied, then, the acousto-optic crystal increases the losses of the laser cavity. Turning off the RF voltage removes this additional loss, and sets the cavity to a high-Q condition.

The advantages of acousto-optic switching are that it is relatively cheap, and insertion losses may be made low by orienting the crystal at Brewster's angle. However the hold-off ratio is low, and hence the technique is usually employed in continuously-pumped Q-switched lasers for which the over-pumping ratio is relatively modest. The switching times achieved are also relatively slow, the switching time corresponding to the time taken for the acoustic wave to propagate out of the region of the cavity mode. Taking the speed of sound in the crystal to be 5 $\mathrm{km\,s^{-1}}$, we can estimate the switching time to be 200 ns for a 1 mm diameter beam. The switching time can be decreased by focusing the cavity mode through the acousto-optic crystal.

Saturable absorbers

A very simple, and quite widely used, method for achieving Q-switching is to insert a saturable absorber into the laser cavity. During the pump pulse the absorber will cause the laser mode to experience a high loss and lasing is prevented. However once the population inversion reaches a high value the absorber will be subjected

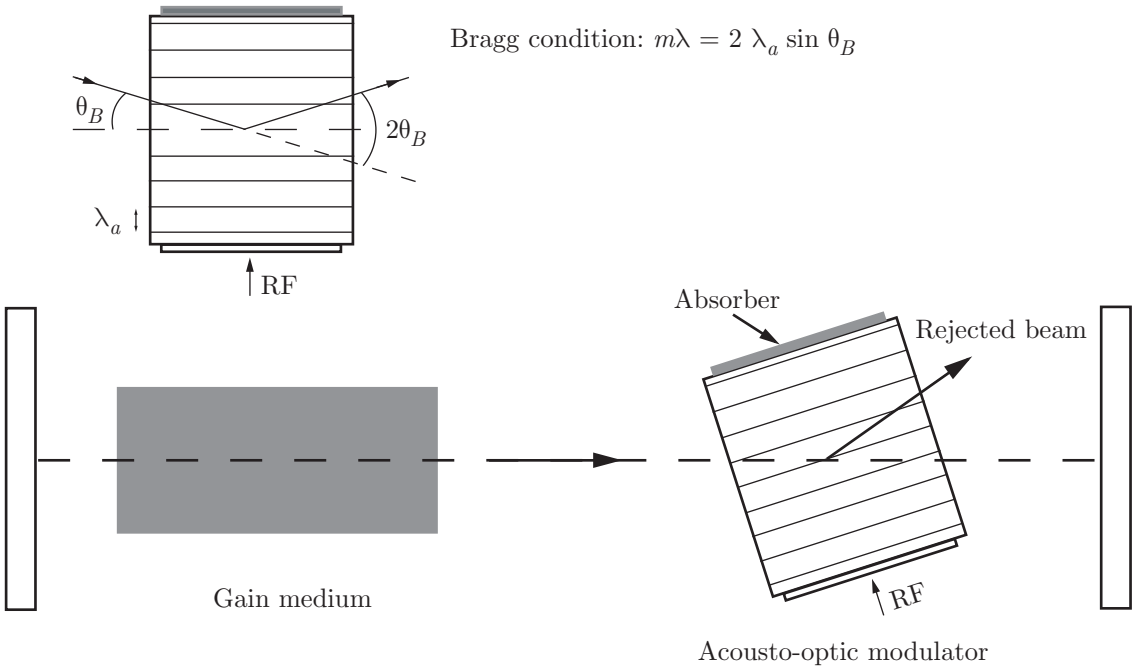

Figure 3.5: Acousto-optic Q-switching. Note that the Bragg angle θ_B has been greatly exaggerated. A value closer to 1° would be more typical. The insertion loss of the acousto-optic crystal can be eliminated (for one polarization) by rotating the crystal out of the plane of the paper through Brewster's angle.

to an ever increasing flux of spontaneous and weak stimulated emission. As population is transferred to the upper level of the absorbing species the absorption starts to saturate, increasing the feedback from the cavity mirror and so increasing the rate of stimulated emission. This positive feedback leads to rapid saturation (or 'bleaching') of the absorber until, at full saturation, it represents essentially zero loss. The cavity is now in a high-Q state, and a giant pulse develops. It is clear that the absorber must have a strong absorption transition at the laser wavelength and also that the lifetime of the upper level must be reasonably long so that the bleached state is maintained for the duration of the Q-switched pulse. On the other hand, the upper level lifetime should be short compared to the desired time between Q-switched pulses. Many types of saturable absorbers have been used, such as solutions of dyes, and doped crystals such as Cr^{4+}:YAG.

Saturable absorbers are a simple (there is no external drive circuit) and cheap way of achieving Q-switching. The relatively slow turn on corresponds to many cavity round-trips, which allows good mode discrimination. As a consequence it is quite straightforward to achieve Q-switched lasing on a single transverse mode of the cavity. The disadvantages are that the shot-to-shot fluctuation in the output energy and in the timing of the output pulse can be relatively high, and that the absorber can degrade over time.

3.2.2 Analysis of Q-switching

An analysis of Q-switching is relatively straightforward, providing we make a few simplifying assumptions: we neglect pumping during the development of the Q-switched pulse, and ignore spontaneous transitions. With these assumptions, the rate equations (3.2) and (3.3) become:

$$\frac{dN_2}{dt} = -N^*(t)\sigma_{21}\frac{I}{\hbar\omega_{\rm L}} \tag{3.23}$$

$$\frac{dN_1}{dt} = +N^*(t)\sigma_{21}\frac{I}{\hbar\omega_{\rm L}}. \tag{3.24}$$

The equations above correspond to the case of **severe bottlenecking** in which the rate of spontaneous decay of the lower laser level is negligible. As such, stimulated emission causes the population to build up in the lower laser level. The rate equation for the evolution of the population inversion during the Q-switched pulse can be found simply by subtraction:

$$\frac{dN^*}{dt} = \frac{dN_2}{dt} - \frac{g_2}{g_1}\frac{dN_1}{dt} \tag{3.25}$$

$$= -\beta N^*(t)\sigma_{21}\frac{I(t)}{\hbar\omega_{\rm L}}, \tag{3.26}$$

where $\beta = 1 + g_2/g_1$.

A second case is also of interest, however: that of an ideal four-level laser. In these systems the lower level decays so quickly that throughout the Q-switched pulse $N_1(t) \approx 0$. In this case $N^*(t) \approx N_2(t)$ and hence the rate equation for the population inversion is identical to that for $N_2(t)$.

It is convenient, therefore, to combine these two cases into one set of equations for the population inversion and the photon density:

$$\frac{dN^*}{dt} = -\frac{\beta}{f_{\rm c}}\frac{N^*}{N^*_{\rm th}}\frac{n}{\tau_{\rm c}} \tag{3.27}$$

$$\frac{dn}{dt} = \left(\frac{N^*}{N^*_{\rm th}} - 1\right)\frac{n}{\tau_{\rm c}}, \tag{3.28}$$

where again we have used $I(t) = n(t)c\hbar\omega_{\rm L}$ and eqn (3.6), and where

$$\beta = \begin{cases} 1 & \text{ideal 4-level} \\ 1 + \frac{g_2}{g_1} & \text{severe bottlenecking} \end{cases} \tag{3.29}$$

The general form of the solution to these equations is shown in Fig. 3.6. Notice that the population inversion density is exactly equal to the threshold value, $N^*_{\rm th}$, when the photon density reaches its peak. Since stimulated emission will continue to occur whilst the photon density falls, the population inversion can be burnt down to a level well below the threshold value $N^*_{\rm th}$.

If we integrate eqn (3.28) over the complete duration of the pulse we find,

$$\int_{n(t=-\infty)}^{n(t=\infty)} dn = \int_{-\infty}^{\infty}\frac{N^*}{N^*_{\rm th}}\frac{n}{\tau_{\rm c}}dt - \int_{-\infty}^{\infty}\frac{n}{\tau_{\rm c}}dt = 0 \tag{3.30}$$

where we have used the fact that the first integral is zero since the initial and final densities of photons in the cavity are both zero.

Now, the rate at which photons are coupled out of the laser is equal to $nV_{\rm c}/\tau_{\rm c}$, where $V_{\rm c}$ is the volume taken up by the oscillating mode in the cavity. Hence the power output from the laser is,

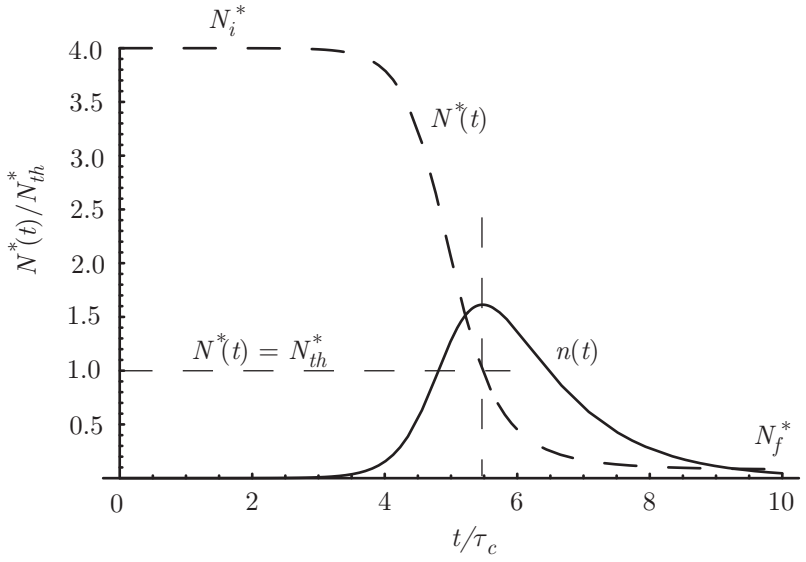

Figure 3.6: General temporal evolution of the population inversion (dashed) and photon (solid) densities during a Q-switched laser pulse for the case $N_i^* = 4N_{\rm th}^*$. The Q-switch is opened at $t = 0$. Notice that it takes several cavity lifetimes for the Q-switch pulse to develop.

$$P = \hbar\omega_{\rm L} V_{\rm c} \frac{n}{\tau_{\rm c}}. \tag{3.31}$$

Applying this result to eqn (3.30), the energy of the output pulse is given by,:

$$\begin{aligned} E &= \hbar\omega_{\rm L} V_{\rm c} \int_{-\infty}^{\infty} \frac{n}{\tau_{\rm c}} dt && (3.32)\\ &= \hbar\omega_{\rm L} V_{\rm c} \int_{-\infty}^{\infty} \frac{N^*}{N_{\rm th}^*} \frac{n}{\tau_{\rm c}} dt && (3.33)\\ &= -\hbar\omega_{\rm L} V_{\rm c} \frac{f_{\rm c}}{\beta} \int_{N_i^*}^{N_f^*} dN^* \qquad \text{(from eqn (3.27))} && (3.34)\\ &= \frac{\hbar\omega_{\rm L}}{\beta} V_g (N_i^* - N_f^*), && (3.35) \end{aligned}$$

where N_i^* and N_f^* are the initial and final population inversion densities, and $V_g = f_{\rm c} V_{\rm c}$ is the volume of the oscillating mode in the gain medium. We therefore can write,

$$E = \eta N_i^* V_g \hbar\omega_{\rm L}, \tag{3.36}$$

where the **energy utilization factor** is defined as,

$$\eta = \frac{N_i^* - N_f^*}{\beta N_i^*}. \tag{3.37}$$

The interpretation of this result is straightforward: The initial number of inverted atoms is equal to $N_i^* V_g$; for an ideal four-level laser transition the number of photons emitted in the pulse is simply $(N_i^* - N_f^*)V_g$; whereas in the case of severe bottlenecking this number is reduced by a factor of β since for these systems each stimulated emission event reduces the population inversion by β. Eqns (3.36) and (3.37) then follow from the fact that each photon carries an energy $\hbar\omega_{\rm L}$.

We may gain further information by dividing eqn (3.28) by (3.27) to give,

$$\frac{dn}{dN^*} = \frac{f_{\rm c}}{\beta}\left(\frac{N_{\rm th}^*}{N^*} - 1\right) \tag{3.38}$$

$$\Rightarrow \int_{-\infty}^{t} dn = \frac{f_{\rm c}}{\beta}\int_{-\infty}^{t}\left(\frac{N_{\rm th}^*}{N^*} - 1\right) dN^*, \tag{3.39}$$

which gives a useful relation between the photon and inversion densities at all times during the Q-switched pulse:

$$n(t) = \frac{f_{\rm c}}{\beta}\left\{N_{\rm th}^* \ln\left[\frac{N^*(t)}{N_i^*}\right] + N_i^* - N^*(t)\right\}. \tag{3.40}$$

After the pulse the photon density is zero, and hence setting $t \to \infty$ in the above yields,

$$N_i^* - N_f^* = -N_{\rm th}^* \ln\left(\frac{N_f^*}{N_i^*}\right) \tag{3.41}$$

$$\Rightarrow \frac{N_i^* - N_f^*}{\beta N_i^*} = -\frac{N_{\rm th}^*}{\beta N_i^*}\ln\left(\frac{N_f^*}{N_i^*}\right). \tag{3.42}$$

We define the **over-pumping ratio** as the ratio of the initial population inversion density to the threshold value:

$$r = \frac{N_i^*}{N_{\rm th}^*}. \tag{3.43}$$

Equation (3.42) may then be rearranged to give a relation between the over-pumping ratio and the energy utilization factor:

$$r = -\frac{\ln(1-\beta\eta)}{\beta\eta}. \tag{3.44}$$

This tells us, for example, that to extract 80% of the energy stored in the initial population inversion in an ideal four-level laser requires pumping the initial population inversion to twice the threshold value. Figure 3.7 plots $\beta\eta$ as a function of the over-pumping ratio. We see that once r reaches approximately 5, virtually all the available energy is extracted from the population inversion.

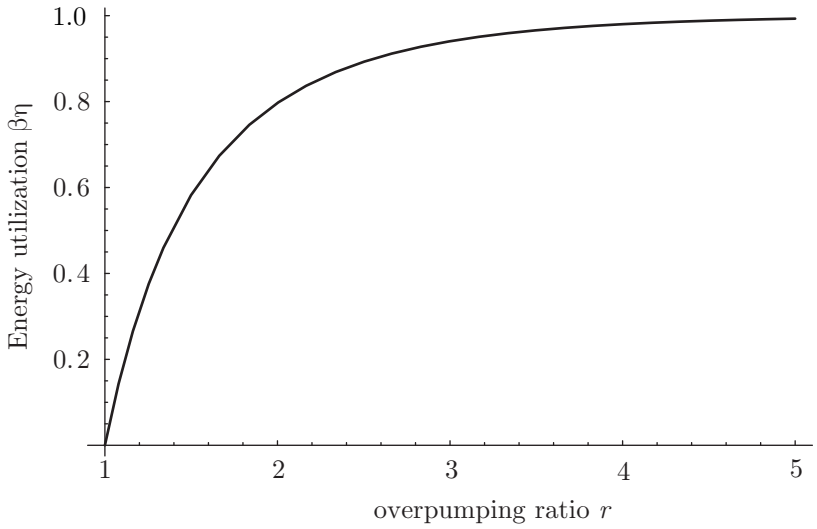

Figure 3.7: The product of the energy utilization factor η and β plotted as a function of the over-pumping ratio r.

Peak output power and pulse duration

The peak density of photons in the cavity occurs when $dn/dt = 0$, or, $N^* = N^*_{\rm th}$. Equation (3.40) then gives the peak photon density as,

$$n_p = \frac{f_{\rm c}}{\beta} N^*_{\rm th} \left(r - 1 - \ln r\right). \tag{3.45}$$

From (3.31) the peak output power is then,

$$P_p = \frac{n_p V_{\rm c} \hbar\omega_{\rm L}}{\tau_{\rm c}} = \frac{N_{\rm th} \hbar\omega_{\rm L} V_g{}^*}{\beta\tau_{\rm c}} \left(r - 1 - \ln r\right). \tag{3.46}$$

We may then estimate the duration $\tau_{\rm Qs}$ of the Q-switched pulse by dividing the pulse energy E by the peak power, to give:

$$\tau_{\rm Qs} = \beta\eta \frac{r}{r - 1 - \ln r} \tau_{\rm c}. \tag{3.47}$$

We note that as $r \to \infty$, $\beta\eta \to 1$ and hence $\tau_{\rm Qs} \to \tau_{\rm c}$. In this limit the Q-switched pulse rises very rapidly owing to the sudden onset of a high rate of stimulated — the population inversion is extracted almost instantaneously — followed by an exponential decrease with a time constant of $\tau_{\rm c}$ as the photons leak out of the cavity.

Lecture 4

Modelocking

4.1 Introduction

Modelocking is an extremely important technique that is used in many laser systems. Its value lies in the ability to generate pulses of laser radiation which have a short duration — typically picoseconds (1 ps $\equiv 10^{-12}$ s) to femtoseconds (1 fs $\equiv 10^{-15}$ s) — and a peak power many orders of magnitude greater than the mean power that can be extracted from the laser medium. Pulses from modelocked lasers have found applications in a wide variety of scientific and technical applications including: imaging techniques such as two-photon laser-scanning fluorescence imaging, ballistic imaging, and optical coherence tomography; ultrafast chemistry, including coherent control of chemical reactions; pump-probe measurements of solid-state materials; and short-pulse laser machining and processing of materials.

4.2 General ideas

We recall that the longitudinal modes of a laser cavity of length L are spaced in angular frequency by an amount,

$$\Delta\omega_{p,p-1} = 2\pi\frac{c}{2L} = \frac{2\pi}{T_\mathrm{c}}, \tag{4.1}$$

where T_c is the time taken for light to travel once round the cavity. In order to avoid clutter, it will be convenient to let $\Delta\omega \equiv \Delta\omega_{p,p-1}$ throughout this lecture.

The right- and left-going waves of a cavity mode may be written in the simplified form,

$$E(z,t) = a_p \exp\left(\mathrm{i}[k_p z \mp \omega_p t + \phi_p]\right), \tag{4.2}$$

where a_p is the amplitude of the wave, ϕ_p is the phase of the wave, the upper (lower) sign refers to waves propagating towards positive (negative) z, and we have ignored the details of the transverse profile of the waves.

Suppose that the laser oscillates simultaneously on N such modes. For simplicity, we will take there to be an odd number of oscillating modes labelled by $p = -P/2 \ldots P/2$. The amplitude of the laser radiation emerging from the output coupler would have the form,

$$E(z,t) = \sum_{p=p_0-P/2}^{p=p_0+P/2} a_p \exp\left(\mathrm{i}[k_p z - \omega_p t + \phi_p]\right), \tag{4.3}$$

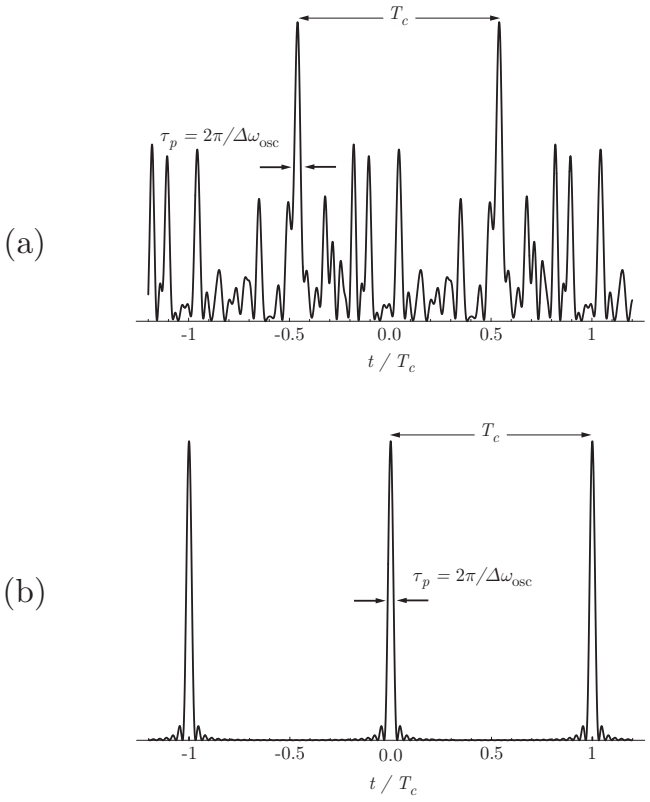

Figure 4.1: Calculated intensity as a function of time emerging from a laser with $N = 61$ modes oscillating simultaneously with equal amplitudes and (a) random (but fixed) phases; (b) identical phases. Notice that for both plots the pattern repeats every cavity round-trip time T_c, and that the sharpest features have the same width. It should also be emphasized that the vertical scales of the two plots are quite different: the peak intensity of the modelocked pulses are approximately N times greater than the mean intensity in (a).

where p_0 is the number of the central longitudinal mode.

Figure 4.1(a) shows a calculation of the output intensity of a laser oscillating simultaneously on 61 modes with equal amplitudes, but with random — though constant — relative phases ϕ_p. The output has two key features,

1. During an interval equal to $T_{\rm c}$ the output contains sharp fluctuations in intensity which are randomly distributed;

2. The sharpest of these peaks has a duration of order $2\pi/\Delta\omega_{\rm osc}$, where $\Delta\omega_{\rm osc} = (N-1)\Delta\omega$ is total bandwidth occupied by the oscillating modes;

3. The pattern is periodic with a period equal to the cavity round-trip time $T_{\rm c}$.

In contrast, Figure 4.1(b) shows the intensity calculated for the case when the modes all have the same phase ($\phi_p = 0$). Now the output consists of a train of short pulses separated by the cavity round-trip time. This situation is known as **modelocking**.

4.2.1 Simple analysis

To gain further insight, let us calculate the output for the case where all the modes have the same amplitude, E_0. The frequencies of the oscillating modes may be written,

$$\omega_p = \omega_{p_0} + p\Delta\omega, \tag{4.4}$$

where ω_{p_0} is the frequency of the centre mode. Similarly, we consider the case in which the phases of adjacent modes differ by a constant amount, $\Delta\phi$, so that we have,

$$\phi_p = \phi_{p_0} + p\Delta\phi. \tag{4.5}$$

From eqn (4.3) the amplitude of the radiation emitted by the laser will then be:

$$E(\boldsymbol{r},t) = E_0 \sum_{p=-P/2}^{p=+P/2} \exp\left[\mathrm{i}(k_p z - \omega_{p_0} t - p\Delta\omega t + \phi_{p_0} + p\Delta\phi)\right]. \tag{4.6}$$

For convenience we consider the amplitude at $z = 0$ and also take the special case[1] in which $\Delta\phi = 0$. Then,

$$E(0,t) = E_0 \exp\left[-\mathrm{i}(\omega_{p_0} t - \phi_{p_0})\right] \sum_{p=-P/2}^{p=+P/2} \exp\left(-\mathrm{i}p[\Delta\omega t]\right). \tag{4.7}$$

The sum appearing in the above is a geometric progression, which may be summed in the usual way to give,

$$E(0,t) = E_0 \exp\left[-\mathrm{i}(\omega_{p_0} t - \phi_{p_0})\right] \frac{\sin\left(\frac{N}{2}\Delta\omega t\right)}{\sin\left(\frac{1}{2}\Delta\omega t\right)}. \tag{4.8}$$

Hence the output takes the form of a (travelling) wave of mean frequency ω_{p_0} multiplied by an envelope function $\frac{\sin\left(\frac{N}{2}\Delta\omega t\right)}{\sin\left(\frac{1}{2}\Delta\omega t\right)}$.

Taking the intensity to be given by $|E(0,t)|^2$ (i.e. ignoring constants of proportionality):

$$I(t) = E_0^2 \frac{\sin^2\left(\frac{N}{2}\Delta\omega t\right)}{\sin^2\left(\frac{1}{2}\Delta\omega t\right)}. \tag{4.9}$$

This is just the function plotted in Fig. 4.1(b).

The temporal behaviour calculated above corresponds to a train of intense modelocked pulses. Notice that the peak intensity of each pulse is equal to $N^2E_0^2$, and hence the peak intensity is a factor of N greater than the mean intensity of N modes oscillating with random phases[2].

Pulses occur at times given by $\Delta\omega t_m = 2\pi m$, and hence the separation of pulses is simply $2\pi/\Delta\omega = T_\mathrm{c}$. The width, τ_p, of the pulses may be taken to be the interval from the peak of a modelocked pulse to the first zero. It is straightforward to show that this is equal to:

[1] It can be shown quite simply that if $\Delta\phi \neq 0$ the peaks of all the modelocked pulses shift in time by a constant value of $\Delta t = \Delta\phi/\Delta\omega$, but that otherwise the train is unchanged.

[2] Note that the mean intensity is the same whether or not the N modes are modelocked or not. Modelocking is an interference phenomenon, and interference neither destroys or creates energy!

$$\tau_p = \frac{2\pi}{N\Delta\omega} = \frac{T_c}{N} \approx \frac{2\pi}{\Delta\omega_{osc}}, \tag{4.10}$$

where $\Delta\omega_{osc} = (N-1)\Delta\omega$ is the (angular) frequency bandwidth covered by the oscillating modes.

It should be clear from the above that if we were able to lock the phase of N longitudinal modes of a laser we could generate a stream of pulses of high peak intensity and short duration. In practice the number of modes that can be locked ranges from several hundred to several thousand, leading to a huge increase in peak power. The duration of the pulses is determined by the frequency width of the laser transition; transitions with broad linewidths can (in principle) generate shorter pulses. In practice pulses with durations from a few picoseconds to a few femtoseconds can be generated; a very active area of current research is the development of techniques related to modelocking to generate pulses in the attosecond range (1 as $\equiv 10^{-18}$ s).

4.2.2 Further general comments

Modelocking techniques can be divided into two classes: **active** and **passive**, as described below.

In the general discussion above we did not consider in any detail the properties of the laser gain medium. As we might expect, the modelocking behaviour of a laser depends on whether the laser transition is homogeneously or inhomogeneously broadened. We have already seen that an inhomogeneously broadened laser transition can oscillate simultaneously on all longitudinal modes for which the round-trip gain exceeds the round-trip loss. The discussion above applies quite well to such lasers, and in such cases the modelocking technique only needs to lock the phases of modes that are already oscillating simultaneously. As a first approximation, the duration of modelocked pulses output from a laser operating on an inhomogeneously broadened transition will be $\tau_p \approx 2\pi/\Delta\omega_D$, where $\Delta\omega_D$ is the inhomogeneous linewidth of the transition.

Homogeneously broadened laser transitions, however, usually only oscillate on a single longitudinal mode (unless spatial holeburning occurs). The modelocking must therefore play a stronger role, and encourage laser oscillation on many modes. As such, the dependence of the pulse duration on the 'strength' of the modelocking will be different from the case of inhomogeneous broadening, and the duration of the modelocked pulses will generally be longer than $2\pi/\Delta\omega_H$, where $\Delta\omega_H$ is the homogeneous linewidth.

4.3 Active modelocking techniques

Active modelocking techniques fall into two classes: **amplitude modulation (AM)** and **frequency modulation (FM)**. We note that we usually require a single modelocked pulse to circulate within the laser cavity. For a linear laser cavity this is achieved by locating the active component near one of the end mirrors, and driving the component at a frequency equal to the longitudinal mode spacing $\Delta\omega$. However, it is possible to generate N such modelocked pulses within the laser cavity by locating the modulator suitably and driving it at $N\Delta\omega$. **Harmonic modelocking** of this type is occasionally used to increase the pulse repetition frequency. In principle it can also be used to generate shorter pulses in homogeneously broadened lasers (see eqn (4.19)).

4.3.1 AM modelocking

Once modelocking is achieved, the intra-cavity radiation takes the form of a short pulse which circulates round the laser cavity. An obvious way, therefore, to achieve modelocking is to introduce some sort of 'shutter' into the cavity which lets through short pulses, but which blocks — or increases the loss of — longer pulses. Radiation which happens to reach the shutter when it is open will pass through with low loss, will experience net round-trip gain, and hence will grow. In contrast, the radiation arriving at the shutter outside this interval will experience a round-trip loss which is greater than the round-trip gain, and so will be absorbed.

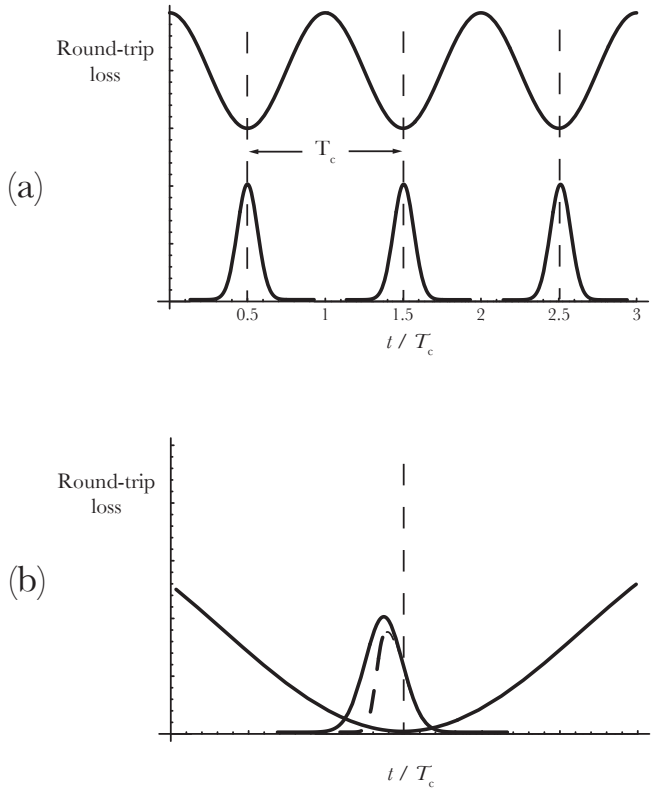

Figure 4.2: Schematic diagram of AM modelocking. In (a) the pulses arrive perfectly synchronized with the shutter and hence experiences minimum loss. The pulse shown in (b) arrives slightly too early, and hence the leading edge is reduced in intensity. The effect is to move the peak of the pulse to later times, closer to synchronization with the shutter.

After many round-trips a pulse will form which propagates round the cavity in synchronization with the shutter, as illustrated schematically in Figure 4.2. The duration of the modelocked pulse so formed is determined by an equilibrium between two competing effects. The shutter tends to decrease the duration of the pulse (and hence increase the bandwidth) since the losses of the leading and trailing edges of the pulse are greater than those experienced by the centre of the pulse. However, the gain medium can only amplify a finite frequency bandwidth which limits the minimum possible duration of the modelocked pulse.

The above description is said to be in the time domain. It is also possible to consider the problem from the perspective of the frequency domain. To do this we suppose that the amplitude transmission of the shutter varies harmonically as,

$$T(t) = T_0 + \frac{\delta}{2}\cos\Delta\omega t. \tag{4.11}$$

Suppose that a wave of angular frequency ω_p is incident on the shutter, such that we may write its amplitude as the real part of,

$$E(t) = E_0 \exp(-\mathrm{i}\omega_p t). \tag{4.12}$$

The amplitude of the transmitted wave will be:

$$E(t) = E_0 \exp(-\mathrm{i}\omega_p t)\left[T_0 + \frac{\delta}{4}\exp(\mathrm{i}\Delta\omega t) + \frac{\delta}{4}\exp(-\mathrm{i}\Delta\omega t\right] \quad (4.13)$$

$$= E_0 T_0 \exp(-\mathrm{i}\omega_p t) + \frac{E_0\delta}{4}\exp(-\mathrm{i}[\omega_p - \Delta\omega]t) + \frac{E_0\delta}{4}\exp(-\mathrm{i}[\omega_p + \Delta\omega]t. \quad (4.14)$$

In other words, the amplitude modulation adds **frequency sidebands** at $\omega_p \pm \Delta\omega$. Since the modulation is driven at the frequency difference between longitudinal modes, the frequencies of the sidebands of a longitudinal mode p lie exactly at the frequencies of the adjacent longitudinal modes $p \pm 1$.

This frequency domain picture provides further insight into how it is possible to achieve modelocking in lasers operating on *homogeneously* broadened transitions. For each longitudinal mode, the amplitude modulation couples energy into the adjacent modes; and it does so with a definite phase. As such, we can consider the sidebands as 'injection seeding' the adjacent longitudinal modes. The system is complex, since each of many modes is coupled to the adjacent modes, but it is clear that the amplitude modulation causes energy to flow from the modes closest to the line centre out to the wings. The amplitude modulation therefore tends to force the laser out of single-mode oscillation, and broaden the frequency spectrum. A steady state is reached when this broadening is balanced by the frequency narrowing arising from the finite frequency response of the gain medium.

We note that for a homogeneously broadened laser transition the round-trip gain of the modes near the line centre is actually greater than the round-trip losses; in the wings of the modelocked pulse the round-trip gain is less than the round-trip loss, as shown schematically in Fig. 4.3. In both cases the difference in energy is made up by energy transfer to or from adjacent modes so that a steady-state is maintained.

Finally we note that for both homogeneously and inhomogeneously laser transitions the system adopts the lowest loss configuration, which corresponds to simultaneous oscillation on a large number of longitudinal modes, phase-locked so as to form a short pulse circulating round the cavity.

Experimental implementation

Amplitude modulation may be achieved by using a Pockels cell in much the same configuration as is used to generate Q-switched pulses. The Pockels cell is driven with a sinusoidal voltage of frequency $\Delta\omega$. Since the amplitude modulation does not need to be 100%, the driving voltage need only be some fraction of the quarter-wave voltage.

It is also possible to use an acousto-optic modulator to implement AM modelocking, using a similar configuration to that used for Q-switching, as shown in Fig. 4.4. However, for modelocking the faces of the acousto-optic block are cut parallel, and the length of the block is cut to be an integer number of half (acoustic) wavelengths long. A standing wave can then be driven in the acousto-optic block, rather than the travelling acoustic wave used to Q-switch. The amplitude of the standing wave passes through zero twice per period, corresponding to windows of minimum loss. Hence, if the block is driven with an angular frequency ω_a, AM modelocking will be achieved if $2\omega_\mathrm{a} = \Delta\omega$.

FM modelocking

Frequency modulation modelocking is achieved by modulating the refractive index of an intra-cavity component at the mode spacing frequency $\Delta\omega$, as illustrated schematically in Fig. 4.5. As we will see, this affects the frequency spectrum of an incident wave and hence the active device is known as a **frequency modulator**.

We consider first the effect on a wave of angular frequency ω of a modulator of length L_mod and time-dependent refractive index $n(t)$. The incident wave will emerge with an amplitude equal to the real part of,

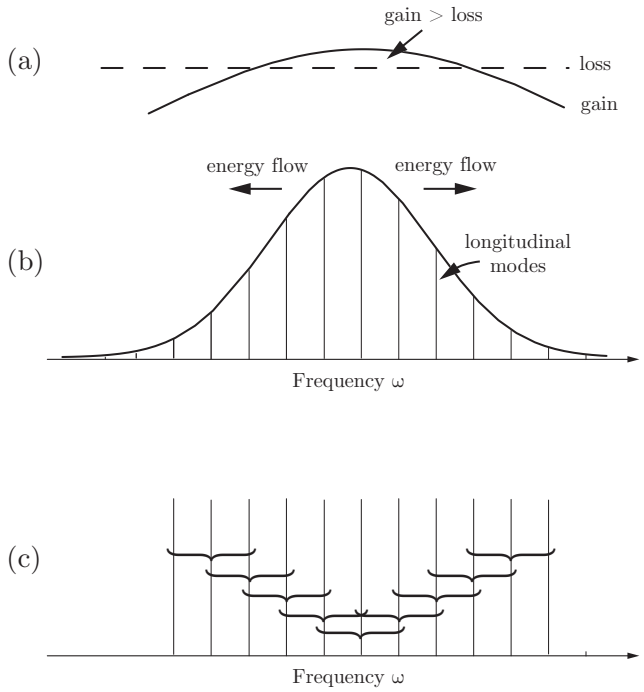

Figure 4.3: Energy flow and coupling of modes in an AM modelocked laser operating on a homogeneously broadened transition. The frequency dependence of the optical gain and loss *in the absence of the amplitude modulation* is shown in (a). As illustrated in (b), the amplitude modulation causes energy to flow from those modes for which the round-trip gain exceeds the round-trip loss to modes in the wings of the gain profile for which the gain is lower than the loss. This energy transfer allows steady-state laser oscillation to be maintained simultaneously on many modes. The amplitude modulation causes each mode to develop sidebands at frequencies equal to those of the adjacent longitudinal modes, and hence couples adjacent modes as illustrated schematically in (c).

$$E(L_{\rm mod}, t) = E_0 \exp\left\{ \mathrm{i}\omega \left[n(t) \frac{L_{\rm mod}}{c} - t \right] \right\}. \tag{4.15}$$

Let us now suppose that the refractive index is modulated harmonically, such that,

$$n(t) = n_0 + \frac{\delta}{2} \sin \Omega t.$$

The transmitted wave will be

$$E(L_{\rm mod}, t) = E_0 \exp\left(-\mathrm{i}\omega t\right) \exp(\mathrm{i}\phi_0) \exp(\mathrm{i}\beta \sin \Omega t), \tag{4.16}$$

where $\phi_0 = \omega n_0 L_{\rm mod}/c$ is the (constant) phase shift experienced by light of angular frequency ω, and $\beta = \omega \delta L_{\rm mod}/2c$ is the amplitude of the modulation in phase introduced by the modulator.

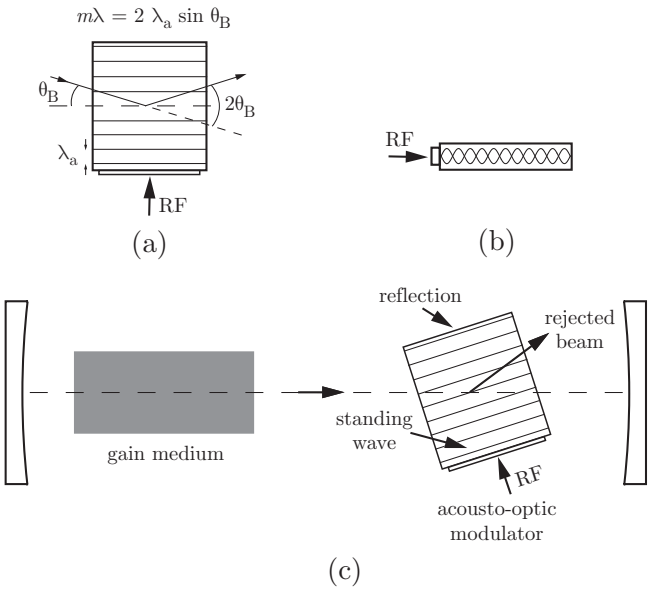

Figure 4.4: Amplitude modulation modelocking with an acousto-optic modulator. In (a) the Bragg condition for reflection of a beam of radiation by incident at an angle θ_B on an acoustic wave of wavelength λ_a. In contrast to the use of acousto-optic crystals in Q-switching, discussed in Lecture 3, in order to achieve modelocking the cavity losses must be modulated at $\Delta\omega$. This is achieved by cutting the acousto-optic crystal so that its two ends are parallel — allowing reflection of the acoustic wave — and separated by an integer of half wavelengths when the modulator is driven at $\Delta\omega/2$. This configuration allows a standing wave to be set-up with an amplitude that passes through zero with a frequency $\Delta\omega$. A typical layout of the optical cavity is shown in (c). Note that the Bragg angle θ_B has been greatly exaggerated. Note also that the insertion loss of the acousto-optic crystal can be eliminated (for one polarization) by rotating the crystal out of the plane of the paper through Brewster's angle.

Now we may use the identity,[3]

$$\exp(\mathrm{i}\beta\sin\Omega t) \equiv \sum_{m=-\infty}^{\infty} J_m(\beta)\exp(\mathrm{i}m\Omega t) \tag{4.17}$$

where $J_m(x)$ is the Bessel function of the first kind of order m.

Thus we may write the amplitude of the transmitted wave as,

$$E(L_{\mathrm{mod}}, t) = E_0\exp(-\mathrm{i}\omega t)\exp(\mathrm{i}\phi_0)\sum_{m=-\infty}^{\infty} J_m(\beta)\exp(\mathrm{i}m\Omega t), \tag{4.18}$$

and we see that the modulator introduces a series of sidebands at frequencies of $\omega \pm m\Omega$, where m is an integer.

Applying these ideas to an intra-cavity modulator driven at the mode spacing frequency $\Delta\omega$, we see that for each mode of frequency ω_p, the modulator generates frequency sidebands at frequencies of $\omega_p \pm m\Delta\omega$ thereby

[3]Consult any text on mathematical methods, for example Riley, Hobson & Bence "Mathematical methods for physics and engineering: a comprehensive guide".

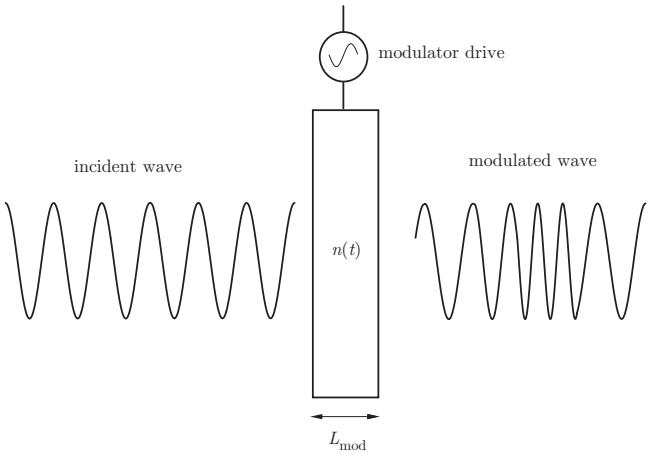

Figure 4.5: Schematic diagram of FM modelocking.

coupling the longitudinal modes in a similar way to amplitude modulation. The width of the pulse generated in this way is determined by the balance between the bandwidth broadening caused by the modulator and the frequency narrowing produced by amplification in the gain medium.

As shown schematically in Fig 4.6, the modelocked pulses are found to be located at the maxima and minima of $n(t)$, when the modulator produces no shift in frequency. This may be understood as follows. The effect of the phase modulation introduced by the frequency modulator is equivalent to modulating the cavity length by driving one of the cavity mirrors sinusoidally. In general a pulse reflected from the moving cavity mirror will be Doppler-shifted in frequency and, since the mirror is driven with a frequency of $\Delta\omega$, when it returns one round-trip later it will be further shifted in frequency by the same amount. Eventually the pulse will be shifted outside the bandwidth of the amplifying transition and will no longer be amplified. However, this progressive frequency shift is minimized for radiation which strikes the mirror at one of the extrema of its motion or, equivalently, if it passes through the modulator near the maxima or minima in $n(t)$.

Notice that there are *two* stable positions for the modelocked pulse. An undesirable consequence of this is that the inevitable disturbances present in a real laser can cause the modelocking to flip between these two states, or even that two modelocked pulses circulate within the cavity.

Experimental implementation

Frequency modulation may be achieved by orienting a Pockels cell with its optical axis parallel with the polarization of the intra-cavity laser radiation. Applying a voltage to the Pockels cell will then change the refractive index experienced by the radiation, but will not alter its polarization state.

4.3.2 Synchronous pumping

Modulating the gain at the longitudinal mode spacing frequency $\Delta\omega$ can result in modelocking in some laser systems by processes analagous to the AM modelocking discussed above. It should be clear that in order for the gain of the laser transition to respond to the modulation, the population inversion must decay in a time which is short compared to the cavity round-trip time.

For example, dye lasers can be modelocked by optically pumping the dye with pulses from a laser which

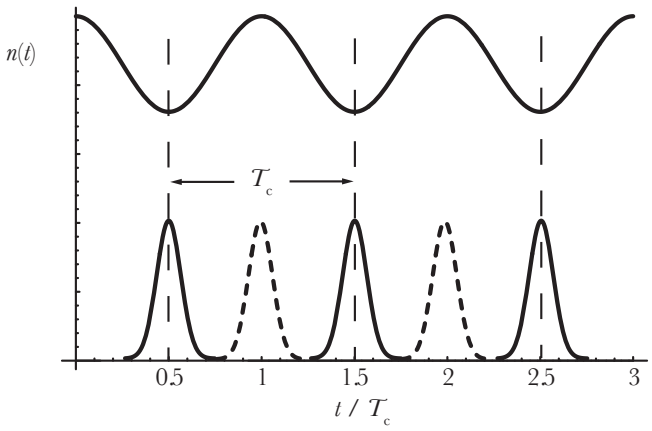

Figure 4.6: Location of the modelocked pulses with FM modelocking.

is itself modelocked[4]. Synchronous pumping can also be achieved in semiconductor lasers by modulating the driving current.

4.3.3 Pulse duration of actively modelocked, homogeneously broadened lasers

For completeness, we note that it can be shown[5] that the duration of the pulses generated by AM modelocking a CW laser operating on a homogeneously broadened laser transition is given by,

$$\tau_p \approx 0.45 \left(\frac{2\alpha_I L_{\rm g}}{\delta} \right)^{1/4} \frac{2\pi}{\sqrt{\Delta\omega_{\rm mod}\Delta\omega_{\rm H}}}, \tag{4.19}$$

where α_I is the (saturated) gain coefficient, $L_{\rm g}$ the length of the gain medium, $\Delta\omega_{\rm mod}$ is the driving frequency of the modulator (usually set equal to $\Delta\omega$, but it may be a harmonic of this), and $\Delta\omega_{\rm H}$ is the homogeneous, linewidth. Notice that the pulse duration decreases with increasing modulation, δ, but only rather slowly. In fact, for most systems the first bracketed term will be close to unity, and hence,

$$\tau_p \approx 0.45 \frac{2\pi}{\sqrt{\Delta\omega_{\rm mod}\Delta\omega_{\rm H}}}. \tag{4.20}$$

A similar result can be obtained for FM modelocking. Notice that since $\Delta\omega_{\rm H} \gg \Delta\omega$ the pulses output by an actively modelocked laser operating on a homogeneously broadened laser will be significantly longer than an inhomogeneously broadened transition of the same linewidth.

4.4 Passive modelocking techniques

Modelocking can also be achieved by inserting into the laser cavity an element with transmission properties which depend on the intensity of the incident radiation. Passive modelocking avoids the need to drive modulators

[4]Obviously care must be taken to ensure that the repetition frequency of the modelocked pump pulses corresponds to the longitudinal mode spacing of the dye laser: in other words the cavity round-trip times must be the same for the two lasers

[5]See, for example, Siegman.

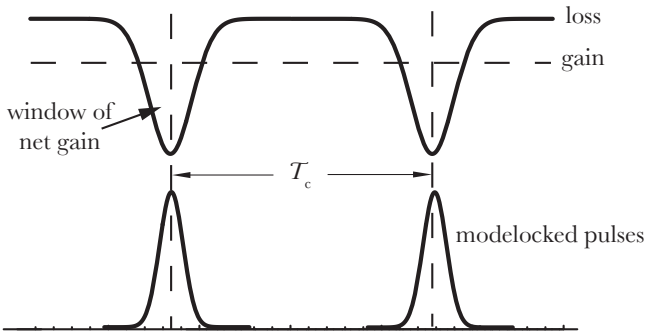

Figure 4.7: Passive modelocking with a fast saturable absorber.

at high frequencies, and is therefore relatively simple to implement practically. Passive modelocking techniques are perhaps of more practical importance than active techniques, and are able to generate the shortest possible laser pulses.

From a theoretical point of view, however, it is more difficult to calculate the pulse duration of the modelocked pulse than in active modelocking, and treatments tend to be specific to the particular case under consideration. In any case, the lower limit of the duration will always be given by approximately by $2\pi/\Delta\omega_{\mathrm{H,D}}$, where $\Delta\omega_{\mathrm{H,D}}$ is the linewidth of the laser transition.

4.4.1 Saturable absorbers

Just as there is for a transition exhibiting optical gain, there are several key parameters for an absorption transition. In particular, the saturation intensity, I_{s}, and the **saturation fluence** (units of $\mathrm{W\,m^{-2}}$), Γ_{s}:

$$I_{\mathrm{s}} = \frac{\hbar\omega}{\sigma_{21}\tau_R} \tag{4.21}$$

$$\Gamma_{\mathrm{s}} = \frac{\hbar\omega}{\sigma_{21}}, \tag{4.22}$$

where τ_R is the recovery time of the transition. We note that the time taken for the absorption to return to its high, unsaturated value once an intense, saturating beam of incident radiation is switched off is of the order of the recovery time.

The saturation intensity is relevant to steady-state conditions, or situations in which the duration, τ, of an incident pulse is much longer than the recovery time. In this case, I_{s} divides two regions: low intensities for which the incident radiation experiences high, unsaturated absorption; high intensities, in which the absorption is strongly reduced by saturation of the transition.

The situation is somewhat different in pulsed conditions. If the duration of an incident laser pulses is short compared to the recovery time the absorption experienced at any time t during the pulse depends only on the intensity integrated over time to that point, i.e. on $\int_{-\infty}^{t} I(t')dt'$. The reason for this is simply that atoms transferred to the upper level of the absorption transition will not decay during the incident laser pulse, and hence the density of atoms in the upper level will depend only on the number of incident photons per unit area, not on the details of the pulse shape. As such, the absorption will start to saturate after a certain number of incident photons per unit area, or, equivalently after a certain incident fluence (energy per unit area). In this case, then, the behaviour is determined by the fluence of the incident pulse relative to the saturation fluence.

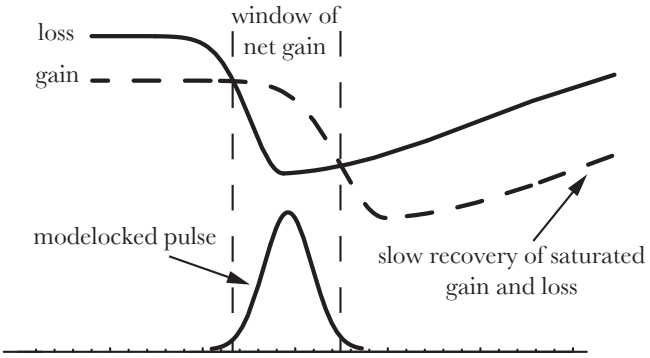

Figure 4.8: Schematic diagram of modelocking with a slow absorber. The interplay of saturation of the gain and loss ensures that net gain only exists in a short window. The saturated gain and loss must recover within a cavity round-trip time T_c.

Fast saturable absorbers

In the context of modelocking, a **fast saturable absorber** is a medium for which the recovery time is very much shorter than the cavity round-trip time, and in fact is short compared to the typical spacing of the 'noise' spikes observed in the non-modelocked, multimode laser (See Fig. 4.3(b)). Typically, then, the recovery time of a fast saturable absorber will be a few tens of picoseconds, or less.

If a fast saturable absorber is placed within a laser cavity, near one of the laser mirrors, the round-trip loss will be lower for intense noise spikes than for low-intensity spikes. Consequently the most intense spike will tend to grow at the expense of the others, and will start to saturate the available gain. The result is a single, intense pulse propagating round the laser cavity: modelocking. We can think of the saturable absorber as acting as a shutter which is activated by the pulse itself, as illustrated in Figure 4.7.

Slow saturable absorbers

Under some circumstances it is also possible to achieve modelocking using a **slow saturable absorber**, for which the recovery time is comparable to — although it must still be less than — the cavity round-trip time.

Figure 4.8 illustrates schematically how modelocking is achieved with a slow absorber. For simplicity we imagine the saturable absorber and the gain medium lie in the same plane. The leading edge of the modelocked pulse will experience a net round-trip loss, owing to the fact that the (unsaturated) absorption is higher than the available gain. As such the front of the pulse will be eroded. As the pulse passes through the absorber, the absorption will gradually decrease as population is transferred to the upper level of the absorption transition. At some point the round-trip loss introduced by the saturable absorber will fall below the round-trip gain, and the pulse will experience amplification. In order for the generated pulse to be short, however, it is essential that shortly after this moment the growing pulse saturates the round-trip gain below the (saturated) round-trip loss so that the trailing edge of the pulse is also attenuated. Further, in order for the system to settle down to a steady-state in which a single, modelocked pulse circulates within the cavity both the gain and the absorption must recover to their initial values in less than a cavity round-trip time.

It should be appreciated, therefore, that modelocking with a slow saturable absorber requires a delicate balance between the energy of the modelocked pulse, and the saturation fluences and recovery times of both the absorber and the gain medium. The requirement that the gain can be returned to its initial value by the pumping mechanism in less than one round-trip time means that the technique is limited to lasers with a relatively short upper level lifetime, such as dye lasers or semiconductor lasers.

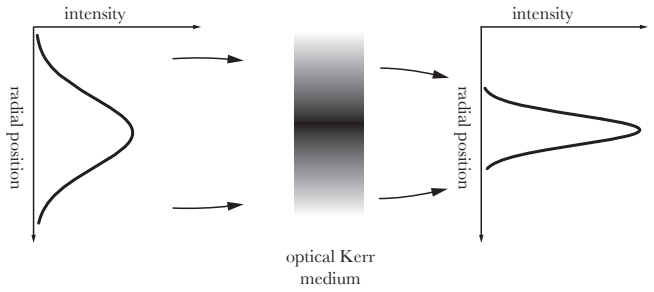

Figure 4.9: Schematic diagram of self-focusing arising from the optical Kerr effect.

4.4.2 Kerr lens modelocking

Kerr lens modelocking (KLM) is one of the most important types of modelocking used today, and is able to generate the shortest optical pulses obtainable directly from a laser oscillator.

As we have already seen, all materials exhibit the Kerr effect, in which the refractive index depends on the square of any applied electric field. If an intense optical pulse is incident on a medium, the **optical Kerr effect** will cause the refractive index to be modified by the electric field of the pulse itself. Since the square of the electric field is proportional to the intensity, I, of the pulse, we may write the refractive index of a material as,

$$n(I) = n_0 + n_2 I. \tag{4.23}$$

It is found that n_2 is positive for nearly all optical materials, with values typically of order 10^{-16} cm^2 W^{-1}.

The optical Kerr effect is responsible for a variety of phenomena, such as self-focusing and self-phase-modulation. The former lies at the heart of Kerr lens modelocking. Figure 4.9 shows the propagation of a beam with a Gaussian transverse profile through an optical Kerr medium. The intensity of the beam will be greater on the propagation axis than in the transverse wings, and hence the refractive index of the medium in the axial region will be greater than in the wings. The optical Kerr medium therefore looks like a thin lens which focuses the beam slightly as it propagates through it.

Figure 4.10 shows schematically how the optical Kerr effect can cause modelocking. The optical Kerr medium is usually the (solid) laser gain medium itself. The laser cavity is designed so that in the absence of any self-focusing the transverse dimensions of the lowest-order cavity mode are comparable to, or larger than, the size of an aperture placed within the cavity. As such, low-intensity radiation experiences a relatively high round-trip loss. Higher intensity radiation, however, will be self-focused by the Kerr effect and so experience lower round-trip losses. Modelocking then arises by essentially the same processes as with a fast saturable absorber: the optical Kerr medium acts as a fast, self-activated shutter.

In order for KLM to be effective, the laser cavity must be operated close to being 'unstable' in order that the weak Kerr lens causes a large change in the transverse extent of the mode.

4.5 Examples of modelocked lasers

4.5.1 CW modelocking

In CW lasers, active modelocking is usually implemented with acousto-optic modulators owing to their lower insertion loss compared to a Pockels cell. The modulator should be placed near one of the end mirrors of the cavity if only a single pulse circulating within the cavity is to be generated.

Some examples are given in Table 4.1.

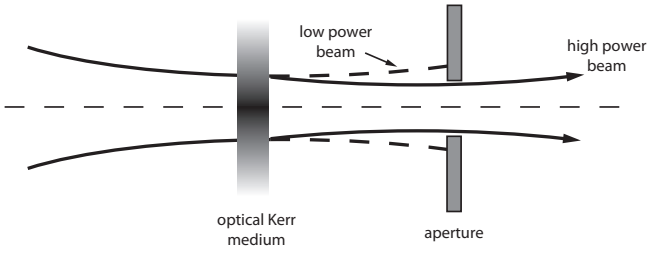

Figure 4.10: Schematic diagram of Kerr lens modelocking.

4.5.2 Pulsed modelocking

The behaviour of pulsed modelocked lasers is determined by the dynamics of the population inversion during the pump pulse. If the lifetime of the upper laser level is short, the duration of the burst of modelocked pulses is determined by the duration of the pumping. However if the lifetime of the upper laser level is long, pulsed modelocking tends to be accompanied by Q-switching. In this case the duration of the burst of modelocked pulses will be approximately equal to the duration of the Q-switched pulse τ_{QS}.

The higher gain found in pulsed lasers means that it is possible to use a Pockels cell to achieve active modelocking, in either AM or FM mode. Passive modelocking of pulsed lasers is achieved with fast saturable absorbers.

Some examples are given in Table 4.1.

Table 4.1: Parameters of some modelocked lasers

Laser	pulsed / c.w.	Method	Pulse duration
He-Ne	c.w.	acousto-optic	1 ns
CO_2	c.w.	acousto-optic	10-20 ns
Ar^+	c.w.	acousto-optic	150 ps
Nd:YAG	c.w.	acousto-optic	100 ps
He-Ne	c.w.	slow sat. abs.	300 ps
GaAs	c.w.	slow sat. abs.	5 ps
Dye	c.w.	slow sat. abs.	25 fs
Dye	c.w.	synch. pumping	1 - 30 ps
Ti:sapphire	c.w.	Kerr lens	10 - 30 fs
Nd:YAG	pulsed	Pockels cell	40 ps
CO_2	pulsed	fast sat. abs.	1 ns
Dye	pulsed	fast sat. abs.	1 ps
Nd:YAG	pulsed	fast sat. abs.	40 ps
Nd:Glass	pulsed	fast sat. abs.	5 ps
Ruby	pulsed	fast sat. abs.	10 ps

Lecture 5

Ultrafast Lasers

5.1 Propagation of short laser pulses in dispersive media

It may be shown that the root-mean-square (RMS) angular frequency bandwidth, $\Delta\omega_{\mathrm{RMS}}$, of a pulse is related to the RMS duration of the pulse, Δt_{RMS}, by,

$$\Delta\omega_{\mathrm{RMS}}\Delta t_{\mathrm{RMS}} \geq \pi. \tag{5.1}$$

In general the value of the **time-bandwidth product** $\Delta\omega_{\mathrm{RMS}}\Delta t_{\mathrm{RMS}}$ depends on:

- the exact shape of the pulse envelope (i.e. Gaussian, square, sech^2 etc.);
- the amount of frequency chirp or phase structure within the pulse.

In addition, of course, different definitions of $\Delta\omega$ and Δt — such as the full-width at half maximum — will yield different time-bandwidth products, although any reasonable definition of these quantities should give a result similar to eqn (5.1).

Short laser pulses, therefore, are associated with a large bandwidth. For propagation *in vacuo*, this presents no difficulties, and the pulse will propagate over arbitrary distances without distortion. However, optical materials are generally dispersive — that is their refractive index varies with frequency — which can lead to distortion of the pulse as it propagates. It is no use generating a short laser pulse if it is stretched and distorted beyond recognition when it reaches the interaction region in an experiment. An understanding of the effect of dispersion, and how it may be controlled, is crucial for both applications and generation of short laser pulses.

In general we may represent an optical pulse by a superposition of harmonic waves. For example, a pulse propagating through a medium is given by the real part of,

$$E(z,t) = \int_0^\infty a(\omega)\exp\left\{\mathrm{i}[k(\omega)z - \omega t]\right\}\mathrm{d}\omega, \tag{5.2}$$

where we take the axis of propagation to be the z axis, and $a(\omega)$ is the amplitude of each angular frequency ω. The wave vector $k(\omega)$ will in general be a function of frequency of the wave, as determined by the **dispersion relation** for the medium. For example, in terms of the refractive index $n(\omega)$ the wave vector is given by,

$$k(\omega) = \frac{\omega}{c}n(\omega). \tag{5.3}$$

In general we can expand the wave vector about its value at some representative central frequency ω_0:

$$k(\omega) = k(\omega_0) + \left.\frac{\partial k}{\partial \omega}\right|_{\omega_0} (\omega - \omega_0) + \frac{1}{2} \left.\frac{\partial^2 k}{\partial \omega^2}\right|_{\omega_0} (\omega - \omega_0)^2 + \dots \tag{5.4}$$

If we retain just the first two terms in this expansion, substitution into eqn (5.2) yields,

$$E(z,t) = \exp\left\{\mathrm{i}[k(\omega_0)z - \omega_0 t]\right\} \int_0^\infty a(\Delta\omega) \exp\left[\mathrm{i}(\Delta\omega)\left(\frac{z}{v\mathrm{g}} - t\right)\right] d(\Delta\omega), \tag{5.5}$$

where $\Delta\omega = \omega - \omega_0$, and the **group velocity** of the pulse is given by,

$$v_\mathrm{g} = \left.\frac{\partial \omega}{\partial k}\right|_{\omega_0}. \tag{5.6}$$

It is easy to see that in this approximation the pulse comprises a **carrier wave** of frequency ω_0, the wavefronts of which propagate at the **phase velocity**

$$v_p = \frac{\omega_0}{k(\omega_0)}, \tag{5.7}$$

multiplied by a **pulse envelope** determined by the integral. Within the approximation we have made here, the pulse envelope has a constant shape and moves at the group velocity. If we had included higher-order terms in the expansion of $k(\omega)$, we would find that the envelope of the pulse would change with propagation distance.

5.1.1 Propagation through an arbitrary system

In determining the behaviour of a short pulse as it propagates between two planes $z = z_1$ and $z = z_2$, it is useful to calculate the total phase accumulated:

$$\phi(\omega) = \int_{z_1}^{z_2} k(\omega) dz \tag{5.8}$$

$$= \phi(\omega_0) + \left.\frac{\partial \phi}{\partial \omega}\right|_{\omega_0} (\omega - \omega_0) + \frac{1}{2} \left.\frac{\partial^2 \phi}{\partial \omega^2}\right|_{\omega_0} (\omega - \omega_0)^2 + \dots \tag{5.9}$$

For a single medium of length L, this expansion is simply equal to eqn (5.4) multiplied by L.

We consider each of the terms above:

- $\phi(\omega_0)$. This is the total phase accumulated at ω_0. Together with the group delay, it determines where the peaks of the carrier wave are located with respect to the envelope of the pulse. As such it becomes important for pulses which comprising only a few optical cycles; techniques for measuring and controlling $\phi(\omega_0)$ are now an active area of research.
- $\left.\frac{\partial \phi}{\partial \omega}\right|_{\omega_0} \equiv \phi^{(1)}(\omega_0)$. This term is known as the **group delay**. To see this, we note that for propagation through a single medium of length L, $\phi^{(1)}(\omega_0) = L/v_\mathrm{g}$, and gives therefore the time taken for the envelope to propagate between the two planes.

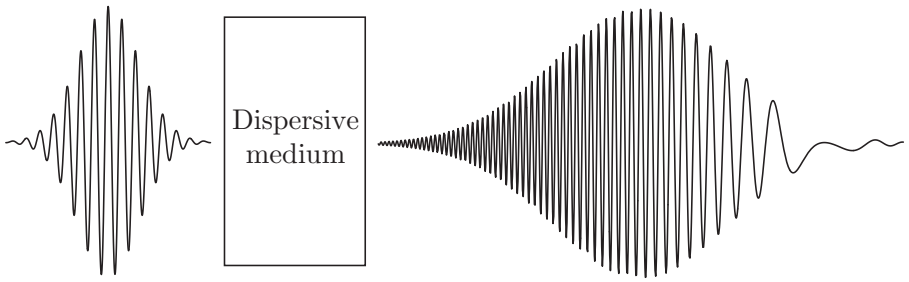

Figure 5.1: Distortion of an initially un-chirped optical pulse by propagation through a dispersive medium with positive GDD. The transmitted pulse is stretched in time, and develops a positive frequency chirp.

- $\left.\frac{\partial^2\phi}{\partial\omega^2}\right|_{\omega_0} \equiv \phi^{(2)}(\omega_0)$. This term is known as the **group delay dispersion (GDD)** or the **quadratic phase**. It is the lowest term in the expansion responsible for distortion of the pulse envelope with propagation. **Positive dispersion** is defined as $\phi^{(2)}(\omega_0) > 0$. The GDD is generally expressed in units of fs^2.
- $\left.\frac{\partial^n\phi}{\partial\omega^n}\right|_{\omega_0} \equiv \phi^{(n)}(\omega_0)$. There are of course higher terms in the expansion: the cubic, quartic, quintic... phase. For the production and propagation of sufficiently short pulses, these will also be important.

Group delay dispersion (GDD)

Let us consider first the frequency dependence of the group delay. Denoting the transit time through the system for quasi-monochromatic waves with frequencies near ω as $\tau(\omega)$, we have from our discussion above,

$$\tau(\omega) = \frac{\partial\phi}{\partial\omega}. \tag{5.10}$$

Expanding about the mean frequency yields,

$$\tau(\omega) = \tau(\omega_0) + \left.\frac{\partial\tau}{\partial\omega}\right|_{\omega_0}(\omega-\omega_0) + \ldots \tag{5.11}$$

$$= \phi^{(1)}(\omega_0) + \phi^{(2)}(\omega_0)(\omega-\omega_0) + \ldots. \tag{5.12}$$

We see that the term in $\phi^{(2)}(\omega_0)$ causes the transit time to increase linearly with frequency. As such, for positive values of the GDD, the high frequency components of the pulse will move to the rear of the pulse; the low frequency components will lead. This will cause the pulse to develop a **positive frequency chirp**, that is for a stationary observer the frequency of the radiation increases as the pulse passes. Further, the difference in group delays will cause the pulse to increase in duration by an amount approximately equal to

$$\Delta\tau_{\text{GDD}} = \phi^{(2)}(\omega_0)\Delta\omega, \tag{5.13}$$

where $\Delta\omega$ is the bandwidth of the pulse. These processes are illustrated schematically in Fig. 5.1

5.1.2 Propagation of Gaussian pulses

Considerable insight may be obtained by applying the general methods discussed above to the special case of an optical pulse with a Gaussian temporal profile. The general conclusions we draw will apply to pulses with arbitrary temporal profile.

Properties of a Gaussian pulse

We consider a generalized Gaussian pulse of the following form,[1]

$$E_{\rm in}(t) = {\rm e}^{-\Gamma t^2}{\rm e}^{-{\rm i}\omega_0 t} \tag{5.14}$$

where the complex Gaussian parameter describing the pulse is given by,

$$\Gamma \equiv a + {\rm i}b. \tag{5.15}$$

Let us first establish the main properties of this pulse. The full-width at half maximum duration of the temporal profile of the pulse *intensity* is easily shown to be,

$$\tau_{\rm p} = \sqrt{\frac{2\ln 2}{a}} = \sqrt{\frac{2\ln 2}{\Re(\Gamma)}}, \tag{5.16}$$

from which we note that the duration of the pulse depends only on $\Re(\Gamma)$.

If $b \neq 0$ the instantaneous frequency changes within the pulse. To see this we note that the phase of the electric field is given by $\psi(t) = -\omega_0 t - bt^2$, and the instantaneous frequency given by,[2]

$$\omega(t) \equiv -\frac{{\rm d}\psi(t)}{{\rm d}t} = \omega_0 + 2bt. \tag{5.17}$$

Thus in general the frequency of the pulse changes linearly with time about the mean value of ω_0; this frequency chirp will be positive if b is positive.

To investigate how such a pulse propagates through an optical system it is necessary to find the frequency spectrum of the pulse by taking the Fourier transform:

$$a(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} E_{\rm in}(t){\rm e}^{{\rm i}\omega t}{\rm d}\omega \tag{5.18}$$

$$\Rightarrow a(\omega) = \frac{1}{\sqrt{2\Gamma}}{\rm e}^{-(\omega-\omega_0)^2/4\Gamma} \tag{5.19}$$

$$= \frac{1}{\sqrt{2\Gamma}}\exp\left[-\Re\left(\frac{1}{4\Gamma}\right)(\omega-\omega_0)^2\right]\exp\left[-{\rm i}\Im\left(\frac{1}{4\Gamma}\right)(\omega-\omega_0)^2\right]. \tag{5.20}$$

It is clear that in general the phase $\psi(\omega)$ of the amplitudes of the frequency components varies quadratically with $(\omega - \omega_0)$ with a curvature determined by $\Im(1/\Gamma)$. The *power* spectrum, which is proportional to $|a(\omega)|^2$, has a full-width at half maximum of,

$$\Delta\omega_{\rm p} \;=\; 2\sqrt{\frac{2\ln 2}{\Re(1/\Gamma)}} = 2\sqrt{2\ln 2}\sqrt{a[1+(b/a)^2]}. \tag{5.21}$$

[1] Note that Siegman writes the pulse in the form $\exp(-\Gamma t^2)\exp({\rm j}\omega_0 t)$ with $\Gamma \equiv a - {\rm j}b$. Consequently equations in Siegman's book may be converted to our notation by making the transformation ${\rm j} \to -{\rm i}$, and vice versa.

[2] Equation (5.14) may be written as $E_{\rm in}(t) = \exp(-at^2)\exp\left[-{\rm i}(\omega_0 + bt)t\right]$. However, it is *not* correct to deduce from this that the effective frequency is $\omega_0 + bt$.

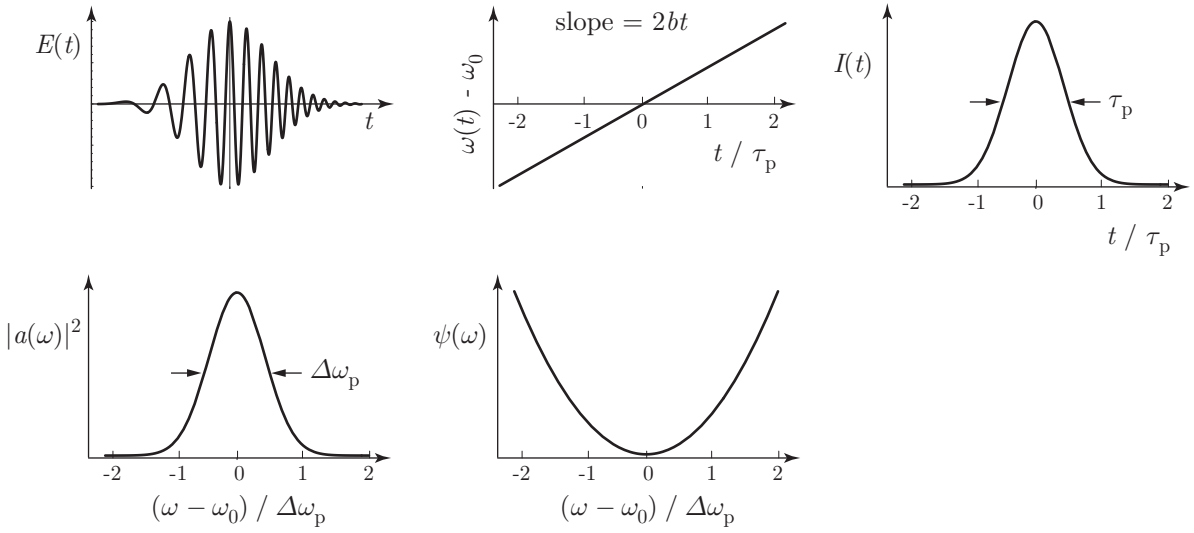

Figure 5.2: Properties of a Gaussian optical pulse with a pulse parameter $\Gamma = a + \mathrm{i}b$ for the case $b > 0$. Shown are the electric field $E(t)$, the instantaneous frequency $\omega(t)$, the temporal profile of the intensity $I(t)$ the power spectrum $|a(\omega)|^2$ and phase $\psi(\omega)$.

In terms of the full-widths at half maxima, the time-bandwidth product of the Gaussian pulse is found from eqns (5.16) and (5.21) to be

$$\Delta\omega_{\mathrm{p}}\tau_{\mathrm{p}} = 4\ln 2\sqrt{1 + (b/a)^2}. \tag{5.22}$$

The time-bandwidth product clearly takes the smallest possible value if $b = 0$, i.e. if the pulse has no frequency chirp or, equivalently, if the phases of the frequency components are all the same.[3] Figure 5.2 summarizes the properties of Gaussian optical pulses of this type.

Propagation of a Gaussian pulse

We now consider the propagation of a Gaussian pulse through an optical system for which the linear and quadratic dispersion are finite, but the higher-order dispersion may be ignored:

$$\phi(\omega) = \phi^{(0)} + \phi^{(1)}(\omega - \omega_0) + \frac{1}{2}\phi^{(2)}(\omega - \omega_0)^2. \tag{5.23}$$

To calculate the form of the output pulse we multiply the frequency spectrum of the input pulse by $\exp[\mathrm{i}\phi(\omega)]$ and take the inverse Fourier transform. As explored in the problems, this turns out to be relatively straightforward, and we find

$$E_{\mathrm{out}}(t) = \sqrt{\frac{\Gamma_{\mathrm{out}}}{\Gamma_{\mathrm{in}}}}\exp[\mathrm{i}(\phi^{(0)} - \omega_0 t)]\exp\left\{-\Gamma_{\mathrm{out}}\left[t - \phi^{(1)}\right]^2\right\}, \tag{5.24}$$

[3] If the widths in eqn (5.22) are replaced with root-mean-square values the factor of $4\ln 2$ is replaced by π. We then see that the case $b = 0$ reduces to the minimum possible value according to eqn (5.1).

where

$$\frac{1}{\Gamma_{\rm out}} = \frac{1}{\Gamma_{\rm in}} - 2{\rm i}\phi^{(2)}. \tag{5.25}$$

We see that the transmitted pulse still has a Gaussian temporal profile, but that the duration of the pulse envelope and the frequency chirp will in general be different from the input pulse since the parameter Γ will be changed. There will also be a relative shift between the peak of the Gaussian pulse envelope — which occurs at $t = \phi^{(1)}$, i.e. it is shifted in time by the group delay — and the carrier wave which is shifted in phase by $\phi^{(0)}$. Since in general $\phi^{(0)}/\omega_0 \neq \phi^{(1)}$ the carrier-envelope offset of the pulse will be altered by the optical system.

Note that there are strong parallels between eqn (5.24) describing the propagation of a Gaussian optical pulse and the equations describing the propagation of a lowest-order Gaussian beam. Comparison of these two situations reveals that the parameter $1/\Gamma$ plays an equivalent role to that of the complex radius of curvature q, and the dimension of time (shifted by $-\phi^{(1)}$) is equivalent to the transverse co-ordinate r. Thus compression (stretching) of a Gaussian pulse are analogous to focusing (defocusing) of a Gaussian beam. Notice too that compression of a chirped Gaussian pulse requires the introduction of a medium with quadratic phase variation in frequency; whereas as focusing a Gaussian beam requires a lens which imposes a phase shift which varies quadratically with the transverse co-ordinates of the beam.

These points are made more obvious by considering the case of a uniform medium. Equation (5.24) may then be written as

$$E_{\rm out}(t) = \underbrace{\exp[-{\rm i}\omega_0(t - z/v_{\rm p})]}_{\text{carrier wave}} \overbrace{\sqrt{\frac{\Gamma_{\rm out}}{\Gamma_{\rm in}}}\exp[-\Gamma_{\rm out}(t - z/v_{\rm g})^2]}^{\text{pulse envelope}} \tag{5.26}$$

since $\phi^{(0)} = k(\omega_0)z$ and the group delay $\phi^{(1)} = z/v_{\rm g}$.

In general the pulse duration and frequency chirp of the optical pulse are altered by transmission through the optical system since the Gaussian parameter Γ is changed. Note, however, that the frequency width of the pulse is *not* altered by the propagation, as must be the case for a linear system.[4] Whether the optical pulse is compressed or stretched in time, and whether the pulse becomes more or less chirped, depends on the relative sign of the initial frequency chirp and the GDD of the optical system.

5.1.3 Nonlinear effects: self-phase-modulation and the B-integral

As we saw in Lecture 4, at high intensities the refractive index of a material becomes nonlinear with the intensity:

$$n = n_0 + n_2 I. \tag{5.27}$$

In propagating a distance L a wave of constant intensity I therefore accumulates a nonlinear phase

$$\phi_{\rm NL} \equiv B = \frac{2\pi}{\lambda_0}\int_0^L n_2 I(z)dz \tag{5.28}$$

in addition to that which would be accumulated by a beam of low intensity. Here λ_0 is the vacuum wavelength of the radiation. The integral above is universally known as the **B-integral** for the propagation.

[4] This may be seen from eqn (5.21); the width of the spectrum is determined by $\Re(1/\Gamma)$ and from eqn (5.25) $\Re(1/\Gamma_{\rm out}) = \Re(1/\Gamma_{\rm in})$.

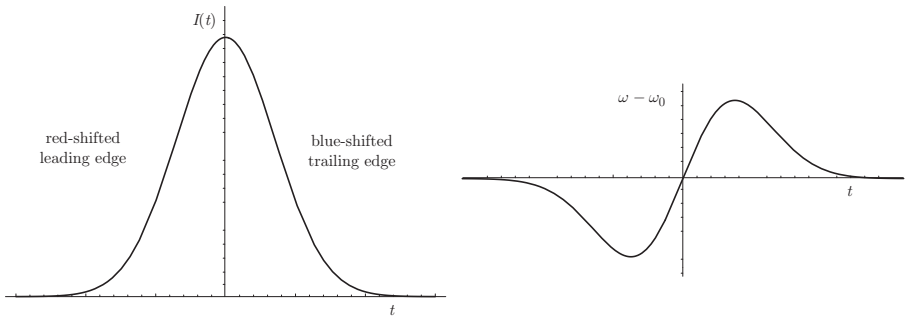

Figure 5.3: Generation of a frequency chirp by self-phase-modulation of an intense optical pulse.

The intensity of a beam of radiation will generally vary with transverse position within the beam, and, as we have seen, this usually leads to self-focusing of the beam. For a laser *pulse* the nonlinear phase will also vary with time during the pulse. Ignoring dispersion, the total accumulated phase of a wave after propagating a distance L into a medium will be given by,

$$\phi = \frac{2\pi}{\lambda_0}\left[n_0 + n_2 I(t)\right] L - \omega t. \tag{5.29}$$

At any point in the pulse, the effective frequency, ω', is given by the (negative) total rate of change of the phase with time:

$$\omega' = -\frac{d\phi}{dt} = \omega - \frac{2\pi}{\lambda_0} n_2 L \frac{\partial I}{\partial t}. \tag{5.30}$$

The nonlinear refractive index therefore causes the phase of the carrier wave in the pulse to vary nonlinearly with time, corresponding to a variation of frequency within the pulse. This process is known as **self-phase-modulation**.

Figure 5.3 shows this process schematically. The leading edge a pulse must have $\partial I/\partial t > 0$ corresponding to a decrease in frequency, or red shift, for materials with $n_2 > 0$. Similarly, the trailing edge of the pulse will be blue-shifted, and as a consequence the pulse will develop a positive chirp. The rate of change of frequency with time is approximately linear over the central part of the pulse, as shown in Fig 5.3.

Self-phase-modulation and self-focusing are nearly always undesirable in a laser system, in which case they must be kept to a minimum. A rule of thumb often employed by system designers is that the total B-integral should be kept below 1 radian.

5.2 Dispersion control

The linear and nonlinear dispersion discussed above causes distortion of short laser pulses as they propagate within a system. In order to overcome this — or at least alleviate it — it is necessary to employ techniques in which additional, controllable dispersion can be introduced. A simple way of achieving this is to pass the pulse through a series of optical components for which the physical (and hence optical) path is different for different wavelengths. In this way controllable, wavelength-dependent phases can be introduced. The dispersion introduced in this way is known as **geometric dispersion**.

Dispersion control is particularly important for modelocked lasers producing very short pulses. In order to generate the shortest possible pulses by modelocking it is necessary for the longitudinal modes must be evenly

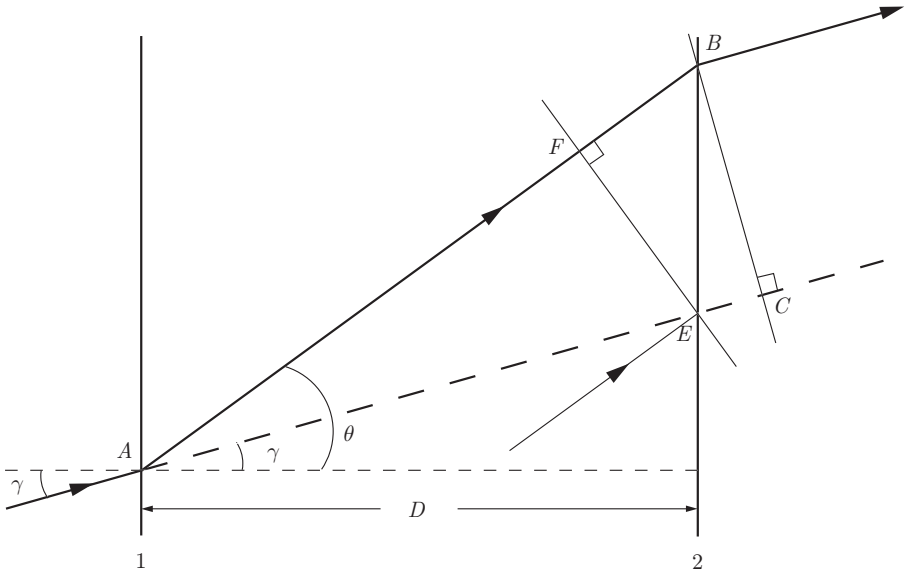

Figure 5.4: Schematic diagram of a general optical system exhibiting geometric dispersion.

spaced in frequency, but the presence of dispersion in the cavity will prevent this. Clearly the problem will become more acute as the duration of the modelocked pulses is reduced, and the associated bandwidth is increased. In practice it is found that dispersion control is necessary to generate pulses shorter than approximately 150 fs. Most materials exhibit positive dispersion, and hence geometric dispersion introduced into a laser cavity needs to be negative (and controllable).

Figure 5.4 shows a general system exhibiting geometric dispersion. Upon reaching the plane 1, light is refracted or diffracted through a wavelength-dependent angle θ. After propagating some distance it is refracted/diffracted again so as to render the direction of propagation parallel to the initial direction. Since θ depends on the wavelength, the optical path, and hence the group delay, will depend on wavelength; the system therefore introduces dispersion.

In calculating the phase accumulated by a given wavelength, we need to define two planes (an input plane and an output plane) through which all waves pass. There is no difficulty in identifying the input plane: it is the plane passing through A, perpendicular to the *incident* ray. However, the output plane needs to be defined more carefully. For example, the plane BC is not a suitable candidate since this plane is different for each wavelength. There is still a lot of flexibility: the output plane could be any fixed plane perpendicular to the rays *leaving* the system. One candidate would be such a plane passing through E.

However, it is more convenient to move the reference output plane back along the undeflected ray to A – i.e. the same plane as the input plane! Using these reference planes the additional optical path travelled by the deflected ray is then equal to the distance $AF - AE$.[5]

Now,

[5]The calculated accumulated phase depends on where the input and output planes are defined to be. However, the calculated GDD (and higher orders) do not depend on these locations; all that will change is the total accumulated phase and the group delay for propagating between the two reference planes, just as we would expect.

$$AE = \frac{D}{\cos\gamma} \tag{5.31}$$
$$AF = AE\cos(\theta-\gamma). \tag{5.32}$$

Hence the optical path difference is given by,

$$OP = -\frac{D}{\cos\gamma}\left[1-\cos(\theta-\gamma)\right], \tag{5.33}$$

and hence the phase introduced is simply,

$$\phi(\omega) = -\frac{\omega}{c}\frac{D}{\cos\gamma}\left[1-\cos(\theta-\gamma)\right]. \tag{5.34}$$

It is a simple matter (in principle) to calculate the GDD etc. by differentiating eqn (5.34), providing the relation between θ and ω is known.

5.2.1 The grating pair

Figure 5.5 shows a ray propagating via a pair of parallel reflection grating. Light of wavelength λ will be diffracted at an angle θ from the first grating according to,

$$m\lambda = d[\sin\theta - \sin\gamma] \quad m = 0, \pm1, \pm2, \pm3, \ldots \tag{5.35}$$

where γ is the incident angle and d the grating spacing. At the second grating the incident angle will now be θ and consequently the second grating will diffract the light at an angle γ if the order of diffraction is m for the first grating and $-m$ for the second. The diffracted ray therefore leaves parallel to the incident ray.

The optical phase introduced by the grating pair is given by eqn (5.34), and this, together with eqn (5.35) allows one to calculate the GDD and other phase terms. For this simple case, however, it is easy to deduce the sign of the GDD. Consider the time taken for light to propagate from A to (say) the position of the mirror or roof prism, which we may take as a convenient reference plane. We know that the time taken for the undeflected light (dashed line) to reach this plane is a minimum, since the path it takes is the same as if the gratings were replaced by mirrors. All other paths must take a longer time, and since the diffraction angle θ increases with wavelength, we conclude that the group delay is smaller for shorter wavelengths (higher frequencies). Hence the group delay decreases with frequency, and the GDD must be negative[6]. It should also be clear that the GDD will increase linearly with the grating separation D, and consequently the grating pair can provide controllable negative GDD. Clearly the different wavelengths leave the grating pair at different heights above the incident ray. Consequently in the exiting beam there is a correlation between position and wavelength, a situation known as **spatial chirp**. In general this is undesirable, and the spatial chirp is undone by reflecting the beam back through the grating pair to form a beam counter-propagating along the axis of the incident. This counter-propagating beam may be selected by allowing the incident beam to enter the system at a small angle to the

[6]Note that the group delay is NOT equal to the optical path difference given by eqn (5.33) divided by the speed of light. To use the optical path we must find the phase introduced by the system and then see how it varies with frequency. Unfortunately this is not as straightforward as the method used in the text to deduce the sign of the GDD: as ω increases the optical path difference decreases, but $\phi(\omega)$ is equal to this path difference multiplied by ω/c and hence it is not obvious whether overall $\phi(\omega)$ increases or decreases with ω. The method above works because we can use our insight to deduce that the group delay, $\phi^{(1)}(\omega)$, decreases with frequency.

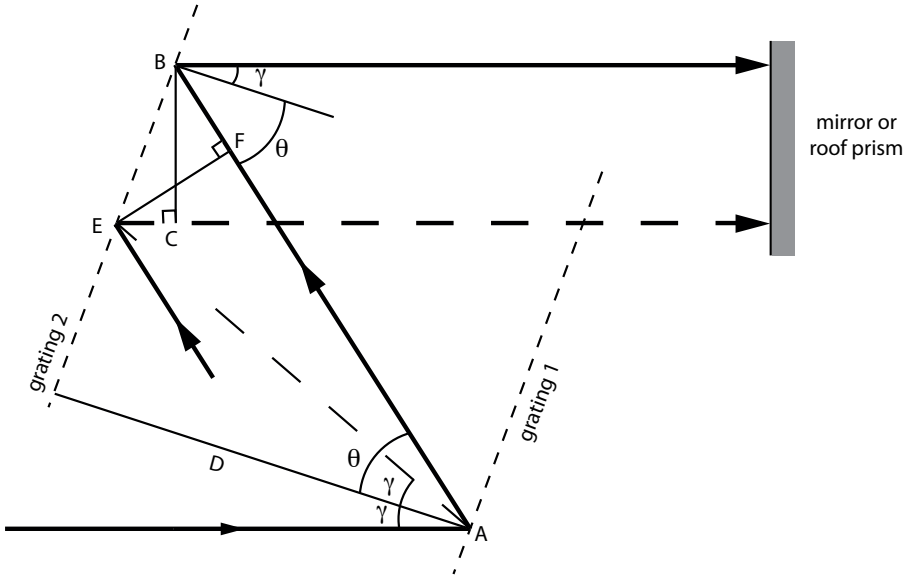

Figure 5.5: A grating pair used to introduce negative GDD.

plane of the page; the exiting beam therefore leaves above the incident beam and may be picked off by a small mirror. Alternatively, and somewhat better, the beam may be returned by reflection from a roof mirror which returns the light parallel to the incident rays, but at a different height. Obviously, double-passing the beam through the grating pair will introduce twice the GDD generated from a single pass.

5.2.2 The prism pair

Dispersion control may also be achieved with the arrangement of prisms shown in Figure 5.6[7]. It is clear that the optical paths are different for different wavelengths, and consequently the arrangement will introduce geometric dispersion. Note that the plane AB is a plane of symmetry and consequently the total GDD is simply twice that introduced by the first prism pair. The advantage, however, of the arrangement of four prisms is that the emerging beam propagates along the same axis as the incident beam. This means that the optical system (in particular a laser cavity) can be aligned *without* the prisms in place, and once this is done the prisms can be introduced without changing the alignment of other components.

It may be shown that for the first two prisms, the optical path and phase difference between a ray refracted at an angle θ and the ray which passes between the apexes of the prisms are also given by eqns (5.33) and (5.34). There is no simple argument for deducing the sign of the GDD in this case: indeed it may be positive or negative. To evaluate the GDD, and other terms, one must use Snell's Law to calculate θ as a function of frequency. This procedure is not difficult, but it is somewhat messy and gives us no particular insight. In fact it is found[8] that the GDD introduced by a prism pair is also negative, provided that l is sufficiently large. The GDD is proportional to $l = D/\cos\gamma$, and hence may be controlled by adjusting l. In practice, however, it is usual to introduce precise control of the GDD by moving prisms II and III perpendicular to their bases, as indicated in Fig. 5.6. This reduces l and introduces more prism material into the path, and therefore increases the GDD.

[7]Note that the diagram given in Svelto appears to be incorrect in that the higher frequency light is refracted through a smaller angle.

[8]This was first shown by R. L. Fork *et al.*, Optics Letters vol. 9, p. 150 (1984).

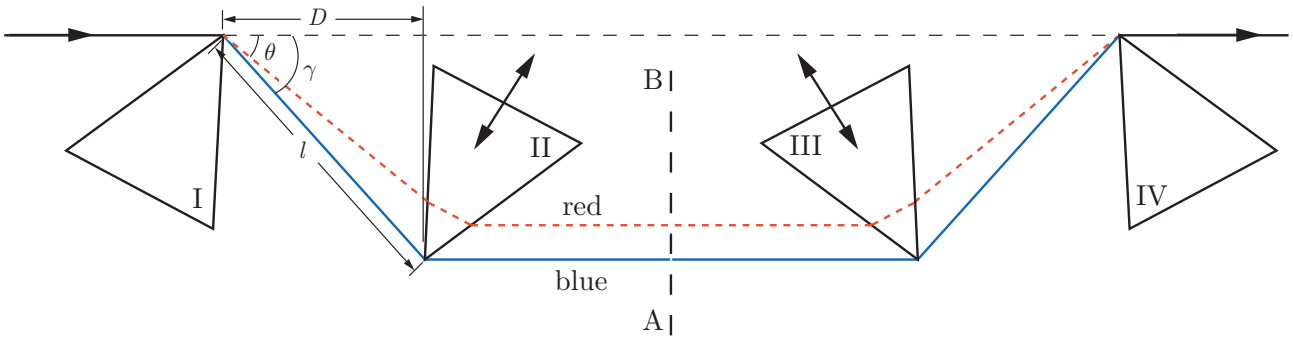

Figure 5.6: A prism pair used to introduce negative GDD. Note that the plane AB is a plane of symmetry; it would be possible to place a mirror here, reflecting rays back through the first two prisms to yield a counter-propagating beam with the same GDD as produced by the four prisms shown.

The prisms are usually oriented at Brewster's angle so as to minimize the insertion loss; and the apex angles of the prisms are chosen so that they operate at minimum deviation (i.e. the incident and exit angles are equal). The GDD introduced by prism pairs is typically much smaller than that provided by grating pairs, but have the advantage that the insertion loss can be much smaller. As such prism pairs are often used to provide dispersion control in modelocked laser oscillators.

5.2.3 Introduction of positive GDD

The two systems above both introduce negative GDD. Clearly we would also like to be able to introduce positive GDD (especially for CPA - see later), but this would appear to be difficult owing to the fact that for both these systems the GDD is negative for a positive separation D.

A clever solution to this problem is to introduce an imaging system between the two elements which can project the first element in front, behind, or at the same position as the second. In this way the effective separation of the elements may be made to be positive, zero, or negative which allows the GDD to be negative, zero, or positive.

In practice this system is usually implemented with diffraction gratings, as shown in Figure 5.7. The two lenses form an image, with unit magnification, of grating $G1$. By adjusting the position of $G1$, its image, $G1'$, may be positioned either side of the second grating $G2$. Notice that the longitudinal magnification is equal to -1, and hence the imaging system changes the orientation of $G1'$ with respect to $G1$.

Many ingenious improvements have been made to this simple idea. More advanced systems replace the lenses with curved mirrors, avoiding problems caused by chromatic aberration. It is also possible to use a folded, single-grating design which automatically ensures alignment of the 'two' gratings as the other components are adjusted.

5.2.4 Chirped mirrors

A recently developed way of introducing controlled GDD is reflection from a **chirped mirror**. A standard 'dielectric mirror' comprises a stack of alternating layers of high and low refractive index material, the thickness of each mirror being equal to a quarter of the wavelength in the medium. In a chirped mirror the thickness of the layers is varied within the stack so that different wavelengths are reflected most strongly at different depths in the stack, as illustrated schematically in Figure 5.8. With careful design and manufacture, it is possible to

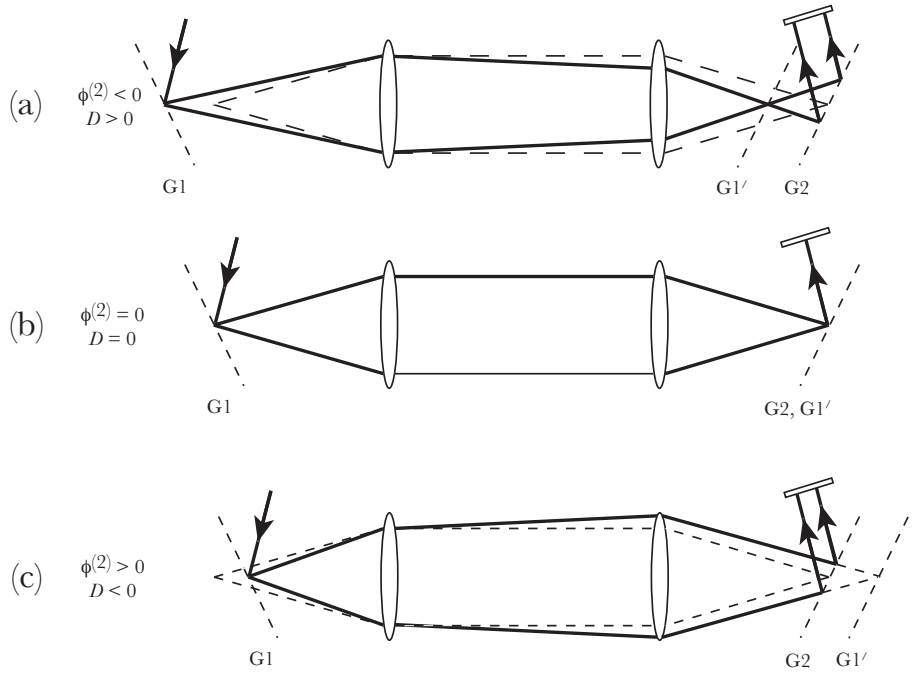

Figure 5.7: Grating configurations giving (a) negative, (b) zero, and (c) positive GDD.

produce any desired GDD – as well as controlled higher-order phase terms. If the GDD produced by one mirror proves insufficient, it may be increased by multiple reflections.

Chirped mirrors have the advantage that they are simple to align and set up, and introduce a minimum of optical elements into the system. A disadvantage is that they must be designed specifically for each application, and are therefore less flexible.

5.3 TW and PW laser systems

Perhaps the supreme example of the application of dispersion control is the generation of laser pulses with peak powers in the terawatt ($1\,\mathrm{TW} \equiv 10^{12}\,\mathrm{W}$) or even the petawatt ($1\,\mathrm{PW} \equiv 10^{15}\,\mathrm{W}$) range.[9]

Imagine trying to generate laser pulses with a peak power of 1 TW. If the pulse had a duration of 1 μs, our 1 TW laser pulse would require an energy of 1 MJ! The first trick, then, in generating high peak powers is to keep the duration of the pulses short; modelocked pulses in the femtosecond to picosecond ranges are used for ultra-high power laser systems.

The laser pulses generated by modelocked lasers typically have very low energies — a few nanojoules is quite common. In order to reach the TW or PW level the pulse energy must be amplified by factors of 10^6 to 10^{10}. If we were to attempt to do this in a simple chain of amplifiers, we would soon reach peak intensities that would cause strong self-focusing, leading to damage of the optical components in the amplifier chain. In practice the intensity to which the pulses can be amplified is often even lower than the limit set by optical damage; at lower intensities self-phase-modulation leads to the development of a complicated phase structure to the pulse which cannot be compensated easily by controlled dispersion. An obvious way to keep the intensity reasonable is to

[9] An excellent review of TW and PW laser systems may be found in S. Backus, Review of Scientific Instruments vol. 69 p. 1207 (1998). Further information on PW lasers systems can be found at the NIF website: www.llnl.gov/nif/.

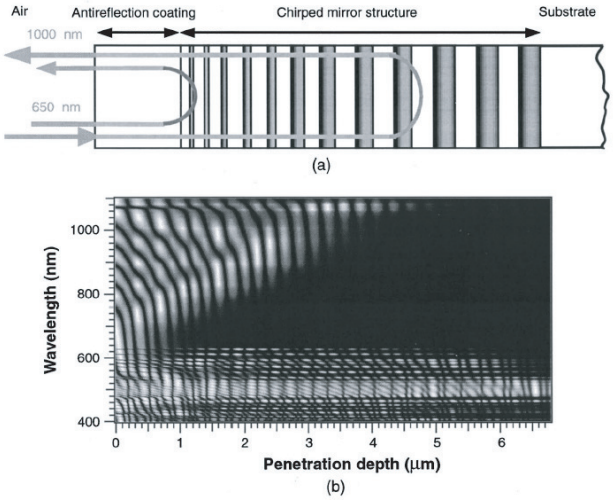

Figure 5.8: Schematic diagram showing: (a) the principle of operation of a chirped mirror; (b) the calculated standing wave pattern in a chirped mirror as a function of wavelength. Note how for this mirror the penetration depth increases with wavelength, giving negative GDD.

increase the diameter of the laser beam. However, this approach quickly requires large diameter laser rods and optics, which are both expensive (if they are available) and unwieldy.

The solution to this problem is **chirped pulse amplification (CPA)**. In this approach the low energy pulses to be amplified to high energy, high peak power pulses in three stages:

1. The pulses are stretched in time by passing them through a pulse stretcher with a large GDD. Stretch factors of order 10^3 to 10^5 are common.

2. The stretched pulses are then amplified in a chain of one or more amplifiers. Since the duration of the stretched pulses is many orders of magnitude longer than the original pulses, they may be amplified to higher energies than in the absence of stretching by a factor approximately equal to the stretch factor.

3. The amplified pulse is then re-compressed to (approximately) the duration of the original pulse.

The CPA process is illustrated schematically in Figure 5.9

The pulse stretcher usually comprises the grating pair illustrated in Figure 5.7(c), set to produce positive dispersion. One reason for this choice is that the material of the following amplifier chain will tend to introduce more positive dispersion, and therefore there is no danger of partially re-compressing the laser pulses as they are amplified, which could lead to higher than expected peak pulse intensities in the laser amplifiers.

5.3.1 Regenerative and multipass amplifiers

In a CPA laser there will typically be several stages of amplification. Immediately after the pulse has been stretched, there will be one or more pre-amplifiers in which most of the gain occurs (perhaps a factor of 10^7). After the pre-amplifiers the energy of the stretched pulses will be sufficiently high that they can extract energy efficiently from one or more power amplifiers. Most of the energy in the final pulse is added by the power amplifiers.

There are two main classes of amplifier used, as illustrated schematically in Fig. 5.10.

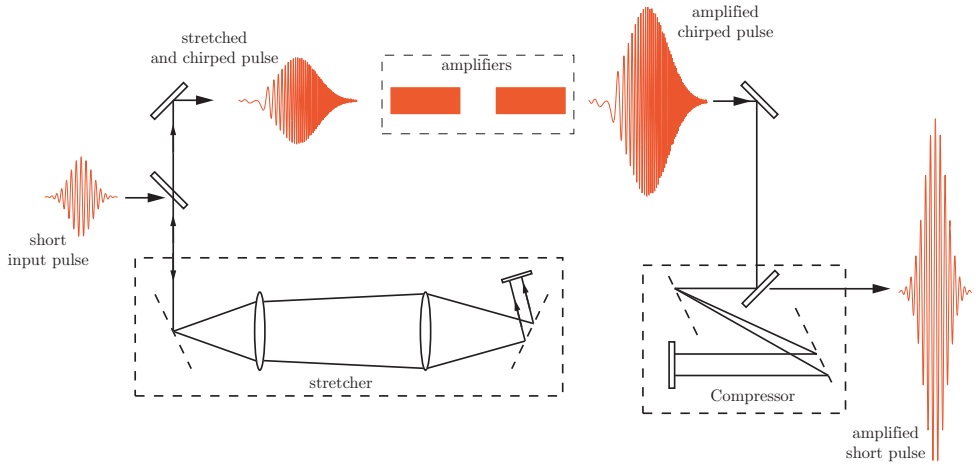

Figure 5.9: Schematic diagram of chirped pulse amplification.

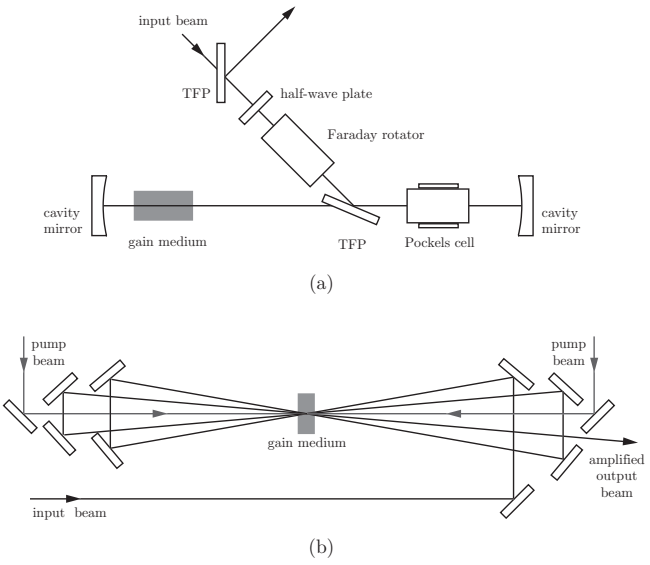

Figure 5.10: Schematic diagram of (a) a regenerative and (b) a multipass amplifier.

Regenerative amplifiers

In a regenerative amplifier the amplification takes place in an optical cavity. Laser pulses are injected into the cavity by applying a voltage pulse to a Pockels cell as shown schematically in Fig. 5.10(a). In the absence of any voltage on the Pockels cell, the incident laser pulses will, after reflection from the cavity end mirror be reflected out of the cavity by the polarizer. However, if a quarter-wave voltage is applied to the Pockels cell, the pulses will return to the intra-cavity polarizer with the orthogonal polarization and hence will be trapped in the cavity and be amplified. After amplification the pulses are switched out by applying a second quarter-wave pulse to the Pockels cell[10].

Since the gain medium is located within an optical cavity, the pumping must be at a level which prevents the single-pass gain from becoming too high, or lasing — or at least strong amplification of spontaneous emission — will occur even in the absence of an injected pulse. As such, in order to extract energy from the gain medium it is necessary for the pulse to undergo many cavity round trips, perhaps of order 20.

The advantages of regenerative amplifiers are that the optical cavity imposes a good transverse mode structure on the amplified pulses. As such, the beam quality should be good. There are several disadvantages, however. Leakage through the polarizers and Pockels cell can give rise to a series of lower-energy pulses propagating ahead of the main pulse. Prepulses of this type can be problematic in many applications. Further, the many round-trips required to extract the energy stored in the population inversion introduces a long length of material into the path of the pulse (100s of millimetres). The consequent linear and non-linear phase structure introduced into the pulse can be difficult to compensate for. Notwithstanding these difficulties, regenerative amplifiers have been used to generate TW pulses with pulse durations as short as 30 fs.

5.3.2 Multipass amplifiers

In a multipass amplifier the pulse is passed through the gain medium several times by reflection from a series of mirrors – sometimes referred to as **angular multiplexing**. Since the gain medium is not located within a cavity the single-pass gain may be much higher without causing too much **amplified spontaneous emission (ASE)**. Multipass amplifiers therefore have the advantage of introducing less material into the path of the pulse, as well as not suffering from prepulses. The disadvantages are that the overlap between the pump laser beam and the amplified laser pulse is not as good as in a regenerative amplifier, and consequently the efficiency will not be as high as in a regenerative amplifier. A typical efficiency for a multipass pre-amplifier is 15%; although multipass power amplifiers may reach 30%, more or less equal to that achieved in a regenerative amplifier.

5.3.3 Low-energy, TW laser systems

A good example of a low-energy TW laser is the system used in the Clarendon Laboratory, which has the following key components:

1. A Ti:sapphire oscillator pumped by a frequency-doubled CW Nd:YLF laser. The oscillator generates a stream of modelocked pulses with a pulse repetition rate of 86 MHz. The duration and energy of each pulse are approximately 35 fs and 10 nJ respectively.

[10]The Faraday rotator is included in the arrangement in order to prevent pulses being reflected back into earlier parts of the laser chain (such as the laser oscillator), since such reflections can cause damage or instabilities. It's inclusion complicates a description of the operation of the regenerative amplifier somewhat, although the essential features are included in the discussion above. Let us suppose that the input pulses are vertically polarized. The half-wave plate is oriented at 45° to the vertical so as to rotate the plane of polarization to the horizontal plane. On on passing through the Faraday rotator the plane of polarization is rotated by (say) +45°. In the absence of any voltage on the Pockels cell, the pulses will be rejected by the cavity and their polarization further rotated by the Faraday rotator in the *same sense*, i.e. another +45°, to give vertically-polarized light. The half-wave plate then flips the polarization to the horizontal plane, and consequently the pulses are reflected by the external polarizer. Trapping laser pulses within the cavity operates as discussed above, i.e. applying a voltage to the Pockels cell rotates the polarization of the laser pulses so that they are transmitted by the intra-cavity polarizer. The pulses are rejected from the cavity by applying a second pulse to the Pockels cell, whereupon they are reflected by the external polarizer in the manner discussed in this note.

2. A pulse stretcher which stretches the pulses to a duration of approximately 300 ps.

3. A subset of the stretched pulses are selected by a Pockels cell and injected into a **regenerative amplifier** operated at a repetition rate of 1 kHz. The regenerative amplifier comprises a 5 mm diameter Ti:sapphire rod pumped by pulses from a diode-pumped, frequency-doubled Q-switched Nd:YAG laser operating at a pulse repetition rate of 1 kHz The regenerative amplifier increases the energy of the pulses to 0.7 mJ.

4. A subset of the pulses from the regenerative amplifier are injected into a **multipass amplifier** comprising a 20 mm diameter Ti:sapphire rod pumped by a flashlamp-pumped, frequency-doubled, Q-switched Nd:YAG laser which delivers approximately 1 J of green pump light per pulse. This amplifier increases the energy of the pulses to 200 mJ.

5. Finally the amplified pulses are compressed by a double-passed grating pair to a duration of approximately 40 fs. Losses in the compressor are dominated by the four reflections from the diffraction gratings, and are such that the compressed laser pulses have an energy of approximately 120 mJ.

6. The compressed pulses therefore have a peak power of 3 TW and a duration quite close to those from the oscillator.

5.3.4 High-energy, PW laser system

In order to reach very high pulse energies (100s of Joules) it is necessary to increase the beam diameter to 100 mm or more. At present it is not possible to grow large-diameter crystals of Ti:sapphire with sufficient optical quality for this purpose. As a consequence, high-energy systems use as the active medium Nd:Glass since the technology for producing optical-quality rods and disks of this material is well developed.

The laser transition in Nd:Glass is inhomogeneously broadened with a large linewidth of approximately 50 THz, and hence in principle can produce modelocked laser pulses with durations as short as approximately 100 fs. In practice, pulses of duration 0.5 - 1 ps can be generated. Modern high-energy 'glass' laser systems amplify pulses from a Ti:sapphire oscillator and pre-amplifier system, the Ti:sapphire being tuned in wavelength to match the peak of the gain of the Nd:Glass laser. Typically a series of multipass rod amplifiers is used for amplification to pulse energies of several tens of Joules. Amplification to higher energies requires the use of larger beam diameters, and hence large-area discs of Nd:Glass are used.

A good example of a PW class laser is the Vulcan laser at the Rutherford Appleton Laboratory, just down the road near Didcot. This was the first laser in Europe to reach the PW level. The laser has the following features:

- Pulses with an energy of 5 nJ, a duration of 120 fs, and a wavelength of 1055 nm are generated by a modelocked Ti:Sapphire oscillator.

- These pulses are stretched by a factor of 4×10^4 to a duration of 5 ns, and then amplified to an energy of order 10 mJ using a new technique known as optical parametric chirped pulse amplification (OPCPA). In this approach the gain medium is replaced by a non-linear, optical parametric amplifier and the pulses are amplified by conversion of energy from the pump laser beam to the wavelength of the pulses. This method avoids the difficulties of gain narrowing and reduces the pedestal level normally associated with regenerative amplifiers.

- The pre-amplified pulses are then introduced into a chain of flashlamp-pumped rod and disk Nd:Glass amplifiers, which amplifies the pulse energy to 670 J.

- Two gold-coated holographic diffraction gratings (940 mm diameter!) with a groove spacing of 480 lines mm^{-1} re-compress the pulse to 650 fs; somewhat longer than the original pulses, but not bad considering that they have been amplified by a factor of 10^{11}!

- After losses caused by reflection at the compressor gratings and beam-steering optics, the energy of the pulses available to the target chamber is approximately 500 J, i.e. a peak power of 0.8 PW. This can be focused to a spot size of order 7 μm diameter, corresponding to a peak focused intensity of order $10^{21}\,\mathrm{W\,cm^{-2}}$. Note that at this intensity the radiation pressure is 11 orders of magnitude greater than atmospheric pressure!

5.4 Applications of ultrafast lasers

At low pulse energies (fraction of a millijoule), ultrafast lasers find a myriad of applications in 'pump-probe' experiments in which an event, such as a chemical reaction, is triggered by a short laser pulse, and the subsequent evolution is studied using a second, short probe laser pulse. The ability to control the phase structure of short laser pulses is opening up the field of 'coherent control' in which, for example, the products of photo-initiated chemical reactions are selected by controlling the precise variation of the electric field within the laser pulse.

At somewhat higher pulse energies (several millijoules), focusing the laser pulse to a beam diameter of a few tens of microns results in a peak laser intensity of $10^{15}\,\mathrm{W\,cm^{-2}}$ and higher. At such intensities the electric field of the electromagnetic wave is comparable, or stronger, than the Coulomb field binding the valence electrons in an atom. A number of applications arise:

- High-harmonic generation. Generation of very high-order (of order 100) harmonics of the driving laser wavelength to produce beams of coherent radiation with wavelengths in the vacuum ultraviolet to soft x-ray spectral region. High-harmonic generation requires relatively low driving intensities of order $10^{15}\,\mathrm{W\,cm^{-2}}$.
- Driving 'table-top' soft x-ray lasers. Typical intensities required for this are 10^{16} - 10^{17} $\mathrm{W\,cm^{-2}}$.
- Plasma accelerators. At intensities of order $10^{18}\,\mathrm{W\,cm^{-2}}$, a laser pulse propagating through a plasma will drive a longitudinal plasma wave that trails the laser pulse in much the same way that a wake follows a boat travelling across a lake. The longitudinal electric fields within the plasma wave can be at least three orders of magnitude greater than those used in conventional particle accelerators, such as those at CERN. For example, recently experiments by groups in Oxford and Lawrence Berkeley National Laboratory demonstrated acceleration of electrons to 1 GeV energy (c.f. the electron rest mass of 0.5 MeV) in an laser-driven plasma accelerator only 33 mm long.

At somewhat higher pulse energies (several Joules), a variety of exotic phenomena become possible:

- Laser-driven fusion. Of particular interest is so-called fast-ignitor inertial confinement fusion, in which a pre-heated core of material is rapidly heated by an intense ultrashort laser pulse to initiate fusion.
- Generation of ultra-intense magnetic fields (up to 10^5 T) during the interaction of intense laser pulses with a solid target.
- Generation of intense, short-pulse proton or heavy-ion beams.
- 'Laboratory Astrophysics'. The extreme temperatures, pressures, and magnetic fields that can be created when a ultra-high intensity laser interacts with matter allows us to recreate (briefly) the conditions found in several types of astrophysical objects.

Lecture 6

Semiconductor & Fibre lasers

6.1 Review of semiconductor physics

As the constituent atoms of a semiconductor crystal are brought together from infinity, the wave functions of the valence electrons begin to overlap. At inter-atomic separations characteristic of the solid state, the wave functions are sufficiently overlapped that the valence electrons are delocalized and can no longer be associated with a particular atom.

The electron wave functions can then be written as **Bloch wave functions**:

$$\psi(\boldsymbol{r}) = u_k(\boldsymbol{r}) \exp(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r}), \tag{6.1}$$

where $u_k(\boldsymbol{r})$ has the periodicity of the crystal lattice.

The energy of the electron is a function of the wave vector $\boldsymbol{k}$. For a *free* electron the energy and momentum of the electron would be given by,

$$\boldsymbol{p} = \hbar\boldsymbol{k} \tag{6.2}$$

$$E = \frac{\hbar^2 k^2}{2m_\mathrm{e}}, \tag{6.3}$$

where m_e is the rest mass of a free electron. However, for electrons moving in the periodic potential of the crystal lattice this relation no longer applies. Deviation from the parabolic dependence of energy on k becomes particularly strong when the wave vector k approaches one of the Brillouin zones[1], which is to be expected since this is just the condition for strong Bragg reflection from the crystal planes. The distortion of the electron energy from the free space value leads to some energies being forbidden and the formation of allowed energy bands with energies $E_n(\boldsymbol{k})$ which are functions of $\boldsymbol{k}$ and an index $n = 1, 2, 3, \ldots$. The formation of the band structure is illustrated schematically in Figure 6.1.

We note that the Bloch wave functions are not momentum eigenstatates[2], and hence $\hbar\boldsymbol{k}$ is not the momentum of the electron. However, the quantity $\hbar\boldsymbol{k}$ does play a role akin to momentum in conservation laws describing the interaction of electrons with fields or phonons. For this reason it is known as the **crystal momentum**.

The number of electron states per band is equal to twice the number of unit cells in the crystal. A semiconductor contains an even number of electrons per unit cell, and consequently at absolute zero all bands are

[1]The first Brillouin zone is defined as the Wigner-Seitz primitive cell of the reciprocal lattice; the Wigner-Seitz primitive cell about any lattice point is the region of space which is closer to that lattice point than to any other.

[2]Fundamentally this relates to the fact that the Hamiltonian is not invariant under arbitrary translations.

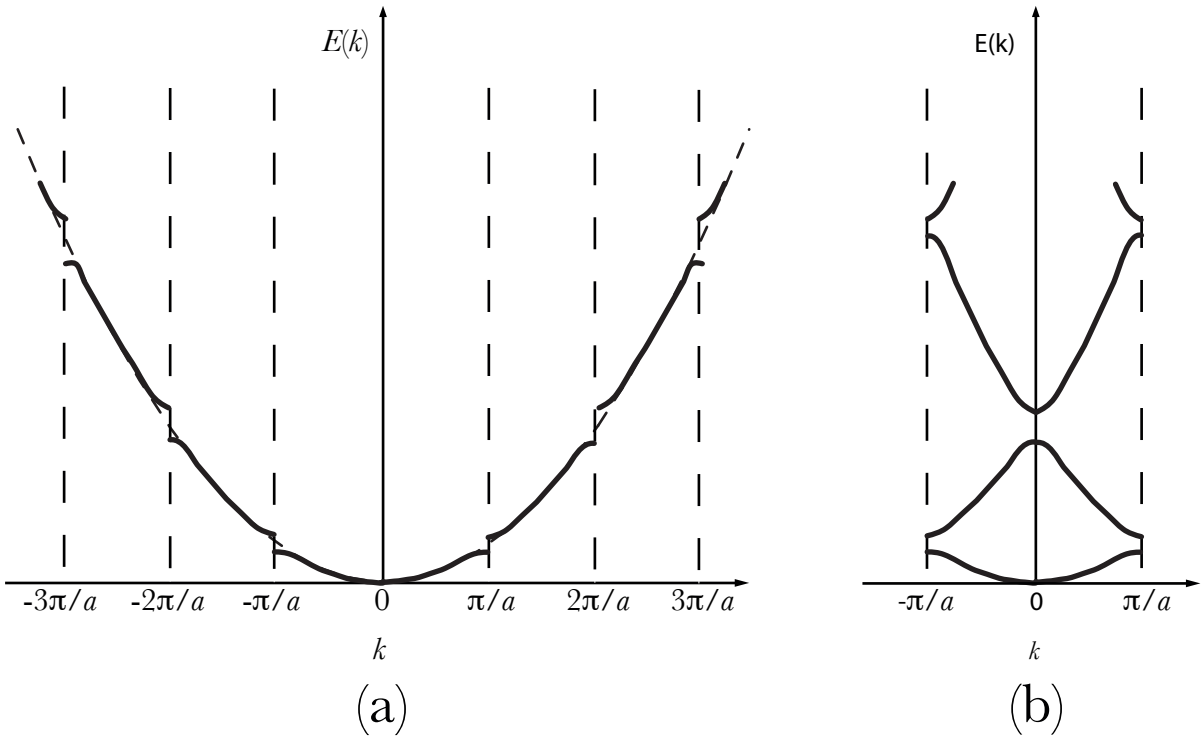

Figure 6.1: Schematic diagram of the formation of band structure in a one-dimensional crystal lattice of lattice spacing a. In (a) is shown the parabolic dependence of the energy $E(k)$ on the wave vector k for a free electron (dashed curve). This dependence is strongly distorted for values of k close to a Brillouin zone, leading to the formation of bands of allowed energies separated by energy gaps. In (b) the bands are shown in the **reduced zone scheme** in which reciprocal lattice vectors are added to (or subtracted from) k so as to bring everything into the first Brillouin zone. This last picture is usually the one used when discussing semiconductors.

either completely full or completely empty. As such, at absolute zero a semiconductor has zero conductivity. What differentiates a semiconductor from an insulator is that the energy gap between the top of the highest full energy band, the **valence band**, and the next highest (empty) band, the **conduction band**, is not too large compared to $k_{\mathrm{B}}T$. Typical values of this **band gap** energy are 0.5 - 3.0 eV, compared to $k_{\mathrm{B}}T \approx 0.025$ eV. Thermal excitation of electrons from the valence band to the conduction band leads to a finite electrical conductivity which is between that of insulators and conductors and increases strongly with temperature.

Effective mass

It may be shown that under a force $\boldsymbol{F}$ an electron in one of the energy bands moves according to,

$$\boldsymbol{F} = \hbar^2 \left(\frac{\partial^2 E_n}{\partial k^2} \right)^{-1} \frac{d\boldsymbol{v}_g}{dt}, \tag{6.4}$$

where v_g is the group velocity of the electron. As such, it is useful to define the **effective mass**:

$$m^* = \hbar^2 \left(\frac{\partial^2 E_n}{\partial k^2} \right)^{-1}. \tag{6.5}$$

Hence the effective mass of the electron depends on the curvature of the band energy $E(k)$. Note that for a free electron, the definition above gives $m^* = m_\mathrm{e}$.

Holes

Figure 6.2 shows a semiconductor in which some electrons have been excited from the valence band to the conduction band to give two partially-filled bands. In describing the partially-filled valence band, it is often easier to focus attention on the relatively few empty electron states than on the many filled states. Empty electron states are known as **holes**, and have the following properties:

- $\boldsymbol{k}_\mathrm{h} = -\boldsymbol{k}_\mathrm{e}$, where $\boldsymbol{k}_\mathrm{h}$ is the wave vector of the hole, and $\boldsymbol{k}_\mathrm{e}$ that of the missing electron.
- $E_\mathrm{h} = -E_\mathrm{e}$, where E_h is the energy of the hole, and E_e is the energy that an electron would have if it occupied the empty state.
- $\boldsymbol{v}_\mathrm{h} = \boldsymbol{v}_\mathrm{e}$
- $m^*_\mathrm{h} = -m^*_\mathrm{e}$. This follows from the first and second points above.
- The hole moves under the influence of external electromagnetic fields as if it had a charge of $+e$.

Hence the holes behave as particles of positive charge, with an energy and mass equal to minus that of the missing electron. Notice that for the partially-filled valence band, the curvature of the electron energy in the region near the top of the band is negative, and hence their effective mass would be negative. In contrast, the holes appear to have a positive effective mass which is another reason for dealing with the hole states rather than the electron states.

6.1.1 Distribution over energy levels

The probability of a level with an energy between E and $E + dE$ being occupied by an electron is given by the Fermi-Dirac distribution:

$$f(E)dE = \frac{1}{\exp([E - E_\mathrm{F}]/k_\mathrm{B}T) + 1}, \tag{6.6}$$

where E_F is the **Fermi energy**. From eqn (6.6) we see that the Fermi energy is that energy for which a level has a probability of 1/2 of being occupied. Alternatively, as shown in Figure 6.3(a), the Fermi energy gives the energy to which the band is filled at $T = 0$.

Imagine that we transfer N_e electrons per unit volume from the valence band into the conduction band. After a few picoseconds the electrons will equilibrate within the conduction band to form a local thermal equilibrium at temperature described by T and a **quasi-Fermi level** E_Fc. Similarly the electrons in the valence band will equilibrate to a thermal distribution described by a quasi-Fermi level E_Fv. Hence the probabilities of a level of energy E being occupied are :

$$f_\mathrm{c}(E) = \frac{1}{\exp([E - E_\mathrm{Fc}]/k_\mathrm{B}T) + 1} \tag{6.7}$$

$$f_\mathrm{v}(E) = \frac{1}{\exp([E - E_\mathrm{Fv}]/k_\mathrm{B}T) + 1}, \tag{6.8}$$

for the conduction and valence band respectively.

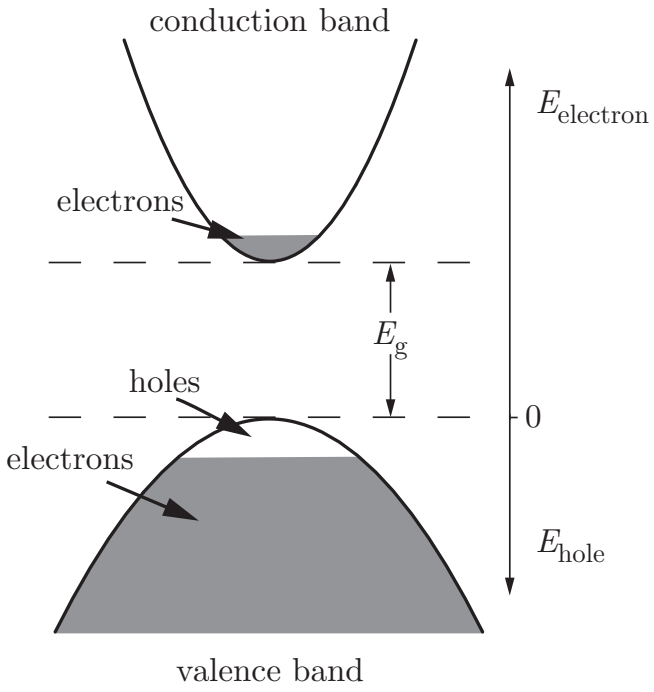

Figure 6.2: Partially-filled valence bands following promotion of electrons from the conduction to the valence band.

6.1.2 Doped semiconductors

In an ultra-pure sample of a semiconductor compound such as GaAs the number of electrons excited across the energy gap up into the conduction band at room temperature is very small. As indicated in Fig. 6.4(a) for such an ultra-pure semiconductor the Fermi energy would lie halfway up the energy gap between the top of the valence band and the bottom of the conduction band, corresponding to the energy for which the probability of occupation is $\frac{1}{2}$.

Higher concentrations of electrons in the conduction band are required for most practical applications. This is achieved by adding carefully controlled concentrations of particular doping species to the bulk semiconductor compound. For example, for the case of GaAs, adding small proportions of Si atoms or Te atoms[3] we introduce a new energy level corresponding to the impurity species which lies just below the conduction band of the host material, as indicated schematically in Figure 6.4. Atoms of these particular added species (Si or Te) have one more electrons in their outer shells than the atoms they replace and can therefore act as donor species by providing a source of electrons which go into the conduction band, because the gap between the impurity level and the conduction band of the host material is small and the impurity atoms are easily ionized at room temperature. This type of material, with an excess of electrons in the conduction band is called **n-type**, and at light doping levels provides a number density of electrons in the conduction band, which although much larger than in the case of the intrinsic undoped material, nevertheless depends strongly on temperature and tends to zero at low temperatures. Under conditions of light doping the Fermi energy lies between the impurity level and the bottom of the conduction band.

[3]Si belongs to group IV and substitutes for atoms originally on Ga sites. Te belongs to group VI and substitutes for atoms originally on As sites. In the case of n-type GaAs, it is Si which is used most commonly.

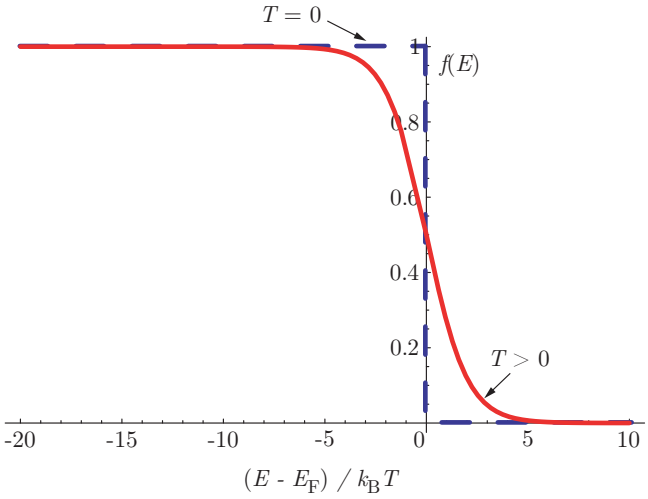

Figure 6.3: The Fermi-Dirac distribution for $T = 0$ (dashed line) and $T > 0$ (solid line).

In contrast if the same n-type material is **heavily** doped the impurity level merges into the conduction band, and the Fermi level moves up to sit close to the top of the occupied states in the conduction band as indicated in Figure 6.4(c).

In an analogous way, by doping GaAs with electron acceptor atoms such as Be, Zn, or C, **p-type** materials with excess holes in the valence band can be produced and at heavy doping levels the Fermi level of such materials moves down into the valence band[4].

Heavily doped semiconductor materials are known as **degenerate** semiconductors. Because they have partially filled bands (like metals) they remain electrically conducting even at very low temperatures.

6.2 Radiative transitions in semiconductors

Suppose that an incident harmonic wave of angular frequency ω causes an electron to make a transition from an upper level of energy E_2 in the conduction band to a level in the valence band of lower energy E_1. If we write the electric field of the incident wave in the form,

$$\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{E_0} \exp\left[\mathrm{i}(\boldsymbol{k}_{\mathrm{opt}} \cdot \boldsymbol{r} - \omega t)\right], \tag{6.9}$$

the perturbation takes the form,

$$H' = e\boldsymbol{E} \cdot \boldsymbol{r}. \tag{6.10}$$

The rate of transitions induced by the perturbation is given by Fermi's golden rule:

$$W_{21} = \frac{\pi^2}{\hbar^2} |H_{21}|^2 \, \delta(\omega - \omega_{21}), \tag{6.11}$$

[4]Be belongs to group IIA and Zn to group IIB, and these atoms substitute for Ga to give rise to excess holes. C belongs to group IVB and substitutes for As.

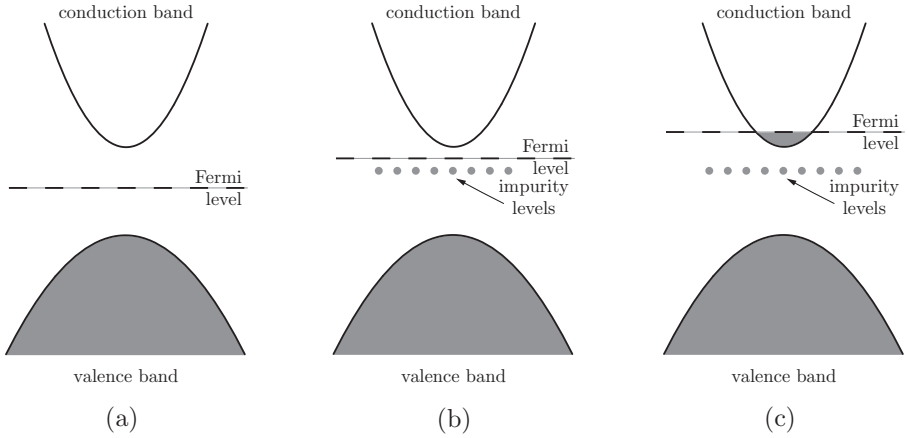

Figure 6.4: Occupancy of conduction and valence bands and position of the Fermi level for: (a) ultra-pure semiconductor; (b) after the introduction of a small density of an electron donor impurity species; (c) after the introduction of a large density of electron donor species.

where,

$$\hbar\omega_{21} = E_2 - E_1, \tag{6.12}$$

and,

$$H_{21} = \int \psi_2^*(\boldsymbol{r}) \left[e\boldsymbol{r} \cdot \boldsymbol{E_0} \exp(\mathrm{i}\boldsymbol{k}_{\mathrm{opt}} \cdot \boldsymbol{r})\right] \psi_1(\boldsymbol{r}) d\tau \tag{6.13}$$

$$= \int u_2^*(\boldsymbol{r}) \left\{e\boldsymbol{r} \cdot \boldsymbol{E_0} \exp\left[\mathrm{i}(\boldsymbol{k}_{\mathrm{opt}} + \boldsymbol{k}_1 - \boldsymbol{k}_2) \cdot \boldsymbol{r}\right]\right\} u_1(\boldsymbol{r}) d\tau. \tag{6.14}$$

We see from eqn (6.14) that the integrand will oscillate rapidly during the integration, and hence be zero, unless

$$\hbar\boldsymbol{k}_2 = \hbar\boldsymbol{k}_1 + \hbar\boldsymbol{k}_{\mathrm{opt}}. \tag{6.15}$$

This, then, is a selection rule for optical transitions between electron levels in a semiconductor; it resembles a statement of conservation of momentum, except that the quantities $\hbar\boldsymbol{k}_{2,1}$ are *crystal* momenta. In the optical region $|k_{\mathrm{opt}}| \ll |k_{1,2}|$ and hence the selection rule becomes,

$$\hbar\boldsymbol{k}_2 \approx \hbar\boldsymbol{k}_1, \tag{6.16}$$

and hence $\boldsymbol{k}$ is unchanged in the transition — the transitions are 'vertical' in our diagrams.

We also note that since the perturbation does not operate on the electron spin s, we also have the selection rule,

$$\Delta s = 0. \tag{6.17}$$

6.2.1 Joint density of states

We now consider a transition between two levels of energy $E_2 = E_c(k_2)$ and $E_1 = E_v(k_1)$ of a given value of $k = k_1 \approx k_2$. In any particular band the density of states with wavevectors between k and $k + dk$ is given by the well known result:

$$g(k)dk = \frac{k^2}{\pi^2}dk. \tag{6.18}$$

We wish to find the density of transitions with a given energy difference $\hbar\omega$. To this end we note that within the parabolic approximation we may write the energies of the upper and lower electron states as,

$$E_2 = E_g + \frac{\hbar^2 k^2}{2m_c^*} \tag{6.19}$$

$$E_1 = -\frac{\hbar^2 k^2}{2m_v^*}, \tag{6.20}$$

where m_v^* is the mass of an electron in the valence band, i.e. $m_v^* = -m_h^*$ such that $m_v^* > 0$.

Hence the energy of the transition is written in terms of k as,

$$\hbar\omega = E_2 - E_1 = E_g + \frac{\hbar^2 k^2}{2}\left[\frac{1}{m_c^*} + \frac{1}{m_v^*}\right] = E_g + \frac{\hbar^2 k^2}{2\mu}, \tag{6.21}$$

where the **reduced mass** is given by,

$$\frac{1}{\mu} = \frac{1}{m_c^*} + \frac{1}{m_v^*}. \tag{6.22}$$

Now that we have established a relationship between the energy of the transition and the wavevector of the electrons involved in the transition we may now use eqn (6.21) to re-write the density of states in terms of the angular frequency of the transition:

$$g(k)dk = \frac{k^2}{\pi^2}dk = g_j(\omega)d\omega \tag{6.23}$$

$$\Rightarrow g_j(\omega)d\omega = \frac{\hbar}{2\pi^2}\left(\frac{2\mu}{\hbar^2}\right)^{3/2}(\hbar\omega - E_g)^{1/2}d\omega, \tag{6.24}$$

where $g_j(\omega)$ gives the density of transitions of angular frequency ω, and is known as the **joint density of states**.

6.2.2 Density of possible transitions

The joint density of states gives the density of transitions of angular frequency ω, but *not all of these transitions can occur* since that requires that the initial state is occupied by an electron and the final state is empty.

For example, the density of possible stimulated emission transitions with frequencies in the interval ω to $\omega + d\omega$ is given by,

$$g_j(\omega)f_c(E_2)[1 - f_v(E_1)]d\omega,$$

i.e. the density of transitions in this frequency interval multiplied by the probability that the upper level is occupied and the lower level is unoccupied. Similarly the density of possible absorption transitions with frequencies in the interval ω to $\omega + \mathrm{d}\omega$ is given by,

$$g_{\mathrm{j}}(\omega) f_{\mathrm{v}}(E_1)[1 - f_{\mathrm{c}}(E_2)]\mathrm{d}\omega.$$

The density of net stimulated transitions with frequencies in the interval ω to $\omega + \mathrm{d}\omega$ is given by the difference of these two expressions:

$$g(\omega)[f_{\mathrm{c}}(E_2) - f_{\mathrm{v}}(E_1)]\mathrm{d}\omega. \tag{6.25}$$

6.2.3 The gain coefficient

We are now in a position to write down the net rate of stimulated emission per unit volume with frequencies between ω and $\omega + d\omega$. From eqns (6.11) and (6.25) we have:

$$R_{SE} = W_{21} g_j(\omega) d\omega [f_{\mathrm{c}}(E_2) - f_{\mathrm{v}}(E_1)]. \tag{6.26}$$

Note that for atoms we would write the above in terms of the population inversion density N^* and optical gain cross-section $\sigma_{21}(\omega)$ as,

$$R_{SE} = N^* \sigma_{21}(\omega) \frac{I}{\hbar\omega} = \alpha(\omega) \frac{I}{\hbar\omega}, \tag{6.27}$$

where $\alpha(\omega)$ is the gain coefficient of the transition. Hence we deduce that the gain coefficient is given by $\alpha(\omega) = \hbar\omega R_{SE}/I$. From eqns (6.26) and (6.14) we find, after tidying and eliminating the intensity using $I = (1/2)\epsilon_0 c n E_0^2$, where n is the refractive index of the medium,

$$\alpha(\omega) = \frac{\omega |D_{21}|^2}{n\epsilon_0 c \hbar^3} (2\mu)^{3/2} (\hbar\omega - E_{\mathrm{g}})^{1/2} [f_{\mathrm{c}}(E_2) - f_{\mathrm{v}}(E_1)]. \tag{6.28}$$

6.2.4 Condition for gain

Notice that the term $[f_{\mathrm{c}}(E_2) - f_{\mathrm{v}}(E_1)]$ plays the role of the population inversion density which appears in the expression for the gain coefficient in atomic or molecular laser transitions. Clearly, in order for there to be positive gain we must have,

$$f_{\mathrm{c}}(E_2) > f_{\mathrm{v}}(E_1) \tag{6.29}$$

$$\Rightarrow \frac{1}{\exp([E_2 - E_{\mathrm{Fc}}])/k_{\mathrm{B}}T + 1} > \frac{1}{\exp([E_1 - E_{\mathrm{Fv}}])/k_{\mathrm{B}}T + 1} \tag{6.30}$$

$$\Rightarrow E_2 - E_1 < E_{\mathrm{Fc}} - E_{\mathrm{Fv}}. \tag{6.31}$$

The left-hand side is simply the photon energy. The photon energy must also exceed the gap energy, and hence we can write the condition for gain as,

$$E_{\mathrm{g}} < \hbar\omega < E_{\mathrm{Fc}} - E_{\mathrm{Fv}}. \tag{6.32}$$

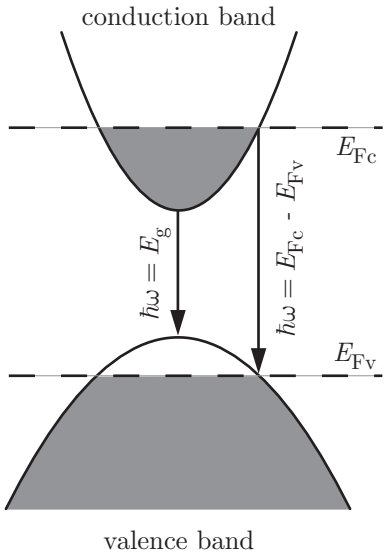

Figure 6.5: Illustrating the condition for gain for the case when $T = 0$.

In order for this condition to be met at any frequency, we must therefore have,

$$E_{\mathrm{Fc}} - E_{\mathrm{Fv}} > E_{\mathrm{g}} \quad \text{Condition for gain.} \tag{6.33}$$

This condition — originally derived in 1961 by Bernard and Duraffourg — holds for all temperatures T, but is particularly easy to understand for the case $T = 0$ as illustrated schematically in Figure 6.5. At absolute zero the gain corresponds to the upper level being occupied whilst the lower level is empty. At finite temperatures the condition for gain is that the probability of a level of energy E_2 being occupied is greater than the probability of a level of energy E_1 being occupied.

6.2.5 Generating gain: the injection density

Gain on optical transitions in semiconductor lasers is generated by raising N electrons per unit volume from the valence band to the conduction band. The density of electrons moved is known as the **injection density**. Since the density of electrons in the conduction band is equal to the density of holes in the valence band we may write,

$$N = \int_0^\infty g_{\mathrm{c}}(E_{\mathrm{c}}) f_{\mathrm{c}}(E_{\mathrm{c}}) dE_{\mathrm{c}} = \int_0^\infty g_{\mathrm{v}}(E_{\mathrm{v}})[1 - f_{\mathrm{v}}(E_{\mathrm{v}})]\mathrm{d}E_{\mathrm{v}}, \tag{6.34}$$

where $g_{\mathrm{c}}(E)$ and $g_{\mathrm{v}}(E)$ are the density of states of energy E in the conduction band and valence band respectively. The density of states may be written in terms of energy by using eqns (6.18), (6.19) and (6.20).

The relations above then determine the positions of the quasi Fermi levels in the conduction and valence bands, and hence the frequency width over which gain can be achieved.

For a given density N, the transcendental eqns (6.34) may be solved to find the quasi-Fermi levels. Figure 6.6 shows the result of one such calculation for GaAs. Notice that for low injection densities the quasi-Fermi levels

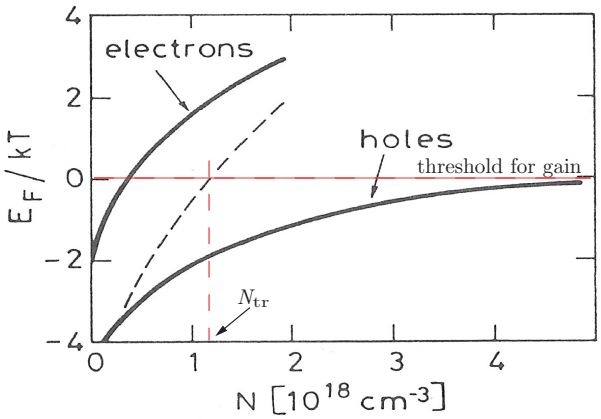

Figure 6.6: Calculation of the variation of the quasi-Fermi levels in the conduction and valence bands of GaAs as a function of the injected electron density N. For the conduction band the energy of the quasi-Fermi level above the bottom of the band is plotted; for the valence band the energy of the holes (which have minus the energy of the corresponding electron state) is plotted. The transparency condition therefore corresponds to the sum of these terms equalling zero, which occurs at an injection density $N_{\rm tr}$ as shown.

lie outside their respective bands. As the injection density is increased the Fermi level in the conduction band rises more quickly than the Fermi level in the valence band drops. Hence the gain becomes positive.

Once the quasi-Fermi levels have been calculated as a function of the injection density, the gain coefficient may be found from eqn (6.28). Figure 6.7 shows the results of such a calculation. Two points are worth noting. The gain coefficient is typically large, some tens of cm^{-1} and hence laser oscillation is possible with very short gain lengths. Second, the bandwidth of the gain is broad and increases as the injection density of electrons is increased. For the example shown, the gain bandwidth is of order 10 THz, corresponding to a 30 nm bandwidth at the operating wavelength of 840 nm.

The gain first becomes positive when eqn (6.33) is met, corresponding to the injection density exceeding a threshold density, known as the **transparency density**, $N_{\rm tr}$. The condition for gain can therefore be written as $N > N_{\rm tr}$, which is analagous to $N^* > 0$ for gain on an atomic transition.

6.3 Homojunction diode lasers

In a homojunction laser, gain is achieved in an intrinsic region of semiconductor located between layers of degenerate p-type and n-type semiconductor, a so-called **p-i-n** junction.

Figure 6.8(a) shows schematically the positions of the conduction bands and Fermi levels at the (rather theoretical!) moment when the three layers are first contacted. Electrons from the n-type material can drift across the intrinsic layer and recombine with holes in the p-type layer; in a similar way, holes drifting across the intrinsic layer can recombine with electrons in the n-type layer. This process sets up an electric field from the n-type to the p-type layer, and equilibrium is reached when the rate of drift of electrons across the intrinsic layer is balanced by the rate at which they are returned by the electric field. The region in which the electric field exists is virtually devoid of charge carriers since they are quickly removed by the field; it is known as the **depletion layer**.

The electric field raises the potential energy of electrons in the p-type layer, and lowers it in the n-type layer,

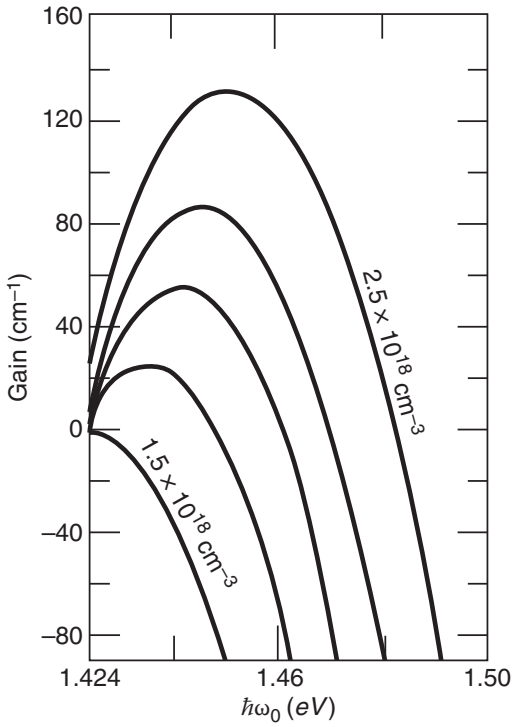

Figure 6.7: Calculated gain coefficient as a function of photon energy for various values of N (in units of 10^{18} cm^{-3}) for GaAs at $T = 300k$.

to give the equilibrium distribution shown in 6.8(b). Notice that under these equilibrium conditions the Fermi energy lies at the same level in all three layers.

If a forward-bias voltage is now applied to the structure, the electric field across the device will be reduced, leading to a distribution of energy levels as shown in 6.8(c). Now there exists a region in the intrinsic layer in which there are electrons in the conduction band and holes in the valence band; gain is possible provided the quasi-Fermi levels obey eqn (6.33), or, equivalently, the density of electrons in the conduction band exceeds the transparency density.

To relate the density of electrons in the conduction band to the diode current, we note that most of the current is carried by the electrons (owing to their greater mobility). If the time for electron-hole recombination in the intrinsic layer is $\tau_{\rm rec}$, then in the steady-state, and in the absence of stimulated emission, the rate of flow of electrons into the intrinsic layer must be balanced by the rate of recombination. Hence,

$$\eta_{\rm i}\frac{I}{{\rm e}V_{\rm g}} = \frac{N}{\tau_{\rm rec}}, \tag{6.35}$$

where $V_{\rm g}$ is the volume of the active region, and the **internal quantum efficiency** $\eta_{\rm i}$ gives the fraction of carriers that recombine radiatively in the active layer — other carriers undergo non-radiative recombination at the junction boundary, or flow round the active region. Hence the threshold current may be written as,

$$I_{\rm thresh} = \frac{1}{\eta}\frac{N_{\rm tr}{\rm e}V_{\rm g}}{\tau_{\rm rec}}. \tag{6.36}$$

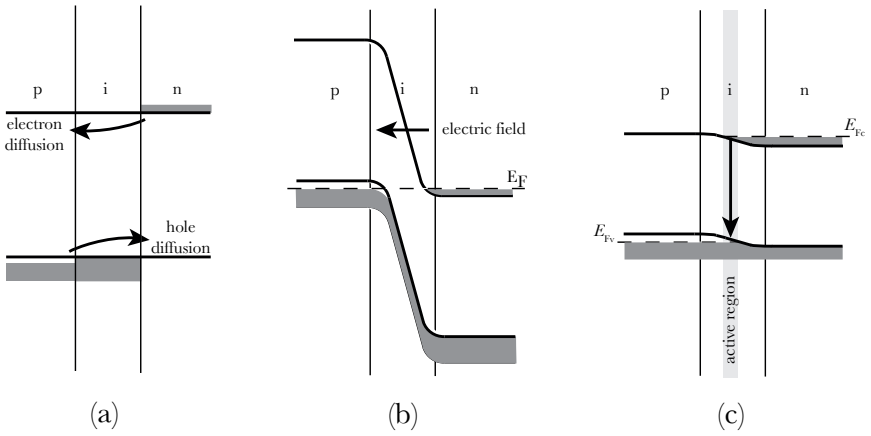

Figure 6.8: Energy levels and Fermi levels in a pin diode: (a) at the (theoretical) moment when the three layers are brought into contact; (b) after the depletion layer has been established; (c) under the influence of a forward-bias voltage.

We see that the threshold current increases with N_{tr} and the volume of the active region.

6.3.1 Structure of homojunction diode lasers

Figure 6.9 shows the structure of a homojunction diode laser. Notice that the gain length is very short, only 200 - 500 μm, and perhaps 100 - 200 μm wide. Optical feedback is usually provided by the uncoated surfaces of the cleaved crystal planes at each end of the device. The high refractive index of GaAs ($n = 3.6$) ensures a Fresnel reflectivity of approximately 32 % which is sufficient for lasing to be achieved.

One of the limitations of homojunction lasers is that the potential barrier preventing electrons from entering the p-type material is very low when the forward-bias voltage is applied. Consequently electrons tend to diffuse into the p-type material, increasing the thickness of the active region beyond that of the depletion layer. In practice diffusion increases the thickness to around 1 μm, much thicker than the thickness of the depletion layer. This increases the threshold current to tens of amps, corresponding to a current density of order 10^5 A cm^{-2}. This high value means that homojunction diode lasers cannot be operated continuously without damaging the structure unless the whole device is cooled to cryogenic temperatures. For this reason, homojunction diode lasers are of historical interest only and all modern diode lasers use heterostructures.

6.4 Double heterostructure lasers

All diode lasers currently used employ a **double heterostructure**, rather than the simple p-i-n geometry discussed above. This structure can reduce the threshold current density by two orders of magnitude (to approximately 10^3 A cm^{-2}) owing to:

- **Photon confinement:** The active layer is surrounded by layers of a higher refractive index. As such the structure acts as a waveguide, and prevents the laser radiation from extending into regions of high loss.
- **Carrier confinement:** The surrounding layers also have large band gaps which confines the current carriers to the active region. The resulting increase in carrier density increases the gain.

tb]

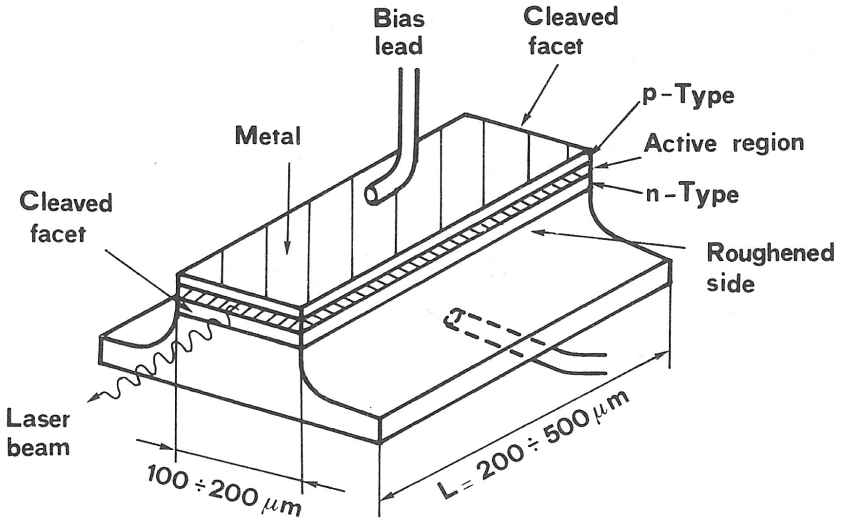

Figure 6.9: Physical structure of a homojunction diode laser.

- **Reduced absorption:** The large energy gap of the surrounding layers ensures that any laser radiation which does extend beyond the active region experiences much lower absorption than it would if the band gap were the same as that in the active layer.

Figure 6.10 shows schematically the structure of heterostructure diode lasers made from (a) AlGaAs and (b) InGaAsP. In each case the active layer (≈ 0.15 μm) is surrounded by thicker layers of n- and p-doped material with a higher band gap. Figure 6.11 shows the variation of the refractive index, laser intensity, and band gap along a line perpendicular to the p-i-n junction under conditions of forward bias. It is clear that the higher band gap of the surrounding material prevents current carriers from moving beyond the active layer. Further, the diode laser radiation is surrounded by layers of higher refractive index and hence the mode is guided.

6.4.1 Forming heterostructures: Material properties

The selection of materials from which to make the cladding layers is a crucial one; not only must their band gap be larger than that of the intrinsic layer, but they must also be capable of being deposited via epitaxial growth techniques on the active region to form crystals free from defects without grain boundaries or regions of dislocation. This means that the alloy used for cladding must have a lattice constant very closely matched to that of the intrinsic material.

As shown in Figure 6.12, if we start with pure GaAs, by adding increasing atomic proportions x of aluminium to form the ternary alloy $Ga_{1-x}Al_xAs$ the lattice spacing is virtually the same whatever the proportion of added Al, even to the extent of replacing all of the Ga by Al to form AlAs since the lattice spacing of GaAs and AlAs are almost identical: 0.564 and 0.566 nm respectively.

This is however a happy circumstance unique to the GaAs system. In general **quarternary compounds** must be employed. For example, in $In_{1-x}Ga_xAs_yP_{1-y}$ the lattice spacing can be matched to the III-V semiconductor InP if $y = 2.2x$. With this ratio fixed, adjusting x then varies the band gap energy, allowing the output of the diode laser to be adjusted between 1150 nm and 1670 nm. This range is particularly useful since

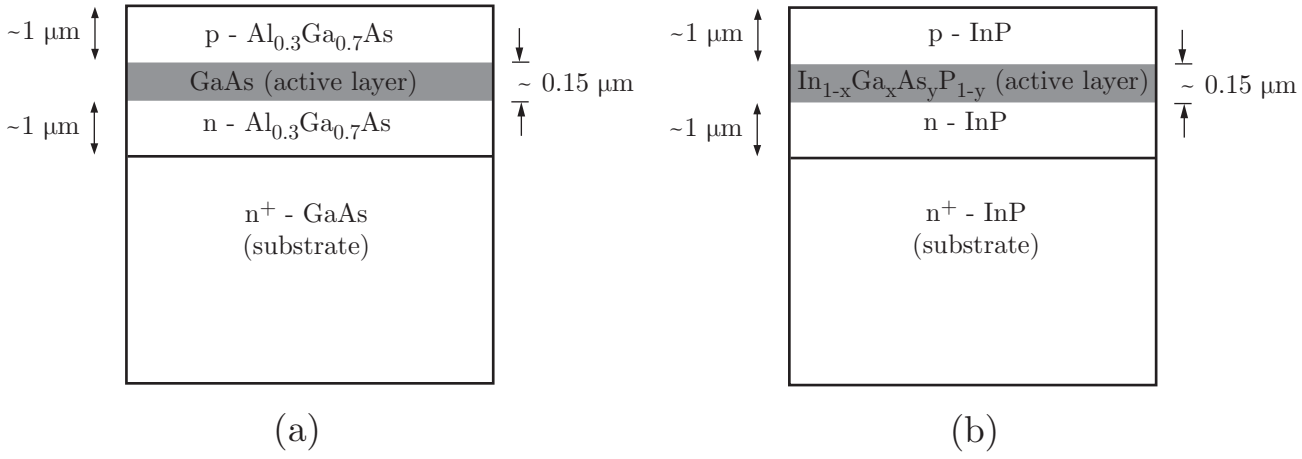

Figure 6.10: Schematic diagram of double heterostructure diode lasers made from: (a) AlGaAs; (b) InGaAsP.

Active region	Material of cladding layers	Wavelength range	Substrate material
GaAs	(AlGa)As	800 - 900 nm	GaAs
GaInP	(AlGa)InP	630 - 650 nm	GaAs
InGaAs	GaAs	900 - 1000 nm	GaAs
InGaAs	InGaAsP	1550 nm	InP
InGaAsP	InGaAsP	1300 - 1550 nm	InP
InAsP	InGaAsP	1060 -1400 nm	InP
InGaAs	InP	1550 nm	InP
InGaN	AlGaN, GaN	370- 460nm	Al_2O_3,GaN

Table 6.1: Diode laser materials and wavelengths of operation.

it covers the second (1300 nm) and third (1550 nm) transparency windows of silica fibres. InGaAsP diode lasers therefore find widespread application in telecommunications.

Table 6.1 lists the properties of some common types of diode laser.

6.5 Geometries of heterostructure diode lasers

Figure 6.13 shows the design of two types of heterostructure diode lasers. In both designs an oxide layer is added immediately under one of the electrode layers to restrict the extent of the current flow, and hence the gain region, in the y-direction. In the x-direction the region of gain is limited to the thickness of the active layer. By controlling the volume of the gain region in this way, the threshold current for achieving laser oscillation is reduced; although, of course, the laser power that can be extracted is also reduced.

For the design shown in Figure 6.13(a), the radiation mode is confined in the x-direction by the lower refractive index cladding layer. In the y-direction the extent of the beam is determined by the width of the gain region, and hence is said to be **gain-guided**. However, the confinement is weak in this direction, and the oscillating mode extends into regions which are not pumped, and so experiences high loss. This problem is overcome in the **buried heterostructure** shown in Figure 6.13(b). Here, the active region is also surrounded in the y-direction by cladding, or 'burying', layers which confine the radiation in that direction as well.

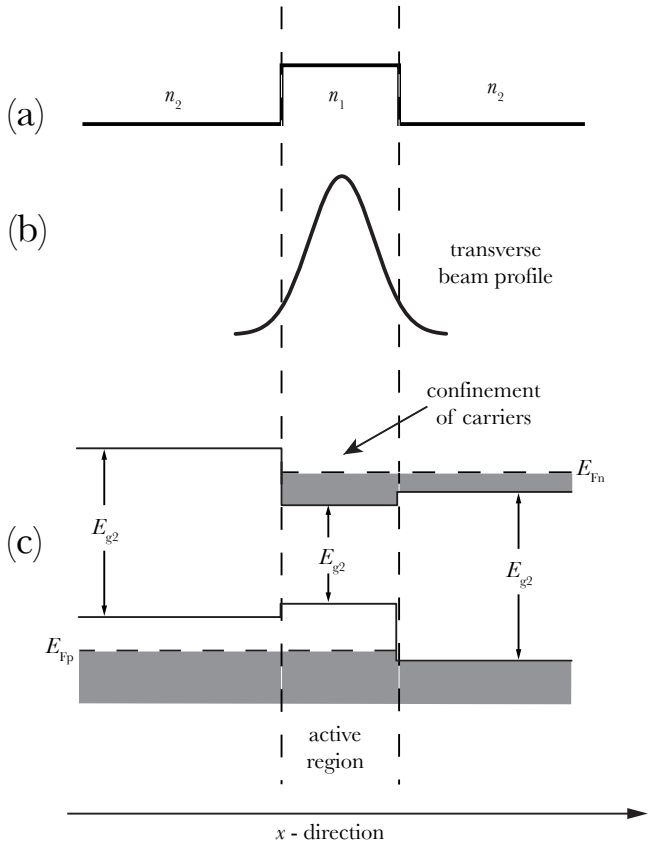

Figure 6.11: Schematic diagram of the variation of: (a) the refractive index; (b) the laser intensity; and (c) the band gap along the x-axis of a double heterostructure diode laser.

6.6 Output characteristics of diode lasers

Diode lasers have very high **slope efficiencies** $dP_{\rm out}/dP_{\rm in} = dP_{\rm out}/(VdI)$, where V is the diode voltage. This can reach approximately 50% may be obtained in commercial devices.

The output power of a single double heterostructure diode laser is limited to powers less than a few hundred mW by over-heating of the diode structure. Much large powers can be generated by manufacturing linear arrays of diode lasers, known as **diode bars**. In turn, diode bars can be stacked to form diode arrays. A wide variety of geometries are available. Diode arrays can give output powers of 1 kW or even higher, and are increasingly used as replacements in flashlamps in solid state lasers such as Nd:YAG, and as the pump laser in fibre lasers.

The frequency spectrum of diode lasers often depends on the driving current. Near threshold the laser mode spectrum has the wide bandwidth characteristic of an LED, but as the excitation current is increased a dominant axial mode establishes itself. This competes more and more effectively for the available inversion as the current increases so that the bandwidth spanned by the oscillating modes decreases as the current increases. Diode lasers can be operated in a single longitudinal mode by providing frequency-selective coupling. For example, a **fibre Bragg grating (FBG)** can be fixed to the laser output; in a FBG the refractive index of the fibre core varies periodically, and hence only those wavelengths satisfying the Bragg condition for reflection are efficiently coupled back into the gain region. Alternatively, wavelength-selective coupling can be provided by modulating

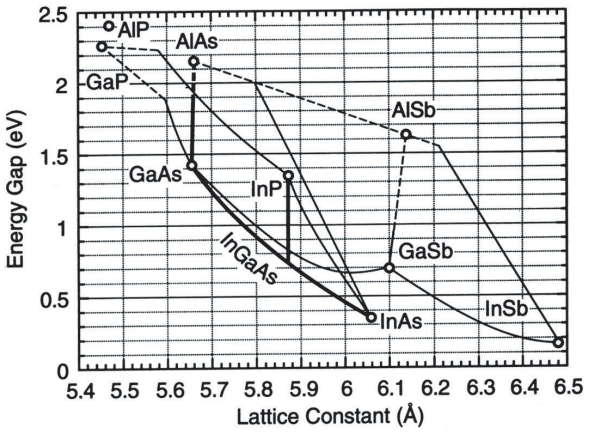

Figure 6.12: The band gap energy versus lattice spacing for some semiconductor compounds.

the thickness of the layers surrounding the diode laser gain region with a "washboard" pattern; for the selected wavelength this provides continuous coupling between the left- and right-going waves in the gain region.

The biggest disadvantage of diode lasers is that they have a high divergence, which reflects the small dimensions of the gain region (typically $L_x = 1 \times L_y = 5$ μm). Further, the divergence is different in the two orthogonal directions ($\theta_x \approx 25^o \times \theta_y \approx 5^o$) which is often undesirable in applications. Techniques have been developed for making the beam more nearly circular, such as the use of prisms to expand the beam in one direction, or anamorphic lenses which have different curvatures in the x- and y-directions. Nevertheless the output beam quality of diode lasers remains poor in comparison to that from other types of laser.

Figure 6.14 shows the typical variation of output power with the diode current. Notice the low threshold current, a few milliamperes, and the sensitivity of the threshold current with the device temperature. This sensitivity reflects the strong dependence on the quasi-Fermi energies on temperature.

6.7 Fibre lasers

Fibre lasers are particularly efficient at converting the often poor beam quality obtained from a diode laser into diffraction-limited laser radiation. In a fibre laser the active ion is doped into the central core of an optical fibre, the ions being pumped by radiation propagating along the fibre axis.

Fibre lasers have several advantages over the traditional rod or disc geometries:

- **Low pump power threshold:** Achievement of the threshold population inversion density requires the pump laser to exceed a threshold intensity. Since the cross-sectional area of an optical fibre is so low, the power required to reach threshold is orders of magnitude smaller than in a rod or disc where the area of the gain medium is at least an order of magnitude larger. In practice threshold pump powers significantly below 1 mW can be achieved. Using the fibre laser geometry it is possible to achieve lasing in lasers with: low gain coefficients; with high threshold pump intensities; and even in three-level laser systems.

- **Guiding of pump beam:** The pump laser radiation is naturally guided by the fibre, allowing very long gain regions to be employed.

- **High efficiency:** The good overlap between the pump and generated laser radiation along the whole length allows efficient conversion of the pump laser energy into laser output.

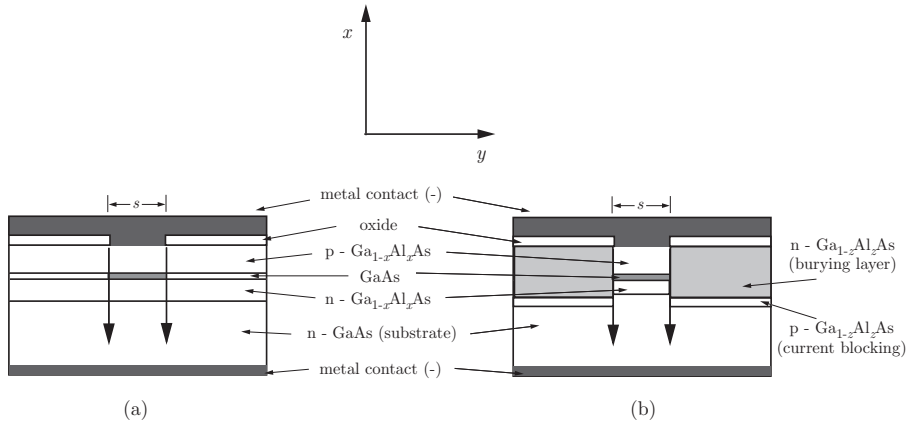

Figure 6.13: Geometries of heterostructure diode lasers: (a) gain-guided laser; (b) buried heterostructure, index-guided laser.

- **Good thermal management:** For any solid state laser system it is important to conduct away pump energy deposited into the gain medium in the form of heat, in order to avoid thermal lensing. The high surface area to volume ratio of the fibre geometry allows heat to be conducted away more effectively.
- **Compact gain medium:** Optical fibres may be coiled, keeping the volume of the laser system small.
- **Robustness:** Optical fibres are robust and much less liable to cracking or other physical damage.

The main disadvantages of fibre lasers are that the volume of the gain region is small. Consequently the energy stored in the gain region will also be small, and hence fibre lasers are not useful for amplifying laser pulses to high energy. Further, the long propagation length through the fibre core can be associated with the development of non-linear loss processes such as stimulated Raman scattering or stimulated Brillouin scattering.

6.7.1 Pumping configurations and fibre materials

In early fibre lasers both the pump radiation and the generated laser radiation were guided in the inner core of an optical fibre (Fig. 6.15(a)). As such the pump laser must be close to diffraction-limited for efficient coupling into the inner core of the fibre, which restricts the use of diode lasers to single-stripe devices.

Figure 6.15(b) shows a solution to this problem: **cladding pumping**. For this geometry the core of refractive index is surrounded by *two* further layers: an outer core of lower refractive index, and a cladding layer of still lower refractive index. The cladding layer will usually be protected by a final protective layer. The pump laser radiation is therefore guided in a multi-mode fashion by reflections at the boundary between the outer core and the cladding layer; with that part of the pump radiation passing through the core contributing to the pumping of the fibre laser. The laser radiation generated in the core can be guided mono- or multi-mode.

The use of cladding pumping allows much greater pump powers to be employed, increasing the output of the fibre laser from less than 1 W to greater than 100 W in a monomode beam and up to 1 kW in multi-mode fibre lasers.

6.7.2 An example

Figure 6.16 shows schematically the design of a high-power Nd:Glass fibre laser. For this system the doped core has a diameter of 10 μm, with a numerical aperture of 0.17. The outer core has a diameter of 400 μm and

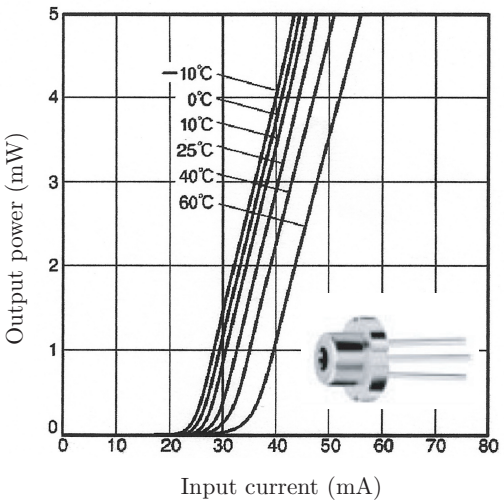

Figure 6.14: Output power of Rohm RLD-78MA double heterostructure diode laser versus current for various operating temperatures. Inset: photo of 5.6 mm diameter packaged diode laser. (From www.rohm.com)

numerical aperture of 0.38. Pump radiation, provided by a diode array operating at 808 nm, is coupled into one end of the fibre through a mirror which has a high transmission at 808 nm, but is highly reflective at the Nd:Glass wavelength. The output coupling mirror is also highly reflective at the pump laser wavelength, but has a transmission of approximately 90% for the Nd:Glass laser radiation. With this design, output powers of 32 W have been achieved for a pump power of 72 W, corresponding to a conversion efficiency of 44%. The output is in the form of a single transverse mode, since although the core supports several low-order modes, the lowest mode has the lowest losses and hence grows at the expense of higher-order transverse modes. Laser oscillation occurs at approximately 1065 nm, with a large spectral width of 15 nm reflecting the large inhomogeneous width introduced by the glass host.

6.7.3 Erbium-doped fibre lasers

Silica optical fibres have transmission windows at 1.3 μm and 1.5 μm. Single-mode fibres used at 1.3 μm have low dispersion and so the distance that signals can propagate before they must be amplified or regenerated is determined primarily by the fibre loss of 0.4 dB km^{-1}. Amplifiers are not available at 1.3 μm, and so signals at this wavelength are regenerated in a regenerative repeater, which: (i) detects and interprets the signal; (ii) corrects errors; (iii) regenerates and retransmits the bitstream. The optimum repeater separation is typically 40 km.

At a wavelength of 1.5 μm conventional fibre has a lower loss (approximately 0.3 dB km^{-1}), but higher dispersion than at 1.3 μm. However in dispersion-shifted fibre, in which material and waveguide dispersion are balanced, the dispersion at 1.5 μm can be reduced to 0.1 ps km^{-1} nm^{-1} compared to 17 ps km^{-1} nm^{-1} for conventional fibre at 1.3 μm.[5] Hence dispersion-shifted fibre can transmit signals some two orders of magnitude

[5]In any optical fibre the dispersion experienced by *each mode* — the intra-modal dispersion — has two contributions: dispersion arising from the dispersion of the materials employed (material dispersion); and dispersion arising from the frequency dependence of the wave vector of each mode (waveguide dispersion), which depends on the geometry of the fibre. In dispersion-shifted fibres the dimensions of the fibre are adjusted so that the total dispersion is low at the desired operating wavelength. In addition, for any fibre different modes have different phase velocities leading to so-called inter-modal dispersion; this can be avoided if only one mode is excited.

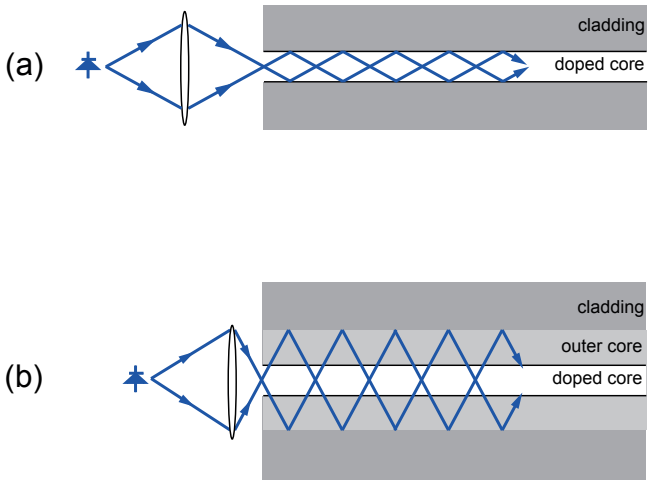

Figure 6.15: Fibre laser pumping configurations: (a) pumping of a conventional fibre; (b) cladding pumping.

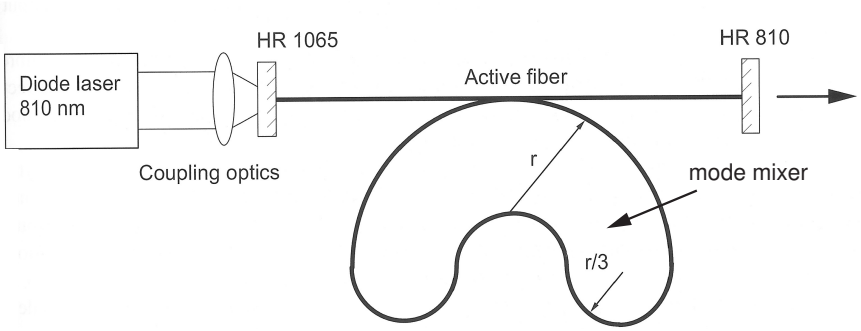

Figure 6.16: Schematic diagram of a high-power Nd:Glass fibre laser.

further — corresponding to distances of tens of thousands of kilometres — before regeneration of the signal is required. Of course, *amplification* of the signal is required more often to overcome the fibre losses; typically amplification is required every few tens of kilometres.

Amplification can be provided by an **erbium-doped fibre amplifier (EDFA)**. The EDFA operates on the $^4I_{13/2} \rightarrow^4 I_{15/2}$ transition at approximately 1550 nm. As discussed previously, pump radiation can conveniently be provided by diode lasers operating at 800, 980, or 1480 nm.

Figure 6.17 shows schematically the configuration of an EDFA. The pump laser radiation is coupled into the amplifier by splicing together a fibre carrying the pump radiation and a fibre coaxial with the amplifier. The amplifying section is a monomode silica fibre doped with around 0.1% erbium, and with a length ranging from metres to hundreds of metres. The single-pass gain is typically of order 10^3. Notice that optical isolators — based on the Faraday effect — are inserted around the EDFA to prevent parasitic laser oscillation arising from the small Fresnel reflections at junctions in the fibre.

EDFAs have been used in the latest 5 Gbit s^{-1} transatlantic fibres which employ amplifiers every 35 km, but only regenerate the signal at the two land stations!

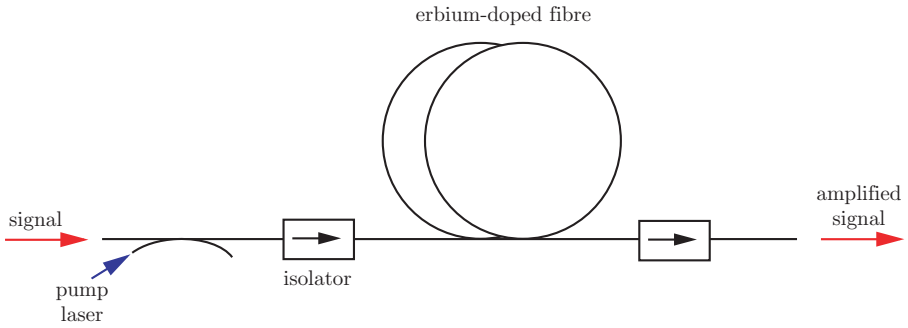

Figure 6.17: Schematic diagram of an EDFA.

Finally we note that the very wide bandwidth (some 30 nm or 3.74 THz) of the erbium transition allows amplification of multiple channels each transmitted at a slightly different wavelength. This technique, **wavelength division multiplexing (WDM)**, could allow, for example, 375 channels of 2.5 Gbit s^{-1} and spaced by 10 GHz to be transmitted within the bandwidth of an erbium amplifier. The implementation of WDM, however, requires the development of suitable laser sources and techniques for de-multiplexing the transmitted signals; a great deal of work is being done in this area.

Lecture 7

Precision frequency control

Here we discuss some of the issues affecting the frequency linewidth of an oscillating laser. We will often give results in terms of the frequency ν (units of Hertz) rather than the angular frequency ω (units of radians per second) since this allows easier comparison with real laser systems.

7.1 Single-mode operation

In principle a laser operating on a homogeneously broadened transition should oscillate on only a single longitudinal mode, but in practice spatial hole-burning means that oscillation can occur on more than one mode. Lasers operating on homogeously broadened transitions will typically operate on many longitudinal modes. In addition, laser oscillation can occur on higher-order transverse modes, which in general have slightly different frequencies.

It is clear that for a laser to achieve the narrowest possible linewidth it should operate on a single longitudinal and transverse mode. Ensuring the latter is straightforward: by introducing a suitable aperture it is possible to ensure that higher-order transverse modes experience sufficiently high losses that they cannot reach the threshold for oscillation.[1]

Several techniques can be used to force a laser to operate on only a single longitudinal mode.

7.1.1 Short cavity

The simplest approach is to ensure that the longitudinal mode spacing $\Delta\nu_{p,p-1} \approx c/2L_\mathrm{c}$ is larger than the range of frequencies, $\Delta\nu_\mathrm{osc}$, for which the laser gain exceeds the threshold value. Assuming that one of the longitudinal modes coincides with the centre of the laser transition, it is easy to show that single-mode operation will occur if,

$$L_\mathrm{c} < \frac{c}{\Delta\nu_\mathrm{osc}}. \tag{7.1}$$

7.1.2 Intra-cavity etalon

An intra-cavity etalon may be used to increase the losses experienced by longitudinal modes adjacent to the mode nearest line centre. The etalon must satisfy two conditions:

[1] See Fig. 8.7 of "Modern Classical Optics" by Geoff Brooker, also reproduced as Fig. 6.12 in H&W.

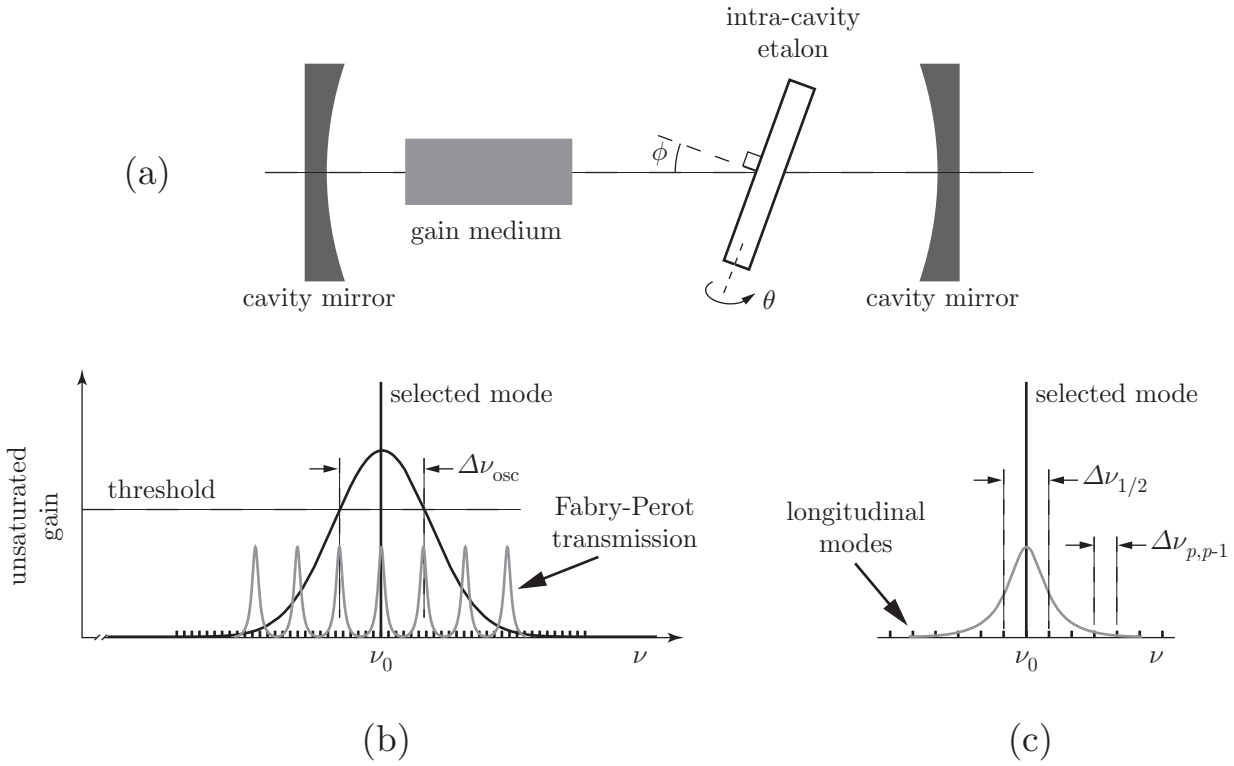

Figure 7.1: SLM operation by using an intra-cavity etalon. (b) Relation between the peaks of the etalon transmission and the laser lineshape when Condition (2) is satisfied. (c) Relation between the FWHM of the etalon and the longitudinal mode spacing when Condition (1) is satisfied.

1. The width of the etalon peaks must be sufficiently narrow to discriminate between adjacent longitudinal modes.
2. The free-spectral range of the etalon must be sufficiently large that adjacent etalon peaks lie outside the range of frequencies for which laser oscillation is possible.

Figure 7.1 shows an idealized arrangement in which the orientation of the etalon has been rotated so that one of its transmission peaks coincides with the longitudinal mode nearest the centre of the laser line.

We may express conditions (1) and (2) as:

$$\frac{\Delta\nu_{1/2}}{2} < \Delta\nu_{p,p-1} \tag{7.2}$$

$$\Delta\nu_{\rm FSR} > \frac{\Delta\nu_{\rm osc}}{2}, \tag{7.3}$$

where $\Delta\nu_{1/2} = \Delta\nu_{\rm FSR}/\mathcal{F}$ is the FWHM of one of the etalon peaks. Using the standard relations for a Fabry-Perot etalon it is easy to show that the thickness d of the etalon must satisfy,

$$\frac{\Delta\nu_{\rm osc}}{2} < \frac{c}{2nd} < \mathcal{F}\frac{c}{L_{\rm c}}, \tag{7.4}$$

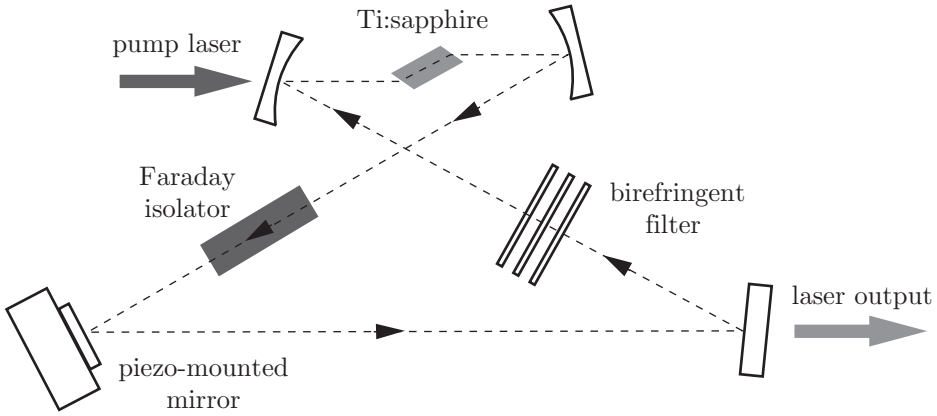

Figure 7.2: Single longitudinal mode tunable Ti:sapphire laser employing a uni-directional ring cavity. Wavelength tuning is achieved by rotating the birefringent filter. One of the cavity mirrors is mounted on a piezoelectric crystal so that the length of the cavity can be adjusted as the wavelength is tuned, ensuring that the same longitudinal mode oscillates as the wavelength is varied.

where n is the refractive index of the material between the etalon plates (it is common to use a solid etalon in this application). Hence we find that,

$$L_{\mathrm{c}} < (2\mathcal{F})\,\frac{c}{\Delta\nu_{\mathrm{osc}}}. \tag{7.5}$$

Thus the use of an intra-cavity etalon allows SLM operation to be achieved in a cavity with a length increased by a factor $2\mathcal{F}$ than with no etalon. It may be shown that for N etalons the maximum length of the cavity can be increased by a factor of $(2\mathcal{F}_1)\,(2\mathcal{F}_2)\,(2\mathcal{F}_3)\ldots(2\mathcal{F}_N)$ above that required for no etalon.

7.1.3 Ring resonators

SLM operation on *homogeneously* broadened laser transitions can be achieved by employing a ring resonator and ensuring that oscillation only occurs in one direction around the ring; this prevents spatial hole-burning. Uni-directional oscillation can be achieved by inserting a Faraday isolator within the ring as shown in Fig. 7.2.

7.2 Output linewidth

7.2.1 Theoretical limit

Suppose that SLM oscillation is achieved. What would be the linewidth of the laser output?

Since the laser cavity can be considered to be a Fabry-Perot etalon, we might expect that the linewidth would be given by the FWHM of an etalon peak, i.e. $\Delta\nu_{1/2} = \Delta\nu_{\mathrm{FSR}}/\mathcal{F}$. However, this is *not* the case since the finite linewidth of the etalon arises from the fact that the amplitude of a wave launched within the etalon decays as it bounces between the mirrors; as such the wave is only 'stored' within the cavity/etalon for a time of order the cavity lifetime τ_{c}, leading to a linewidth of order $1/\tau_{\mathrm{c}}$. For an oscillating laser, however, the losses are balanced by the gain so, in effect, the cavity lifetime becomes infinite. This would suggest that the laser linewidth should be zero.

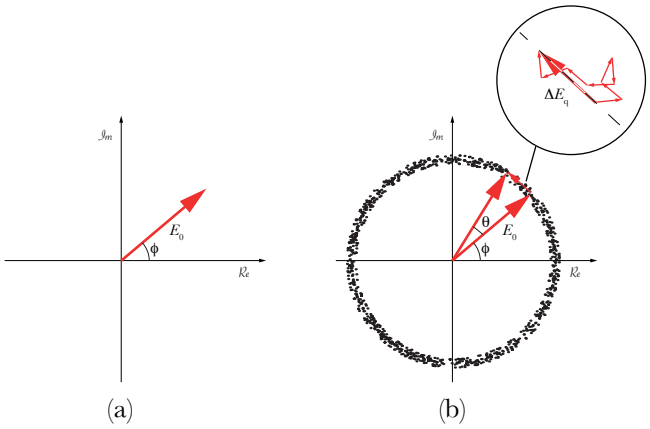

Figure 7.3: Origin of the Schawlow-Townes linewidth. (a) In the absence of spontaneous emission, if a laser oscillates on a single cavity mode the electric field may be written as $E(t) = E_0 \exp(-\mathrm{i}[\omega_\mathrm{L} t - \phi])$ where ϕ is an arbitrary constant phase. This may be represented by a constant vector in an Argand diagram, as shown, where the axes correspond to $\Re\left[E(t)\exp(\mathrm{i}\omega_\mathrm{L} t)\right]$ and $\Im\left[E(t)\exp(\mathrm{i}\omega_\mathrm{L} t)\right]$. The effect of spontaneous emission is shown in (b). After each spontaneously emitted photon is emitted into the oscillating mode, the tip of the electric field vector moves in a random direction by an amount E_sp. The dots shown in (b) show the positions occupied by the tip of the electric field vector after a very large number of photons have been emitted spontaneously. The inset in (b) shows the effect on the electric field vector of the emission of 8 spontaneously emitted photons, and in particular the net displacement, ΔE_θ, along a direction, $\hat{\boldsymbol{\theta}}$, orthogonal to the original electric field vector. Notice that the electric field vector of each spontaneously emitted photon has a different component along $\hat{\boldsymbol{\theta}}$, and consequently on average the displacement ΔE_θ is given by the result of a one-dimensional random walk (along $\hat{\boldsymbol{\theta}}$) of varying step size. The resultant displacement causes the orientation of the electric field vector to rotate through an angle θ, as shown.

In fact the fundamental limit to the laser linewidth arises from the role of spontaneous emission within the laser cavity. As we will see, the effect of spontaneous emission is a slow drift of the phase of the laser output, which leads to a finite frequency width.

Supose that an oscillating mode within the cavity contains $\bar{n}$ photons. The energy of the mode is proportional to $\bar{n}$, or, equivalently, to E_0^2 where E_0 is the amplitude of the electric field of the mode. Hence we have the relation $E_0 \propto \sqrt{\bar{n}}$.

If the cavity lifetime is τ_c then photons are lost from the cavity at a rate $\bar{n}/\tau_\mathrm{c}$; these form the output beam. Since photons produced by stimulated emission are perfectly coherent with each other, in the absence of spontaneous emission the amplitude and phase of the electric field at any point in the cavity would be constant and could be written in the form $E(t) = E_0 \exp\left(-\mathrm{i}\left[\omega_\mathrm{L} t - \phi\right]\right)$. The amplitude of this field could be represented by a constant vector in the complex plane, as shown in Fig. 7.3.

Occasionally a photon will be emitted into the cavity spontaneously. This will have a random phase relative to that of the stimulated photons, leading to a small shift in the phase (and also the amplitude, but this is small) of the oscillating mode. Spontaneous emission therefore causes the phase of the mode to change randomly by small amounts so that the phase of the mode diffuses over time with a characteristic time τ_ST, corresponding to a frequency width $\Delta\omega_\mathrm{ST} = 1/\tau_\mathrm{ST}$.

We now estimate τ_ST and hence the **Schawlow-Townes linewidth**. Since $E_0 \propto \sqrt{\bar{n}}$ we can write,

$$\frac{E_\mathrm{sp}}{E_0} = \frac{\sqrt{1}}{\sqrt{\bar{n}}}. \tag{7.6}$$

If N photons are emitted spontaneously into the mode the change θ of the phase of the mode is

$$\theta = \frac{\Delta E_\theta}{E_0}, \tag{7.7}$$

where ΔE_θ is the is the component of the vector sum of the electric fields of the N photons along the $\hat{\boldsymbol{\theta}}$ direction.

The value of ΔE_θ depends on the phases of the spontaneously emitted photons. Since these are uncorrelated, we may use the results of the theory of a one-dimensional random walk of variable step size. These state that for N steps in the random walk — i.e. after the emission of N spontaneous photons — the mean amplitude and mean square amplitude along $\hat{\boldsymbol{\theta}}$ are given by:

$$\overline{\Delta E_\theta} = N\overline{E_{\theta\mathrm{sp}}} \tag{7.8}$$

$$\overline{\Delta E_\theta^2} = N\overline{E_{\theta\mathrm{sp}}^2}, \tag{7.9}$$

where $\overline{E_{\theta\mathrm{sp}}}$ and $\overline{E_{\theta\mathrm{sp}}^2}$ are the mean amplitude and mean square amplitude along $\hat{\boldsymbol{\theta}}$ of a single spontaneously emitted photon. Since for spontaneously emitted photons E_{sp} is the same, but their phase can take any value with equal probability, it is easy to show that

$$\overline{E_{\theta\mathrm{sp}}} = 0 \tag{7.10}$$

$$\overline{E_{\theta\mathrm{sp}}^2} = \frac{E_{\mathrm{sp}}^2}{2}, \tag{7.11}$$

and hence that after N photons have been emitted into the mode:

$$\overline{\theta} = 0 \tag{7.12}$$

$$\overline{\theta^2} = \frac{\overline{\Delta E_\theta^2}}{E_0^2} = \frac{N}{2}\left(\frac{E_{\mathrm{sp}}}{E_0}\right)^2. \tag{7.13}$$

If photons are spontaneously emitted into the mode at a rate R_{sp}, then after a time τ the root mean square value of θ will be

$$\theta_{\mathrm{rms}} = \sqrt{\overline{\theta^2}} = \sqrt{\frac{R_{\mathrm{sp}}\tau}{2\bar{n}}}, \tag{7.14}$$

where we have used eqn (7.6).

We can say that the uncertainty in the phase of the mode is significant once $\theta_{\mathrm{rms}} \approx \pi/2$, and hence we may write,

$$\sqrt{\frac{R_{\mathrm{sp}}\tau_{\mathrm{ST}}}{2\bar{n}}} \approx \frac{\pi}{2},$$

from which we deduce,

$$\Delta\omega_{\mathrm{ST}} = \frac{1}{\tau_{\mathrm{ST}}} \approx \left(\frac{2}{\pi}\right)^2 \frac{R_{\mathrm{sp}}}{2\bar{n}}. \tag{7.15}$$

To find the rate $R_{\rm sp}$ at which photons are emitted spontaneously into the mode we note that the rate of spontaneous emission into a mode is equal to $1/\bar{n}$ of the rate of stimulated emission from the upper laser level. Further, under steady-state conditions the rate of stimulated emission from the upper level must equal $\bar{n}/\tau_{\rm c}$, i.e. the rate at which photons leave the cavity. Hence $R_{\rm sp} = 1/\tau_c$ and so,

$$\Delta\omega_{\rm ST} \approx \left(\frac{2}{\pi}\right)^2 \frac{1}{2\bar{n}\tau_{\rm c}}. \tag{7.16}$$

It is convenient to write $\bar{n}$ in terms of the output power P of the laser. Since the cavity loses photons at a rate $\bar{n}/\tau_{\rm c}$, we may write $P = (\bar{n}/\tau_{\rm c})\hbar\omega_{\rm L}$ and hence,

$$\Delta\omega_{\rm ST} \approx \frac{2}{\pi^2}\frac{\hbar\omega_{\rm L}}{P}\frac{1}{\tau_{\rm c}^2} = 8\frac{\hbar\omega_{\rm L}}{P}\Delta\nu_{\rm c}^2 \tag{7.17}$$

$$\Rightarrow \Delta\nu_{\rm ST} \approx \frac{4}{\pi}\frac{h\nu_{\rm L}\Delta\nu_{\rm c}^2}{P}. \tag{7.18}$$

A more complete derivation shows that,

$$\Delta\nu_{\rm ST} = \left(\frac{N_2}{N^*}\right)_{\rm th} \frac{\pi h\nu_{\rm L}\Delta\nu_{\rm c}^2}{P}, \tag{7.19}$$

where $(N_2/N^*)_{\rm th}$ is the ratio of the population density in the upper laser level to the population inversion density under the operating conditions of the laser. Typically this ratio is close to unity.

7.2.2 Practical limitations

The Schawlow-Townes linewidth is usually extremely small (very often much less than 1 Hz!). In the absence of active frequency-locking techniques the linewidth of even a single-mode laser will be much greater than $\Delta\nu_{\rm ST}$.

A change in cavity length will change the frequency by an amount,

$$\frac{\Delta\nu_p}{\nu_p} = -\frac{\Delta L_{\rm c}}{L_{\rm c}}. \tag{7.20}$$

A wide variety of processes (vibration, thermal effects, changes in ambient pressure etc.) can cause the cavity length to change, and hence the output frequency to drift or change.

To give just one example, if the coefficient of thermal expansion of the material on which the cavity mirrors is mounted is $\alpha_{\rm T}$, then a change in temperature of ΔT gives,

$$\frac{\Delta\nu_p}{\nu_p} = -\alpha_{\rm T}\Delta T. \tag{7.21}$$

For invar (a steel alloy with very low coefficient of thermal expansion of $\alpha_{\rm T} \approx 10^{-6}$) then to keep $\Delta\nu_p$ below, say, 1 MHz requires temperature stabilization to better than 2 mK for a laser operating in the visible region of the spectrum.

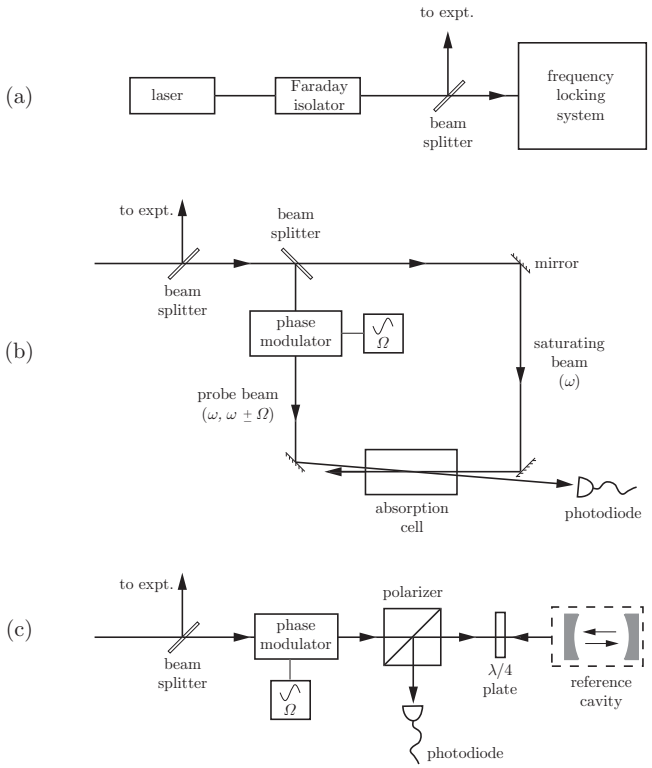

Figure 7.4: (a) Schematic diagram of the arrangement used to frequency lock a laser. (b) Frequency-locking to a Doppler-free absorption line. (c) Frequency-locking to a reference cavity.

7.3 Frequency locking

Under practical conditions in order to achieve the narrowest linewidths, and to prevent drift, it is necessary to lock the frequency of a laser to a stable reference frequency. In principle this is straightforward:

1. The frequency ν of the laser to be locked is compared with some fixed frequency $\nu_{\rm ref}$.
2. Any difference in frequency is converted to an error signal, which increases with the magnitude of the frequency error and differs in sign according to the sign of $(\nu - \nu_{\rm ref})$.
3. The frequency of the oscillating mode is adjusted — usually by changing the length of the laser cavity — so as to minimize the magnitude of the error signal, thereby minimizing the frequency error.

This is usually achieved by picking off a small portion of the laser beam and modulating its frequency so that phase-sensitive detection (PSD) techniques can be employed. Figure 7.4 shows how this is achieved in the general case, and for the cases of locking to an absorption line or an external cavity.

Form of error signal

We now consider the form of the error signal for the general case of a frequency-modulated wave passing through (or reflecting from) a reference system with a frequency-dependent *amplitude* response $a(\omega)$ which is strongly peaked in the region of the reference frequency $\omega_{\rm ref}$. The amplitude of the wave leaving a frequency modulator with a time-dependent amplitude transmission $\exp(\mathrm{i}\beta \sin \Omega t)$ may be written as,

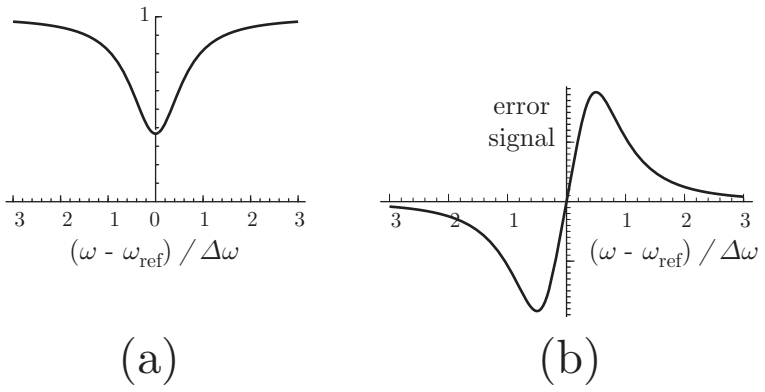

Figure 7.5: Frequency-locking to a Lorentzian response function centred at $\omega_{\rm ref}$ with a full-width at half maximum $\Delta\omega$. (a) the amplitude response function $a(\omega)$; (b) the amplitude of the error signal $S_\Omega(t)$ as a function of the output frequency ω of the laser.

$$E(t) = E_0 \exp(-\mathrm{i}\omega t) \exp(\mathrm{i}\phi_0) \sum_{m=-\infty}^{\infty} J_m(\beta) \exp(\mathrm{i}m\Omega t), \tag{7.22}$$

where β is proportional to the phase modulation introduced by the modulator. If the amplitudes of the sidebands other than $m = \pm 1$ may be ignored, we have

$$E(t) = E_0 \exp(\mathrm{i}\phi_0) \left\{ -J_1(\beta) \exp\left[-\mathrm{i}(\omega+\Omega)t\right] + J_0(\beta) \exp(-\mathrm{i}\omega t) + J_1(\beta) \exp\left[-\mathrm{i}(\omega-\Omega)t\right] \right\}, \tag{7.23}$$

where we have used the fact that $J_{-1}(\beta) = -J_1(\beta)$. After passing through the reference system the amplitude of the radiation is

$$E(t) = E_0 \exp(\mathrm{i}\phi_0) \left\{ -J_1(\beta) a(\omega+\Omega) \exp\left[-\mathrm{i}(\omega+\Omega)t\right] + J_0(\beta) a(\omega) \exp(-\mathrm{i}\omega t) + J_1(\beta) a(\omega-\Omega) \exp\left[-\mathrm{i}(\omega-\Omega)t\right] \right\}, \tag{7.24}$$

It may be shown that for a square-law detector, i.e. one which detects a signal $S(t) \propto E(t)E*(t)$, the signal at the modulation frequency Ω is given by,

$$S_\Omega(t) \propto \Omega \left|a(\omega)\right| \left. \frac{\partial \left|a(\omega)\right|}{\partial \omega} \right|_\omega \cos \Omega t. \tag{7.25}$$

As shown in Fig. 7.5 the **error signal** is proportional to the derivative of $|a(\omega)|$, and hence the sign of the error signal depends on whether ω is greater or less than the reference frequency $\omega_{\rm ref}$.

7.3.1 Locking to atomic or molecular transitions

Locking to atomic or molecular transitions provides the best long-term stability and gives highly reproducible results: the frequency reproducibility is typically found to be 1 part in 10^{11}.

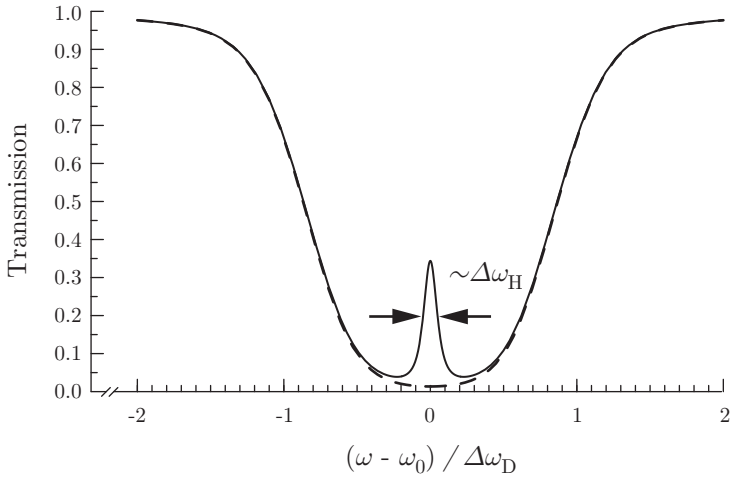

Figure 7.6: Calculated frequency dependence of the transmission of a weak probe through a sample of Doppler broadened atoms with (solid line) and without (dashed line) a counter-propagating saturating beam with the same frequency as that of the probe beam. Note that the main absorption has a frequency width of order the Doppler width $\Delta\omega_{\rm D}$, but the spike of reduced absorption has a width approximately equal to the homogeneous width of the atoms, $\Delta\omega_{\rm H}$. For the purposes of illustration, in this calculation $\Delta\omega_{\rm H}$ was set equal to $0.1\Delta\omega_{\rm D}$; in practice the ratio of the homogeneous and to inhomogeneous linewidths would usually be very much smaller.

The most convenient transitions suitable as reference frequencies occur in atoms or molecules in the gaseous state, and consequently it is necessary to use **Doppler-free** techniques to eliminate Doppler broadening, as illustrated in Fig. 7.4(b). In this approach the laser beam is divided into a strong 'saturating' beam and a weak 'probe' beam. The saturating and probe beams pass through the absorption cell in an exact, or near, counter-propagating geometry, and the intensity of the transmitted probe beam is measured by a photodiode. This is the same arrangement used in saturation absorption spectroscopy.

In order to determine the form of the error signal we must find the frequency response $a(\omega)$ of the reference system, which in this case is the frequency dependence of the amplitude transmission of the probe beam. Now, *in the absence of the saturating beam* the transmitted amplitude will be large for frequencies ω such that $|\omega - \omega_0| \gg \Delta\omega_{\rm D}$, and it will be small for detunings $|\omega - \omega_0|$ comparable to or small compared to $\Delta\omega_{\rm D}$. This behaviour is shown in Fig. 7.6.

Since they propagate in opposite directions, the saturating beam and the probe beam usually interact with different velocity classes. As such, for frequencies such that $|\omega - \omega_0| \gg \Delta\omega_{\rm H}$ the absorption of the probe beam is unaffected by the presence of the saturating beam. However, for frequencies $\omega \approx \omega_0 \pm \Delta\omega_{\rm H}$, both beams interact with the same ($v_z \approx 0$) class of atoms and hence the absorption experienced by the probe beam is reduced. This leads to a very narrow feature in the transmission of the probe beam, as shown in Fig. 7.6 which can be used to lock the frequency to ω_0.

7.3.2 Locking to an external cavity

Locking to the resonant frequency of an external cavity provides better short-term stability than locking to an atomic transition since the error signals are large and hence have a high signal-to-noise ratio. There are many ways in which this can be done, but the most widely used is the **Pound-Drever-Hall technique** illustrated schematically in Fig. 7.4(c).

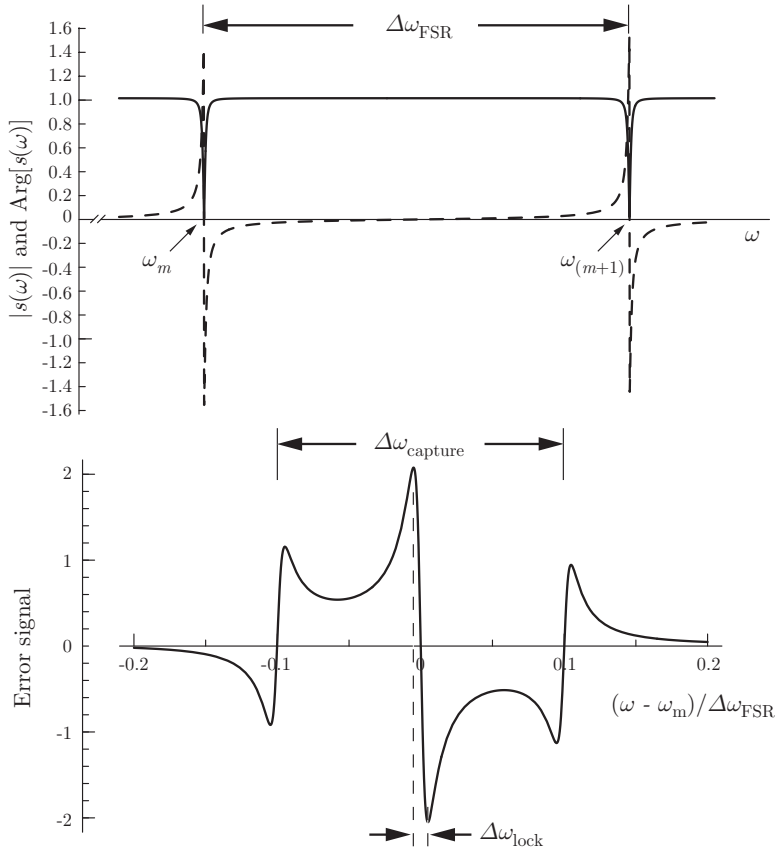

Figure 7.7: The Pound-Drever-Hall technique. (a) shows the amplitude (solid line) and phase (dashed line) of the beam *reflected* from a cavity as a function of frequency, where ω_m and $\omega_{(m+1)}$ are two adjacent resonant frequencies of the cavity. (b) shows the error signal as a function of frequency near one of the resonant frequencies. Notice that the 'capture' range of the error signal is a significant fraction of the free spectral range, $\Delta\omega_{\rm FSR}$, of the cavity, but that the error signal has a slope of very large magnitude near the resonant frequency.

The Pound-Drever-Hall technique utilizes the frequency-dependent *reflection* from a high-finesse cavity.[2] The advantage of using the signal reflected from the cavity is that the error signal has a much wider 'capture' range than that which would be obtained by monitoring the transmitted beam.

To see this we note that the cavity acts as a Fabry-Perot etalon with a transmission which is sharply peaked. As shown in Fig. 7.5(b) when used in transmission the resulting error signal would only extend over a frequency range of order the cavity linewidth $\Delta\omega_{\rm c}$, and this must be narrow for accurate locking. This behaviour may also be considered in the time domain; a narrow linewidth cavity corresponds to a long cavity lifetime $\tau_{\rm c}$ and consequently the transmitted beam does not respond to fluctuations in the beam which occur in a time significantly shorter than $\tau_{\rm c}$, or, equivalently, to frequency shifts greater than of order $1/\tau_{\rm c} \approx \Delta\omega_{\rm c}$.

The beam reflected from a reference cavity behaves differently since it is comprised of a component reflected directly from the first mirror plus leakage of radiation circulating within the cavity. The response function is more complex, and leads to an error signal which has a very wide capture range, yet a narrow central feature which can be used to lock the frequency to a tight tolerance. Fig. 7.7(a) shows the amplitude and phase of the

[2] In this application cavity finesses up to 2×10^5 may be obtained by using 'super-polished' mirrors.

signal reflected from a reference cavity. It may be seen that the reflected amplitude is approximately unity apart from sharp dips in the region of the resonance frequencies of the cavity; this behaviour is to be expected from our understanding of the *transmission* of a Fabry-Perot etalon. Fig. 7.7(b) shows the error signal derived from a beam modulated at a frequency Ω. The error signal is large, and has the correct sign, for frequencies within approximately $\pm\Omega$ of the resonance frequency. At the same time the central feature has a much smaller width, of order $\Delta\omega_{1/2} = \Delta\omega_{\mathrm{FSR}}/\mathcal{F}$, where $\mathcal{F}$ is the finesse of the reference cavity, allowing narrow frequency-locked linewidths to be obtained. This approach can be used to achieve extremely narrow linewidths if the reference cavity is isolated against mechanical and thermal fluctuations and drift.