6.013 Lecture 23: Lasers[1]

A. Overview

This lecture reviews the basic principles, operation, and applications of laser amplifiers and oscillators, followed by a few examples. This discussion circumvents quantum-mechanical derivations and proceeds directly to physical consequences, namely that atoms and molecules exist in discrete energy states and that they can change state by absorbing or emitting a photon of frequency $f = \Delta E/h$ [Hz], where ΔE [J] is the energy difference between the initial and final state, and h is Planck's constant (6.625×10^{-34} [Js]). Amplification of electromagnetic waves occurs when a stream of photons triggers coherent emission of additional photons from a medium having more atoms or molecules in the relevant upper energy state than are in the lower state. Oscillation occurs if these amplified waves are trapped by reflections in the amplification region so that they intensify until saturation is reached. Laser processes can occur in solids, liquids, or gases.

Lasers (Light Amplification by Stimulated Emission of Radiation) amplify electromagnetic waves or oscillate in bands ranging from radio frequencies, where they were developed (and called masers for "Microwave Amplification"), to ultraviolet wavelengths and x-rays. The physical principles are similar at all wavelengths, though the details may differ.

Laser amplifiers and oscillators generally utilize the energy levels of atoms in gases or solids, or transitions between the conduction and valence bands in semiconductors. For example, optical fiber communications systems today commonly use Erbium-doped fiber amplifiers (EDFA's) that amplify ~1.4-micron wavelength signals having bandwidths up to ~4 THz. Semiconductor, gas, and glass fiber amplifiers are also used to communicate within single pieces of equipment and for local fiber or free-space communications.

Lasers also generate coherent beams of light used for measuring distances and angles (e.g. lidar is analogous to radar), for recording and reading data from memory devices such as CD's and DVD's, and for cutting, welding, and shaping materials including even the human eye. Laser pointers have been added to pocket pens while higher-power industrial units can cut steel plates several inches thick. Weapons and laser-driven nuclear fusion reactions require still higher-power lasers. Peak laser powers can exceed 10^{15} watts by instantly dumping the energy stored in the upper energy states of an ensemble of atoms or molecules. This can be compared to the total U.S. electrical generating capacity, which is $\sim5\times10^{11}$ watts, three orders of magnitude less. The electric field strengths within a focal spot of <100-micron diameter can strip electrons from atoms and accelerate them to highly relativistic velocities within a single cycle of the radiation.

---

[1] Content contributed by A. Bers, E. Ippen, and D. Staelin, April, 2002.

B. Basic Physics of Optical Fiber Amplifiers

The basic principles of optical fiber amplifiers are well illustrated by a typical EDFA comprising an optical fiber doped with erbium atoms, as illustrated in Figure 23-1. Erbium has many energy states, two of which (labeled 1 and 2 in the figure) can readily be excited so as to have an inverted population, i.e. more atoms in the upper state than in the lower one. If the energy levels of these two states are E_1 and E_2, then the frequency f [Hz] of amplification is given by:

$$E_2 - E_1 = hf \ \ [J] \tag{1}$$

where h is Planck's constant, 6.625×10^{-34} [JHz^{-1}]. The energy for transferring erbium atoms to their upper state comes from the "pump" lasers surrounding the fiber that send light into the erbium-doped amplifying fiber core at angles less than the critical angle.

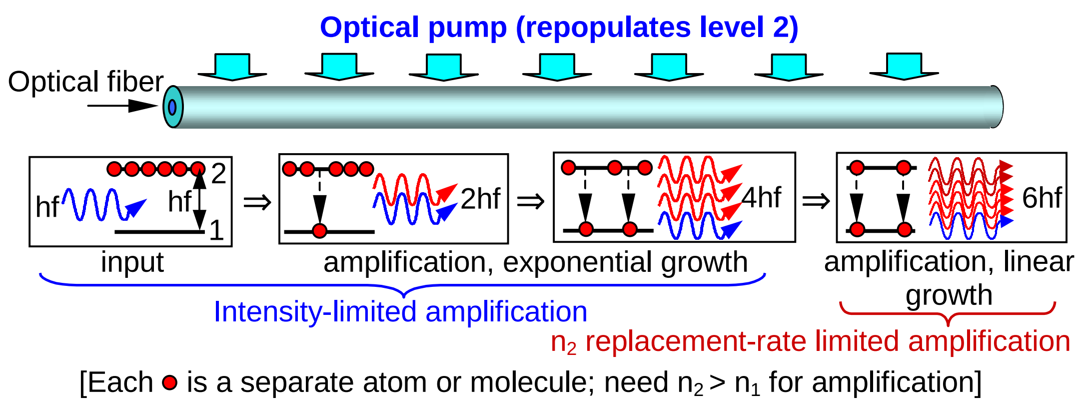


Figure 23-1. Optical fiber amplifier with exponential and linear growth

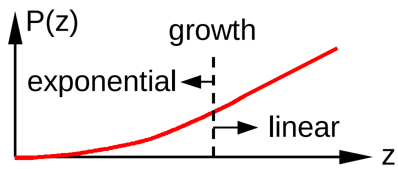
A propagating input photon of frequency f has energy hf [Joules] and can stimulate one of the atoms in the upper state E_2 to decay to a lower state E_1, thus releasing a second photon that is always in phase with the first, as suggested in the figure; both in turn stimulate two more to yield a total of four. The growing power in the wave equals Nhf, where N is the number of coherent photons comprising that beam. This amplification process then continues exponentially until the wave is so strong that the pump can no longer keep the upper state more populated than the lower state, so the photon absorption and emission events for these two states nearly balance. At this point the wave continues to amplify linearly with distance, limited by the atoms per meter the pump can excite per second. That is, the power increase, watts/meter, added to the propagating beam by the amplifier is limited to $N_p hf$, where N_p is the number of erbium atoms raised to level 2 by the pump per meter per second.

Figure 23-2 suggests how the signal power P(z) increases as it propagates in the z direction down the amplifying fiber. In the exponential region the power is:

$$P(z) \cong P_{in}\, e^{(g - \alpha)z} \qquad (2)$$

where g is the gain (nepers m^{-1}) and α is the absorption coefficient; P_{in} is the input signal power at z = 0. In the linear region the signal power is $P(z) \cong P_{in} + Dz$, where P_{in} is the input signal power at the beginning of the linear region, z is distance into that region, and D is the linear amplifier gain (Wm^{-1}).

Figure 23-2. Power growth in laser fiber amplifiers



The rate of change in the population density n of atoms per meter in state 2 is:

$$dn_2/dt \cong -An_2 - B(n_2 - n_1) \;\; [m^{-1}s^{-1}] \qquad (3)$$

where An_2 is the rate of spontaneous decay to other atomic states, Bn_2 is the rate of stimulated emission producing a photon coherent with the stimulus, and Bn_1 is the rate of photon absorption due to stimulated transitions from state 1 to state 2. B is proportional to the input photon flux F [photons s^{-1}], and therefore to the wave power P [W]. The power gained by the incoming wave is approximately $B(n_2 - n_1)hf$ [Wm^{-1}] where the power of that incoming wave is Fhf Watts. The An_2 spontaneously emitted photons are not coherent with the input beam and radiate uselessly in all directions.

The probability of spontaneous decay between energy states 2 and 1 is:

$$A_{21} = 2\omega^3|D_{ij}|^2/h\varepsilon c^3 \;\; [s^{-1}] \qquad (4)$$

where D_{ij} is a quantum mechanical dipole moment, electric or magnetic, specific to each pair of states i,j. The ω^3 dependence of A has a profound effect on maser and laser action. For example, any maser or laser must excite enough atoms to level 2 to equal the sum of the stimulated and spontaneous decay rates. But if the spontaneous decays increase with ω^3 and dominate, then the pump energy must increase with the product of the number of excitations required, ω^3, and the energy of each excitation, $h\omega/2\pi$; that is, pump power requirements increase very roughly as ω^4, making construction of x-ray or gamma-ray lasers extremely difficult without exceptionally high pump powers. Conversely, at radio wavelengths the spontaneous rates of decay are so small that exceedingly low pump powers suffice, as they sometimes do in the vast darkness of interstellar space.

In well-designed lasers atoms are removed very rapidly from level 1 by processes that are discussed later, leaving that state constantly empty. Otherwise the number density n_1 would increase and the term $B(n_2 - n_1)$ in Equation (3) would diminish, thus reducing amplifier gain.

C. Pumping of Lasers

Laser amplification can occur only when the population of the upper state n_2 exceeds that of the lower state n_1, as suggested by Equation (3). But Equation (3) also applies to changes in level-1 populations. Therefore, in the limit where A becomes negligible compared to B, a strong pump can only produce equilibrium between the two populations n_1 and n_2. Therefore exciting a laser with a pump signal tuned to the desired pair of levels ($f = (E_2 - E_1)/h$), as suggested in Figure 23-3a, cannot by itself produce laser amplification because $n_2 < n_1$. Other methods must be used; the most common involve three or four energy states.

Figure 23-3. Two-, three-, and four-level lasers

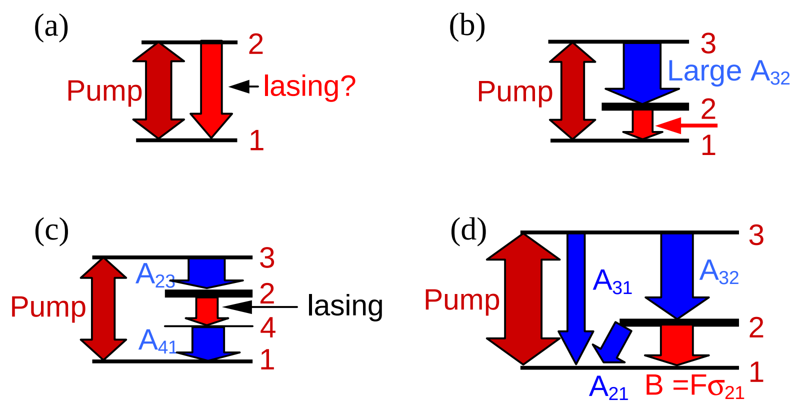


Figure 23-3b illustrates a three-level laser system where the pump maintains equilibrium between n_1 and n_3, while a large spontaneous decay rate A_{32} keeps n_3, and therefore n_1, near zero. The result is the nearly ideal case where the upper state is fully occupied and the lower state is empty. Figure 23-3c illustrates a four-level laser where large values for A_{23} and A_{41} produce a similar result, nearly empty levels 1, 3, and 4, and a heavily populated level 2. Four-level lasers are more common than three-level systems because of the greater number of available transitions and frequencies. Most lasers use the ground state of the atom or molecule as level 1; the ground state is that state into which all spontaneous transitions eventually would lead in the absence of any excitations by either photons or collisions.

Figure 23-3d suggests how the spontaneous transition rates A_{ij} can lower the power efficiency η of any laser. The intrinsic power efficiency η_i is the limit imposed by the ratio of the signal frequency f_s to the pump frequency f_p; no more than one additive signal photon can be produced per effective pump photon, and their energy ratios are hf_s/hf_p. Therefore the power efficiency η_i due to this frequency difference is simply f_s/f_p. We may also define a B/A efficiency η_B associated with the loss due to A_{21} stealing energy from the stimulated emission involving levels 2 and 1, and an A/A efficiency η_A due to spontaneous emission between levels 3 and 1 bypassing the desired path driven by A_{32} (see Figure 23-3). These efficiencies are:

$$\eta = \eta_i \eta_B \eta_A \quad (5)$$

$$\eta_I = f_s/f_p \quad (6)$$

$$\eta_B = B_{21}/(A_{21} + B_{21}) \quad (7)$$

$$\eta_A = A_{32}/(A_{31} + A_{32}) \quad (8)$$

The total power efficiency of most lasers ranges between one and 60 percent, and that efficiency typically drops sharply for wavelengths shorter than ~0.5 microns due to declining values for η_B as a result of the difficulties of obtaining high pump powers.

D. Energy States and Sources in Lasers and Masers

Erbium atoms in glass have electronic states associated with quantized electron orbits around those atoms, and state transitions simply consist of transitions of an electron between one orbit and another. In semiconductor lasers the state transitions occur between the conduction and valence bands within the thin transparent junction between the p- and n-type semiconductors, as suggested in Figure 23-4. At the ends of the flat p-n junction (in a lateral direction) are parallel mirrors that partially trap the laser energy, forming an oscillator. Such p-n junction lasers are pumped simply by forward-biasing the diode so that electrons thermally excited into the n-type conduction band can diffuse into the active region of the junction. In the active region photons can stimulate emission, yielding amplification and ultimately oscillation. Such diodes are readily voltage modulated to produce digital pulse streams carrying information at data rates above 100 Mbps.

Figure 23-4. Semiconductor laser

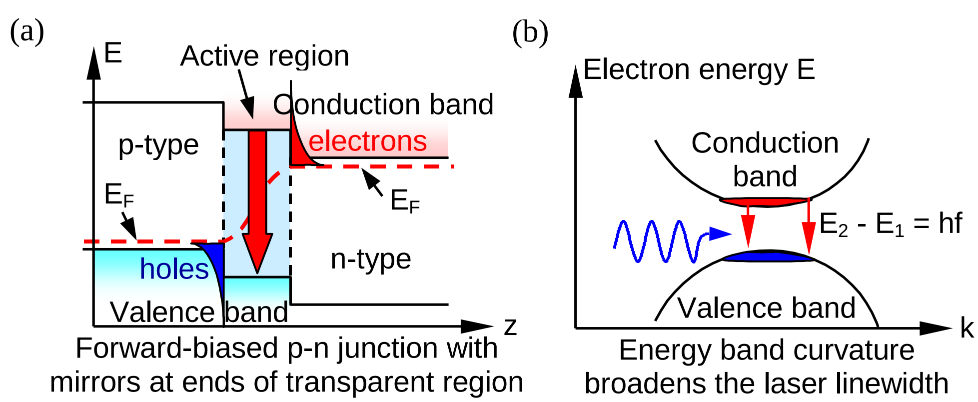


Figure 23-4a shows a side view of the diode, where the vertical axis is electron energy and the horizontal axis is position z through the diode from the p to n sides. The exponentials suggest the energy distributions of the holes and electrons in the valence and conduction bands, respectively. The Fermi level E_F is analogous to sea level for electrons; it tilts up toward the right because of the voltage drop placed across the diode. Figure 23-4b plots electron energy E versus the magnitude of the k vector for electrons (quantum approaches treat electrons as waves characterized by k), and suggests why diode lasers can have broad bandwidths. Incoming photons can stimulate any electron in the conduction band to jump to any empty level (hole) in the valence band, and both of these bands have significant energy spreads ΔE, where $\Delta f \cong \Delta E/h$ [Hz].

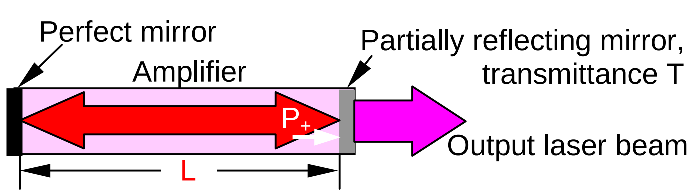
Early lasers operated in the microwave region and were called masers. Among the first were two-level NH_3 molecular beam vacuum systems that overpopulated the upper energy level by electrically guiding freshly excited upper-energy-state ammonia molecules to a different location than for the lower-state molecules. This cumbersome system was soon superceded by solid-state masers that incorporated crystals seeded with impurity atoms having suitable sets of energy levels. The first optical lasers were gaseous (e.g. helium-neon and carbon dioxide), and used electrically driven atomic collisions in mild vacuum to excite the upper energy levels. Later, ruby and glass lasers seeded with appropriate impurities became more common, as did semiconductor lasers of various types. The most powerful lasers use chemical or even nuclear explosions to produce energetic gases that can amplify at specific wavelengths.

Many types of astrophysical masers exist in low-density interstellar gases containing H_2O, OH, CO, and other molecules. They are typically pumped by strong radiation from nearby stars or by collisions occurring in shock waves. Sometimes the lasers radiate radially from stars, amplifying starlight, and sometimes they radiate tangentially along linear circumstellar paths that have minimal relative Doppler shifts. Laser or maser action can also occur in darkness far from stars as a result of molecular collisions. The detailed frequency, spatial, and time structures observed in astrophysical masers offer unique perspectives on a wide range of phenomena.

E. Laser Oscillators

Laser amplifiers oscillate nearly monochromatically if an adequate fraction of the amplified signal is reflected back so as to be amplified further. For example, consider the laser pictured in Figure 23-5.

Figure 23-5.
Laser oscillator



In this case perfectly parallel mirrors are placed at both ends of a laser amplifier, separated by L meters. Such a system will oscillate if the roundtrip gain at any frequency exceeds unity, where absorption and imperfect mirrors account for most lost. One of the mirrors transmits a fraction T of the power P_+ incident upon it, where T might be ~0.1.

Amplifiers at the threshold of oscillation are usually in their exponential region, so the condition for oscillation becomes:

$$P_+(1 - T)\, e^{2(g - \alpha)L} \geq P_+ \tag{9}$$

which implies $e^{2(g - \alpha)L} \geq (1-T)^{-1}$. Generally g is designed to be as high as practical, and then L and T are chosen. The pump power must be above the minimum threshold that yields $g > \alpha$.

The output power from such an oscillator is simply $P_{out} = TP_+$ watts, where $P_{out} = \eta P_{pump}$ and η is defined, for example, by Equation (5). Therefore:

$$P_+ = P_{out}/T = \eta P_{pump}/T \to \infty \quad \text{as} \quad T \to 0 \tag{10}$$

That is, P_+ can approach extremely high values if $T \cong 0$. This is the basis of the Q-switched laser, which uses $T \cong 0$ to build the level of P_+ and then switches T abruptly to ~1, dumping P_+ at power levels so great as to threaten survival of the device. If the pulses build quickly in the oscillator while $T \cong 0$, then little heating need occur prior to the output pulse; voltage breakdown then becomes a limiting factor. The name Q-switched follows from the fact that the Q of the resonator is abruptly switched from a very high value (all energy being trapped inside) to a very small value as the pent up energy escapes instantly.

The resonant frequencies of such a laser are controlled by the natural line frequency and linewidth, and also by the resonant frequencies of the TEM mirror cavity resonator. If the mirrors are perfect conductors, then the length L of the cavity must be an integral number of half wavelengths. That is, $m\lambda/2 = L$, where the wavelength λ is typically shorter than the free-space wavelength due to the index of refraction N of the laser material.

Figure 23-6. Resonances of a cavity laser.

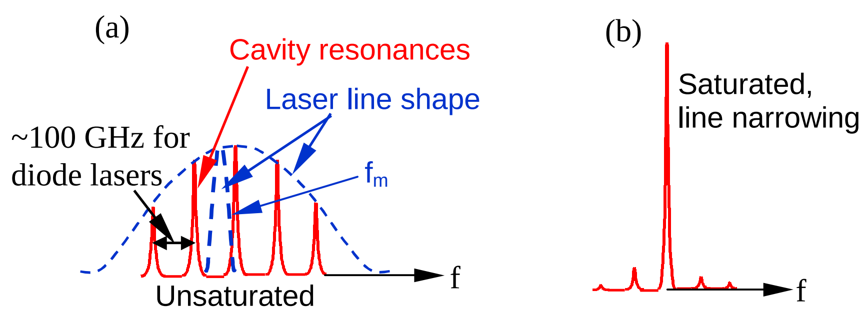


Figure 23-6a illustrates the natural line shape of a typical laser diode, within which the various cavity resonances can be observed. In general,

$$\lambda_m = 2L/m = c/f_m N \tag{11}$$

where these resonant frequencies $f_m = mc/2LN$ might be spaced at $c/2LN \cong 100$ GHz for a laser diode having $L \cong 1$ mm and $N = 1.5$. If the natural line shape is narrow compared to the spacing between cavity resonances f_m, also illustrated in Figure 23-6a, then the cavity might require tuning before the oscillator operates. If the amplifier has substantial gain in the exponential regime, then the line shape determines which laser resonance f_m has the most gain, and that resonance alone will dominate, as suggested in Figure 23-6b. If the gain is sufficient, the laser line can become nearly monochromatic and narrow compared to Δf, the natural width of the preferred cavity resonance.

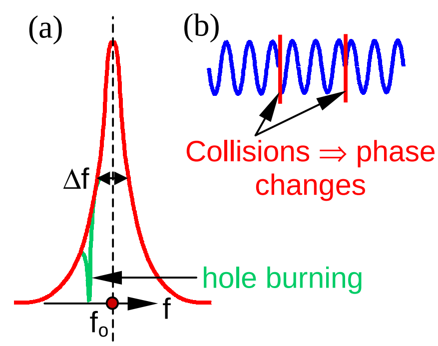
F. Laser Output Spectrum

The output spectrum of a laser oscillator depends only on its spectral gain and saturation characteristics, while the output of a laser amplifier also depends on its input spectrum. The gain characteristics of a laser are determined by its stimulated emission "B" coefficient, defined by Equation (3). In general, B is proportional to the incoming photon flux F (photons/second) and to the atomic cross-section σ_{ij} [m^2] for stimulated emission between the two given energy levels i,j. The frequency dependence of B_{ij} is:

$$B_{ij} \propto F\sigma_{ij} \propto F\, g_{ij}(f)\, A_{ij}/f^3 \qquad (9)$$

where $g_{ij}(f)$ is the spectral line shape for the i,j transition and A_{ij} is the probability of spontaneous emission [s^{-1}].

Figure 23-7. Laser spectral line shape



The shape of the spectral line associated with the i,j transition depends in part on the distribution of energy levels within the i and j groups. For example, the laser diode in Figure 23-4 has a spread in energy associated with the distribution of energy levels occupied by electrons in the conduction band and holes in the valence band. Similarly, the atoms in an EDFA are subject to slightly different local electrical fields due to the random nature of the glassy structure in which they are imbedded. This results in each atom having slightly different (fixed) energies for each level and in the ability of an EFDA to amplify over a relatively large bandwidth, much larger than the bandwidth of any single atom. In contrast, many lasers have no such line broadening induced by local inter-atomic fields, and other mechanisms control the line shape.

In gases the width of any spectral line is also controlled by the frequency of collisions. Figure 23-7b illustrates how an atom or molecule with sinusoidal time variations in its dipole moment might be interrupted by collisions that randomly reset the phase. An electromagnetic wave interacting with this atom or molecule would then see a less pure sinusoid. This new spectral characteristic would no longer be a spectral impulse, i.e., the Fourier transform of a pure sinusoid, but rather the transform of a randomly interrupted sinusoid, which has the Lorentzian line shape illustrated in Figure 23-4a. Its half-power width is Δf, which is approximately the collision frequency divided by 2π. The limited lifetime of an atom or molecule in any state due to the probability A of spontaneous emission results in similar broadening, where $\Delta f \cong A/2\pi$; this is called the intrinsic linewidth of that transition.

The spectral behavior of an amplifier can be different in its exponential and linear regions. In the exponential region governed by $P(z) \cong P_{in}\, e^{(g - \alpha)z}$, small differences in g(f) are exponentially magnified in high-gain amplifiers. In the linear region where $P(z) \cong P_{in} + Dz$, small differences in D(f) have only a linear effect.

In the saturated region of amplifiers where each atom interacts across the full bandwidth of the input signal, different signals compete for gain because the total power available is limited by the pump power. However all signals experience the same gain reduction as more are added because the power added to each signal frequency is proportional to B for that frequency, and B is proportional to the input photon flux F for each signal. Therefore the power added is proportional to the input power at each frequency. If one of the signals is very powerful and varies, however, the resulting time variations in the gain experienced by the other signals may be unacceptable. In the exponential growth (unsaturated) region of amplifiers, the upper level population n_2 is not impacted much by any of the signals, so all signals experience similar gains independent of the others.

In other amplifiers the active atoms or molecules have bandwidths narrower than that of the amplifier as a whole, and different subpopulations amplify different portions of the total band. In this case a strong signal that saturates one portion of the total amplifier band will not reduce the gain experienced by signals at frequencies that interact only with other atomic subpopulations. Strong signals can, however, sharply lower the gain for other signals at the same frequency, a process sometimes called "hole burning" because a "hole" appears in the gain spectrum, as suggested in Figure 23-7a. Sometimes such holes have lifetimes longer than the pulse that created them, depending on the repopulation rate for n_2.