6.014 Lecture 19:
Lasers[1]

A. Overview

This lecture reviews the basic applications, principles, and operation of laser amplifiers and oscillators, followed by a few examples. The principal simplification in this discussion is to replace all quantum-mechanical derivations with their basic results, namely that atoms and molecules exist in discrete energy states with some probability distribution, usually Boltzmann (exponentially declining with increasing energy). Transitions of an atom or molecule between two of its states often involves absorption or emission of photons of frequency $f = \Delta E/h$ [Hz], where ΔE [J] is the energy difference between the two states and h is Planck's constant (6.625×10^{-34} [Js]).

Lasers (Light Amplification by Stimulated Emission of Radiation) are generally used as light amplifiers or oscillators, and have operated from radio frequencies, where they were developed, to ultraviolet wavelengths and x-rays. The physical principles are essentially the same at all wavelengths, although some of the details may differ.

Currently Erbium doped fiber amplifiers (EDFA's) are widely used to amplify repeatedly optical signals of bandwidths up to ~4 THz as they propagate in glass fibers over intercontinental distances; for lack of demand most of this installed potential bandwidth is not currently used, however. They are also used in local areas and in single pieces of equipment for fiber or free-space communications. They further generate coherent beams of light used for measuring distances and angles (e.g. lidar is analogous to radar), and for recording and reading data from memory devices such as CD's and DVD's.

Higher-power units are increasingly used in manufacturing to mark or cut materials; even steel plates several inches thick are cut this way. Weapons and laser-driven fusion reactions require still higher-power lasers. Peak 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.

B. Basic Amplifier Physics

A typical EFDA comprises an optical fiber doped with erbium atoms. Erbium has many energy states, two of which (labeled 1 and 2 in L19-2) have an inverted population, i.e. more atoms in the upper state than in the lower one. A photon of frequency f has energy hf [Joules] and can stimulate one of the atoms in the upper state to decay to a

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

lower state, thus releasing a second photon that is in phase with the first, as suggested in Fig. L19-2. This amplification process then can continue exponentially, until the incoming wave is so strong that the upper level quickly equilibrates with the lower level, so the net absorptions and emissions of photon for these two states nearly balance.

This tendency for incoming photons to trigger state transitions up or down is governed by the equation for the rate of change in upper-state population n_2 (atoms m^{-3}):

$$dn_2/dt = -[An_2 + B(n_2 - n_1)] \quad (1)$$

where A is the Einstein "A" coefficient for state 2, which is its rate of spontaneous emission to any other state, and the B coefficient is the probability of stimulated emission (Bn_2 events occur $m^{-3}s^{-1}$) and absorption (Bn_1 events occur $m^{-3}s^{-1}$). B is proportional to the intensity of the ambient radiation, as shown below. The values of both A and B depend on the particular pair of states of interest, and are often represented as A_{ij} and B_{ij}. The total density of net stimulated power emission [Wm^{-3}] is simply:

$$P = hf\, d(n_2 - n_1)/dt \quad [Jm^{-3}s^{-1}] \quad (2)$$

Figure 19-3a illustrates typical energy levels of an electron in an isolated hydrogen atom. Its lowest, or "ground", state is -13.6 electron volts, which means that it requires a 13.6-e.v. photon to free the electron entirely (to "ionize" the atom) so that the electron leaves the nuclear proton behind. A large number of electronic states, increasingly dense at the higher energies, exists between the ground state and ionization. These states can be split by local magnetic or electric fields, or even by the orientation of the nuclear spin and its associated magnetic dipole moment (up or down). Molecules like water vapor store energy not only in their electronic configuration, but also in vibrational, bending, and rotational motions. Transitions between electronic, vibrational or bending, and rotational energy levels typically fall in the visible and ultraviolet, infrared, and microwave regions, respectively, although states commonly combine all three energy storage mechanisms.

In a cubic meter of material in thermal equilibrium there are n_i atoms or molecules in state i of energy E_i, where:

$$n_i/n_j = e^{-(E_i - E_j)/kT} \quad (3)$$

where T is the equilibrium temperature and k is Boltzmann's constant, where $k = 1.38\times10^{-23}$ [JK^{-1}]. In solids, liquids, and dense gases the frequent interatomic interactions or collisions establish thermal equilibrium at the kinetic temperature T_k of the medium. If a medium is suffused with sufficient radiation intensity that the level populations are determined primarily by the radiation, then the relative populations of two levels are characterized instead by the "radiation " temperature T_{rad}. For example, the interstellar medium is so sparse that collisions seldom control level populations; they are often determined instead by weak stellar radiation and A.

The spontaneous rate of emission of a single atom A_{ij} between states i and j depends on the quantum mechanical dipole moment $|D_{ij}|$ (electric or magnetic) linking those two states:

$$A_{ij} = k^3|D_{ij}|^2 2/3h\varepsilon \ [s^{-1}] \tag{4}$$

Thus the probability distribution of lifetimes of an atom in state i is exponential, and its time constant relative to i,j transitions is $\tau_A = A^{-1}$. Since $\tau_A \propto k^{-3} \propto \omega^{-3}$, ultraviolet and x-ray state lifetimes are extremely short, so these states must be replenished frequently in order for stimulated emissions to dominate the photon output.

The B coefficient (related to the "Einstein B coefficient", but including the radiation intensity) is:

$$B_{ij} = Ig_{ij}(f)|D_{ij}|^2 N\pi^2 2/3h^2 c\varepsilon \ \ [s^{-1}] \tag{5}$$

where I is the radiation intensity $|\overline{E}|^2/2\eta_o$ [Wm^{-2}], $g_{ij}(f)$ is the line shape (its integral is unity), and N is the index of refraction of the medium. For many simple systems the line shape is "Lorentzian" and given by:

$$g_{ij}(f) = [2/\pi(\Delta f)]/[1 + 4(f - f_o)^2/(\Delta f)^2] \ \ [Hz^{-1}] \tag{6}$$

The physical significance of the line-width Δf [Hz] is that it is the full width of g(f) at its half-power points.

C. Simple Lasers

Although lasers can utilize any number of levels, they usually rely principally on 2, 3, or 4. Here we illustrate the principles with three levels, as suggested in Figure L19-6. By strongly illuminating the laser material with radiation tuned to the energy gap $hf = E_3 - E_1$, the stimulated emission and absorption between levels 1 and 3 can nearly equalize those two populations. If the material system is well chosen, the spontaneous decay into level 2 can be very rapid, so soon nearly all atoms reside in that state alone. Incident photons with $hf = E_2 - E_1$ can then stimulate emission that is amplified as the waves propagate through the laser. The total amplification within that path length then depends on its length and on B_{21}. The same figure also illustrates a four-level laser, commonly used in erbium-doped fiber amplifiers (EDFA's).

If we assume that most atoms are in state 2 of the 3-level laser, then its population n is governed by (1), including the effects of the pump:

$$dn/dt = R - n(A + B) = R - (n/\tau_A) - I_{\#}\sigma n \tag{7}$$

where R = Repopulation rate s^{-1} of n, $I_{\#}$ = incoming photons $m^{-2}s^{-1}$, and σ = stimulated-emission cross-section [m^2] $atom^{-1}$.

In the steady state, dn/dt = 0, so we can solve (7) to find:

$$n = R\tau_A/(1 + \tau_A I_\# \sigma) = R\tau_A/(1 + I_\#/I_{\#sat}) \quad (8)$$

where $I_{\#sat} = (\tau_A \sigma)^{-1}$. The significance of $I_{\#sat}$ is that when the radiation intensity $I_\#$ exceeds this level, the equilibrium pumped population n is already halved due to its depletion by the stimulated transitions to level 1.

If we assume the stimulated emission rate B >> A, then the increase in $I_\#$ is:

$$dI_\#/dz \cong I_\# \sigma n \quad (9)$$

which has the simple solution:

$$I_{\#out} = I_{\#in} e^{\sigma n z} = I_{\#in} e^{gz} \quad (10)$$

where g is the laser amplifier gain (nepers m^{-1}). Thus the gain g is:

$$g = \sigma n = \sigma R\tau_A/(1 + I_\#/I_{\#sat}) \quad (11)$$

which approaches zero for $I_\# >> I_{\#sat}$. This gain $g = \sigma n$ is limited by the maximum achievable value of n, which occurs when all relevant atoms are in the upper state of the desired transition.

The pump power must be sufficient to produce R repopulations per second of level 2, and even with perfect efficiency this requires pump power $P_{pump} > hf_pR$ [Wm^{-1}], where f_p is the pump frequency. Since the population density n_2 [m^{-1}] typically varies along the length z of an amplifier, so does the local gain g(z). If we approximate n(z) using (8), and pump power by its 100-percent efficiency bound hf_pR, then we find the gain:

$$g(z) \cong \sigma P_{pump}/[hf_pA(1 + I_\#/I_{\#sat})] \quad (12)$$

and the total amplifier gain at any frequency, end-to-end, is:

$$I_{out}/I_{in} = e^{\int g(z)dz} \quad (13)$$

At the beginning of the amplifier the gain g(f) [m^{-1}] is constant and the signal grows exponentially. Later the signal $I_\#$ reaches such an amplitude that g(z) diminishes, although it remains positive; the signal amplitude then grows more linearly with distance.

Such amplifiers become more difficult when the A term is large relative to B. This happens most commonly at large values of ω, and therefore k, since $A \propto \omega^3$. Figure L19-6a shows how large values of A_{31} and A_{21} could short-circuit the pumping of level 2, thus reducing gain. Even if $A_{32} >> A_{31} + A_{21}$, the pump must not only pump more atoms per second into level 3, but these photons are more energetic. The lack of suitable pumps

is a major challenge for ultraviolet and x-ray lasers. Successful approaches often involve intense transient electrical discharges and collisional or chemical excitation mechanisms. In practice it is often easier to produce ultraviolet (UV) laser light by doubling the frequency of visible laser light in nonlinear media. Total UV system efficiencies of ~5 percent and more (wall socket to output beam) have been achieved this way.

The line shape of the emission or absorption associated with any particular energy transition depends on the details of the associated energy levels. Most lines (resonances) resemble the line illustrated in Figure L19-6b. Its width at half-maximum is Δf Hz. The common Lorentzian line shape $g_{ij}(f)$ is given by (6). The line width Δf has a minimum value determined by $\Sigma_j A_{ij}$, or the state lifetime; $\Delta f \cong \tau_A^{-1}$.[2] Usually this is overwhelmed by other line-broadening mechanisms such as collisions that perturb the energy level and alter its associated phase.

The nature of the line broadening affects the behavior of a laser. Laser transitions that exhibit homogeneous broadening incorporate atoms, each of which has a line width equal to that of the entire laser amplifier. In this case a strong stimulating signal in one narrow band can depopulate n (in the saturation region) and reduce the gain for other signals that begin weaker than the first. That is, the strong bands get stronger at the expense of the weaker ones. This fact is widely used to narrow the linewidth of laser output below that of the Δf intrinsic to that transition. Alternatively, amplifiers employing spectral lines that are inhomogeneously broadened tend to amplify all bands more equally. In this case each atom has a narrow linewidth Δf dependent on local circumstances, and a center frequency that could be anywhere across a much larger bandwidth. This center frequency differs for atoms that have different Doppler shifts or that experience different interatomic or macroscopic electric or magnetic fields. Helium-Neon gas lasers are an example. A strong signal can then steal strength (deplete n) only from a narrow band (Δf) nearby, but not over the entire band of the amplifier.

D. Laser Oscillators

Lasers that are closed at their ends so that amplified signals are trapped inside can amplify spontaneously produced photons to produce self-sustaining oscillations. This requires that the net gain experienced by the signal as it bounces back and forth across the length L of the amplifier is positive, producing exponential signal growth until saturation limits are reached. If one of the two end mirrors has a power transmission coefficient T, allowing some laser light to escape, then the net round-trip gain $e^{2(g-\alpha)L}$ (since there are also internal losses $\alpha(m^{-1})$) has to be increased to accommodate those losses to the external world. This constraint is:

$$(1 - T)e^{2(g-\alpha)L} \geq 1 \qquad (14)$$

[2] The spectrum $g_{ij}(f)$ is proportional to the Fourier transform of the ensemble of finite-duration sinusoids associated with oscillations at that frequency. Long-duration (large τ_A) states (sinusoidal pulses) yield narrow spectral lines. If collisions interrupt or alter the phase of these sinusoids, that too broadens Δf, where $\Delta f \cong \tau_c^{-1}$, where τ_c is the mean-free-time to collision.

The resonant frequency(s) of a laser must lie within the pumped bandwidth. If the line is homogeneously broadened, the resonant CW laser frequency will be the one that both 1) satisfies the resonant constraint that there must be an integral number of half-wavelengths along the axis of the laser, and 2) has the highest round-trip gain. If the line is inhomogeneously broadened, then the laser can oscillate at several frequencies simultaneously, generally separated by more than the local Δf, and only at frequencies that also satisfy the resonance constraint:

$$m\lambda_m/2 = L = mc/2f_m \Rightarrow f_m = mc/2L \tag{15}$$

The separation between resonant frequencies is thus c/2L Hz. For example, a table-top gas laser might be 1 meter long, so the spacing is ~150 MHz, whereas a semiconductor laser with refractive index $N \cong 3$ and $L \cong 0.5$ mm would have resonances separated by ~100 GHz.

E. Examples of Lasers

Some of the first lasers, or masers (Microwave Amplification by Stimulated Emission of Radiation), were discovered in space. Now many such space molecules have been found. Interstellar molecules are better candidates than atoms for maser action because they have many more energy levels spaced closely near the ground state than do atoms. Close spacing yields the low frequencies and large values of τ_A needed to build up large values of n in low energy environments like space. Many such molecules are located in HII regions of stellar formation where young hot stars provide intense ultraviolet light that pumps upper energy levels; the energy then cascades down the energy level diagram like an irregular waterfall, preferentially overpopulating certain energy levels that can then amplify. Stellar envelopes are another such environment, where gas clouds circling the star may share the same Doppler shift along paths generally tangential to the star and perhaps several stellar radii away (see Figure L19-9). These very long path lengths can amplify random noise at the far end of the path to levels easily seen by terrestrial radio telescopes. Alternatively, the bright star provides both the pump and the initiating radiation at the maser frequency, and the maser emission is more nearly radial. Interstellar clouds can similarly amplify, but generally with less intensity because of the weaker pumping.

The first masers used hydrogen and ammonia gases where the atoms or molecules were released in partial vacuum by a low temperature source. Then the upper-state molecules were electromagnetically separated by fields that applied forces to the particles that depended on the particle energy state.

Solid-state lasers like glass fibers or ruby crystals are pumped by flash lamps or CW semiconductor lasers, while semiconductor lasers directly inject electrons or holes into a p-n junction where they can support stimulated emission across the bandgap, as illustrated in Figure L19-10b. Because the energy levels occupied in the conduction and valence band are distributed, these homogeneous lasers exhibit broad bandwidths.