2 Masers and Lasers

January 31, 2008

When radio waves were first set up by Hertz and detected at a distance, it gave a striking confirmation of Maxwell's electromagnetic theory, and of its prediction that longer wavelength electromagnetic radiation would travel through space at the same speed as light and the invisible infrared and ultraviolet radiation. However, there are important differences between typical radio sources and typical light sources of the traditional sort, which are:

1. It is fairly easy to make a radio transmitter with a sharply defined and stable frequency, while even the best traditional light sources are broadened by the limited time-scale of an atomic or molecular state, and by the Doppler broadening caused by the motion of the atoms
2. A radio transmitter emits radiation from a coherent motion of electrons over the whole antenna, and this motion is driven by coherent motion of the charges in what is basically an LC (inductance–capacitance) circuit. The conventional light source, whether the sun, a star, a candle, an incandescent bulb, or a fluorescent source, is the result of incoherent emission of light from a large number of independent sources of atomic or molecular size.
3. The quanta of light were readily detectible once Einstein pointed out the significance of the measurements of the photoelectric effect. The energies of quanta of radio waves were very much smaller, and were much harder to detect.

For a long time there was a big gap between the longest wavelengths in the infrared and the shortest available radio waves. By 1940 it had become a military necessity to develop radar in the centimetric wavelength range, and Randall and Boot developed the cavity magnetron, which was compact enough to power an airborne radar set, and generated short enough wavelengths to locate a hostile airplane accurately. In this way we got the technology to power our microwave ovens.

It was the development first of *masers* (Microwave Amplification by Stimulated Emission of Radiation) and then of *lasers* (Light Amplification by Stimulated Emission of Radiation) that extended the existence of coherent sources right into the ultraviolet spectrum. The fundamental theoretical idea was contained in a 1917 paper by Einstein on the relation between absorption rates and emission rates of radiation by matter, which is discussed in the next section.

As late as the early 1970s it was being argued that, despite the extravagant claims made for lasers and masers, they had not actually proved useful. Such a claim was drawn to my attention by my father-in-law. My reply to him was that since many of my colleagues were doing experiments with the help of lasers that would have been impossible without them, sooner or later engineers and inventors would follow their example, and find applications of more general interest. The past thirty five years have supported my optimism.

2.1 Absorption and emission rates of radiation

In a remarkable paper published in 1917, Einstein derived a relation between the rate at which an atomic or molecular system radiates energy, and the rate at which it will absorb energy from an external electromagnetic field of the appropriate frequency. This argument is based on the *principle of detailed balance*, which evolved from Einstein's earlier work on Brownian motion. This principle states that, in thermal equilibrium, the flow of energy along each possible path in one direction is precisely balanced by the flow in the opposite direction. This is rather like the economics of a barter economy, in which the value of goods received by A from B are precisely balanced by the value of goods received by B from A. It is not obvious that the principle of detailed balance is true, but it is related to the principle of *time reversal invariance*, which is true for most, but not all, microscopic processes.

Einstein's argument considers a set of atoms of one particular type, each with levels i, j of energy ϵ_i, ϵ_j. In classical electromagnetic theory, the rate at which a system gains or loses energy from an electromagnetic field is proportional to the intensity of the field, the square of the amplitude of the field. In addition, we know that atomic and molecular systems in a higher energy state can radiate a photon and go down to a lower energy state. For energy to be conserved in the transition from the state j to the state i with the emission of a photon of frequency ν we must have $\epsilon_j - \epsilon_i = h\nu > 0$. The intensity of the field in the mode of frequency ν is proportional to the number N_ν of phonons in the mode. The transition rate R_{ij} from i to j with the absorption of a photon and the transition rate R_{ji} from j to i with emission of a photon can then be written as

$$R_{ij} = I_{ij} n_i N_\nu \ , \quad R_{ji} = I_{ij} n_j N_\nu + S_{ji} n_j \ , \tag{1}$$

where I_{ij} gives the rate per atom and per photon of the transitions induced by the field, and S_{ij} gives the rate of spontaneous transitions per atom from the higher energy state. Einstein's principle of detailed balance tells us that these two rates must be equal when the system is in equilibrium, so we have $R_{ij} = R_{ji}$ when, in accordance with the Boltzmann distribution for the atomic states,

$$n_i \propto \exp(-\epsilon_i/k_B T) \ , \quad n_i/n_j = \exp[(\epsilon_j - \epsilon_i)/k_B T] \ . \tag{2}$$

and N_ν has its equilibrium value for a mode of frequency ν in a cavity. When we put the detailed balance condition together with Eqs. (1) and (2), we get

$$S_{ji} = [I_{ij} \exp(\epsilon_j - \epsilon_i)/kT) - I_{ji}] N_\nu = [I_{ij} \exp(h\nu/kT) - I_{ji}] N_\nu \ . \tag{3}$$

When this is combined with the Planck formula for cavity radiation, which, for a single mode of frequency ν, is

$$N_\nu = \frac{1}{\exp(h\nu/kT) - 1} \ , \tag{4}$$

it gives

$$I_{ij} = I_{ji} = S_{ji} \ , \tag{5}$$

so the coefficients of the spontaneous and induced emission, and of induced absorption, are all the same. Einstein actually used a more careful version of this argument as a derivation of the Planck distribution for thermal radiation, as well as to derive the relation between the rates of spontaneous and induced emission.

2.2 Stimulated emission of radiation

Nearly forty years after Einstein discovered this theoretical result, Townes realized that it provided a way of persuading a collection of atoms to oscillate together and to emit their radiation coherently. Townes was an expert on the physics and applications of microwaves, like many other scientists of his age, so he worked out how to apply this idea to microwaves, and the maser was developed.

There are three essential features of any such device. The first is that a population of atoms or molecules predominantly in a single excited state should be set up. The second is that any photon emitted by one of the atoms should have a high probability of exciting emission of photons from other atoms, and these other photons in their turn can cause further emission. The result of this will be that the early photons emitted trigger a coherent emission from all the participating atoms. The third feature is that there must be a way for the emitted coherent light beam to escape from the region occupied by the atoms before it is reabsorbed. Such a device can either be *pulsed*, in which case the atoms in excited states descend to lower energy states before they are raised to higher energy states again, or *continuous*, for which there is a continuous pumping up of atoms to the high energy state while the radiation is continuously being emitted. The maser was rather easier to develop than the laser, because it was easier to develop good traps for wavelengths of the order of centimeters than for wavelengths of the order of fractions of a micron. Townes and his collaborators decided to study the transition in ammonia with a wavelength of 1.25 cm, firing a beam of excited ammonia molecules across a resonant cavity that trapped radiation of that wavelength. They had to reduce the losses in their cavity so that the amplification produced by the maser action would overcome the natural losses in the cavity. The beam was selected in such a way that the wanted excited state of the molecules predominated in the beam.

There are two important characteristics of maser and laser action, which distinguished physicists found counterintuitive and, initially, unbelievable. In his autobiography Townes says that both Niels Bohr and von Neumann told him that the maser could not be doing what he said it was doing, although von Neumann only took a quarter of an hour to convince himself that it was really OK. The first is that it acts as an amplifier of incident radiation, so that more radiation comes out at the laser frequency than goes in. The second feature, which was a big surprise to those who had not thought about it, was that the spectral width of the emitted radiation was much less than would be obtained from the individual molecules. These are features that can be seen in natural laser action from astronomical sources. Instead of the broad dark lines seen in the spectrum of the sun (the lines are due to absorption by the solar atmosphere of light emitted from the surface), where lasers occur there are narrow bright lines in the spectrum. The narrowing occurs because it is the *average* velocity of the molecules that determines their coherence with the wave, just as it the average velocity of electrons in an antenna that determines the coherence of a radio signal. Similarly, it does not matter when individual atoms or molecules enter and leave the beam, but it the overall size of the collection of molecules that determines the coherent response.

2.3 Solid state and gas lasers

The first working laser was produced by Ted Maiman in 1960, at Hughes Research Labs. It was based on a synthetic ruby, which as an aluminum oxide crystal, with inclusions of chromium ions (Cr^{3+}), which give the characteristic red color of the gem. The chromium impurities are found at random positions in an Al_2O_3 crystal. A flash of light is focussed onto the crystal, and this excites many of the ions to an excited states, which then decay, non-radiatively, to a long-lived state that decays by laser action to the ground state, emitting photons of 694.3 nm wavelength. The ruby crystal is made as a long thin rod with flat ends; light is trapped in the rod by reflection from the sides of the rod and emitted through the ends. This is a pulsed laser.

The first gas laser was invented by Ali Javan and Bill Bennett at Bell Labs, also in 1960. This operates with a mixture of helium and neon. Current passed through the gas excites the helium atoms, which transfer their energy to a matching long-lived state of similar energy in the neon atoms. The photons are trapped in a narrow cylindrical region with transparent ends, and laser action produces light at 632.8 nm wavelength. This laser can operate continuously, as was one of the lasers used early for barcode scanning.

2.4 Semiconductor lasers

Semiconductors can be used as lasers, and, with the help of semiconductor technology developed for other purposes, their construction and operation is relatively simple. Energy can be supplied to a semiconductor by passing an electric current, and, just as in a light-emitting diode, a fair part of the energy supplied can go to exciting electrons from the valence band to create electron-hole pairs. After these pairs have been created the electrons will lose energy by phonon emission or scattering to go to the bottom of the conduction band, while the holes will also lose energy to the phonon system and go to the top of the conduction band (a *hole* has its lowest energy state when it displaces one of the highest energy electrons, just as a bubble in water has its lowest energy when it is at the top of the water).

Now we need to discuss the difference in behavior between a *direct* band-gap semiconductor, like GaAs, and an *indirect* band-gap semiconductor like Si or Ge. In Eq. (1) of Chapter 1 the electron wave function was written as

$$\psi(x) = \cos(kx) f_k^\alpha(x) \quad \text{or} \quad \sin(kx) f_k^\alpha(x)\,,$$

or, more conveniently, in its three-dimensional complex form

$$\psi(x) = [\cos(\mathbf{k}\cdot\mathbf{r}) + i\sin(\mathbf{k}\cdot\mathbf{r})] f_{\mathbf{k}}^\alpha(\mathbf{r})\ .$$

Here the vector $\mathbf{k}$ gives the spatial variation of the wave function from one lattice site to another, and $\hbar\mathbf{k}$ gives the momentum of the electron. Here $\hbar$ is Planck's constant over 2π, from which the coffee bar in the Physics-Astronomy Building gets its name. In most of the standard semiconductors with a diamond-like structure (*zinc blende structure* for the III-V and II-VI materials), the top of the valence band is at $\mathbf{k} = 0$, so that is where the holes settle down when they have lost their excess energy. In a direct band-gap semiconductor such as GaAs the lowest point of the conduction band is also at $\mathbf{k} = 0$, so that an electron at the bottom of the conduction band can recombine with a hole at the top of the valence band to give a photon whose energy is equal to the band gap. Momentum is conserved in this process because the energy of a photon in the visible or in the near infrared is negligible compared with the energy carried by electrons and holes. Coherent emission of photons whose energy is equal to the band gap can give laser action.

In silicon and germanium the situation is quite different, because the lowest energies in the conduction band have values of $\mathbf{k}$ that correspond to quite large values of momentum, and such electrons can only recombine with $\mathbf{k} = 0$ holes at the top of the valence band if the energy goes to a phonon as well as to a photon, so that the momentum as well as the energy is conserved. In such a recombination the photon will have a variable energy which is less than the energy of the band-gap. This is not only an inefficient way of producing light, but it is hopeless for a laser, since the photons are spread over a wide frequency (energy) range, so cannot be coherent.

Gallium arsenide has an energy gap of 1.43 eV, and this corresponds to a wavelength of 867 nm, in the infrared, which may be fine as a flashlight for a night vision device, but does not give visible light, and might even be inconvenient for scanning your groceries. The energy gap can be increased without destroying the quality of the crystal by replacing a proportion of the gallium atoms by aluminum atoms, to get a solid of the formula $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$, colloquially known as AlGas. This remains a direct gap semiconductor up to $x = .38$, which gives a gap of more than 1.8 eV, corresponding to a wavelength of around 660 nm This is used as the (thin) central slab of a sandwich between two slabs of GaAs. One side of the sandwich is p-doped, the other n-doped, and the difference in refractive index of the two semiconductors has the effect of trapping most of the light in the region where the gap is largest. This method produces a quite efficient semiconductor laser at the lower frequency end of the visible spectrum, and Kroemer and Alferov received the 2000 Nobel Prize in Physics for their respective contributions to this development.

There was a long gap between the development of red semiconductor lasers and of blue lasers, and it was only a few years ago that the difficulties were overcome by Shuji Nakamura, of the Nichia Corporation. He received a bonus of 20,000 Yen for his development of the GaN laser, moved to UC Santa Barbara in 1999, and successfully sued his former employers for an increase of the bonus by a factor of 10^4.

2.5 Natural lasers and masers

Small lasers and masers generally need careful design to trap the radiation for long enough for laser action to occur. In a dense gas cloud in the neighborhood of a star gravity can trap the gas of excited ions, and the region of its confinement can be large enough for laser action to occur without trapping of the radiation. Such lasers or masers show up because of their narrow and intensely bright emission lines.

2.6 Applications of lasers and masers

The scientific uses of lasers and masers have been obvious for a long time. Their high intensity, sharp frequency, and strong directionality give them obvious uses for making measurements which would otherwise be impossible. High precision spectroscopy can be used to measure the speeds of molecules, to get atomic time standards, or to measure lengths accurately. The lunar ranging device, which uses the transmission of a very narrow laser beam from the earth, its reflection from a fixed mirrors on the moon, and its detection a couple of seconds later on the earth was an early spectacular use of lasers. At a more mundane level, scanning of bar-codes in shops, use of a laser pointer in a slide presentation, or use of lasers to control agricultural or road-building machinery from a distance are well publicized applications. Laser printers and scanners, laser readers for CDs are widely used applications.